DIGITAL TRANSFORMATION OF MEDICATION LABEL IDENTIFICATION IN THE OPERATING THEATRE: TECHNOLOGICAL EVOLUTION

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Supervisor

Dr Samaneh Madanian

Dr Minh Nguyen

30 April 2026

By

Hoang Nam Nguyen

School of Engineering, Computer and Mathematical Sciences

Abstract

Medication errors remain a significant challenge in healthcare, often resulting in adverse patient outcomes. Traditional identification methods, such as barcode scanning and RFID/NFC, though effective in certain contexts, fall short in handling the complexities of curved surfaces and multi-angle visual data associated with cylindrical medication containers. This research explores the potential of computer vision techniques to address these limitations by proposing novel methods for unwrapping and stitching medication labels captured from multiple angles.

The study investigates two approaches: Image Stitching Followed by OCR and OCR Followed by Text Stitching, each designed to enhance the accuracy and efficiency of medication identification. The proposed methods aim to tackle challenges such as label distortion, varying orientations, and inconsistent label designs. A multi-camera system is developed to capture images from different perspectives, followed by the implementation of algorithms for unwrapping and stitching. The performance of the proposed methods is evaluated based on their ability to extract accurate medication information and integrate seamlessly into healthcare workflows.

The results demonstrate that both methods offer high accurate medication identification results. This research contributes to advancing automated medication identification, ultimately reducing errors and improving patient safety. The findings provide a foundation for integrating computer vision into practical healthcare systems, delivering a scalable and reliable solution for medication identification.

Keywords: medical identification, cylindrical distortion, image stitching, OCR

Contents

Abstract **2**

Attestation of Authorship **19**

Artificial Intelligence Declaration **20**

Publications **21**

Acknowledgements **22**

1 Introduction **23**
1.1 Background 25
1.2 Motivation 26
1.2.1 Medication Error: Risks and Realities 27
1.2.2 Factors Contributing to Medication Errors 30
1.2.3 Role of Technology 31
1.2.4 CV in Medication Identification 33
1.3 Research Objective 34
1.4 Contribution of This Thesis 35
1.5 Structure of the Thesis 36

2 Literature Review **39**
2.1 Introduction 40
2.2 Review Methodology 42
2.2.1 Search Strategy 42
2.2.2 Literature Characteristics 43
2.3 Manual Process in Medication Administration and Identification . . . 47
2.3.1 Limitations of Manual Processes in General Healthcare Settings 48
2.3.2 Limitations of Manual Processes in Operating Theatres and Anaesthesia Settings 49
2.3.3 The Role of Technology in Reducing Medication Errors 50
2.4 Barcode/QR code 51
2.4.1 Applications of Barcode/QR Code 52
2.4.1.1 Anaesthesia/Operating Theatres 54
2.4.2 Advantages of Barcode/QR Code 55

2.4.3 Limitations of Barcode/QR Code . . . 57
2.5 RFID/NFC . . . 60
2.5.1 Application of RFID/NFC . . . 61
2.5.1.1 Operating Theatre and Anaesthesia Settings . . . 62
2.5.2 Advantages of RFID/NFC . . . 63
2.5.3 Disadvantages of RFID/NFC . . . 65
2.6 AI and CV . . . 67
2.6.1 Applications of CV Systems . . . 68
2.6.2 Advantages of CV . . . 70
2.6.3 Disadvantages of CV . . . 72
2.7 Comparative Analysis of Medication Identification Technologies . . . 79
2.7.1 Cost and Scalability . . . 79
2.7.2 Accuracy and Error Prevention . . . 82
2.7.3 Efficiency and Workflow Integration . . . 85
2.8 Research Gaps in Medication Identification Technologies . . . 88
2.8.1 Limited Research about Hybrid Technologies . . . 88
2.8.2 Challenges in Real-World Implementation . . . 89
2.8.3 Lack of Real-Time Processing and Scalability . . . 89
2.8.4 Cost and Resource Barriers . . . 90
2.9 Conclusion . . . 90

3 Research Methods **92**
3.1 Introduction . . . 93
3.1.1 Objectives and Scope . . . 93
3.1.2 Research Problem . . . 94
3.1.3 Research Approach . . . 95
3.2 Research Design . . . 97
3.2.1 Methodology Overview . . . 97
3.2.2 Overall System Architecture . . . 102
3.3 System Setup and Data Acquisition . . . 103
3.3.1 Dual-Camera System . . . 103
3.3.2 Experimental Platforms . . . 106
3.3.3 Software Architecture and Implementation . . . 106
3.3.4 Data Collection and Annotation . . . 107
3.3.4.1 Drug Set Composition . . . 108
3.3.4.2 Component Training Datasets . . . 108
3.3.4.3 End-to-End System Evaluation Dataset . . . 110
3.3.4.4 Dataset Summary . . . 111
3.4 Medication Database Foundation . . . 114
3.4.1 Baseline Database Requirement . . . 114
3.4.2 Motivation for Database Enrichment . . . 114
3.4.3 Query and Matching Logic . . . 115
3.5 Evaluation Protocol and Metrics . . . 116
3.5.1 Experimental Setup and Testing Protocol . . . 116

3.5.2 Intermediate Performance Metrics 117
3.5.3 System Identification Metrics 118
3.6 Conclusion 119

4 Approach 1: Visual Reconstruction of Medication Labels via Image Stitching 122
4.1 Introduction 124
4.2 Related Work 125
4.2.1 Cylindrical Label Reconstruction Under Partial Visibility and Distortion 125
4.2.1.1 Geometry-Based Approaches and Deployment Barriers 126
4.2.1.2 Learning-Based Methods and Practical Considerations 126
4.2.1.3 Research Gap and Design Rationale 127
4.2.2 Segmentation for Cylindrical Label Isolation 130
4.2.3 Ellipse Detection Design: Contour-Based vs. Learning-Based Approaches 131
4.2.4 Cylindrical Unwrapping and Classical Feature-Based Stitching 132
4.3 Proposed Image Stitching Pipeline 134
4.3.1 Dataset Acquisition 135
4.3.1.1 Evaluation Proxy and Laboratory Setting 135
4.3.1.2 Image Acquisition 136
4.3.1.3 Dataset Annotation 136
4.3.2 Label Segmentation and Geometric Modeling 138
4.3.2.1 YOLOv8-Based Instance Segmentation 138
4.3.2.2 Ellipse Estimation Strategies 140
4.3.3 Traditional Contour-Refined Ellipse Fitting 141
4.3.3.1 Contour Extraction 141
4.3.3.2 Keypoint Localization and Horizontal Point Sampling 143
4.3.3.3 Vertical Line Extension and Ellipse Point Sampling 145
4.3.3.4 Ellipse Parameter Fitting 146
4.3.4 Deep Learning Based Ellipse Detection 148
4.3.4.1 Synthetic Dataset Creation 148
4.3.4.2 Network Architecture 151
4.3.4.3 Loss Function 153
4.3.5 Cylindrical Label Unwrapping 155
4.3.5.1 Mesh Generation 156
4.3.5.2 Geometric Transformation 157
4.3.6 Image Registration and Stitching 159
4.3.6.1 Feature Detection and Description 159
4.3.6.2 Feature Matching and Geometric Alignment 160
4.4 Experimental Evaluation and Results 161
4.4.1 Experimental Setup and Dataset 161
4.4.2 Model Training and Hyperparameters 162
4.4.3 Evaluation of Label Segmentation (YOLOv8-s) 163

4.4.4 Ellipse Estimation Results - Traditional vs. Deep Learning Comparison . . . 165
4.4.5 Cylindrical Distortion Correction Results - Unwrapping Evaluation . . . 168
4.4.6 Image Stitching Results - Multi-view Fusion Performance . . 170
4.4.7 Task-Specific Evaluation: Text Extraction Quantity . . . 173
4.5 Limitations and Future Work . . . 175
4.6 Conclusion . . . 177

5 Approach 2: Textual Reconstruction of Medication Labels via Text Stitching 180
5.1 Introduction . . . 182
5.2 Related Work . . . 184
5.2.1 OCR in Healthcare and Curved Surfaces . . . 184
5.2.2 Domain Adaptation and Fine-Tuning for OCR . . . 185
5.2.3 Text Alignment and Stitching . . . 186
5.3 Proposed Method . . . 187
5.3.1 Fine-tuning PaddleOCR for Medication Label Recognition . . 188
5.3.2 Dataset Collection and Annotation . . . 188
5.3.2.1 PaddleOCR Pipeline Architecture . . . 190
5.3.3 Alignment of Text Fragments . . . 191
5.3.4 Text Fragment Merging . . . 193
5.4 Experimental Results . . . 197
5.4.1 PaddleOCR Fine-tuning Evaluation . . . 197
5.4.1.1 Quantitative Results on Validation Set . . . 197
5.4.1.2 Qualitative Evaluation and Failure Mode Analysis . 199
5.4.2 Synthetic Dataset for Text Stitching Evaluation . . . 201
5.4.2.1 Synthetic Dataset Generation . . . 201
5.4.2.2 Quantitative Evaluation . . . 204
5.4.2.3 Qualitative Evaluation . . . 206
5.4.3 Application to Real Medication Container Images . . . 208
5.4.3.1 Quantitative Evaluation on Real Dataset . . . 209
5.4.3.2 Qualitative Results: Two-View Stitching . . . 211
5.4.3.3 Qualitative Results: Multi-View Stitching . . . 213
5.5 Conclusion . . . 215

6 LLM-Assisted Construction of a Medication Database Using Retrieval-Augmented Generation 219
6.1 Introduction . . . 220
6.1.1 Background and Motivation . . . 220
6.1.2 Problem Statement . . . 222
6.1.2.1 Limitations of Reconstruction Outputs . . . 222
6.1.2.2 Why Reconstruction Outputs Cannot Serve as a Final Database . . . 223

6.1.2.3 Objective . . . 223
6.2 Related Work: LLMs in Information Extraction . . . 224
6.2.1 LLMs for Structured Extraction and Schema Mapping . . . 224
6.2.2 Retrieval-Augmented Knowledge Integration . . . 225
6.3 Proposed Database Construction Pipeline . . . 226
6.3.1 Overall Pipeline and System Flow . . . 226
6.3.2 LLM Agent for Data Structuring and Normalization . . . 227
6.3.2.1 Text Consolidation and Error Correction . . . 228
6.3.2.2 Schema Mapping and JSON Generation . . . 229
6.3.3 Knowledge Integration: Retrieval and Enrichment . . . 231
6.3.3.1 Knowledge Base Construction . . . 231
6.3.3.2 Two-Stage Retrieval Process . . . 233
6.4 Experimental Results . . . 236
6.4.1 Overview . . . 236
6.4.2 Experimental Setup . . . 236
6.4.3 Quantitative Evaluation . . . 239
6.4.3.1 Field Completeness . . . 240
6.4.3.2 Critical Field Accuracy . . . 242
6.4.4 Case Studies: System Behavior and Outcomes . . . 242
6.4.4.1 Case 1: Successful reconstruction under multi-view input. . . . 243
6.4.4.2 Case 2: Recovery of missing context via RAG. . . . 245
6.4.4.3 Case 3: Incorrect entity reconstruction under ambiguous OCR input. . . . 245
6.4.4.4 Case 4: Retrieval-induced enrichment with misaligned context. . . . 248
6.5 Web Application Development and Clinical Workflow Integration . . . 248
6.5.1 Clinical Workflow and System Integration . . . 250
6.5.2 Workflow Overview . . . 251
6.5.2.1 Medication Image Upload and Automatic Processing 252
6.5.2.2 Manual Verification and Editing . . . 252
6.5.2.3 Database Storage and Retrieval . . . 253
6.5.3 Front-end Design . . . 254
6.5.4 Backend and Database Architecture . . . 255
6.5.5 Implementation Details . . . 256
6.6 Conclusion . . . 257
6.7 Limitations . . . 258
6.8 Future Directions . . . 259

7 Experimental Results **261**
7.1 Introduction . . . 262
7.2 Experimental Setup . . . 263
7.2.1 Dataset Description . . . 263
7.2.2 Experimental Setup . . . 265

7.2.2.1 Hardware . . . 266
7.2.2.2 Evaluation Metrics . . . 267
7.2.3 Method 1: Image Stitching Setup . . . 268
7.2.4 Method 2: Text Stitching Setup . . . 269
7.3 Quantitative Results . . . 271
7.3.1 Top-1 and Top-5 Accuracy . . . 271
7.3.2 Top-1 Similarity Scores and Confidence Analysis . . . 273
7.3.3 Speed and Computational Efficiency . . . 275
7.4 Qualitative Results and Failure Mode Analysis . . . 277
7.4.1 Method 1: Image Stitching Failure Characteristics . . . 277
7.4.2 Method 2: Text Stitching Robustness and Localized Failures . . . 279
7.5 Comparative Analysis and Discussion . . . 280
7.5.1 Accuracy and Reliability . . . 281
7.5.2 Computational Efficiency and Deployment Scalability . . . 282
7.5.3 Trade-offs in Method Selection . . . 282
7.5.4 Limitations of the Current Evaluation . . . 282
7.6 Conclusions . . . 283

8 Prototype Development for Medication Identification System 286
8.1 Introduction . . . 287
8.2 Design Rationale: Approach Selection . . . 288
8.3 Technology Stack and Component Selection . . . 290
8.3.1 Optical Character Recognition: PaddleOCR . . . 290
8.3.2 Programming Language: Python . . . 290
8.3.3 User Interface: PyQt6 . . . 290
8.3.4 Image Processing: OpenCV . . . 291
8.3.5 Local Data Storage: SQLite . . . 291
8.3.6 Dual-Camera Hardware Setup . . . 292
8.3.7 Software Architecture and Processing Pipeline . . . 293
8.3.7.1 Class and Sequence Diagrams . . . 296
8.4 User Interface Design . . . 297
8.5 Key Functional Features and Implementation Details . . . 299
8.5.1 Dual-Angle Coverage and Text Merging . . . 299
8.5.2 Concurrent Processing via Multi-Threading . . . 300
8.5.3 Configurable Processing Rate Through Frame Skipping . . . 301
8.5.4 Medication Matching via String Similarity . . . 301
8.5.5 Model Flexibility . . . 302
8.6 Summary and Future Directions . . . 302

9 Conclusions and Future Work 304
9.1 Summary of Research Contributions . . . 305
9.1.1 Image-Based Label Reconstruction . . . 305
9.1.2 Text-Based Label Reconstruction . . . 306
9.1.3 Medication Database Construction Using AI-Based Processing 307

9.1.4 Experimental Evaluation Under Realistic Conditions 308
9.1.5 Prototype System Design and Implementation 308
9.2 Key Research Findings and Their Implications 309
9.3 Limitations 310
9.4 Future Research Directions 311
9.4.1 Expanding Dataset Diversity 312
9.4.2 Advanced OCR Enhancement and Domain Adaptation 312
9.4.3 Multi-Camera and Sensing Approaches 312
9.4.4 Clinical Validation and Real-World Deployment Studies . . . 313
9.4.5 Advanced Matching and Knowledge-Based Reasoning 313
9.5 Final Remarks 314

References **316**

Appendices **329**

List of Tables

1.1 Medication error reports from varied healthcare settings. 29

2.1 Combined Search String for Technology and Medication Safety in Medication Identification. 43
2.2 Combined Search String for AI-related Technology in Medication Identification. 44
2.3 Inclusion and Exclusion Criteria for Literature Review 44
2.4 Comparison of Studies on CV and DL-based Medication Identification 75
2.5 Comparison of Medication Identification Technologies: Cost and Scalability 80
2.6 Comparison of Medication Identification Technologies: Accuracy and Error Prevention. 83
2.7 Comparison of Medication Identification Technologies: Efficiency and Workflow Integration. 86

3.1 Comprehensive summary of datasets used in the study, from component training through end-to-end evaluation. 111
3.2 Dataset Overview 113

4.1 Summary of related works on cylindrical label reconstruction 128
4.2 Performance of the YOLOv8-s model on the label segmentation task. 163
4.3 Quantitative comparison of ellipse estimation methods. 165
4.4 Text extraction quantity gain per container (mean ± standard deviation) 175

5.1 Datasets used for OCR training and fine-tuning 197
5.2 Quantitative performance of the pretrained baseline versus fine-tuned PaddleOCR detection model on the 200-image medication label validation set. 198
5.3 Average performance metrics for the text stitching algorithm on the synthetic dataset. 205
5.4 Single-view OCR and proposed stitching method on 40 medication label image pairs. 210

6.1 Standardized schema fields for medication database records. 229
6.2 Field-level completion rates across 15 medication containers with summary statistics. 241

6.3 Critical field accuracy evaluated on 15 medication labels using manually annotated ground truth. 243

7.1 Top-1 and Top-5 accuracy comparison between Method 1 (Image Stitching) and Method 2 (Text Stitching) on 1,060 test samples. 271
7.2 Statistical summary of Top-1 similarity scores for both methods. Scores range from 0 to 1, with 1.0 indicating perfect match confidence. 273
7.3 Execution time breakdown for Method 1 and Method 2 on Machine 1 (high-performance system). Times are reported in milliseconds, averaged across all 1,060 test samples. The "(2 ×)" notation indicates processing of both camera views. 276
7.4 Execution time breakdown for Method 2 on Machine 2 (portable edge-device system). Times are reported in seconds, averaged across all 1,060 test samples. 277
7.5 Comparative failure mode characteristics of Method 1 and Method 2. Method 1 exhibits catastrophic failures (unable-to-predict), while Method 2 shows graceful degradation. 280
7.6 Comprehensive comparison of Method 1 and Method 2. Method 2 demonstrates clear advantages in accuracy, robustness, speed, and clinical deployment feasibility, making it the preferred approach for operating theatre environments. 281

8.1 System Component Roles 288

List of Figures

2.1 Article selection flowchart 45
2.2 Publication trends of medication identification techniques. 46
2.3 Example of a medication container with a (a) Barcode and (b) QR code (Biotium, 2023). 51
2.4 (a) Basic RFID system (TTelectronics, 2023) and (b) a medication container with RFID tag (Insights, 2023). 60

3.1 Examples of Cefazolin AFT and Bacteriostatic Water for Injections, showing that a single image does not capture the complete label information. 94
3.2 System architecture of the proposed dual-camera POC and its integration with label reconstruction, database enrichment, and evaluation modules. 95
3.3 Flow chart of the two main label reconstruction pipelines proposed in this study. 98
3.4 The process of enriching the medication database, from multiple medication images to a comprehensive, verified knowledge base. 100
3.5 Example images were taken to capture the full label of the vial *DBL Rocuronium Bromide Injection*. 101
3.6 The overall process of medication identification, from dual-image capture to final medication ID. 102
3.7 Comparison between single-camera and dual-camera setups. 103
3.8 Examples of single- and dual-camera captures: (a) Single Vecure Vial, (b) Dual Vecure Vials with overlapping views, (c) Single Metoprolol IV, and (d) Dual Metoprolol IV images with overlapping views. . . . 104
3.9 The camera mount allows the camera to move around and adjust its distance to the medication container. 105
3.10 Examples of medication containers used in this research. The dataset reflects real clinical variability in label appearance, size, curvature, and text density. 108
3.11 Illustration of the image capture and annotation workflow. 109

4.1 Flowchart of the proposed label reconstruction pipeline. 125
4.2 Illustration of the reconstruction pipeline for cylindrical medication labels. 134

4.3 The cylindrical container image acquisition workflow. 137
4.4 Examples from the multi-view medication containers dataset. 138
4.5 Examples of medication bottle with annotations. 138
4.6 Dual path ellipse detection for cylindrical label unwrapping. 140
4.7 The flow chart of the proposed Contour-Based Ellipse Fitting method 141
4.8 Examples are shown as follows: (a) the input drug images and (b) the corresponding binary masks generated by the YOLO segmentation model. (c) displays the approximated contour from the segmented mask with dimensions 148×2, while (d) shows the simplified contour, reduced to 23×2 using the Ramer-Douglas-Peucker algorithm. 142
4.9 Illustration of the sampling process: (a) keypoint localization, (b) horizontal point sampling, (c) vertical line intersection with the contour, and (d) extraction of ellipse points. 143
4.10 Example of ellipse fitting: (a) sampled points and fitted ellipses, and (b) aligned ellipse results on the drug image. 146
4.11 Illustration of an ellipse with defined parameters $(c_x, c_y, r_x, r_y, \theta)$ . . . 149
4.12 Some examples of the generated dataset used in this research. Left: binary cylinder mask; Right: cylinder mask overlaid with the ground truth ellipse parameters used for geometric validation. 151
4.13 The FCOS network architecture. C3, C4, and C5 represent the feature maps from the backbone network, while P3 through P7 denote the feature levels in the FPN used for final predictions. 153
4.14 The mesh generation process: (a) Input image with segmented mask generated by the YOLO instance segmentation model (b) Top and bottom ellipses of the label detected using the FCOS ellipse detection model (c) Points sampled on top and bottom half ellipses of the bottle label and (d) The final mesh constructed by connecting sampled points. 156
4.15 Illustration of the cylindrical unwrapping process, while every quadrangle are being applied a perspective transformation to convert the region to a plat plane. 157
4.16 Illustration of the keypoint extraction process from unwrapped labels. 159
4.17 Demonstration of feature correspondence and geometric alignment between unwrapped labels during the registration and stitching process. 160
4.18 Qualitative results of YOLOv8-s label segmentation under varying conditions. Predicted masks are shown as overlays. 164
4.19 From left to right, segmentation mask overlay, traditional contour-based fitting, DL-based regression, and the reconstructed cylindrical representation from the DL ellipse result 166
4.19 From left to right, segmentation mask overlay, traditional contour-based fitting, DL-based regression, and the reconstructed cylindrical representation from the DL ellipse result. 167
4.20 Examples of the cylindrical label unwrapping process applied to various medication vials. The results show accurate distortion correction and clear text preservation in central regions. 169

4.21 Results of stitching two unwrapped images: (a) left image, (b) right image, and (c) the stitched result. 171
4.22 Examples of complete stitched labels containing comprehensive drug information (Part 1). 172
4.22 Examples of complete stitched labels containing comprehensive drug information (Part 2). 173
4.23 The label gap issue occurs when the camera faces the transparent back of the vial, leading to fragmented segmentation and unstable cylindrical reconstruction. 176

5.1 Overview of the proposed text-stitching pipeline (Approach 2) for medication label reconstruction from dual-camera images. 184
5.2 Examples of captured drug images. Multiple perspectives are recorded for each container to accommodate text curvature and label distortion. 188
5.3 Example of text-region annotation using LabelMe. 189
5.4 Workflow of the PaddleOCR text detection and recognition pipeline. . 190
5.5 Example of LCS-based overlap detection between OCR outputs s_1 = `um Chlo` and s_2 = `Calcium Ch`. The LCS identifies the overlapping portion `um Ch`. 196
5.6 Some OCR results from the validation dataset. 200
5.7 Example images from the synthetic dataset: (a) full text image, (b) left half, and (c) right half. 202
5.8 Images and corresponding OCR results in the synthetic text dataset. . 203
5.9 Examples of text stitching from fragmented OCR outputs. The fragments (in blue) are combined to produce complete sentences (in bold). 206
5.10 Comparison between processed results (left) and ground truth (right) . 207
5.11 Illustration of OCR processing and text merging: (a, b) left and right images, (c, d) corresponding OCR results, and (e) stitched text result. . 210
5.12 Illustration of OCR processing and text merging: (a, b) left and right images, (c, d) corresponding OCR results, and (e) stitched text result. . 211
5.13 A visual representation of medication label stitching results. (a-f) Images of the drug container captured from different angles, and (g) the corresponding textual information with visual boundaries for clarity. . 213
5.14 A visual representation of medication label stitching results. (a-f) Images of the medication container captured from different angles, and (g) the corresponding textual information with visual boundaries for clarity. 214

6.1 Overview of the query-based medication identification pipeline. Text extracted from dual-view images is used solely as a query to retrieve the correct drug identity from a medication database. 221
6.2 Automated Medication Data Extraction and Refinement/Enrichment Architecture. 226

6.3 LLM agent architecture implemented using LangGraph: (1) Text Consolidation node merges OCR fragments and corrects systematic errors, (2) Schema Extraction node generates structured JSON with 12 medication fields, and (3) RAG Refinement node validates and enriches fields against authoritative sources. 227
6.4 Prompt used for medication text merging. 228
6.5 Prompt used for structured medication information extraction. 230
6.6 Examples of Medsafe medication datasheets used as external reference sources in this study. 232
6.7 Two-stage retrieval process consisting of global search followed by targeted refinement. The framework emphasises preserving label-derived information, limiting corrections to cases supported by consistent reference evidence, and using retrieved content as supplementary context for incomplete fields. 233
6.8 Prompt for medication information refinement with RAG. 235
6.9 Workflow graph of the LLM agent implemented using LangChain and LangGraph. 237
6.10 Example of multi-view OCR aggregation and structured extraction under noisy and fragmented input conditions. 244
6.11 Example of retrieval-augmented enrichment for Solu-Cortef. The initial LLM extraction contains limited clinical context, while the RAG-enhanced output incorporates additional prescribing information such as indications, precautions, and administration details. 246
6.12 Example of incorrect entity reconstruction under ambiguous OCR input, where fragmented text leads to partial or inaccurate drug name normalization. 247
6.13 Example of retrieval-augmented enrichment where additional fields are introduced based on external context. While core attributes remain unchanged, new fields such as diluent and storage instructions are added without explicit support from the original OCR-derived content. 249
6.14 Schematic overview of the medication database pipeline, flow from multi-angle image acquisition, through automated processing utilizing OCR, LLM, and RAG refinement, to human expert validation and final storage in MongoDB for clinical retrieval. 251
6.15 Users can upload multiple images of drug containers for processing. The backend pipeline processes these images using OCR, LLM, and RAG to generate structured drug information. 252
6.16 Users can manually verify and edit extracted drug data before storing it in the database. This step ensures data quality and accuracy. 253
6.17 The database view displays all stored drugs, supporting quick search, query, and retrieval of structured drug information. 254

7.1 Sample images from the experimental dataset. Each medication ID is represented by two images captured from different angles (left and right views) using a dual-camera setup. 264
7.2 Overview of Method 1: Image stitching followed by OCR for drug label identification. The pipeline emphasizes geometric reconstruction prior to text extraction, enabling a consolidated visual representation of the medication label. 268
7.3 Overview of Method 2: OCR followed by text stitching. By operating in the text domain, this pipeline avoids geometric reconstruction, reducing computational cost and sensitivity to misalignment. 270
7.4 Distribution of prediction outcomes for Top-1 and Top-5 evaluation. Categories include correct predictions, incorrect predictions due to database mismatch, and unable-to-predict cases caused by upstream pipeline failures (e.g., unsuccessful image stitching). 272
7.5 Distribution of Top-1 similarity scores for Method 1 (Image Stitching) and Method 2 (Text Stitching). Method 2 produces substantially higher scores for correct matches and exhibits tighter clustering. 274
7.6 Comparison of Top-1 similarity score distributions. Method 2 exhibits higher mean, tighter interquartile range, and greater separation between correct and incorrect predictions. 275
7.7 Representative failure cases of Method 1, illustrating artifacts caused by image stitching and geometric unwrapping errors.. 278
7.8 Representative failure cases of Method 2, illustrating performance degradation under poor label quality and overlapping text. Left: Metaraminol ampoule with faded, low-contrast printing; Right: Clonidine ampoule with densely packed and overlapping text on a curved surface, leading to ambiguous OCR outputs. 279

8.1 Creality Nebula Camera selected for the dual-camera setup. The camera features a short focal length enabling close-up capture of medication labels and manual focus adjustment for precise label imaging. 292
8.2 3D design of the prototype hardware, created in Tinkercad, showing the arrangement of the bottle holder, dual adjustable camera stands, and camera mounts for multi-angle label capture. 293
8.3 Physical prototype constructed using a laser cutter. The hardware provides stable positioning for medication containers and flexible camera positioning to capture labels from complementary viewing angles. . . 294
8.4 Class diagram showing the modular architecture of the prototype. Arrows indicate dependencies between classes. 296
8.5 Sequence diagram illustrating the interaction between system components during a typical medication identification workflow. The diagram shows the progression from user interaction through image capture, OCR processing, database matching, and result presentation. 298

8.6 Main application window showing dual live camera feeds (left and right), the top-ranked medication identification result with confidence score (center panel), and merged label text (bottom panel). The interface is designed for rapid comprehension in high-pressure clinical environments. 299
8.7 Configuration interface enabling users to adjust system parameters including OCR model selection, confidence threshold for automatic acceptance of results, and frame-skipping rate controlling the frequency of identification attempts. 300
8.8 Comparison of single-threaded and multi-threaded processing architectures. In single-threaded execution (left), image acquisition is blocked during OCR, causing frame drops and lag. In multi-threaded execution (right), acquisition and processing run concurrently, improving responsiveness and stability. 301

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

30 April 2026

Artificial Intelligence Declaration

Artificial Intelligence Guidelines:

Artificial Intelligence (AI) tools, including generative AI tools approved for use by AUT (AI Tool Register), can be used as learning tools when preparing to write your research proposal or research component. AI tools also have capability to assist with data modelling, data analysis and data visualisation. However, the research component you submit for examination must be substantively your own work. AI tools cannot solely be used to generate content when writing or creating an artwork/artefact, as this constitutes plagiarism. The formal guidelines for appropriate use of AI on the AI Hub must be followed. Also refer to the Using Artificial Intelligence in your Research: Self-Assessments, Checklists, and Sample Text, for examples and sample text.

If you do plan to use AI tools, this declaration should be submitted with the initial research proposal, Confirmation of Candidature, and when you submit your research component for examination. You must complete the template below for each AI output used detailing the following:

- **Chapters where AI was used in any way**
- **Purpose of AI Use**
- **AI Tool(s) Used**
- **Prompts Used**
- **Post-AI Processing Methods**

Specify the chapter number(s) in your research proposal/research component where AI was used.	
Chapter number(s):	All chapters
Briefly explain why the AI tool was used (e.g., idea generation, summarization, data analysis, image creation, etc.).	
Purpose of AI Use:	Assist with language refinement, clarity improvement, grammar correction, and restructuring of sentences. AI was not used to generate original research ideas, experimental design, or results.
Clearly state the AI tools used.	
AI Tool(s) Used:	ChatGPT
Prompts or queries entered into the AI tool (only the instructive portion of the prompt is required). Include all relevant variations.	
Prompts Used:	Examples of prompts include: - "Revise this paragraph for clarity and academic tone." - "Reduce repetition while preserving meaning." - "Improve grammar and readability of this section." - "Rewrite this sentence to be more concise."
A summary or description of the AI-generated output. (Optional: attach output in an appendix.)	
Output Received:	Revised text suggestions, improved grammar, and alternative phrasing for clarity and conciseness.
Describe how the AI output was used, edited, or transformed.	
Post-AI Processing Methods:	All AI-generated outputs were reviewed, edited, and validated by the author. Technical content, methodology, and experimental results were independently developed and verified.

You may use this free text box to explain the intent for the use of the AI generated content. *For example, rather than stating what prompt(s) were used, you may articulate the purpose for which the prompt was used (supported by examples) here:*
AI tools were used strictly as assistive writing aids and not as a source of original research content. All key contributions, including problem formulation, methodology design, experimentation, and analysis, were conducted independently by the author.

STUDENT DECLARATION					
By signing you are confirming that the AI use stated in the table(s) above are accurate and follow AUT's recommended guidelines detailed in the AI Hub and Postgraduate Handbook.					
Student Name:	Hoang Nam Nguyen	**Signature:**		**Date:**	30/04/2026

Publications

Nguyen, H. N., Madanian, S., Nguyen, M., Merry, A., & Parry, D. (2025). Digital Transformation of Medication Identification: Technological Evolution. Studies in health technology and informatics, 329, 1684-1685.

Acknowledgements

Completing this PhD journey marks an incredible milestone for our family in New Zealand. This path has been filled with moments of challenge, stress, and joy. Beyond my academic research, I was fortunate to collaborate with several organization and institutions, including SaferSLEEP, Parking Premium, N3T, the University of Auckland, and Auckland Transport. The wonderful people I worked alongside across these projects provided me with invaluable professional and personal experiences in New Zealand. Above all, this journey allowed me to watch my young son grow every day into a happy, strong, and independent individual.

I would like to express my sincere gratitude to my advisors, Dr. Samaneh Madanian and Dr. Minh Nguyen, for their invaluable guidance, patience, and continuous support throughout this endeavor.

Special thanks go to the Department of Computer and Information Science at the Auckland University of Technology for providing the computational resources, hardware infrastructure, and collaborative environment that made this work possible.

Finally, I extend my deepest gratitude to my family and friends for their enduring encouragement, love, and unwavering support throughout this journey.

Chapter 1

Introduction

This chapter establishes the research context, objectives, and significance of this study. It highlights the challenges of accurate and efficient medication identification in healthcare environments, particularly in high-risk settings like anesthesia. The chapter also presents the need for a robust, Computer Vision (CV) -based solution, outlines the research aims and key contributions, and provides an overview of the thesis structure.

Medication errors remain a persistent threat to patient safety, with particularly severe consequences in anesthesia due to the potency and rapid action of medications used in operating theatres (Merry & Wahr, 2021). Errors in prescribing, dispensing, or administration can lead to immediate and serious harm, often stemming from human and operational challenges (Litman, 2018). Addressing these vulnerabilities requires innovative solutions that reduce reliance on manual medication verification in time-sensitive environments.

Existing tools such as barcode scanning improve safety in controlled settings but struggle to perform in the dynamic, high-pressure conditions of anesthesia (Jelacic et al., 2015; Dunn & Anderson, 2019). This limitation underscores the need for a more adaptive and automated approach. To address this gap, the research focuses on leveraging CV to enhance real-time medication identification, reducing errors and

improving patient safety.

The primary objective of this study is to develop a prototype of CV-based tool to mitigate medication errors by providing real-time, automated verification, ensuring accurate medication identification in healthcare settings. The introduction also provides an overview of medication error patterns, existing technological solutions, and the proposed system's design and impact.

1.1 Background

According to the National Library of Medicine Medical Subject Heading, a medication error is an "error in the prescription, dispensing, or administration of a medication with the result that the patient fails to receive the correct medication or the indicated proper dosage" (Wheeler & Wheeler, 2005). For example, prescribing errors occur when a healthcare provider makes a mistake while writing a prescription, including choosing the wrong medication, or prescribing an incorrect dosage. Dispensing errors happen at the pharmacy level when a pharmacist or technician incorrectly fills a prescription. This might involve giving a patient the wrong medication due to similar medication names, or dispensing the wrong dosage. In addition, administration errors occur when a medication is given incorrectly to a patient. This can happen in hospitals, nursing homes, or even at home when a patient misuses the medication.

Medication errors occur across all areas of healthcare, including hospitals, outpatient clinics, nursing homes, and pharmacies. Their persistence reflects a complex interplay of human factors, such as fatigue or oversight, and systemic issues, including inadequate training and resource constraints (Kelly, Harrington, Matos, Turner & Johnson, 2016; Warnakulasuriya et al., 2022). The consequences of medication errors can range from minor inconveniences to severe patient harm (Jelacic et al., 2015), placing a dual burden on healthcare systems through compromised safety and increased costs (Maydana et al., 2017).

This research focuses specifically on medication errors in operating theatre, a field where the risks are uniquely amplified. Anesthetic medications, such as those used in the operating room are potent and fast-acting, demanding precise dosing and timing. In these high-stakes environments, even small mistakes - whether in calculating a dose or selecting a medication can lead to severe outcomes (T. A. Bowdle et al., 2019). The administration of anesthesia, often performed under time constraints and while

managing multiple tasks, increases the risk of errors, making it a crucial focus for safety enhancements (Jelacic et al., 2015).

Efforts to reduce medication errors have involved technologies such as barcode and Radio Frequency Identification (RFID) scanning, which help improve medication verification. While these methods have shown promise, they have limitations, such as being vulnerable to damage (Koppel, Wetterneck, Telles & Karsh, 2008) or electronic interference (Fram, Rivlin & Beredjiklian, 2020). These shortcomings highlight the need for innovative solutions to enhance medication safety, particularly in anesthesia. This thesis addresses this need by exploring the potential of CV to improve accuracy and reliability in high-risk scenarios.

1.2 Motivation

Medication errors pose a significant risk to patient safety, particularly in fast-pace environment such as the operating room. Reducing such errors is essential to improving patient outcomes and ensuring safer surgical procedures.

Current common approaches to mitigating medication errors including barcode code scanning and RFID systems, which help verify medications before administration. While these methods have improved accuracy in medication dispensing and administration (Barakat & Franklin, 2020; Dunn & Anderson, 2019; Decia et al., 2017; Martinez Perez, Dafonte & Gómez, 2018), they come with limitation. For instance, barcodes rely on manual scanning, and not all medication containers have standardized barcode labels (Xie, Kalia, Strudwick & Lau, 2019). RFID systems require specialized infrastructure, making widespread adoption challenging in all healthcare settings (Martinez Perez et al., 2018).

To overcome these limitations, CV-based medication identification has emerged as a promising alternative. Different solutions have been proposed including classification,

object detection or optical characteristic recognition (Aghavali et al., 2024; Jin et al., 2022; Gromova & Elangovan, 2022). These methods provide a contactless and efficient solution, reducing reliance on manual scanning and minimizing the risk of human error.

This research introduces a CV-based medication identification system designed to significantly improve medication safety in the demanding operating room environment. The proposed system leverages dual-camera image capture, Optical Character Recognition (OCR) for text recognition, and advanced text-matching to accurately identify medications before administration. By integrating this real-time, automated analysis into anesthesia workflows and targeting critical stages of medication administration, the solution aims to mitigate human oversight errors common under high-pressure conditions (Rodriguez-Gonzalez et al., 2015), thereby reducing medication errors, enhancing efficiency, and improving overall patient safety in surgical settings.

1.2.1 Medication Error: Risks and Realities

Medication errors can involve various factors, such as Adverse Drug Events (ADEs), overdose, underdose, incorrect medication selection, and administration errors. For instance, ADEs refer to harmful or unintended reactions caused by medication use, including allergic reactions, side effects, or medication interactions (Bates et al., 1995). Additionally, overdose occurs when a higher-than-required dose is administered, potentially leading to toxic effects or organ damage, while underdose results from a lower-than-required dose, which may reduce the medication's effectiveness and lead to treatment failure. In anesthesia, medication errors are not uncommon which compromised patient health and safety. For instance, in a study conducted at a 1,046-bed tertiary care academic medical center in Boston, MA, data were collected from November 2013 to June 2014. Among 277 observed operations and 3,671 medication administrations,

5.3% involved a medication error or ADE, with the vast majority (79.3%) being preventable, and a significant portion (79.3% of errors and 47.2% of ADEs) having the potential for or resulting in patient harm, often of serious or life-threatening severity (Nanji, Patel, Shaikh, Seger & Bates, 2016).

Medication errors pose a significant threat to patient safety, occurring at various stages from prescribing to administration. The risks are particularly high in anesthesia due to the potency of medications such as fentanyl and propofol, where even small miscalculations can lead to catastrophic consequences. Studies reveal that factors like human fatigue, complex workflows, and time pressure further increase the likelihood of errors (Wheeler & Wheeler, 2005; Al-Ahmadi, Al-Juffali, Al-Shanawani & Ali, 2020). In addition to immediate patient harm, these errors erode trust in healthcare systems and contribute to rising medical costs, with estimated annual costs ranging from $20 billion for all adverse events to an additional $35.7 to $45 billion specifically for hospital-acquired infections (Rodziewicz, Houseman, Vaqar & Hipskind, 2024). Given the limitations of current safeguards, this research is motivated by the urgent need for more effective solutions to reduce these risks.

Table 1.1: Medication error reports from varied healthcare settings.

Source of Information	**Main Observations**	**Consequences**
Prospective hospital study (Gariel et al., 2018).	Medication errors (2.6%) in paediatric anaesthesia are linked to opioids, antibiotics, incorrect dosing (67.5%), dilution mistakes (26%), and longer procedures (>120 min, AOR: 4).	Medication errors can lead to adverse events, especially in younger patients.
Japanese error database (Suzuki et al., 2021)	Medication errors in the operating room frequently involve incorrect dose (170 cases) and incorrect medication (152 cases), with key factors including pressor medications, anesthesia-inducing medications, lack of knowledge, and preparation errors.	Patients experienced adverse effects, including potential life-threatening reactions, delayed treatment, increased complications, and prolonged recovery. Errors also led to increased stress and liability for medical staff.
National review (Tariq RA, 2022)	6.5 errors per 100 admissions; 1.5 million preventable ADEs annually in the U.S.; dispensing errors often due to look-alike medications.	1 in 854 inpatient deaths linked to errors; $3.5 billion/year in extra hospital costs for medication-related injuries.
National economic study (Elliott et al., 2021)	237 million errors annually; 34% of significant errors in primary care prescribing; 26% of errors potentially moderate harm.	£98.5 million and 1,708 lives lost yearly (avoidable); worst-case scenario: £1.6 billion and 22,303 lives.
Retrospective GP records (Leitch et al., 2021)	Medication-related harm is common in general practice, affecting 10.8% of patients, with 21.2% of cases being moderate or severe..	Higher age, more consultations, and multiple medications increase risk, with certain medication types causing the most harm.

To better understand the scope of the issue, Table 1.1 highlights some key findings on medication errors across various healthcare settings. Gariel *et al.* (Gariel et al., 2018) reported a 2.6% error rate in pediatric anesthesia, mainly due to incorrect dosing (67.5%) and dilution mistakes (26%), with longer procedures increasing risk. Suzuki *et al.* (Suzuki et al., 2021) found that operating room errors frequently involved incorrect dosing (170 cases) and medication selection (152 cases), often linked to pressor and anesthesia medications. Tariq *et al.* (Tariq RA, 2022) estimated 6.5 errors per 100 U.S. hospital admissions, leading to 1.5 million preventable ADEs and $3.5 billion in extra costs annually. Elliott *et al.* (Elliott et al., 2021) identified 237 million errors per year in the U.K., with 26% posing moderate harm, costing £98.5 million and causing 1,708 avoidable deaths. A recent study in New Zealand (Leitch et al., 2021) found medication-related harm in 10.8% of general practice patients, with risks increasing with age, multiple consultations, and polypharmacy.

The severe implications of medication errors, particularly within the environment of operating theatres, highlight a critical patient safety challenge. As these errors persist despite current safeguards, there is an urgent need for robust, technologically driven solutions that can enhance accuracy, reduce human error, and ensure timely identification and administration of medications in healthcare settings.

1.2.2 Factors Contributing to Medication Errors

In the operating theatre, healthcare professionals may experience emotional stress and fatigue due to the high-stakes nature of surgical procedures, the need for constant precision, time pressure, and the responsibility of ensuring patient safety (Rice, Glass, Ogle & Parsian, 2014). Long working hours, complex decision-making, and the potential for life-threatening consequences further contribute to their mental and physical strain

(Wheeler & Wheeler, 2005). This fast-paced environment may make surgeons and anesthesiologists overwhelmed with heavy workload. Stress can affect the concentration and performance, resulting in an increased risk of errors (Wheeler & Wheeler, 2005). For example, they may occasionally select the incorrect ampule/syringe or incorrect dosage during the medication process (Tariq RA, 2022). This type of incident can occur more frequently when medication containers are placed close to each other or have similar appearances (Bryan, Aronson, Williams & Jordan, 2021). Moreover, poor communication, such as such as unclear instructions or incomplete transfer of information can lead to error in the prescribing, dispensing or administering of medications (Al-Ahmadi et al., 2020). These communication breakdowns can happen between healthcare professionals or healthcare professionals and patients.

Medication errors can also result from system-related factors at various stages, including unclear prescribing protocols, inadequate medication storage practices, and insufficient safeguards in medication management systems (Litman, 2018). For instance, healthcare facilities without modern electronic prescribing system may rely on handwritten prescriptions, which can be prone to misinterpretation (Albarrak, Al Rashidi, Fatani, Al Ageel & Mohammed, 2014). Lack of standardization can also lead to confusion and potential mistakes, thereby increasing the risk of medication errors. Moreover, complex medication regiments involving multiple medications, specific timing and various doses increase the changes of medication errors, as health care provides and patient may find difficulty to follow of the medications (Tariq RA, 2022).

1.2.3 Role of Technology

Technology has significantly contributed to reducing medication errors, particularly in medication identification, through solutions such as barcode scanning, RFID, and CV. Barcode and RFID systems enhance medication tracking by verifying medication

identity and dosage, reducing errors in the dispensing and administration process (Decia et al., 2017; Yoo, Hwang & Jheon, 2016; Uy, Kury & Fontelo, 2015). CV also enables the automated recognition of medication labels and packaging, helping to minimize human errors in fast-paced, high-pressure healthcare environments (Ou, Tsai, Zhou & Wang, 2020).

However, these technologies have their limitations. While electronic prescribing has addressed handwriting issues, and barcode scanning has improved accuracy in general healthcare settings, the fast-paced and complex workflow of the operating room may limit the effectiveness of these solutions. For example, barcode scanning systems require a clear line-of-sight to the barcode, which slows down the process and makes it inefficient in the fast-paced environment of the operating room (Poon et al., 2010). Some medications do not have barcodes, and varying barcode standards across different medication manufacturers can cause compatibility issues, further complicating the identification of medications under time pressure (Hachesu, Zyaei & Hassankhani, 2016; Strykowski, Hadsall, Sawchyn, VanSickle & Niznick, 2013). These limitations make existing technologies less effective in anesthesia care, emphasizing the need for more efficient and adaptable solutions.

While current technologies help mitigate errors, they have limitations in addressing deeper systemic issues such as workflow strain and the cognitive load on healthcare professionals. The ongoing occurrence of medication identification errors, despite the widespread adoption of these technologies, highlights the need for a more effective solution. There is a clear need for new tools that can assist healthcare professionals in reducing errors, particularly in high-pressure environments like anesthesia care. This research aims to explore advanced technologies that provide real-time, adaptive support for medication identification, addressing the unique challenges faced by healthcare professionals.

1.2.4 CV in Medication Identification

Among emerging technologies, CV stands out as a powerful tool for improving medication safety, particularly in anesthesia/operating theatre. By leveraging techniques such as image recognition and object detection, CV can identify medications, verify dosages, and detect discrepancies in real time, capabilities essential for high-pressure environments like the operating theatres (Jin et al., 2022; Tran, Richardson, Chen & Lin, 2022). It can address common errors such as misidentifying look-alike vials (Kittisorayut, Nonsiri, Kamin & Makdee, 2024) and verifying doses for patients (Palenychka, Lakhssassi & Palenychka, 2018), where precision is critical. The increasing attention on CV in medication identification highlights their potential to streamline medication verification.

However, implementing a robust CV-based medication identification system presents unique challenges, particularly with cylindrical medication containers. The curved shape of these containers obstructs a full view of the label, making it difficult to extract key information such as medication names, expiry dates, and dosage details. A single image often fails to capture the entire label, necessitating a multi-angle imaging approach. While this enhances scanning accuracy, it also introduces additional complexities:

1. Managing variations in scale, aspect ratio, and orientation of cylindrical containers.

2. Correcting distortions and misaligned text caused by curved surfaces.

3. Ensuring consistent alignment of label regions across multiple images.

4. Developing algorithms that integrate multi-image data into a structured and accurate format.

To address these challenges, this research explores two CV-based approaches for medication identification: **Image Stitching followed by OCR** and **OCR followed**

by Text Stitching. The following chapters provide an in-depth examination of both methods, evaluating their effectiveness in overcoming the limitations of cylindrical medication packaging and identifying the most robust solution for medication identification.

1.3 Research Objective

This research aims to develop and evaluate a CV-based system for extracting information from cylindrical medication containers. The focus is on addressing challenges caused by curved surface distortion, partial visibility, and OCR noise, with the goal of improving the reliability of medication identification in high-risk clinical environments such as operating theatres. By enhancing the robustness of visual information extraction of drug images, the system seeks to support safer and more accurate medication recognition and reduce the risk of medication errors.

To achieve this objective, the following research questions are explored:

1. What are the key challenges in automated medication identification in clinical environments such as operating theatres, and how can CV-based systems address these challenges?

2. How can multi-view imaging improve the reliability of medication label information extraction from cylindrical containers?

3. How can text-based reconstruction from OCR outputs improve robustness and reduce geometric reconstruction errors in medication label understanding?

4. How can language models combined with retrieval-augmented enrichment transform noisy OCR text into structured medication records, and what are the limitations in terms of completeness and external dependency?

These research questions guide a systematic investigation of medication information

extraction in this study, focusing on robustness to cylindrical surface distortion, reduction of text extraction errors, and improvement of medication identification accuracy.

1.4 Contribution of This Thesis

This research contributes to the field of medication identification through the development of a robust, real-time system using CV techniques. The key contributions are as follows:

- Development of a dual-camera acquisition system supporting multi-view capture of cylindrical medication labels, enabling the reconstruction approaches investigated in this thesis.
- Development and empirical comparison of two candidate label-reconstruction approaches: image stitching followed by OCR, and OCR followed by text stitching, with the latter selected for the final prototype based on experimental evidence (Chapter 7):
  - ***Image Stitching*** *Followed by OCR*: This approach first stitches images from the two cameras into a single, comprehensive view of the medication bottle, followed by OCR processing for text extraction.
  - *OCR Followed by* ***Text Stitching***: This method independently processes each camera's image using OCR, then stitches the extracted text to form a more complete label description.
- **Text Matching and Database Retrieval**: Text similarity measures, such as Levenshtein distance and cosine similarity, were implemented to match extracted text against the medication database, ensuring accurate medication identification.

- **Offline and fast medication identification**: The system is designed to operate fully offline, leveraging a lightweight on-device OCR model for fast and accurate text recognition. This design avoids reliance on third-party cloud-based services (e.g., Google OCR), making the system suitable for deployment in operating theatres where internet connectivity, data privacy, and real-time responsiveness are critical.

- **Prototype Development**: A working prototype was developed, integrating multi-camera input handling, OCR processing, and database retrieval. The system provides visual feedback, displaying top candidates with accuracy scores to indicate successful medication identification.

These contributions advance the practical application of CV and OCR technologies in healthcare, providing a more reliable and efficient solution for medication identification, with significant potential to improve patient safety and reduce medication errors in real-world medical settings.

1.5 Structure of the Thesis

This thesis is structured around two distinct phases of the proposed medication identification system. The first phase, covered in Chapters 4 and 5, concerns real-time label reconstruction and identification: image acquisition, cylindrical label reconstruction, and database matching. These chapters present two alternative approaches to label reconstruction: image stitching followed by OCR, and OCR followed by text-domain stitching, developed independently and compared empirically in Chapter 7, with the superior approach selected for the final prototype (Chapter 8). The second phase, covered in Chapter6, concerns offline, construction-time preparation of the medication reference database using LLM- and RAG-assisted processing. This database-construction phase

is independent of, and does not execute during, real-time identification.

This thesis is organized into several chapters, each focusing on different aspects of the research. Apart from **Chapter 1**, which introduces the background and motivation, the subsequent chapters are structured as follows:

- **Chapter 2**: Reviews state-of-the-art medication identification methods, including barcode/QR code systems, RFID/NFC technologies, and CV–based approaches, to establish the research baseline and identify gaps in existing solutions.
- **Chapter 3**: Outlines the research methodology, detailing the design, implementation, and evaluation of the multi-camera acquisition system, as well as the OCR and database retrieval pipelines used for medication identification.
- **Chapter 4**: Introduces and evaluates the first proposed approach, *Image Stitching Followed by OCR*, which reconstructs a unified label image from multiple camera views prior to text extraction.
- **Chapter 5**: Presents and analyzes the second proposed approach, *OCR Followed by Text Stitching*, in which OCR is applied independently to each camera image and the extracted text is subsequently fused for medication identification.
- **Chapter 6**: Describes the construction of a verified medication database by extracting and structuring information from multi-view medication images, leveraging Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG).
- **Chapter 7**: Details the experimental evaluation of both Image Stitching and Text Stitching methods, including quantitative metrics such as accuracy and processing speed, qualitative performance analysis, and a comprehensive comparative discussion.

- **Chapter 8**: Presents the design and implementation of a prototype medication identification system, covering system architecture, hardware configuration, software components, and performance optimization strategies. mn
- **Chapter 9**: Summarizes the key contributions of the thesis toward developing a dual-camera medication identification system aimed at enhancing patient safety, and discusses the study's limitations and directions for future research.

Throughout this thesis, the following terms are used consistently: "approach" refers to one of the two candidate label-reconstruction methods developed in Chapters 4 and 5; "pipeline" refers to the sequence of processing steps within a given approach or within the overall system; "phase" is reserved specifically for the distinction between offline, construction-time processing (database construction, Chapter 6) and real-time identification processing (Chapters 4, 5, and 8).

Chapter 2

Literature Review

This chapter provides a comprehensive review of existing medication identification methods, including barcode/Quick Response (QR) code, Radio Frequency Identification (RFID)/ Near-field Communication (NFC), and Computer Vision (CV) technologies. It explores the principles, benefits, limitations, and real-world applications of each method, followed by a comparative analysis to highlight the gaps in current research. Finally, it identifies challenges in the integration and scalability of these technologies, offering insights into areas for further development in medication identification systems.

2.1 Introduction

The integration of advanced technologies into healthcare has revolutionized medication administration and identification, reducing medication errors and enhancing patient safety. This chapter presents a narrative review of current medication identification technologies, offering a comprehensive overview of traditional and modern methods, such as barcode and RFID systems, alongside emerging approaches in CV. By analyzing these technologies, their strengths, weaknesses, and real-world applicability are examined to identify gaps and opportunities for future research.

Common key technologies in medication identification have been reviewed in this chapter, each contributing to improved accuracy, efficiency, and safety in healthcare settings. Barcode and QR systems enable rapid and accurate identification of medications by scanning unique codes on packaging, reducing the risk of errors during dispensing and administration (Jelacic et al., 2015; Ti, Chen & Wu, 2020). RFID and NFC technologies provide contactless tracking and verification, allowing for real-time monitoring of medication inventory and ensuring the right medication reaches the right patient (Martinez Perez et al., 2018; Uy et al., 2015). Meanwhile, CV, as part of broader artificial intelligence (AI) approaches such as deep learning (DL) and machine learning (ML), enables advanced capabilities like automated recognition of medication labels without relying on barcodes or RFID tags.

We employed a structure search strategy to select relevant articles, ensuring a thorough examination of existing literature. The distribution of publications in the medication identification domain was assessed, highlighting trends and focus areas within this research field. By analyzing the evolution of these technologies, the review identifies promising advancements and technology adoption across healthcare applications.

Additionally, this review prioritizes applications in operating theatres due to their

unique safety demands, characterized by a time-sensitive and high-consequence environment, particularly during anaesthesia administration.

Operating theatres are high-stakes environments where minor medication errors can have severe consequences. The fast-paced nature of surgical procedures underscores the importance of robust and reliable medication identification systems. However, research on theatre-specific applications of these technologies is very limited, necessitating a reliance on general healthcare studies to infer potential applications. This review assesses the relevance of barcode, QR code, RFID, NFC, and CV technologies, evaluating their benefits and limitations in medication identification, since these represent the primary categories of automated identification and verification systems currently implemented in healthcare to address medication errors. Specifically:

- Barcode and QR code systems serve as the foundational and most widespread existing technologies, offering basic, cost-effective digital medication identification.
- RFID and NFC technologies represent the next evolutionary step in automated identification, specifically addressing some of the key drawbacks of barcode systems, including the need for line-of-sight and their relatively low robustness.
- CV techniques represent the cutting-edge and most promising frontier in this domain. They move beyond simple data encoding by enabling complex visual recognition, offering the potential to overcome the limitations of both barcode and RFID systems.

Recently, the rapid evolution of CV techniques has presented unprecedented opportunities to overcome the limitations of traditional barcode and RFID methods. However, their application to the specific visual challenges of medication identification remains an emerging field. Therefore, a narrative review is essential to critically assess the actual capabilities of technology-assisted medication identification in healthcare. By doing so,

we can better understand the complexities that even modern CV approaches may not yet fully address, such as the need for robust multi-view solutions for medication identification, which this work aims to explore. Bridging these gaps will provide valuable insights into how such technologies can enhance medication administration practices in operating theatres and broader healthcare environments, ultimately improving patient safety and surgical outcomes.

2.2 Review Methodology

2.2.1 Search Strategy

To identify relevant studies for this review, a systematic search strategy was used as shown in Table 2.1, the search was conducted in scientific databases including PubMed, IEEE Xplore, and Scopus, to cover the intersection of technology applications and healthcare.

The search keywords were carefully chosen to address medication identification methodologies, including "barcode," "QR code," "RFID," "NFC," "computer vision," "deep learning," "machine learning." Additionally, the specific term "convolutional neural network" was included to target a highly relevant subset of studies, as this model is central to many modern image-based medication identification systems. These technology keywords were then combined with domain-specific terms such as "drug administration," "medication administration," "drug identification," "medication error," and "patient safety." Boolean operators were applied to connect these keywords, ensuring that the retrieved studies addressed both the technology and its clinical application in medication management.

To align with the focus of this research on operating theatres, additional keywords such as "operating theatre," "surgery," "surgical," "perioperative," and "anaesthesia"

Table 2.1: Combined Search String for Technology and Medication Safety in Medication Identification.

Search String
("Barcode*" OR "Bar-code*" OR "QR Code*" OR "RFID" OR "NFC" OR "Computer Vision" OR "Deep Learning" OR "Machine Learning" OR "Convolutional Neural Network") AND ("drug administration" OR "drug identification" OR "medication administration" OR "medication identification" OR "drug safety") AND ("operating theatre" OR "operating room" OR "surgery" OR "surgical" OR "perioperative" OR "anesthesia" OR "anaesthesia" OR "hospital*")

were incorporated to identify studies directly related to medication administration in surgical settings. Given the limited availability of theatre-specific research, broader healthcare term "hospital" was also included to gather insights applicable to operating theatres.

However, the search string proved insufficient for identifying relevant ML, DL, and CV methodologies, as its healthcare-specific terminology, such as "operating theatre" and "hospital" narrowed the scope of the results. Consequently, many general AI studies related to medication identification were excluded. This limitation underscores the need for a revised and more inclusive search strategy that omits restrictive healthcare-specific terms and incorporates broader keywords such as "medication classification," and "drug recognition." Such an approach would enable the identification of a wider range of AI techniques with potential applications in healthcare, as shown in Table 2.2. This approach ensured a comprehensive review while prioritizing the unique challenges of medication administration in operating theatres.

2.2.2 Literature Characteristics

A total of 236 papers were included in the final review, spanning the years 2015 to 2025, reflecting a growing interest in technology-driven solutions for medication

Table 2.2: Combined Search String for AI-related Technology in Medication Identification.

Search String
("Computer Vision" OR "Deep Learning" OR "Machine Learning" OR "Convolutional Neural Network") AND ("drug identification" OR "medication identification" OR "drug classification" OR "pill identification" OR "drug recognition" OR "medication recognition")

Table 2.3: Inclusion and Exclusion Criteria for Literature Review

Criterion	**Inclusion Criteria**	**Exclusion Criteria**
Topic Relevance	Studies directly addressing medication identification, medication verification, medication safety, or the reduction of medication errors in healthcare settings.	Studies not focused on medication identification/safety (e.g., medication discovery, treatment efficacy).
Technological Focus	Research on Barcode, QR Code, RFID, NFC, AI, ML, DL, CV, applied to medication identification.	Studies on technologies not listed (e.g., purely electronic health records).
Study Design / Publication Type	Peer-reviewed journal articles, conference papers, review articles, book chapters.	Non-academic websites or news articles, pre-prints, dissertations/theses.
Language	Publications available in English.	Publications not available in English.
Publication Date	Studies from 2015 onwards to reflect modern technology and current healthcare practices.	Studies before 2005, unless they are foundational works still relevant today.
Availability	Full text accessible.	Full text not accessible through AUT library.

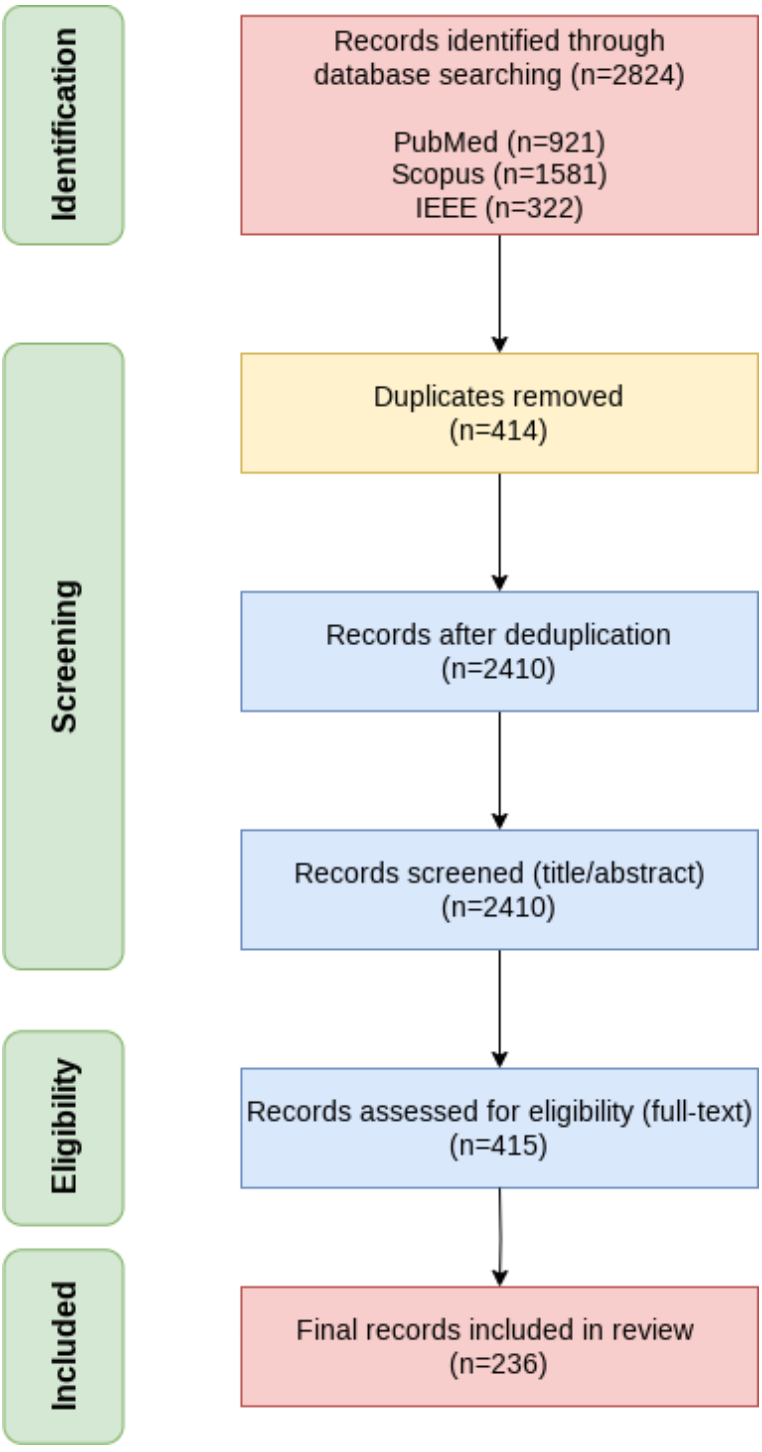


Figure 2.1: Article selection flowchart

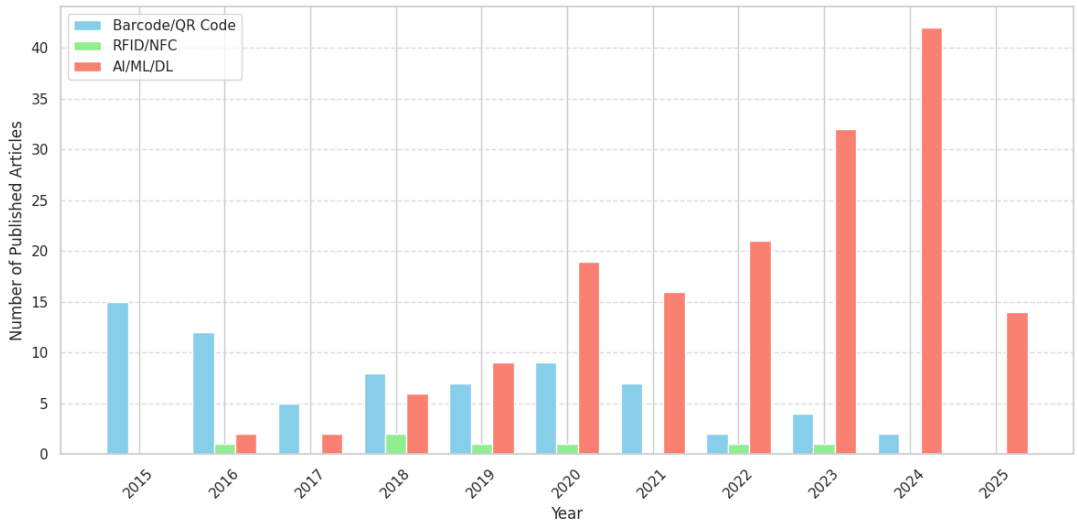


Figure 2.2: Publication trends of medication identification techniques.

identification (see Fig. 2.1). The selection of these 236 papers followed the PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses framework to ensure a transparent review process. The PRISMA flow diagram is illustrated in Fig. 2.1. Initially, 2824 records were identified through database searches using the search strategy outlined in Table 2.1. After removing 414 duplicates, 2410 records remained for screening. During the screening phase, 1995 records were excluded based on titles and abstracts, as they did not meet the inclusion criteria (e.g., studies not related to medication administration, non-English papers, or pre-2015 publications as in Table 2.3), leaving 415 records. In the eligibility phase, 179 full-text articles were excluded due to insufficient focus on the target technologies or lack of relevance to medication identification, resulting in 236 papers included in the final review.

Publication trends in Fig. 2.2 show a notable increase in research activity from 2019 onward, particularly in the use of CV and AI, driven by advancements in ML and DL techniques. This surge reflects growing recognition within the research community that traditional medication identification methods, while foundational, have inherent limitations, especially in high-stakes environments such as operating theatres. The trend

also suggests a shift toward more adaptive and intelligent systems aimed at improving the efficiency of medication identification and enhancing patient safety.

2.3 Manual Process in Medication Administration and Identification

The manual process of medication administration and identification in healthcare settings involves multiple steps that rely heavily on human effort. These processes can vary significantly across contexts such as operating theatres, emergency departments, and general inpatient care.

In general inpatient settings, the process typically includes prescribing (e.g., handwritten prescriptions by doctors), transcribing (recording into a medication administration record), dispensing (retrieving medications from a pharmacy), preparing (e.g., ward-based mixing with double-checks), and administering medications to patients, often followed by manual documentation in nursing notes or electronic health records (EHRs) (Mulac, Mathiesen, Taxis & Granås, 2021; Khan, Khan & Kothandan, 2016).

Specifically, in operating theatre and anaesthesia settings, the manual process is tailored to the unique demands of surgical environments and specific surgery operations but remains heavily reliant on human. Nurses and anaesthetists typically follow a multi-step procedure: preparing medications, labeling syringes manually, verifying medication details by checking labels against patient records, administering medications, and monitoring patient responses (Zadvinskis, Smith & Yen, 2018; Fox, Pedersen & Gumpper, 2015). These processes, while foundational to healthcare, are prone to challenges that could compromise patient safety and efficiency, such as selecting the incorrect medication or administering an incorrect dose.

2.3.1 Limitations of Manual Processes in General Healthcare Settings

The manual process of medication administration across healthcare settings faces significant challenges that increase the risk of medication errors. Human error is a primary concern, with common mistakes such as administering the wrong medication, dose, patient, route, or timing, often caused by miscommunication, misinterpretation, or subjective judgment (Al-Ahmadi et al., 2020). For example, data from the National Health Service (NHS) indicates that 15.6% of medical incidents between 2005 and 2010 were due to either missed or delayed medication (Ti et al., 2020). In another study conducted in 2008 across hospitals in the United States, found that 34% of medication errors occurred during the administration phase, with only 2% being intercepted as near misses (Decia et al., 2017).

Transcription errors can occur during prescription recording, often arising from illegible handwriting or the misinterpretation of abbreviations (Truitt, Thompson, Blazey-Martin, Nisai & Salem, 2016). Administration errors, such as dose omissions, incorrect frequency, or misidentification of patients, account for 20% of all medication errors in tertiary care settings. This finding is based on a prospective observational study conducted between March 2020 and February 2021 at a tertiary care teaching hospital in South India (Parthasarathi et al., 2021).

In addition to low accuracy, manual processes are time-consuming, for example when nurses must make multiple trips to the medication room to retrieve undispensed medications, disrupting workflow and potentially compromising patient safety (Mulac et al., 2021). Similarly, reviewing electronic medical records (EMRs) for patient risk factors, such as their allergies lab data, can be slow and error-prone due to the large volume of information that needs to be processed (Mathioudakis et al., 2021).

Moreover, inconsistent or incomplete documentation also poses significant challenges in healthcare, particularly in adverse drug reaction (ADR) surveillance (Jasim, Kak, Ageed & Zeebaree, 2023). Repetitive electronic documentation can also contribute to distractions and errors, such as omitting medications or entering incorrect orders (Zadvinskis et al., 2018).

Finally, environmental factors also contribute to errors. Chaotic environments, particularly in emergency departments (EDs), increase the risk of medication errors due to verbal orders, incomplete patient details, and rapidly changing patient loads (Owens, Palmore, Penoyer & Viers, 2020). Therefore, implementing structured protocols and decision-support systems can help mitigate these challenges and enhance patient safety.

2.3.2 Limitations of Manual Processes in Operating Theatres and Anaesthesia Settings

In anaesthesia settings, the risks are exceptionally high, making errors particularly threatening. Errors such as administering the wrong medication or dose can have catastrophic consequences, including delayed emergence, or even death (Jelacic et al., 2015; An & Jiang, 2025). Similarly, the highly subjective nature of current patient monitoring can lead to dangerous over- or under-sedation, resulting in severe complications such as hypotension or cardiovascular collapse (Dutt & Saadeh, 2023). These incidents underscore a fundamental vulnerability within the manual medication verification/identification process, where a single mistake can directly cause life-threatening patient harm. Common causes of these errors, such as syringe or vial swaps, often come from human factors challenges like reliance on handmade labels, sub-optimal storage, or deceptive visual similarity of medications (T. Bowdle et al., 2018; Patel & Loveridge, 2015).

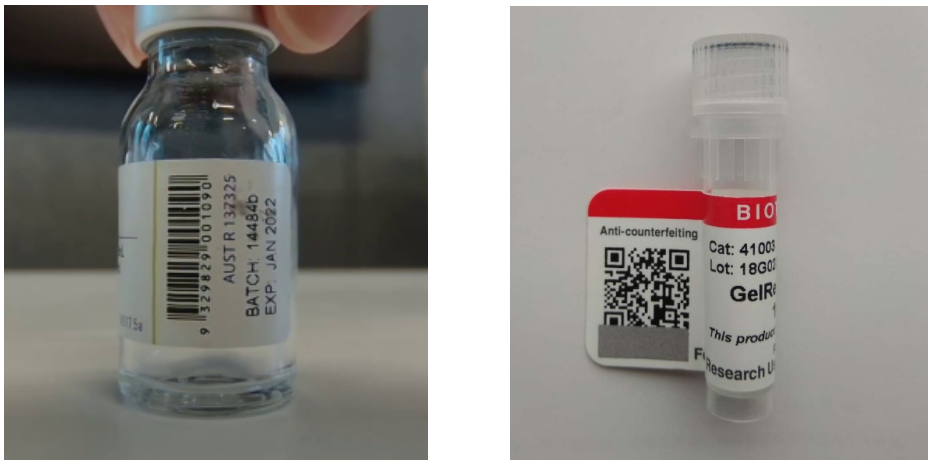
Compounding these risks, the demanding operational environment in the operating

theater exacerbates the risk of errors. Time-sensitive procedures require efficiently, however, slow manual verification and inefficient documentation process can delay medication administration. This not only compromise patient care, resulting in patient discomfort or infection, but also potentially prolong the process of identifying and addressing medication errors (Uy et al., 2015). Such inefficient workflows reveal a fundamental need for speed and a critical imperative for safety protocols. In addition, the frequent interruption and competing priorities in the operating theatre, such as managing multiple patients or responding to emergencies, further disrupt workflows and heighten the risk of errors (Burkoski et al., 2019).

Variability in medication preparation and administration methods across healthcare providers can also lead to inconsistencies and potential errors in anesthesia settings (Jelacic et al., 2015). This lack of standardization presents a significant systemic risk, introducing inconsistencies that make effective error detection and prevention more challenging. In conclusion, these processes deeply dependent on human, environmental, and systemic factors, highlight critical issues in medication identification within anesthesia settings. Despite best intentions, the system remains highly vulnerable to errors with potentially devastating consequences.

2.3.3 The Role of Technology in Reducing Medication Errors

To mitigate the risks of medication errors associated with manual medication identification and administration, healthcare settings are increasingly adopting technologies such as Barcode Medication Administration (BCMA) systems, RFID, and AI-powered CV to improve safety and accuracy (Dunn & Anderson, 2019; Martinez Perez et al., 2018; J. Ye, 2023). BCMA ensures the "five rights" (right patient, medication, dose, route, and time) of medication administration by verifying patient and medication details through barcode scanning (Mulac et al., 2021), while RFID enables contactless tracking to



(a) Barcode (b) QR code

Figure 2.3: Example of a medication container with a (a) Barcode and (b) QR code (Biotium, 2023).

maintain sterility, and AI algorithms to automate the medication administration process, significantly reducing the likelihood of human error (Bouzillé et al., 2019).

2.4 Barcode/QR code

Barcode technology is a method of automatic identification and data capture that uses a series of parallel lines of varying widths and spacing to encode information in a machine-readable format, as shown in Fig. 2.3 (a). Its application in healthcare has been instrumental in reducing medication administration errors and improving patient safety (Mulac et al., 2021; Decia et al., 2017). Barcodes in healthcare typically encode essential information such as patient identification numbers, medication names, dosages, and expiration dates, which can be scanned using handheld or integrated devices to verify and document the administration process (Burkoski et al., 2019; Ameer, Dhillon, Peters & Ghaleb, 2015).

The adoption of barcode technology in healthcare has surged over the years. According to the 2015 American Society of Health-System Pharmacists (ASHP) national

survey, 93.7% of U.S. general and children's medical-surgical hospitals adopted BCMA systems to verify patient identity and electronically check administered doses, a significant rise from just 1.5% in 2002 (Pedersen, Schneider & Scheckelhoff, 2016). This growth underscores the role of barcode technology in enhancing patient safety, especially in high-risk environments like operating theatres and anesthesia, where medication errors can have serious consequences (T. Bowdle et al., 2018; J. Ye, 2023).

Unlike barcode, a QR code is a two-dimensional barcode (Fig. 2.3 (b)) that able to store information in a squared-shaped grid, including text, URLs, and other information. QR codes offer several advantages compared to traditional barcode, such as larger data storage and more efficient scanning from any orientation (C.-H. Lin, Tsai, Tsai, Wen & Hu, 2012; Uzun & Bilgin, 2016).

2.4.1 Applications of Barcode/QR Code

Barcode technology plays a crucial role in verifying medication administration and preventing errors. For instance, Macias et al. (Macias, 2018) observed 2,912 medication administrations using barcode verification across 715 patients in the onco-hematology day unit at Príncipe de Asturias University Hospital in Madrid, Spain. Additional studies reported barcode scanning rates of 71% for medications, 91% for scannable doses, and 80% for patient ID wristbands across two medical wards in a 700-bed hospital in Norway (Mulac et al., 2021).

Beyond administration, barcode technology is also essential in medication dispensing and preparation. Many hospitals use barcoded unit-dose plastic bags that contain critical information such as medication name, dosage, expiry date, lot number, and national article identifier. These barcoded bags are linked to patient data and processed via Computerized Physician Order Entry (CPOE) systems, ensuring accurate dispensing. Before administration, the barcode is scanned to confirm the correct match between the

medication and the patient. For example, in a prospective uncontrolled before-and-after study conducted across six clinical wards in a Dutch university hospital from 2018 to 2020, barcode technology was applied in 3,191 medication administrations (3,120 analyzed), reinforcing its role in streamlining medication management (Jessurun, Hunfeld, Van Rosmalen, Van Dijk & Van Den Bemt, 2021). Another study, the 2013 ASHP national survey on informatics, found that nearly 75% of hospitals in the US were using BCMA, exceeded the use of any one system alone and of paper-based systems (Fox et al., 2015).

BCMA can also be integrated with other digital healthcare systems such as electronic Medication Administration Records (eMAR), CPOE, and Automated Dispensing Cabinets (ADCs) to create closed-loop medication management systems and reduce errors. For instance, it can be integrated into EHR systems to give clinicians access to accurate, real-time medication information. Specifically, CPOE systems allow clinicians to input medication orders directly into the EHR, helping to eliminate errors caused by illegible handwriting or manual data entry (J. Ye, 2023).

Another before-and-after study by Truitt et al. (Truitt et al., 2016) from 2008 to 2011 examined the effects of implementing BCMA and eMAR technologies on the incidence of ADEs at a 400-bed academic medical center using incident reports. The study found that ADEs decreased from 0.26% to 0.20%, and the rate of transcription errors dropped from 0.089% to 0.036%. Furthermore, in a comparative study of three hospitals in Queensland, Australia, the implementation of a closed-loop Electronic Medication Management System (EMMS), comprising CPOE, ADCs, unit-dose packaging, BCMA, and eMAR, significantly reduced medication turnaround times to 35 minutes compared to 120 minutes in paper-based systems.

While traditional linear barcodes have driven these advancements, the capabilities of two-dimensional QR codes are also being explored to further enhance medication safety. QR codes come with greater data storage capacity and ease of scanning via common

smartphone devices, offer unique application in healthcare settings (C.-H. Lin et al., 2012). For dispensing, QR codes can encode comprehensive medication details, such as batch numbers and expiry dates, facilitating rapid and accurate retrieval of medication information (C.-H. Lin et al., 2012). At the administration phase, nurses can scan QR codes on medication labels using hospital-issued devices or even personal smartphones to verify the five rights efficiently (Uzun & Bilgin, 2016).

Beyond provider-centric uses, QR codes are also implemented through innovative patient-facing methods. For instance, they are implemented for medication administration where patients scan QR code transparencies printed on each medicine package (Ti et al., 2020). Specifically, the smartphone obtains the QR code transparency associated with the medicine and when these QR code transparencies are stacked, an image sufficient to recognize the patient can be viewed on the smartphone screen. This allows patients to confirm that they have received all the medications prescribed by their doctor, reducing the risk of receiving the wrong medication.

Furthermore, QR codes' ability to link to external resources, such as patient education materials, adverse event reporting forms, or instructional videos, provides a powerful means of enhancing patient empowerment and adherence, extending beyond basic identification. As prescription medication use increases with age, and with a globally aging population, improving medication safety for outpatients is becoming increasingly critical (Tseng & Wu, 2014).

2.4.1.1 Anaesthesia/Operating Theatres

In anaesthesia settings, systems like the Codonics Safe Label System (SLS) are widely utilized in operating theatres for syringe labeling and barcode scanning, ensuring accurate medication identification before administration (Jelacic et al., 2015). A study conducted in an operating theatre demonstrated that integrating the Codonics SLS with the Smart Anesthesia Manager system significantly improved medication safety,

resulting in a 41% reduction in reported errors between 2014 and 2015 at the University of Washington Medical Center (T. Bowdle et al., 2018).

During surgery in the intraoperative phase, barcode scanning is performed before medication administration, often integrated with anaesthesia information management systems. For example, Khan et al. (Khan et al., 2016) conducted a randomized controlled trial with ten anesthetists to evaluate a prototype device that enforces barcode scanning for medication administration. The study found that the device significantly increased compliance with safe medication administration procedures from 36.3% (control) to 100% (intervention) in simulated cases.

Finally, in the postoperative and recovery phases, barcode systems are used to verify medications as part of BCMA workflows, ensuring accurate medication administration after surgery. Nurses utilize wireless barcode scanners to scan medication barcodes, linking the process to electronic records for documentation. A qualitative study in a Midwestern academic medical center involving 19 nurses and 30 interviews explored the effectiveness of barcode technology in postoperative settings, highlighting its role in improving medication safety, enhancing workflow efficiency, and reducing documentation errors (Zadvinskis et al., 2018).

2.4.2 Advantages of Barcode/QR Code

Barcode/QR technology offers substantial benefits in medication administration, improving patient safety and operational efficiency across healthcare settings. It helps minimize medication errors (Truitt et al., 2016), enhance workflow efficiency (Zadvinskis et al., 2018), improve compliance (Khan et al., 2016), and provide a cost-effective solution for medication administration (Carver, Burgess, Cooper, Elders & Kramer, 2018).

Firstly, BCMA plays a crucial role in minimizing medication errors by ensuring accurate medication administration and verification. For example, Owens et al. (Owens

et al., 2020) observed a community hospital's emergency department and found that the medication administration error rate decreased from 2.96% to 0.76% after BCMA implementation. QR codes also contribute to this improvement; for instance, Ti et al. (Ti et al., 2020) proposed a secure method where patients scan QR code transparencies on medicine packages to verify prescribed medications, which enhances self-administration safety. These studies demonstrate that barcode and QR code systems are fundamental tools in patient safety, directly mitigating the risk of medication errors across diverse clinical settings.

Secondly, both barcode and QR code systems improve workflow efficiency by streamlining the medication administration process, reducing the time spent on manual data entry and verification. Zadvinskis et al. (Zadvinskis et al., 2018) conducted a study in an academic medical center in 2011, showing that barcode scanning integrates seamlessly into healthcare workflows by automating patient and medication identification, reducing manual documentation, and minimizing interruptions, thereby allowing nurses to spend more time on patient care. In high-stakes settings like operating theatres, barcode systems reduce cognitive load by automating verification, thereby improving safety and efficiency. For instance, Khan et al. (Khan et al., 2016) proposed a prototype device requiring barcode scanning of syringe labels before medication administration to reduce errors in anaesthesia. In a study involving 10 anaesthetists from Singapore General Hospital, the device was shown to improve workflow by standardizing the medication verification process and ensuring full compliance with safe administration procedures. QR codes, easily scanned by smartphone, can also enable nurses to quickly scan uni-dose medication labels and patient wristbands, automating verification and ensuring the right medication is given to the right patient, thereby reduces time spent on manual checks and improving safety and efficiently of medication administration workflow (Decia et al., 2017). This evidence suggests that the value of barcode and QR code technologies lies not only in error prevention but also in creating standardized,

efficient, and less cognitively demanding workflows, resulting in a safer environment for both clinicians and patients.

Finally, barcode technology is cost-effective due to the affordability of scanners and the simplicity of printing barcodes. Printing barcodes on labels or packaging is inexpensive, requiring only standard printers. In theater settings, affordable scanners and easy barcode printing make the technology cost-effective. For instance, compact scanners integrated into workflows come with minimal expense (T. A. Bowdle et al., 2019). The widespread use of smartphones means that QR code scanning often requires no specialized equipment, further reducing implementation costs for healthcare facilities, making advanced verification accessible even for smaller clinics or patient-centric applications. This affordability is a critical factor in the widespread adoption and sustainability of barcde/QR code in healthcare settings. These systems have the potential to raise the baseline standard of medication safety across a diverse range of healthcare institutions.

2.4.3 Limitations of Barcode/QR Code

Although barcodes/QR codes offer numerous benefits in medication verification and identification, their implementation in the healthcare industry also presents challenges. These include technical issues (Zadvinskis et al., 2018; Koppel et al., 2008), cost-related barriers (Sakowski & Ketchel, 2013; Dunn & Anderson, 2019), and compliance and workflow challenges (Grailey et al., 2024; T. A. Bowdle et al., 2019). Various studies have highlighted these concerns, emphasizing the need to address these challenges for optimal barcode system integration.

Firstly, while barcode technology has the potential to significantly prevent medication errors by automating the verification of patient identity and medication details, this potential is limited by gaps in system design and coverage that are not directly tied to

compliance issues. For instance, BCMA systems may not verify multi-dose medication containers, leaving room for potential errors in dosage or administration (Zadvinskis et al., 2018). Both linear barcodes and QR codes are also susceptible to physical damage, such as scratches, smudges, or creases, which can render them unreadable and compromise the integrity of the data (Koppel et al., 2008). In addition, while QR codes offer enhanced data storage and smartphone compatibility, their capacity to link to external content introduces specific security risks, such as unauthorized access to sensitive information if not managed within a secure healthcare system (Upton, Olsson-Brown, Marshall & Sacco, 2017). These limitations highlight that while barcode/QR code technology can reduce certain types of errors, it is not a comprehensive solution. It fundamentally fails to address the full complexity and variability of the medication administration process, particularly for scenarios that fall outside of its designed scope.

Secondly, the potential of barcode and QR code technology to significantly prevent medication errors is critically limited by a combination of inflexible design and frequent technical failures. In operating theatres, for example, not all implantable devices have scannable barcodes or are in the database, particularly custom or last-minute devices, reducing the system's comprehensiveness and leaving room for errors (Wilson et al., 2020). These findings highlight the limitations of a one-size-fits-all barcode/QR code approach in specialized environments, where the technology struggles to accommodate the dynamic and variable nature of clinical workflows. Compounding these design gaps are frequent technical challenges that compromise effectiveness. For instance, system malfunctions (Saleem, 2023), slow scanning times (Ho & Burger, 2020), and failure to scan properly (van der Veen, Taxis & van den Bemt, 2023) often require multiple attempts, which wastes time, frustrates users, and results in inefficient workflows (Xie et al., 2019). Barcode and QR code scanners can also misread or misinterpret data, potentially leading to errors (Mulac et al., 2021), while poor interface design can cause confusion (Huang & Gramopadhye, 2016; Zadvinskis et al., 2018). These technical

flaws show a critical disconnect between the technology and user needs, often making the system itself a source of friction and frustration that can inadvertently create new risks.

Thirdly, while individual components of barcode technology, such as scanners and printing barcodes or QR codes, are relatively inexpensive and easy to produce, the implementation and maintenance of a comprehensive barcode system in healthcare settings present significant cost barriers. For instance, a study conducted by Sakowski et al. (Sakowski & Ketchel, 2013) in 2013 estimated that the cost of implementing and operating BCMA, including electronic pharmacy management and medication repackaging, over five years was $40,000 (range: $35,600 to $54,600) per BCMA-enabled bed and $2,000 (range: $1,800 to $2,600) per harmful error prevented. In anesthesia settings, the cost of implementing barcode systems, such as ADCs or automated anesthesia carts, presents a significant barrier, particularly for community hospitals and private practitioners (Dunn & Anderson, 2019). While the use of ubiquitous smartphones for QR code scanning can reduce hardware costs, this is often offset by the significant investment required for secure software development, backend integration, and ongoing maintenance. This financial burden underscores a significant practical barrier, limiting the widespread adoption of the technology by many facilities and potentially hindering improvements in patient safety.

Finally, although barcode technology is designed to improve compliance by providing a systematic and automated process for verifying patient identity and medication details, practical challenges often lead to low compliance rates and workarounds that reduce its effectiveness. For example, nurses may scan patient notes instead of wristbands, which increases the risk of administering medication to the wrong patient (Barakat & Franklin, 2020). Inadequate training, perceived negative impacts on workflow, poor equipment, and staff resistance to new systems all contribute to low scanning rates, with

some wards reporting compliance as low as 0–20% (Grailey et al., 2024). In operating theatres, anesthesia providers may not consistently scan syringe barcodes before administration, with initial compliance rates as low as 25% in a pilot study conducted by Bowdle et al. (T. A. Bowdle et al., 2019). Another study found that only 36.3% of medication administrations in a control group complied with scanning protocols, compared to 100% in an intervention group, highlighting the gap between potential and actual compliance (Khan et al., 2016). A safety system is only effective if clinicians find it easy and reliable to use, and these studies reveal that these compliance gaps, which are likely to extend to any QR code-based system, highlight a critical disconnect between the technology's intended use and its real-world application.

2.5 RFID/NFC



(a) (b)

Figure 2.4: (a) Basic RFID system (TTelectronics, 2023) and (b) a medication container with RFID tag (Insights, 2023).

RFID technology, including NFC, is a wireless method for identifying and tracking objects or individuals using radio waves (Finkenzeller, 2010). An RFID system consists of tags (with a microchip and antenna), readers (which emit radio waves to communicate with tags), and a backend system for data processing, as shown in Fig. 2.4 (a). Both RFID and NFC use radio waves for tagging and tracking, but NFC operates at 13.56

MHz and is limited to short-range communication (up to 10 cm), making it ideal for secure, close-proximity tasks. In healthcare, RFID/NFC technology is employed to enhance patient safety, improve operational efficiency, and ensure accurate tracking of resources by enabling real-time identification and monitoring of patients, staff, medications, and equipment (Martinez Perez et al., 2018; Yoo et al., 2016).

The adoption of RFID/NFC technologies in healthcare has steadily increased, particularly in medication administration (Uy et al., 2015). Market analyses indicate significant growth, with the global healthcare RFID market valued at USD 16.95 billion in 2016 and projected to expand at a compound annual growth rate of 7.7% between 2017 and 2023 (Profetto, Gherardelli & Iadanza, 2022). While barcode technology has seen a higher adoption rate, RFID/NFC continues to expand due to its advantages in real-time tracking and automated data capture. RFID/NFC is often used alongside barcode technology to provide complementary benefits, such as real-time tracking and error prevention, in various hospital areas including pharmacies, blood banks, and operating theatres (Yoo et al., 2016; Decia et al., 2017)

2.5.1 Application of RFID/NFC

In general healthcare settings, RFID/NFC technology is applied to streamline medication administration, improve patient identification, and enhance hospital operations across multiple departments. For instance, RFID systems enhance patient safety by enabling real-time tracking of medication, specimens, and materials while ensuring accurate documentation through cross-checking patient information with medical orders. RFID wristbands and closed-loop medication administration (CLMA) systems further improve efficiency and prevent medical errors (Yoo et al., 2016).

Another application involves the use of NFC wristbands as part of a Hospital Management System (HMS), where the wristbands store patient-specific data securely

for quick and efficient identification and authentication (Macriga, Samyukthaa & Yogavarshini, 2023). In addition, RFID/NFC technologies are implemented during the care process of patients in an Intensive Care Unit (ICU), specifically for identifying the medication to administer, health personnel, and the patient through NFC tags, which reduces medication errors and improves safety.

2.5.1.1 Operating Theatre and Anaesthesia Settings

In operating theatre and anaesthesia settings, RFID/NFC technology plays a crucial role in addressing the unique challenges of surgical environments, where rapid and accurate identification and tracking are essential for patient safety. For instance, Naik et al. (Naik, Bhuatara & Chougala, 2019) proposed the use of RFID for medication identification during the medication delivery phase in operating theatres. In this system, RFID tags on medications are linked to a database of Received Signal Strength Indication (RSSI) signatures unique to each bed location. This application is particularly beneficial in busy surgical environments with multiple patients, as it helps reduce the risk of wrong-patient medication errors.

While long-range RFID is highly effective for these real-time tracking applications, NFC offers distinct advantages for close-proximity tasks. Due to its short range, NFC acts as a security feature by ensuring the user is scanning the intended item, thereby reducing the risk of accidental or unintended verification. RFID and NFC are often used in tandem, with RFID complementing barcode systems by enabling real-time, room-wide tracking of patients, medications, and medical equipment (Martinez Perez et al., 2018). NFC, on the other hand, is applied in operating theatres to facilitate medication management across multiple stages, including prescribing, administration, and review (O'Connell et al., 2016). A mobile application can utilize these integrated technologies to locate patients and medications, ensuring the correct medication is available at the right time during anaesthesia (Decia et al., 2017). This integration enhances

both efficiency and patient safety by minimizing delays and preventing medication mismatches.

Additionally, the efficiency and ease of use of NFC technology were demonstrated in a system tested with 89 students (36 medical, 10 nursing, and 43 pharmacy), particularly during the administration phase. All participants reported high levels of technological familiarity, with each indicating daily use of either a smartphone or a computer (O'Connell et al., 2016). Furthermore, RFID technology in operating theatres can involve the use of implanted RFID microchips for patient identification, allowing medical staff to scan the chips and access medical records during preoperative or postoperative stages. This approach supports accurate medication order verification, improving overall medication safety in surgical settings (Fram et al., 2020).

2.5.2 Advantages of RFID/NFC

RFID/NFC technology plays a crucial role in enhancing medication administration by improving the accuracy, efficiency, and safety of medication delivery, especially within the high-stakes environment of the operating theatre. These technologies enhance patient safety (Naik et al., 2019; Yoo et al., 2016), streamline medication administration processes (Macriga et al., 2023; Martinez Perez et al., 2018), and improve care coordination by enabling accurate identification, real-time tracking, and automated verification (Macriga et al., 2023; Decia et al., 2017), all of which are essential for preventing errors in surgical and anaesthesia care.

RFID/NFC technology could significantly reduces medication errors in medication administration, a critical concern in both general healthcare and operating theatres. For instance, a study on hospital robotic delivery using RFID achieved a 92% success rate in medication identification without specialized hardware (Naik et al., 2019). Another study conducted in a simulated hospital ward demonstrated that an NFC system reduced

prescribing errors to a mean of 0.22 per simulated round, compared to 4.09 errors with a paper-based system (O'Connell et al., 2016). In operating theatres, RFID can be integrated with barcode systems for medication administration, enabling convenient and accurate documentation of care practices by cross-checking patient information with medical orders, thereby reducing the risk of medication errors (Yoo et al., 2016). These applications are particularly beneficial in busy surgical environments like anaesthesia settings, as they help reduce the risk of wrong-patient medication errors, a severe threat to patient safety during surgery.

In addition, RFID/NFC technology streamlines medication administration workflow in anaesthesia by automating the verification process, reducing the administrative burden on healthcare professionals, and enabling real-time medication tracking (Macriga et al., 2023). A study from Seoul National University Bundang Hospital demonstrated that RFID/NFC technology improves care quality and patient safety by streamlining medication delivery, minimizing nurses' workload, and allowing for greater patient interaction (Yoo et al., 2016). Additionally, RFID/NFC ensures real-time location tracking of medications, making sure the right medications are available when needed, which enhances workflow efficiency (Decia et al., 2017). NFC further enhances medication administration by combining reading and writing capabilities into a single mobile device, enabling more efficient task management for healthcare professionals (Martinez Perez et al., 2018). It also supports computerized tracking of admissions, care plans, vital signs monitoring, prescriptions, and medication administration.

Beyond reducing medication errors and enhancing operational efficiency, RFID/NFC technology provides crucial benefits related to security and accountability that are particularly vital in anaesthesia settings. For instance, NFC's inherent security features provide a secure method for verifying patient and medication data, minimizing the risk of unauthorized access to sensitive information (Macriga et al., 2023). Similarly, RFID systems can enhance safety by sending alerts when unauthorized individuals or objects

enter restricted zones, such as operating rooms, helping maintain a controlled environment for medication administration and surgical care (Decia et al., 2017). Furthermore, this technology promotes accountability by tracking medication usage, which helps reduce medication waste and prevents the unauthorized use of controlled substances, a major concern in anaesthesia (Decia et al., 2017; Yoo et al., 2016). These specialized security and accountability measures, combined with high user acceptance, such as the 90% of students who expressed enthusiasm for using NFC systems for medication administration, highlight the technology's strong potential for widespread adoption in surgical units (O'Connell et al., 2016).

2.5.3 Disadvantages of RFID/NFC

Despite its benefits, RFID/NFC technology faces several limitations that can impact its effectiveness, adoption, and safety in medication processes, such as technical limitations (Naik et al., 2019; Decia et al., 2017), cost and implementation challenges (Martinez Perez et al., 2018), user-related issues (Uy et al., 2015), and safety concerns (Fram et al., 2020).

Specifically, RFID/NFC technology in medication administration is subject to several technical limitations that can affect its reliability and accuracy. For example, RFID systems for automated medication delivery proposed by Naik et al. have a narrow reading range, which can limit their effectiveness in detecting and delivering medications to the correct bed location (Naik et al., 2019). Similarly, the CAMACUA system, which integrates RFID for real-time tracking, suffers from limited accuracy when objects are closer than 1 meter from room entrances, and it may produce false negatives (failing to detect objects in the corridor) or false positives (misplacing objects) (Decia et al., 2017). These technical issues can disrupt medication administration workflows and potentially compromise patient safety by leading to medication errors.

These findings reveal that despite their promise, RFID/NFC systems are not infallible and can suffer from a fragile reliability that undermines their effectiveness in high-stakes environments.

In addition, the high upfront cost and complexity of implementing RFID/NFC technology present significant challenges for medication administration. For instance, the NFC passive tags and mobile devices necessary for the system come with a high initial cost. Specifically, the NFC tags cost 0.10 euros per unit, while mobile phones equipped with NFC capabilities are priced around 150 euros each. Although large-scale implementation could reduce the cost per unit over time, the upfront expense remains a concern (Martinez Perez et al., 2018). Additionally, the adoption of RFID for medication administration in operating theatres is hindered by low adoption rates, with costs and negative perceptions frequently cited as major barriers to implementation (Uy et al., 2015). Additionally, integrating RFID/NFC systems with hospital workflows can be challenging, requiring training for personnel. During the transition, both the traditional paper system and the new system should coexist for smooth implementation (Martinez Perez et al., 2018). This highlights a gap between potential and accessibility, as high costs and complexity limit adoption to larger institutions, creating disparities in patient safety across healthcare settings.

Furthermore, the use of RFID/NFC technology in medication administration introduces several safety concerns that can impact patient outcomes. The radio waves used for communication between RFID readers and tags can cause electromagnetic interference with other medical equipment in the operating theater, leading to inaccurate readings and an increased risk of medication errors (Haddara & Staaby, 2018). For example, implanted RFID microchips, sometimes used for patient identification, may interfere with imaging procedures such as ultrasound or CT scans. They can also increase the risk of infection or become damaged during surgery (Fram et al., 2020). These risks, combined with the technical challenges of maintaining secure and reliable

systems, add complexity to their integration into clinical workflows. This presents a significant paradox: a technology designed to enhance safety may, in some cases, introduce new and unforeseen risks that must be carefully weighed against its benefits.

2.6 AI and CV

AI methods, including ML, DL, and CV, had significantly transformed medication identification by providing innovative solutions that enhanced patient safety, optimized medication management, and improved clinical outcomes. While previous technologies relied on physical tags or labels, AI and CV moved beyond these limitations by enabling the precise identification and administration of medications based on their visual appearance. In high-stakes settings like operating theatres, AI leveraged advanced models such as Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) to provide a new level of automated verification and accuracy, representing the next evolution in medication safety protocols.

The AI/CV-based medication identification process typically involved four main stages: data collection and preprocessing, feature extraction, model training, and prediction. The first step involved collecting relevant data, such as images of medicine blister packs or pill imprints (Warnakulasuriya et al., 2022). These images were then preprocessed to standardize input formats for CV models. Common preprocessing steps included resizing, padding, normalization, and segmentation to isolate the medication packaging from the background using techniques such as edge detection (Chiu, 2024).

Next, CNNs were employed to extract meaningful features from the preprocessed images. For example, dual CNN models were used to process images of blister packs to detect edges, shapes, and textures (Warnakulasuriya et al., 2022). These networks generated feature maps that were downsampled via pooling layers to reduce dimensionality while retaining essential patterns. Popular CNN architectures such as VGG16,

Inception-V3, and ResNet50-V2 were also utilized to capture distinct features of both original and generic medications (Laohakijvitoon & Nilwong, 2024).

In the third step, CV models were trained on labeled datasets to learn how to classify or identify medications. For instance, a CNN was trained on a dataset of 6,240 images spanning 26 medication types, including both original and generic versions (Laohakijvitoon & Nilwong, 2024). During training, the model adjusted its parameters to effectively distinguish among various medications. Transfer learning techniques were also employed to enhance model performance by leveraging pre-trained networks (Laohakijvitoon & Nilwong, 2024; Vencer, Marqueses, Espineli, Caballero & Caballero, 2024). Optimization algorithms such as Adam were typically used during training to minimize loss functions like categorical cross-entropy (Silva, e Silva, Orlando & Bezzon, 2024).

Finally, once trained, the CV model classified or recognized the identity of a medication. During inference, the CNN processed a new blister pack image, passing extracted features through fully connected layers to generate a probability distribution over possible medication classes and selecting the class with the highest probability (Warnakulasuriya et al., 2022). For real-time applications, object detection models such as YOLOv8 detected and classified medications in video feeds, outputting bounding boxes and class labels to assist with verification during surgery (D. Zhu, Huang, Yung & Ip, 2024). Additionally, DNNs such as SST-CNN were applied to predict both medication type and concentration from spectral data, ensuring safe and accurate medication administration (Z. Zhu et al., 2024).

2.6.1 Applications of CV Systems

In healthcare settings, CV applications for medication identification address critical needs by reducing medication errors and enhancing patient safety. CV-based systems

are uniquely suited to the operating theatre's fast-paced environment, where the high risk of human error from factors like look-alike, sound-alike (LASA) medications can have severe consequences. For example, a DL-based medication identification model using YOLO was deployed in a medical center in Taiwan to verify LASA medications, significantly reducing medication errors and improve patient safety (Ting, Chung, Chen, Chiu & Hsieh, 2020). Similarly, another YOLO-based system implemented in Thailand aimed to enhance medication safety by mitigating errors caused by LASA medications (Kittisorayut et al., 2024). Beyond LASA identification, recent advances in pill recognition leveraged CNNs, Graph Neural Networks (GNNs), and smartphone-based systems to improve accuracy and minimize medication errors in both pharmacy and home settings (Heo, Kang, Lee, Jeong & Kim, 2023; Nguyen et al., 2023; Dang et al., 2024).

CV-driven systems also play a vital role in supporting medical providers by streamlining their medication workflow, ultimately boosting efficiency and reducing cognitive load. The technology automates essential tasks such as medication identification, counting, and verification, which frees up personnel to focus on the patient. For example, Tavakoli et al. (Tavakoli et al., 2025) proposed a CV-based system for identifying and counting pharmaceutical items, streamlining workflow efficiency, providing timely data for informed decision-making, improving inventory processes, and achieving high accuracy in medication identification and counting. In another study, a CV system combining YOLOv8 and the Voronoi Diagram algorithm was employed to automate unit-dose repackaging in hospitals. The detection network, trained with 8,856 labeled images of blister packs and pills, achieved a 99.0% mean average precision (mAP). This high accuracy significantly reduced manual labor and enhanced the medication management process (Tani, Norberto, da Silva & Durand-Petiteville, 2024).

While pill classification is a related application within the broader medication identification space, it addresses a different problem - tablet/pill appearance rather

than label text. The literature on image stitching, OCR, and text detection directly underpinning the methods developed in this thesis is reviewed in detail in Chapters 4 and 5.

2.6.2 Advantages of CV

In healthcare, CV technologies offer significant advancements over other medication identification methods like barcodes and RFID, which are limited by their reliance on reading pre-encoded data from barcode or physical tags.

Firstly, CV significantly enhances accuracy and reduces medication errors. While technologies like barcodes and RFID rely on reading static, pre-encoded data and are prone to misreads if tags are damaged or mislabeled, CV systems can intelligently interpret rich visual information to provide a more robust and versatile form of verification (Heo et al., 2023; Palenychka et al., 2018). This is achieved through the use of CNNs, which were trained on extensive datasets of medication images. These models could accurately identify and differentiate medications by analyzing subtle visual features such as shape, color, size, and the imprints on its surface. For example, an EfficientNetB3-based CV system was developed to identify medications and vitamins, helping to prevent counterfeit products and ensure treatment efficacy, achieving 99% accuracy on a dataset collected from the open-source Kaggle platform (Verma, Chauhan & Kumar, 2024). Similarly, the ClinicYA system employed a Mask R-CNN algorithm combined with K-means clustering to classify pills based on shape and color, achieving 99.27% accuracy and providing a fast and accurate solution for pill identification (Srikamdee, Suksawatchon, Suksawatchon & Werapan, 2022). Moreover, unlike barcodes/RFID, CV systems were able to identify medications in complex scenarios (e.g., multiple medications in one image, variable lighting, medication rotations), making them suitable for practical applications. (Ou et al., 2020; Suksawatchon, Srikamdee,

Suksawatchon & Werapan, 2022).

Secondly, CV approaches are more adaptable because they are not dependent on a specific label or tag and able to recognize a wide range of medications based on their appearance alone. This flexibility allows for the identification of a greater variety of medications and is particularly useful for new or generic medications that may not have standardized coding. For instance, Xia et al. (Xia, 2024) proposed using Near-Infrared Spectroscopy to develop intelligent medication identification models that can classify medications accurately and efficiently, even in large quantities. Additionally, DL models classify medications based on composition, usage instruction, enabling applications beyond simple identification. For instance, Vu et .al (Vu, Pham & Le, 2025) implemented a medication classification system based on medication composition and usage description using NLP techniques, such as text preprocessing, word embedding, and CV models, to classify medications.

Finally, CV-based solutions demonstrate scalability and adaptability in medication identification. Specifically, CV models, trained on large datasets, was able to identify and adapt to new medications through transfer learning (Gala, Gandhi, Gandhi & Sawant, 2021; Geoffrey et al., 2022), unlike static barcode/RFID databases. In addition, CV generalizes across diverse condition, such as low light and damage packaging, reducing the need for controlled environments required by barcode/RFID scanners. For instance, the mobile app ClinicYa, which used CV and K-means clustering to identify pills by shape and color, accurately recognized the medication even under suboptimal real-world conditions (Srikamdee et al., 2022). In another study, CV models such as YOLOv3 and Mask R-CNN were used for pill shape recognition in unconstrained images, including noise and poor illumination (Suksawatchon et al., 2022).

2.6.3 Disadvantages of CV

While CV offers significant advantages over traditional medication identification methods such as barcodes and RFID, it also introduces disadvantages and challenges.

One major drawback is the high initial and ongoing costs associated with CV solutions. Developing and maintaining these systems requires substantial investment in computational resources, continuous model updates, data curation, and algorithm refinement to adapt to new medications or datasets (Aghavali et al., 2024; Nguyen et al., 2023). This goes beyond a financial obstacle, highlighting a fundamental difference in technological frameworks. Large CV models require robust computational infrastructure (e.g., GPUs, cloud computing) and reliable network connectivity for real-time processing, which may not be feasible in resource-limited settings like a small operating theatre. In contrast, barcode and RFID systems rely on static, standardized databases, can be operated with minimal infrastructure using handheld scanners or readers, which typically result in lower long-term operational costs.

In addition, CV models are complex and require training. CV models, particularly DL systems, demand large volumes of high-quality, annotated data for effective training and accurate performance. The process of compiling these labeled datasets is both time-consuming and labor-intensive, often requiring domain expertise to ensure precision and consistency in annotations (Li et al., 2020). This heavy reliance on data creates a paradox: the same thing that makes CV powerful also makes it fragile. For example, even systems with high accuracy in controlled settings can fail in real-world use due to factors like lighting changes, blurry images, or overfitting. In contrast, barcode and RFID systems operate using pre-encoded and standardized data formats, eliminating the need for such complex data preparation. Furthermore, training CV models can take significant time and specialized expertise in model architecture and hyperparameter tuning. The training time depending on the dataset and model's architecture (Xia,

2024). Beyond the time requirement, model development involves complex tasks such as selecting suitable architectures, tuning hyperparameters, and evaluating performance through iterative experimentation, which require in-depth knowledge of ML, since the model's performance may be sensitive to the choice of hyperparameters (Mohd Yusof, Muda, Pratama & Carbo-Dorca, 2022). Furthermore, CV systems often need retraining to adapt to new medications, packaging, or environmental conditions, whereas barcode/RFID databases are easily updated with new codes (My, Le, Vo & Hoang, 2025). In comparison, barcode and RFID systems are considerably simpler to deploy, as they utilize standardized data encoding and reading formats that do not require model training or optimization.

Although CV models can achieve high identification accuracy, there remains a risk of misidentification and potential medication errors. Challenges such as LASA medications, lighting variations, or poor image quality can compromise model performance. For instance, Ting et al. (Ting et al., 2020) reported cases of misidentification caused by aluminum foil wrapping on the back of blister packages, where the textual patterns on the reverse side led to incorrect recognition. Barcode/RFID systems are less prone to such errors, as they rely on unique codes or tags, assuming proper labeling. Another concern is overfitting or limited dataset diversity, which can reduce the accuracy of AI systems and compromise patient safety (E. K. Lee et al., 2017; Vu et al., 2025). Verma et al. (Verma et al., 2024) reported instances of misidentification in their study on medication and vitamin classification using the EfficientNetB3 model. Specifically, fish oil was incorrectly classified as Neozep and, in two cases, as Bioflu. In contrast, barcode/RFID methods do not rely on visual interpretation and thus are less susceptible to such errors.

Moreover, CV models, particularly DL architectures such as CNNs and RNNs are often considered opaque, making it challenging for healthcare professionals to interpret or trust their decisions. This lack of transparency is a critical concern in safety-sensitive

applications. In contrast, barcode/RFID systems offer clear, unambiguous outputs by matching scanned codes to standardized database entries. These systems are inherently transparent, relying on physical tags. Furthermore, CV systems are susceptible to bias, such as over-representing commonly prescribed medications, which can skew predictions. For example, Nguyen et al. (Nguyen et al., 2023) reported that their DCN-based pill identification model exhibited bias due to the higher prevalence of certain pills in prescription data. Similarly, Silva et al. (Silva et al., 2024) highlighted that the limited availability of real-world data can lead to biased datasets, reducing the generalizability of CV models. This data scarcity also hampers the effective evaluation of model performance in practical healthcare settings, an lead to skewed predictions, unlike the neutral, standardized nature of barcode/RFID data.

In conclusion, while CV-based medication identification systems show great potential, their current limitations present a major barrier to widespread adoption. Recent studies (Table 2.4) demonstrate that AI/CV algorithms can achieve high accuracy and F1/mAP scores across diverse datasets, yet challenges such as controlled testing environments, misidentifications, and limited real-world validation remain. There is a pressing need for new approaches that go beyond accuracy alone, as the ultimate goal is to reduce medication errors across healthcare settings, including operating theatres. Future CV systems must be affordable, data-efficient, and transparent to ensure they can be reliably implemented. By addressing these critical requirements, the full potential of CV technology can be realized, not just as a technical advancement, but as a vital tool for improving medication safety and patient outcomes.

Table 2.4: Comparison of Studies on CV and DL-based Medication Identification

Study	**Algorithms**	**Evaluation Metrics**	**Limitation**
(Heo et al., 2023)	YOLOv5, ResNet-32, RNN	Accuracy: 78%	Limited data for lowercase/symbols, reduced small-dataset performance, no deep learning segmentation.
(Nguyen et al., 2023)	Graph Transformer Network	mAP@50: 94.7	Missed detection for occluded pills, accuracy drop with incomplete graphs, high retraining cost for new classes.
(H. Lee & Youm, 2021)	R-CNN (Inception-V3)	Accuracy: 92.7%	Tested only in lab with limited pill types; needs real-world validation and added features.
(Ting et al., 2020)	YOLO v2	F1: 95.99% (back-side)	Limited to blister packs, struggles with similar pill shapes, long training times, full retraining for new drugs.

Continued on Next Page

Table 2.4 – Continued from Previous Page

Study	Algorithms	Evaluation Metrics	Limitation
(Spina et al., 2025)	YOLO v4	Accuracy: 99.2%	Single-center design, small sample size, no diverse population testing.
(Chiu, 2024)	CNN, Edge Detection Processing	Accuracy: 96%	Reduced segmentation accuracy when pill edges are close; needs more trials.
(D. Zhu et al., 2024)	YOLOv8	Precision: 95.1%, mAP@50: 87.4%	Misses detections in cluttered/low-light settings; needs better feature detection or auxiliary imaging.
(Vu et al., 2025)	NLP, PhoBERT, BiLSTM	Accuracy: 95%	Used controlled dataset; needs real-world testing and advanced models.
(Dhayanithi et al., 2024)	MobileNet	Accuracy: 98.01%	Needs real-world testing and broader validation before deployment.
(Pilania, Kumar & Singh, 2023)	CNN	Accuracy: 92.4%	Tested only under controlled conditions; may fail with real-world lighting/angles/backgrounds.

Continued on Next Page

Table 2.4 – Continued from Previous Page

Study	Algorithms	Evaluation Metrics	Limitation
(Dang et al., 2024)	YOLOv8	mAP@50: 99.5%	Needs larger pill database, better recognition for new pill types, improved usability.
(Suksawatchon et al., 2022)	YOLOv3, Mask R-CNN	F1: 99.5% (Mask R-CNN)	Image quality, shadows, and improper pill arrangement reduce accuracy; no color/imprint recognition.
(Nguyen, Nguyen, Pham, Nguyen & Nguyen, 2022)	CNN, GNN	F1: 83.15% (ResNet-34)	Depends on knowledge graph quality; noisy edges or missing modules reduce performance.
(Y.-S. Lin, Tsai, Chang, Shih & Cheng, 2020)	Faster RCNN (ResNet101)	Accuracy: 99.03%	Limited dataset, controlled backgrounds; may lack robustness in real-world use.
(Ou et al., 2020)	FPN + Inception-ResNet v2	Accuracy: 82.1%	Misidentification with visually similar pills (color/shape/imprints).

Continued on Next Page

Table 2.4 – Continued from Previous Page

Study	Algorithms	Evaluation Metrics	Limitation
(Verma et al., 2024)	EfficientNetB3	Accuracy: 99%	Minor misclassifications despite high overall accuracy.

2.7 Comparative Analysis of Medication Identification Technologies

We provide a comparative analysis of the three primary medication identification technologies: barcodes/QR codes, RFID/NFC, and CV. By evaluating these methods against a common set of criteria, we can identify their respective strengths, weaknesses, and suitability for different healthcare environments. The comparative analysis will focus on three key categories essential to medication administration:

- **Cost and Scalability:** The financial investment required for both initial setup and long-term maintenance, as well as the ease with which each system can be expanded to a larger scale.
- **Accuracy and Error Prevention:** The core ability of each technology to correctly identify medications and proactively reduce the risk of human error throughout the workflow.
- **Efficiency and Workflow Integration:** How each system streamlines clinical tasks and integrates seamlessly into existing hospital processes.

2.7.1 Cost and Scalability

Table 2.5: Comparison of Medication Identification Technologies: Cost and Scalability

Technology	**Cost**	**Scalability**
Manual Methods	Low; minimal infrastructure required, but high potential costs from human errors	High; suitable for any healthcare setting, universally applicable
Barcode/QR Codes	Low; inexpensive scanners and printing infrastructure	High; widely adopted, easily integrated across diverse clinical environments
RFID/NFC	Moderate to high; requires investment in tags, readers, backend systems, and maintenance	Moderate; practical for high-risk or high-volume settings, but less feasible for smaller facilities
CV	High; demands high-performance computing, cloud services, and large labeled datasets	Low; limited adoption due to technical requirements and specialized expertise, best suited for well-resourced institutions

This section provides a unified analysis of two key implementation factors: cost, which encompasses the total expenditure associated with implementing and maintaining a technology, including expenses for hardware, software, staff training, data management, and ongoing maintenance, and scalability, defined as a technology's ability to be effectively expanded across diverse healthcare settings, from small clinics to large hospitals. This combines evaluation highlights the fundamental trade-offs between initial investment and the potential for widespread adoption.

The manual method of medication identification is a low-cost and highly scalable option, as it does not require any technological infrastructure and is suitable for any healthcare setting. However, this method's reliance on human makes it vulnerable to errors caused by fatigue and cognitive bias, giving it the lowest cost-effectiveness due to the high potential for error-related expenses. Therefore, despite its universal applicability, the manual method requires technological assistance to function as a crucial safety net.

Barcode and QR code systems are widely recognized as the most cost-effective and scalable solution for medication identification. Their initial investment is minimal, requiring only inexpensive hardware such as scanners along with readily available printing infrastructure (T. A. Bowdle et al., 2019). This low-cost, standardized model facilitates seamless integration into existing clinical workflows across diverse settings, from pharmacies and wards to operating theatres. In surgical environments in particular, compact and affordable scanners make the technology both practical and economical. These advantages have contributed to its widespread adoption, with 93.7% of U.S. general and children's hospitals implementing barcode-based medication administration by 2015 (Pedersen et al., 2016).

In contrast to barcodes, RFID/NFC systems offer only moderate scalability due to their higher initial and ongoing costs, as well as more complex infrastructure requirements. The significant upfront investment in tags, readers, and backend systems,

along with additional expenses for installing readers at multiple checkpoints, integrating with hospital information systems, and maintaining compatibility, creates a high total cost of ownership. These technologies provide valuable advantages such as advanced automation, real-time tracking, and high accuracy, but their benefits are often only justifiable in high-risk or high-volume clinical settings. For smaller or resource-limited facilities, the need for ongoing maintenance, device calibration, and tag replenishment poses significant barriers compared to the low-cost, widely scalable barcode system.

CV systems currently exhibit the lowest scalability due to their substantial costs and demanding technical requirements. Deploying these solutions requires high-performance computing resources (e.g., GPUs and cloud infrastructure), reliable network connectivity, and significant investment in data curation. Large, high-quality labeled datasets are essential for training DL models not only make implementation expensive but also limit adoption beyond well-resourced institutions. While CV holds the potential to reduce long-term operational costs by minimizing errors and streamlining workflows once fully integrated, its reliance on advanced infrastructure and specialized expertise constrains scalability in smaller or resource-limited healthcare settings.

Overall, as summarized in Table 2.5, barcode/QR systems currently offer the best balance of cost and scalability, RFID/NFC provides moderate scalability at higher cost, and CV remains the least scalable option, though its promise may grow as infrastructure and affordability improve.

2.7.2 Accuracy and Error Prevention

Table 2.6: Comparison of Medication Identification Technologies: Accuracy and Error Prevention.

Technology	**Accuracy**	**Error Prevention**
Manual Methods	Variable; prone to human error due to fatigue, distraction, and cognitive bias	Low; no automated verification, high risk of administering wrong medication or dosage
Barcode/QR Codes	High when properly implemented; relies on scanning physical codes linked to databases	High; reduces transcription and documentation errors, but effectiveness depends on barcode integrity
RFID/NFC	High under optimal conditions; automated, non-line-of-sight verification	High; real-time verification, integrated in closed-loop systems, but can be limited by signal interference or tag damage
CV	High in ideal conditions; visual identification can catch errors other systems miss	High; can detect misplacements or look-alike medications, though still vulnerable to LASA medications, poor image quality, and dataset limitations

Accuracy in medication administration refers to a system's ability to correctly identify and verify medications, minimizing misidentification errors that could compromise patient safety. Error prevention, on the other hand, reflects a technology's capacity to proactively reduce human-based mistakes, such as administering the wrong medication or dosage. Together, these two metrics are critical for evaluating the reliability and clinical impact of medication identification systems.

Manual methods are inherently prone to error. While human expertise is vital, reliance on visual inspection and manual documentation leaves room for errors due to factors like fatigue, distraction, and cognitive bias, especially in high-pressure environments like operating theatres. This results in variable accuracy and a low potential for error prevention, as the absence of automated verification significantly increases the risk of administering the wrong medication or dosage.

Barcode/QR code systems enhance both accuracy and error prevention through standardized, machine-readable formats. When properly implemented, these systems provide reliable medication verification by linking scanned data to a central database, as seen with BCMA. This process significantly reduces the risk of human error in transcription and documentation. However, the system's effectiveness is contingent on physical integrity; damaged or obscured barcodes can lead to misreads, which can still result in medication errors.

RFID/NFC systems offer enhanced accuracy by providing automated, non-line-of-sight tracking. These technologies enable real-time verification and can be integrated into closed-loop systems to significantly reduce medication errors. Their high accuracy under optimal conditions is a significant strength. Nonetheless, their application can be limited by potential signal interference with medical equipment, a risk of infection from tags, or physical damage during procedures.

CV systems show the highest potential for both accuracy and error prevention. CV models have demonstrated over 99% accuracy in ideal conditions, performing visual

identification that can catch errors other systems would miss. For example, a CV system can detect that a pill is incorrect even if it's placed in the right package. However, this high accuracy is not absolute. CV models are still vulnerable to challenges from LASA medications, poor image quality, and dataset limitations.

In conclusion, as seen from Table. 2.6, manual methods are highly susceptible to human error, resulting in variable accuracy and low error prevention. In contrast, barcode and RFID technologies provide a significant leap in reliability by automating verification, achieving high accuracy and error prevention in controlled conditions. However, CV systems offer the highest potential, leveraging visual intelligence to perform robust identification that can detect errors other systems would miss. Despite this, all technologies have limitations, underscoring the need for a multi-layered approach that combines automated verification with human oversight.

2.7.3 Efficiency and Workflow Integration

Table 2.7: Comparison of Medication Identification Technologies: Efficiency and Workflow Integration.

Technology	**Workflow Method**	**Efficiency & Integration**
Manual Processes	Relies on visual checks, manual transcription, and paper documentation.	Low Efficiency. Time-consuming and prone to human error, creating a fragmented and labor-intensive workflow.
Barcode/QR Codes	Uses a scan-based system (BCMA) for data capture and verification.	High Efficiency. Streamlines medication administration by automating data entry, reducing charting errors, and saving time.
RFID/NFC	Provides automated, non-line-of-sight identification of multiple items simultaneously.	Very High Efficiency. Eliminates the need for individual scanning, enabling faster processes and real-time inventory management.
CV	Uses passive, continuous visual monitoring of medications without manual input.	Highest Potential. Offers an uninterrupted, intelligent safety check that integrates seamlessly, allowing clinicians to focus entirely on the patient.

We compare how each technology streamlines medical identification tasks and integrates seamlessly into existing hospital processes.

Manual processes are inherently inefficient and labor-intensive. They rely on visual checks, manual transcription, and paper-based documentation, which are time-consuming and prone to errors. This lack of integration with digital health records creates a fragmented workflow, slowing down the medication administration process, especially in high-volume or emergency situations like operating theatres.

Barcode technologies significantly improve efficiency by automating data capture and verification. For instance, BCMA systems streamline workflow by enabling nurses and pharmacists to quickly scan a patient's wristband and the medication, thereby linking the medication to the patient's digital health record (Parthasarathi et al., 2021). QR codes extend this functionality by enabling verification through either a smartphone or a dedicated scanner (Ti et al., 2020). Together, these approaches reduce manual entry, minimize charting errors, and save valuable time.

RFID/NFC technologies further advance workflow integration by eliminating the need for line-of-sight scanning. They enable real-time medication tracking and automated verification, reducing administrative burden and ensuring timely delivery of the correct medications. Unlike barcode systems, which are limited to one-way identification, NFC also supports read–write operations on a single mobile device, facilitating computerized tracking of patient data, prescriptions, and administration tasks. These capabilities make RFID/NFC more versatile and better suited for enhancing both patient safety and operational efficiency.

CV offers a highly efficient and seamless solution for medication identification by eliminating the need for physical RFID tags or barcode/QR code, which can slow workflows and require standardized labeling. Instead, CV systems can directly analyze medication containers or images in real time, identifying medications based on

visual features such as shape, color, and imprints. This enables faster, hands-free verification, and supports parallel processing of multiple medications in a single frame. By integrating into existing digital systems, CV enhances clinical workflows through automated documentation, reduces manual scanning steps, and minimizes medication errors, ultimately improving both safety and efficiency in healthcare environments.

In summary, manual methods are labor-intensive and error-prone, while barcode/QR code systems automate verification and integrate with digital records. RFID/NFC adds real-time tracking and bidirectional data capabilities, further reducing workload. CV offers the most flexible, hands-free, real-time identification of multiple medications, even under challenging conditions. Table 2.7 summarizes the strengths and limitations of these technologies for efficiency and workflow integration.

2.8 Research Gaps in Medication Identification Technologies

Despite the significant advancements in medication identification technologies such as barcodes/QR codes, RFID/NFC, and CV, several critical research gaps remain. These gaps highlight the need for further exploration to enhance the accuracy, usability, and reliability of these systems in real-world healthcare environments.

2.8.1 Limited Research about Hybrid Technologies

Current medication identification systems often operate in isolation, relying on specific technologies like barcodes, RFID, or CV. However, these methods come with advantages and drawbacks, there is limited research on integrating these modalities to create hybrid systems that combine their strengths. For example, the robustness of CV in recognizing diverse packaging could complement the reliability of RFID in environments with

limited visibility. Exploring multi-modal frameworks that seamlessly integrate these technologies could significantly improve performance and adaptability.

2.8.2 Challenges in Real-World Implementation

While many studies have demonstrated the efficiency of these technologies in controlled environments, their performance in real-world scenarios remains underexplored. Factors such as poor lighting, occluded or damaged labels, varying packaging designs, cylindrical bottle shapes that distort text, and environmental interference (e.g., electromagnetic noise affecting RFID) present challenges that need to be addressed. Additionally, the accuracy of some DL methods, such as OCR systems often degrades when processing text distorted by curved surfaces or poor print quality, further complicating medication identification. Future research should focus on developing systems that can robustly handle these practical limitations, including improving OCR algorithms to address text distortions and irregular packaging.

2.8.3 Lack of Real-Time Processing and Scalability

Real-time identification is crucial in dynamic settings like operating theatres and emergency rooms. Technologies such as CV, which rely on computationally intensive processes, may experience delays in processing. Additionally, ensuring scalability to manage large datasets, such as hospital-wide medication inventories, presents a significant challenge. Addressing these issues will require the development of lightweight algorithms, optimization of hardware utilization, and the implementation of distributed processing methods.

2.8.4 Cost and Resource Barriers

The implementation of advanced technologies, particularly RFID and CV, can be cost-prohibitive for smaller healthcare facilities or low-resource settings. Studies often neglect to address these barriers or propose cost-effective solutions. Research on affordable and scalable systems, such as low-cost RFID tags or edge computing devices for vision models, is essential to ensure widespread adoption.

2.9 Conclusion

This chapter provided a comprehensive review of existing literature on medication identification technologies, covering a range of approaches including barcodes/QR code, RFID/NFC, and CV. The discussion highlighted the principles, benefits, and limitations of each method, alongside their applications in medication identification. Barcode/QR code systems emerged as cost-effective and widely adopted solutions, while RFID/NFC offered advanced capabilities such as contactless scanning and enhanced data storage but faced challenges like cost and electromagnetic interference. CV methods, with their ability to analyze visual data in complex scenarios, showcased significant promise, yet encountered challenges in real-world implementations due to computational demands and environmental constraints.

A comparative analysis emphasized the unique strengths and weaknesses of each technology, demonstrating their suitability for various use cases. The chapter also identified critical research gaps, such as limited exploration of hybrid systems, challenges in real-world deployments, the need for real-time scalability, and barriers related to cost and resources. Addressing these gaps will require multidisciplinary efforts to develop robust, scalable, and cost-effective solutions capable of meeting the demands of dynamic healthcare environments.

This review serves as a foundation for future research aimed at advancing medication identification technologies, ensuring accuracy, reliability, and safety in diverse and practical healthcare settings.

Chapter 3

Research Methods

Achieving high-accuracy identification on cylindrical medication containers is fundamentally challenged by label distortion, occlusion, and the limited perspective provided by a single image. To overcome these constraints, this thesis presents Computer Vision (CV)-based dual-camera approaches designed to capture more comprehensive visual information from medication labels. By capturing two images from different perspectives of the container, the system was developed through a comparative evaluation of two label reconstruction methods: (1) an image stitching approach, which computationally segments, unwraps, and merges raw visual data before applying Optical Character Recognition (OCR); and (2) a text stitching approach, which prioritizes efficiency by performing OCR on the captured images first and subsequently aligning the resulting text fragments.

Additionally, to enhance clinical safety beyond simple name recognition, a robust and enriched drug database was constructed using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to integrate and verify information against authoritative sources. The methodology concludes in the design and development of a Proof-of-Concept (POC) application that provides a highly accurate, efficient, and validated solution for real-time medication verification.

3.1 Introduction

3.1.1 Objectives and Scope

The primary objective of this research is to design, implement, and evaluate a highly accurate and reliable CV-based system for medication identification within operating theatres. The proposed system aims to enhance patient safety and minimize the risk of medication errors resulting from drug misidentification. By ensuring precise medication recognition, supporting timely clinical decision-making, and strengthening safety in critical environments such as operating theatres and anaesthesia, the system ultimately contributes to improved patient outcomes and more efficient medication management/identification. The scope of this work encompasses:

- Developing and comparing two distinct cylindrical label reconstruction approaches: one based on visual label reconstruction from dual images (Image Stitching) and the other on textual label reconstruction using combined OCR outputs (Text Stitching).
- Creating a robust and verifiable medication database for information retrieval using advanced LLM and RAG techniques to enhance clinical safety.
- Integrating these components into a functional POC designed for real-time identification of cylindrical medication labels in high-stakes clinical environments such as operating theatres. The acceptance threshold is $> 97\%$ accuracy, a performance benchmark established in consultation with our industrial partner, SAFERSleep (SAFERsleep Ltd., n.d.), to ensure reliable medication identification.



Figure 3.1: Examples of Cefazolin AFT and Bacteriostatic Water for Injections, showing that a single image does not capture the complete label information.

3.1.2 Research Problem

Automated identification systems are hampered by significant technical challenges inherent to medication containers and the rapid decision-making required in anaesthesia. In particular, many medication containers used in operating theatres have cylindrical shapes, which introduce geometric distortion when captured by cameras. This curvature causes text stretching and compression across the label surface, especially toward the edges of the image, making reliable OCR more difficult.

Another major challenge arises from label occlusion and limited field of view when images are captured from a single viewpoint. Because medication labels wrap around cylindrical containers such as vials and ampoules, a single image often captures only a portion of the label. As a result, text required for accurate medication identification may be partially visible or entirely missing. For example, Fig. 3.1 illustrates images captured from two different medication containers, demonstrating that a single image frequently provides only incomplete textual information from the label.

In addition to these geometric and visibility challenges, the clinical environment imposes strict performance requirements. Operating theatres involve rapid decision-making and time-sensitive workflows, meaning that any automated identification solution must achieve high recognition accuracy while operating at sufficient speed to integrate seamlessly into clinical practice. These constraints make the design of reliable medication identification systems particularly challenging in anaesthesia settings.

3.1.3 Research Approach

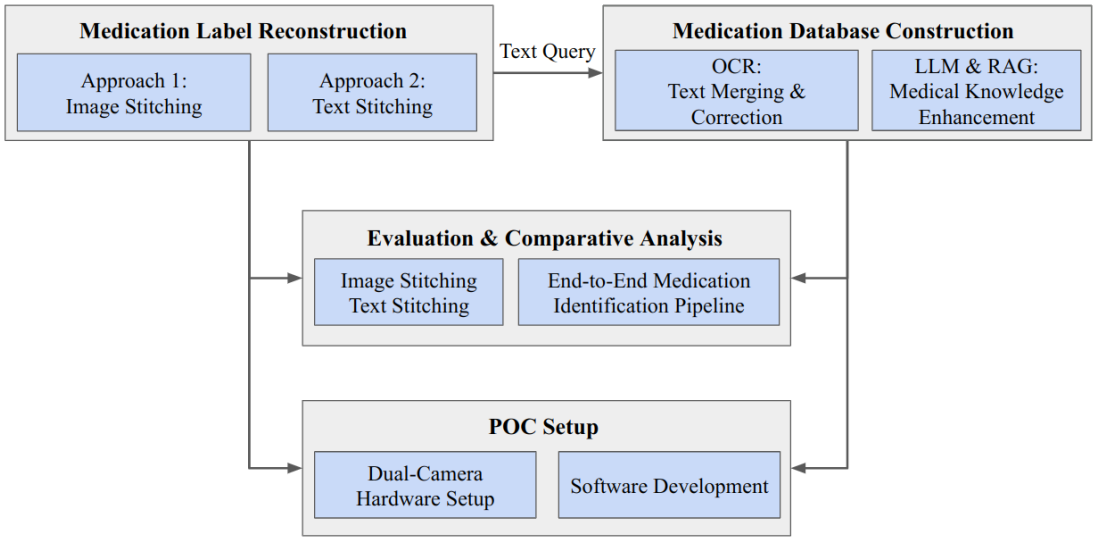


Figure 3.2: System architecture of the proposed dual-camera POC and its integration with label reconstruction, database enrichment, and evaluation modules.

We adopted an iterative, multi-stage approach to address the challenges of accurate medication identification. Each stage was designed to ensure both technical feasibility and practical relevance in operating theatre environments. The overall methodology integrates CV and OCR techniques with knowledge-based data enrichment to achieve high reliability and clinical usability. The framework consists of four main stages: Medication Label Reconstruction, Medication Database Construction, Evaluation and Comparative Analysis, and POC System Setup, as illustrated in Fig. 3.2.

Medication Label Reconstruction This stage focuses on recovering more complete label information from images captured by a dual-camera setup, as cylindrical medication containers often cause label distortion and partial visibility from a single viewpoint. To address these limitations, two complementary reconstruction strategies were developed. The first approach, *Image Stitching*, reconstructs a larger visual representation of the label by geometrically aligning and merging images from multiple viewpoints, which is effective when sufficient visual overlap exists between the captured images.

The second approach, *Text Stitching*, operates directly on OCR outputs by aligning and merging textual fragments extracted from different views, potentially processes faster and allowing label information to be reconstructed even when geometric alignment is unreliable.

The two methods were designed to address different failure modes. Image Stitching may degrade when feature matching fails due to limited overlap, motion blur, or reflections on the label surface, whereas Text Stitching may be affected by OCR errors or fragmented text recognition. A detailed description of both methods and their comparative analysis is provided in Chapters 4 and 5.

Medication Database Construction Conducted in parallel with the label reconstruction process, this stage focuses on building a structured and queryable medication knowledge base to support reliable medication identification. At the early stage of this research, a dataset of medication labels was constructed by collecting images of anaesthesia drugs. OCR was applied to extract textual information from the images, and the resulting text fragments from multiple views were merged to reconstruct more complete label content. The extracted text was then manually verified and consolidated to create an initial set of medication entries, including key attributes such as drug names, dosage units, and expiration dates. These verified entries formed the foundation of a structured medication database.

In the second stage, LLMs and RAG techniques were employed to reduce manual effort by correcting OCR merging results and to enrich the database using authoritative pharmaceutical documentation. Relevant information was retrieved from trusted medical references and regulatory sources to complement and verify the extracted label data. This process enabled the addition of contextual metadata and improved the consistency and completeness of the medication entries. The resulting database serves as the reference source for medication identification and is deployed through a web-based interface that supports user verification, manual updates, and contextual information

retrieval.

Evaluation and Comparative Analysis The proposed reconstruction methods were systematically evaluated to assess their effectiveness and practical suitability in operating theatre. In particular, the Image Stitching and Text Stitching approaches were compared under different hardware configurations to analyse trade-offs between recognition accuracy, processing speed, and system requirements.

The evaluation was conducted at multiple levels of the pipeline. Image-based components were assessed using standard CV metrics, such as mean Average Precision (mAP), to measure detection and segmentation performance. The quality of the OCR and text reconstruction processes was evaluated using text-based metrics, including character-level similarity and text recognition accuracy. Finally, the end-to-end medication identification pipeline was evaluated using identification accuracy and computational performance metrics, such as processing time, to determine the feasibility of real-time deployment. A detailed description of the evaluation methodology and experimental results is provided in Chapter 7.

POC Setup A functional POC system was set up using a dual-camera configuration, integrating image acquisition, label reconstruction, OCR, and the medication database into a single application. This setup enabled end-to-end testing and real-time validation of the proposed approach under conditions representative of an operating theatre.

3.2 Research Design

3.2.1 Methodology Overview

Fig. 3.3 illustrates our overall method for developing a medication identification system. Due to the challenges posed by cylindrical medication containers, which negatively impact OCR performance, this research initially explored combining two images to

reconstruct a more complete medication label. However, after conducting experiments to assess accuracy and speed of this method, consulting with SAFERsleep (SAFERsleep Ltd., n.d.), and considering the high-stakes nature of anesthesia and operating theater settings, we decided to explore a new OCR-based solution that prioritizes reliability and robustness. Details of the comparison of these two approaches are provided in Chapter 7.

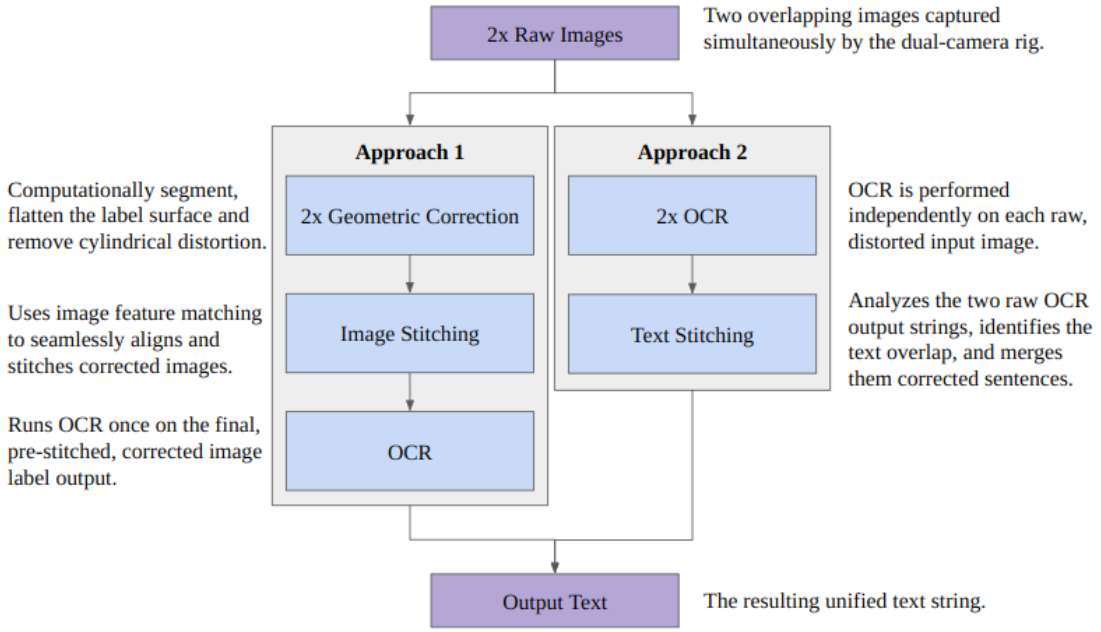


Figure 3.3: Flow chart of the two main label reconstruction pipelines proposed in this study.

Phase 1: Image Stitching (Approach 1) The details of this high-fidelity reconstruction technique are illustrated in Fig. 3.3 (Approach 1). This phase involves computational pre-processing steps, including medication label segmentation, ellipse detection, geometric unwrapping to correct for cylindrical distortion, and registration/stitching of the corrected label images, followed by OCR applied to the resultant flat image. A more detailed description of the implementation and evaluation is provided in Chapter 4.

Image Stitching is particularly well suited to medication identification in anaesthesia and operating theatre settings, where containers vary widely in shape and curvature,

complicating text alignment and readability (Gromova & Elangovan, 2022). By generating a geometrically accurate reconstruction, this method provides a consistent visual reference that enhances interpretability and supports the robustness of the identification system.

Phase 2: Text Stitching (Approach 2) The methodological divergence between image-based and text-based fusion is illustrated in Fig. 3.3 (Approach 2). While Phase 1 aims for high-fidelity label reconstruction, its reliance on geometric correction, image stitching, and multiple supporting models introduces higher computational cost, potentially limiting suitability for real-time or resource-constrained deployment. These constraints motivated the exploration of a more streamlined alternative.

Phase 2 therefore adopts a text-centric strategy, shifting multi-view fusion from the image domain to the text domain. OCR is applied directly to the raw dual-camera inputs without explicit geometric reconstruction, and the resulting text fragments are aligned and merged using a text stitching algorithm to form a more complete label representation. By eliminating the need for label segmentation, ellipse detection, and image registration, this approach substantially reduces computational complexity while maintaining reliable label reconstruction.

This phase investigates the feasibility of near real-time medication identification in anaesthesia environments, where both speed and reliability are critical. A more detailed description of the algorithm and its evaluation is provided in Chapter 5, with particular emphasis on the trade-offs between processing efficiency and final medication identification accuracy, addressing the computational overhead observed in Phase 1.

Phase 3 (Chapter 6): Database Enrichment Accurate medication identification relies not only on extracting text from labels but also on having a reliable reference database against which the text can be matched. Without a comprehensive and structured database, even perfectly recognized text may lead to misidentification due to missing entries, inconsistencies, or variations in label formats.

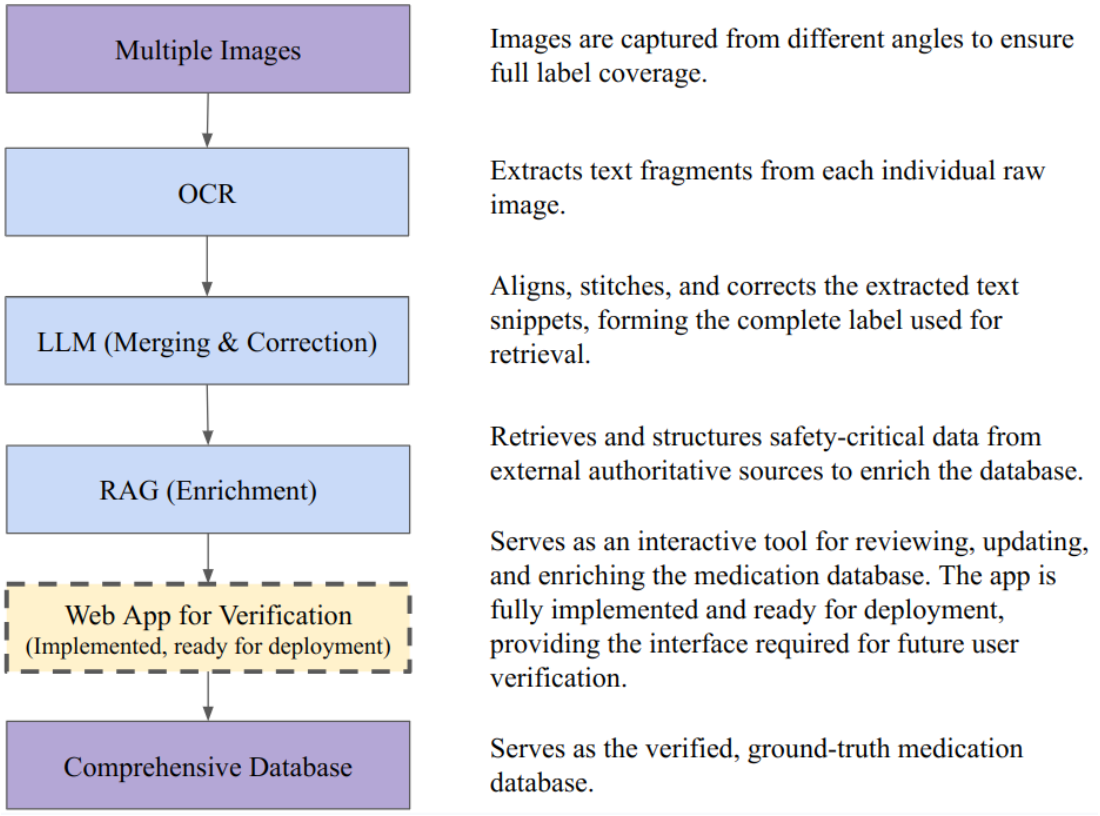


Figure 3.4: The process of enriching the medication database, from multiple medication images to a comprehensive, verified knowledge base.

An initial medication database was constructed to support the identification process by collecting medication label images and extracting their textual information using OCR. All extracted label information was manually verified and stored in a structured database to serve as the reference for querying. During medication identification, the OCR outputs generated from the Image Stitching or Text Stitching approaches are used as query inputs and matched against this database to determine the corresponding medication.

Beyond the baseline system, the medication database is further enhanced using LLMs. LLMs are not used directly in the real-time identification pipeline due to latency and reliability constraints in operating theatre environments. Instead, they are applied offline to refine OCR outputs and improve the quality of the stored database entries.

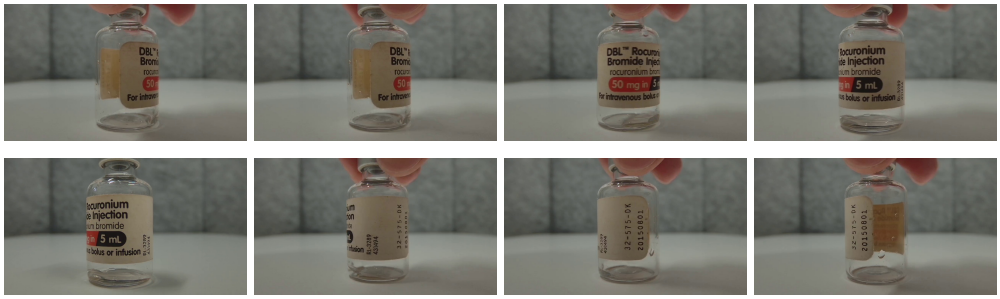


Figure 3.5: Example images were taken to capture the full label of the vial *DBL Rocuronium Bromide Injection*.

Additionally, while the initial database is constructed from information extracted directly from medication labels, these labels typically contain only essential identification details. Additional clinically relevant information such as side effects, contraindications, and storage instructions may not appear on the physical label. The retrieval component and local medication database are maintained and queried entirely offline. The LLM generation step (GPT-4o mini, accessed via AP) requires network connectivity and is therefore not part of the offline operation; it is used only during construction-time database enrichment, not during real-time identification.

Fig. 3.4 illustrates this process. First, to ensure full label coverage, multiple images of each medication are captured (e.g., Fig. 3.5), from which text fragments are extracted using OCR. These fragments are then aligned, stitched, and refined with the assistance of a LLM to produce a coherent label representation. RAG is subsequently applied to query external authoritative sources and enrich the database with additional safety-critical information, including warnings, dosing guidance, and full chemical names. By leveraging verified regulatory sources and LLM-based text enrichment, this approach aims to produce an accurate and comprehensive medication database that supports safer and more informed identification in clinical environments

The enriched database is delivered through a web application designed for healthcare professionals, enabling user verification, manual updates, and the addition of new entries. Although the system has been fully developed and is ready for deployment, it has not

yet been used in a clinical setting. As part of this research, the application demonstrates the potential for creating a comprehensive, verified ground-truth medication database that can improve both identification accuracy and clinical safety.

3.2.2 Overall System Architecture

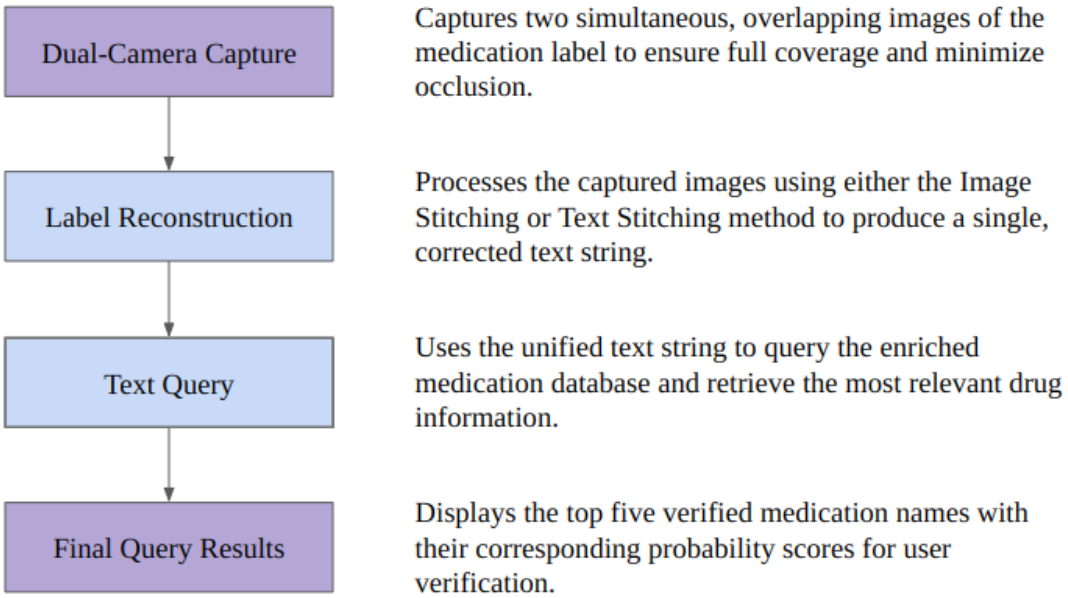


Figure 3.6: The overall process of medication identification, from dual-image capture to final medication ID.

The overall system architecture is illustrated in Fig. 3.6, outlining the sequential data processing pipeline from raw image capture to the identified medication. The process begins with dual-camera acquisition to capture complementary views of the medication label, followed by reconstruction through either the Image Stitching (Chapter 4) or Text Stitching (Chapter 5) method to obtain a unified text output. This text is then used to query the LLM/RAG-enriched medication database (Chapter 6), which retrieves and verifies the most relevant drug information. Finally, the system presents the top-ranked identification results to the user, completing an end-to-end workflow for medication verification.

Rather than returning a single detected medication ID, the system outputs a ranked list of top-k candidates, each accompanied by a confidence score. This design accounts for possible OCR inaccuracies, label occlusions, or incomplete text capture that could otherwise lead to misidentification. By presenting multiple high-probability matches, the system enables healthcare professionals to verify the correct medication manually, significantly reducing the risk of false positives in critical environments such as operating theatres. Additionally, the ranked output provides transparency into the system's confidence and supports ongoing performance evaluation and refinement of the identification pipeline.

3.3 System Setup and Data Acquisition

3.3.1 Dual-Camera System

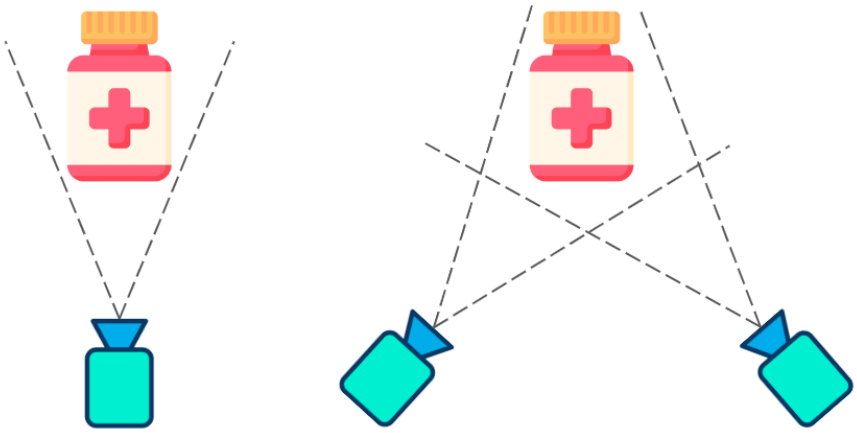
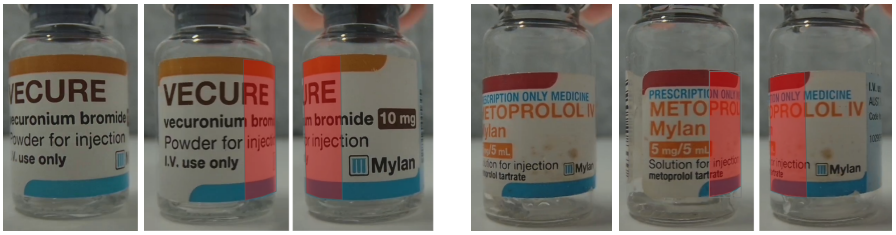


Figure 3.7: Comparison between single-camera and dual-camera setups.

We adopted a dual-camera system to address the limitations of single-view imaging for cylindrical medication containers, as shown in Fig. 3.7. By capturing two overlapping perspectives, the setup expands the field of view and compensates for label curvature, thereby mitigating the risk of missing critical information due to occlusion.



(a) (b) (c) (d)

Figure 3.8: Examples of single- and dual-camera captures:
(a) Single Vecure Vial, (b) Dual Vecure Vials with overlapping views,
(c) Single Metoprolol IV, and (d) Dual Metoprolol IV images with overlapping views.

This configuration ensures more comprehensive coverage of the medication label, enabling the extraction of richer information and limiting the chance of omitting important details.

The dual-camera configuration was selected as the minimum viewpoint count required to test whether a second viewing angle recovers label text that is occluded or undetectable from a single view, while keeping system cost and hardware complexity low. A higher camera count could further improve coverage of highly curved or fully wrapped labels but was considered out of scope for this phase of the research; this trade-off is discussed as future work in Section 9.4.3.

The camera separation angle (typically 30-45°) and crop margin (5-10%) were set based on practical configuration during system setup rather than derived from a dedicated calibration study. These values were chosen to ensure sufficient overlap between the two camera views while maintaining adequate coverage of the full label area; however, no systematic sweep of separation angle or margin against downstream accuracy was performed. This is identified as a limitation (Section 9.3), and a calibration study is proposed as future work (Section 9.4.3).

As illustrated in Fig. 3.8, neither the single- nor dual-camera captures can individually provide complete information about the medication labels due to curvature and occlusion. However, by combining the overlapping regions from both camera views,

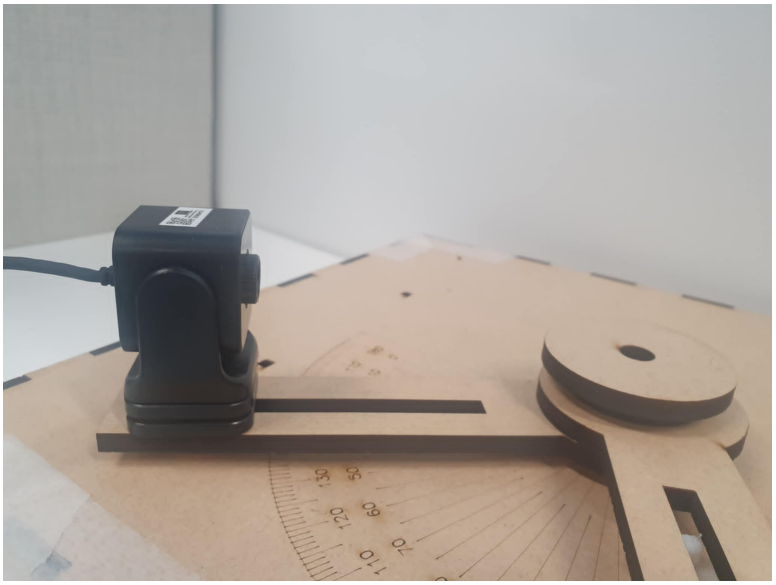


Figure 3.9: The camera mount allows the camera to move around and adjust its distance to the medication container.

more comprehensive and reliable label information can be reconstructed, enhancing the accuracy of medication identification.

The dual-camera approach was selected as a practical compromise between information completeness and system complexity. Compared to more elaborate multi-camera configurations, it maintains a relatively simple design that enables synchronized image capture and reliable text extraction without imposing excessive computational demands. This setup supports near real-time operation in anaesthesia/operating theatre, using standard hardware resources. Furthermore, its moderate hardware requirements make the system more cost-effective and accessible for healthcare settings with limited budgets, while still providing sufficient visual coverage for accurate medication identification.

Specifically, the camera physical configuration used in this research consists of two key components. The system utilizes two USB Creality Nebula Cameras, each providing a resolution of 1920×1080 and equipped with a manually adjustable focal length. This configuration ensures consistent image quality across captures and provides flexibility for fine-tuning the focus based on the container's size and distance. The cameras

are mounted on a manually adjustable dual-camera fixture using custom-fabricated mounts that allow precise adjustment of the relative angle between the two cameras (see Fig. 3.9). This systematic adjustability was critical during the data collection and calibration phases to establish optimal viewing overlap for both the Image Stitching (Approach 1) and Text Stitching (Approach 2) methodologies.

3.3.2 Experimental Platforms

To manage the distinct computational demands of the research, two primary platforms were utilized:

1. **High-Performance Training Platform**: Provided by the Auckland University of Technology, this platform is equipped with an NVIDIA RTX 2080 Ti GPU and was used for CV/OCR model training, complex image stitching experiments, and running resource-intensive LLM models.

2. **Deployment and Testing Platform**: A conventional laptop with an Intel Core i5 processor and 8GB of RAM, used for final efficiency evaluation and application testing, simulating the low-resource environment.

3.3.3 Software Architecture and Implementation

The software architecture was designed to ensure efficient, synchronized data acquisition and reliable medication identification. It was implemented with a focus on modularity, flexibility, and real-time performance, Details of the architecture and implementation are provided in Chapter 8. The main components are outlined below:

Programming Environment: Developed using the Python ecosystem for its extensive support of CV and DL frameworks. Key libraries include OpenCV for image acquisition and processing, PyTorch for model inference, and PyQt for graphical user interface development.

System Design: Implemented a multi-threaded architecture to handle high-speed, synchronized input from the dual-camera setup. The design ensures continuous data flow from physical cameras to computational pipelines with minimal latency. Additional flexibility is provided through adjustable processing frequency (e.g., skip-frame control) to balance speed and accuracy according to operational needs.

Core Modules: The system consists of three core modules. The OCR Engine is a deep learning based text recognition module for accurate extraction of text from dual images and supports model selection, allowing users to switch between different OCR models for testing or performance optimization. The Database Lookup module performs offline searches in the medication database to retrieve medication identifiers and key attributes from the recognized text. Finally, the User Interaction Module provides real-time visual feedback on the identified medication and allows users to configure operational parameters, such as model type and skip-frame settings.

Operational Modes: Designed as a compact, self-contained application with no dependency on the internet or third-party services (e.g., Google OCR). The system supports fully offline operation, is optimized for real-time inference with low-latency display, and features a fully customizable design to suit different deployment requirements.

3.3.4 Data Collection and Annotation

The success of CV models critically depends on a comprehensive and high-quality dataset. To capture the variability present in real-world operating theatres, we adopted a systematic data collection and annotation protocol designed to maximize diversity while ensuring reproducibility and minimizing data leakage.

3.3.4.1 Drug Set Composition

The medication dataset was constructed from a diverse collection of 53 medication types commonly used in anaesthesia across New Zealand hospitals. These samples, comprising 83 individual drug containers, were provided by Auckland City Hospital (Healthpoint, n.d.) and SAFERsleep (SAFERsleep Ltd., n.d.). The dataset encompasses significant variations in label material, size, color, curvature, and text density, as illustrated in Fig. 3.10. To ensure robust model performance, each container was imaged from multiple perspectives. By capturing around 50 images per container, a total dataset of 4,122 images was generated. This structural diversity allows for a comprehensive evaluation of the system's resilience to challenging environmental factors, including glare, occlusion, and label curvature.

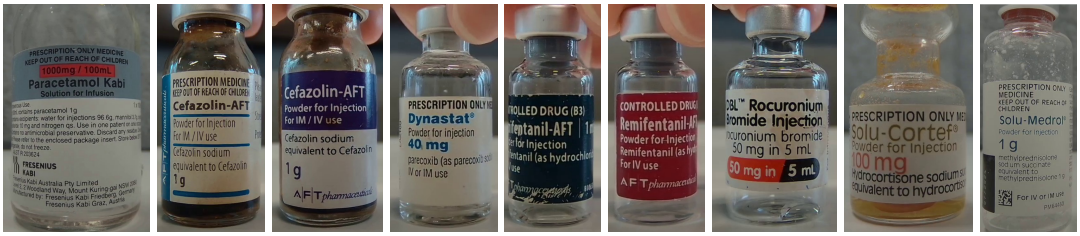


Figure 3.10: Examples of medication containers used in this research. The dataset reflects real clinical variability in label appearance, size, curvature, and text density.

3.3.4.2 Component Training Datasets

Figure 3.11 illustrates the image acquisition and annotation workflow. Images were captured using a dedicated camera setup under controlled variations. To account for different label curvatures and potential partial occlusions, containers were imaged from multiple camera positions.

Label Segmentation Dataset: A total of 4,122 images were annotated for the medication label segmentation task. The dataset was partitioned into training and validation sets using an 80/20 split. To ensure the integrity of the evaluation and prevent

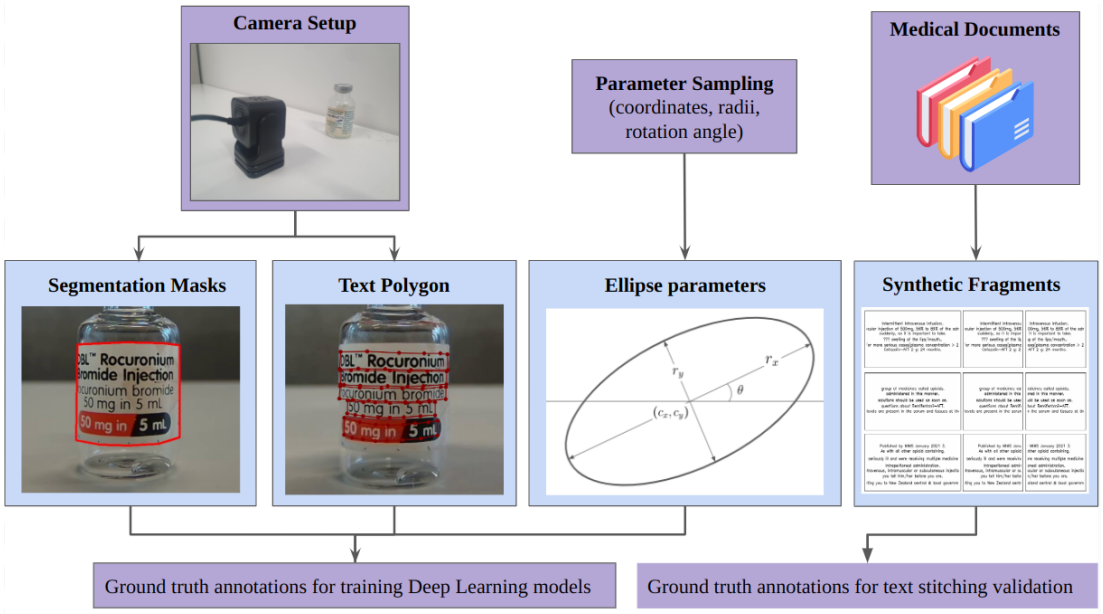


Figure 3.11: Illustration of the image capture and annotation workflow.

data leakage, the dataset was split at the drug level rather than the image or container level. This approach ensures that all containers of the same medication appear in only one split, avoiding situations where visually similar containers could inadvertently appear in both training and evaluation sets. This approach ensures that images of the same physical drug container do not appear in both the training and testing sets, thereby providing a more rigorous assessment of the system's ability to generalize to new, unseen medication samples.

Text Detection Dataset: For the text detection task, 700 images from the clinical drug dataset were manually annotated. The data was partitioned into 500 training and 200 validation images, with the split performed at the drug-level. Also, we combine this domain-specific clinical images with two public datasets - CTW1500 (Yuliang, Lianwen, Shuaitao & Sheng, 2017) and ICDAR2015 (Karatzas et al., 2015) with 1,500 samples each, to improve generalization across varied fonts. This combination ensures the model is resilient to the complex text layouts and distortions typical of curved medication labels.

Synthetic Ellipse Dataset: To facilitate geometric unwrapping in the image stitching process after obtaining the label segmented mask, a synthetic ellipse dataset was generated. Estimating the cylindrical shape of the container is critical, as it allows the system to mathematically reverse the perspective distortion of the medication container label. Each synthetic ellipse is parametrized by center coordinates, radii, and rotation angle. Cylindrical label masks are simulated by combining two half-ellipses representing the top and bottom projections. To ensure robustness against real-world segmentation imperfections, parameters and visible arcs are randomly varied, with added noise and distortions to mimic imperfect mask boundaries. We created 11,000 and split 10,000 for training and 1,000 for validation with different random parameters to improve generalization.

Text Stitching Development Dataset: Finally, to develop and evaluate the text stitching method, a synthetic dataset of 5,000 samples was generated to simulate dual-camera label capture. The data generation pipeline involved rendering high-fidelity text images sourced from medical documentation, which were then split into two overlapping halves to simulate the dual-camera perspectives. This approach simulates the dual-camera acquisition process, generating fragmented text samples that mimic the overlapping views of cylindrical medication labels. By providing a ground-truth reference, this synthetic dataset enables the systematic evaluation of the stitching algorithm's performance in reconstructing more complete label information from disparate source fragments.

3.3.4.3 End-to-End System Evaluation Dataset

In addition to component-specific training datasets, a dedicated end-to-end evaluation dataset is constructed to assess the complete medication identification pipeline under realistic conditions. The dataset is designed to ensure comprehensive coverage while maintaining strict separation from all training, tuning, and validation data.

The evaluation dataset consists of cylindrical medication containers captured using a dual-camera setup, providing two complementary viewpoints per container. This configuration supports realistic scenarios involving partial visibility, perspective variation, and label curvature.

The dataset includes 53 distinct medication types, with each type represented by 20 image pairs, resulting in a total of 1,060 samples. All classes are equally represented, ensuring no class imbalance in evaluation metrics. Each image pair is annotated with a corresponding DrugId as ground truth for end-to-end evaluation.

Importantly, the same physical medication containers are used as in earlier data collection; however, all evaluation images are independently acquired and are not reused from any training, tuning, or validation pipelines. This ensures that there is no overlap in image data across pipeline stages, even though the underlying objects are shared.

3.3.4.4 Dataset Summary

Table 3.1 and 3.2 provides a comprehensive overview of all datasets utilized in this research. The data pipeline is designed to address the specific challenges of medication identification through a combination of high-fidelity clinical images and specialized synthetic data, proceeding from component-level training through integrated end-to-end evaluation.

Table 3.1: Comprehensive summary of datasets used in the study, from component training through end-to-end evaluation.

Dataset	**Purpose**	**Type**	**Size**	**Leakage Prevention**
Label Segmentation	Training	Real-world	4,122	Drug-level split
Text Detection	Training	Hybrid	3,700	Drug-level split
Synthetic Ellipse	Training	Synthetic	11,000	Parameter variation
Text Stitching	Development	Synthetic	5,000	Algorithmic validation
End-to-End Evaluation	**Evaluation**	**Real-world**	**1,060**	**Distinct from train datasets**

The end-to-end evaluation dataset (1,060 samples) is completely distinct from

all component training datasets, ensuring that system performance metrics reported in Chapter 7 reflect genuine end-to-end capability under realistic clinical conditions rather than component overfitting. This separation is critical for validating the practical deployment readiness of the medication identification system in operating theatre environments.

Table 3.2: Dataset Overview

Dataset	**Purpose**	**Type**	**Size**	**Split**
Label Segmentation	Training (YOLOv8-s)	Real-world	4,122 images	80/20, drug-level split
Text Detection	Training	Hybrid	3,700 images	700 clinical (500 train/200 val) + 1,500 CTW1500 + 1,500 ICDAR2015
Synthetic Ellipse	Training (FCOS)	Synthetic	11,000 masks	10,000 train/1,000 val
Text Stitching Development	Algorithm dev	Synthetic	5,000 samples	Simulated dual-view split
OCR Fine-tuning	Training	Real-world	700 images	500 train/200 val, drug-level
Text Stitching Validation	Evaluation	Real-world	40 pairs	10 medication types × 4
End-to-End System Evaluation	Final evaluation	Real-world	1,060 samples	(53 types × 20) Fully held out, drug-level

3.4 Medication Database Foundation

3.4.1 Baseline Database Requirement

After extracting text from the medication label, the identification process continues by searching within the medication database. Specifically, a functional database containing key reference fields is essential for accurate medication matching. The core fields include drug name, manufacturer, and dosage information, which together form the baseline data required for reliable identification. These fields serve as the direct targets for text queries generated by the reconstruction pipelines (Image Stitching and Text Stitching).

3.4.2 Motivation for Database Enrichment

A simple medication name-matching approach is insufficient for reliable medication identification, as label variations, incomplete text, and ambiguous abbreviations can lead to errors. To improve accuracy, it is essential to utilize all available information from the medication label, which requires incorporating corresponding details into the database. However, manual entry is time-consuming and prone to human error.

To address this, we construct a more complete label representation by capturing multiple images of the same container and employing a LLM to stitch, refine, and correct the extracted text segments. This automated text reconstruction ensures comprehensive label coverage, reducing manual effort while enhancing the robustness of database entries. Furthermore, the database is enriched with contextual information from authoritative external references (e.g., indications, side effect, storage instructions) using a RAG approach. This enables cross-validation of OCR-derived queries, improving verification accuracy and reducing the risk of medication misidentification.

In clinical settings, medication databases are typically updated through controlled

and verifiable workflows to ensure reliability and patient safety. When a new medication or label variation is introduced, images are first captured and processed using the OCR-based extraction pipeline. The extracted text from multiple views is then merged and corrected by LLMs, followed by enrichment with additional information from authoritative sources using a combination of RAG and LLMs. The resulting entries are subsequently reviewed and verified by authorized personnel. Once validated, the medication can be safely added to the database and used for identification in operating theatre.

3.4.3 Query and Matching Logic

The final step of the identification process involves querying the enriched database using the reconstructed text string from the CV pipeline (Image Stitching and Text Stitching). The matching mechanism employed a fuzzy-matching algorithm (such as the Levenshtein distance or Jaccard similarity) to compare the potentially noisy OCR output against the ground-truth names in the database. This logic allows for robust identification even when the extracted text contains minor errors, returning a ranked list of possible drug IDs based on similarity score (Top-1, Top-5).

Medication matching between the stitched OCR text and the reference database is performed using Levenshtein edit distance, defined as the minimum number of single-character insertions, deletions, or substitutions required to transform one string into another. For two strings a and b, the normalized similarity score is computed as:

$$similarity(a, b) = 1 - (Levenshtein(a, b)/max(len(a), len(b))) \quad (3.1)$$

Candidate medications are ranked by descending similarity score, and the top-k candidates are returned as predictions. This function is used throughout the results analysis in 7.

3.5 Evaluation Protocol and Metrics

3.5.1 Experimental Setup and Testing Protocol

Shared Experimental Setup To ensure a fair and rigorous comparison between the two medication identification methods (Image Stitching and Text Stitching), a controlled testing protocol was established. Testing was conducted under controlled scenarios designed to replicate real-world conditions. These scenarios included predefined degrees of rotation and lighting variations to evaluate system performance under different viewing angles and illumination settings. This shared experimental setup ensured consistency across all tests and allowed for fair comparison between the Image Stitching and Text Stitching approaches.

The research employed two computational environments: the High-Performance Training Platform (PC with NVIDIA RTX 2080 Ti) was used for model training and computationally intensive development (Approach 1), while system evaluations and efficiency comparisons were benchmarked on the Deployment/Testing Platform (Conventional Laptop, i5, 8GB RAM), which served to simulate the real-time operational constraints of the target edge device. A comprehensive description of the experimental setup and methodology is provided in Chapter 7.

Approach 1 Setup (Image Stitching) This methodology is characterized by a sequential pipeline of interconnected models and processing steps designed to transform raw, cylindrical label captures into accurate, machine-readable data. The process begins with a deep learning model for label segmentation, followed by an ellipse detection algorithm that calculates the geometric parameters necessary for correction. Then CV techniques including perspective transformation and image registration are used to facilitate the unwrapping of the segmented label and the stitching of disparate fragments. The output of this pipeline is a single, flattened image, which is then processed by an

OCR model to extract the medication label information.

Approach 2 Setup (Text Stitching) This methodology was optimized for low-latency processing by prioritizing efficient, direct text extraction from raw dual-camera inputs. The approach relies on a high-accuracy, fine-tuned OCR model applied directly to the captured images to retrieve text fragments from each perspective. These outputs are subsequently processed by a custom text stitching algorithm, which leverages sequential overlap logic to intelligently merge the two OCR results into a single, comprehensive label representation. By avoiding complex geometric unwrapping steps, this method significantly reduces computational overhead while maintaining high reconstruction fidelity.

3.5.2 Intermediate Performance Metrics

These metrics assess the quality and accuracy of the label data recovered by the two reconstruction methods (Image Stitching and Text Stitching) before the final database query step.

Image Metrics These metrics quantify the precision of the initial CV steps essential for geometric correction in Approach 1. Specifically, we use Mean Average Precision (mAP) to evaluate the performance of both the segmentation model (for label boundary extraction) and the detection model (for ellipse parameter estimation). mAP provides a comprehensive measure by accounting for both precision and recall across different Intersection over Union (IoU) thresholds, allowing a balanced assessment of detection accuracy and completeness (Everingham, Van Gool, Williams, Winn & Zisserman, 2010; T.-Y. Lin et al., 2014). A high mAP indicates that the geometric correction module can reliably isolate and model the curved label region, which is critical for achieving accurate unwrapping and minimizing distortion in subsequent text extraction and stitching stages.

Text Metrics These metrics directly assess the quality and reliability of the OCR output and reconstructed text string, which are critical for accurate medication identification. Since the ultimate objective is to recover readable and semantically correct text from curved and partially occluded medication labels, evaluation at both the character and word levels is necessary. Character-level metrics capture fine-grained recognition errors introduced by OCR and geometric distortion, while word-level metrics better reflect clinically meaningful mistakes that may alter dosage or drug interpretation. Accordingly, the following metrics are employed, as they are widely adopted in OCR and document analysis literature (R. Smith, 2007; Resnik & Lin, 2010):

- Precision: Measures the proportion of correctly recognized text among all text identified by the system, indicating how well the OCR avoids false positives.
- Recall: Assesses the proportion of correctly recognized text among all text present in the ground truth, reflecting the OCR's ability to capture all relevant information.
- F1 Score: The harmonic mean of precision and recall, providing a balanced evaluation of recognition accuracy and completeness.
- Character Error Rate (CER): Evaluates errors at the character level (insertions, deletions, substitutions). It provides fine-grained insight into where small mistakes occur (e.g. a missed letter or mis-read digit).
- Word Error Rate (WER): Measure the error rates at the word level, which matters more from a clinical perspective: misrecognizing a word (e.g. "mg" vs "ml") can have serious consequences.

3.5.3 System Identification Metrics

These metrics evaluate the accuracy and efficiency of the complete end-to-end medication identification pipeline, including OCR processing and database retrieval. To

provide a comprehensive assessment, both identification accuracy and computational efficiency are measured. Together, these metrics reflect the system's ability to correctly identify medications while maintaining real-time responsiveness, including:

Top-1 Identification Accuracy: Measures the proportion of cases where the system's first (topmost) prediction matches the correct drug label. This metric evaluates the system's ability to make accurate, definitive predictions.

Top-5 Identification Accuracy: Reflects the proportion of cases where the correct drug label appears within the top 5 predictions provided by the system. This metric high- lights the system's capacity to narrow the search space effectively, offering useful alternatives even when the top prediction is incorrect.

Inference Latency: The total time from dual-camera image capture to the final identification result displayed to the user (crucial for real-time viability).

3.6 Conclusion

This chapter established a comprehensive research methodology designed to address the challenges of medication identification on cylindrical containers.The core of this approach is founded on the dual-camera hardware setup, which fundamentally addresses label occlusion and perspective distortion by capturing complementary views around the container.

Methodologically, this chapter introduces a comparative framework between two distinct pipelines designed to address complementary aspects of the medication identification problem. The first, Image Stitching (Approach 1), focuses on high-fidelity geometric reconstruction and visualization of curved labels, enabling detailed examination of text alignment and completeness. The second, Text Stitching (Approach 2), explores a lightweight, low-latency alternative aimed at achieving comparable interpretability with reduced computational demand. This iterative design allows systematic

evaluation of trade-offs between fidelity and efficiency, supporting the identification of techniques that may be feasible for deployment on devices with limited computational resources. Furthermore, the integration of the LLM and RAG framework extends the study beyond basic text recognition, introducing a knowledge-enrichment process to enhance contextual understanding and reliability of the final identification output.

Collectively, this methodology prioritizes not only high accuracy but also the efficiency, cost-effectiveness, and offline operability required for successful integration into anaesthesia/operating theatre environments, ultimately advancing patient safety by enhancing the reliability of medication recognition systems.

From Methodology to Execution

Chapter 3 established the overall research methodology, including the dual-camera imaging concept and the design of two cylindrical label reconstruction pipelines. It defined the Image Stitching approach (Phase 1) as a method designed to reconstruct label geometry through geometric correction and dual-view image fusion.

The following chapter, Chapter 4, translates this conceptual framework into execution. It presents the implementation of the Image Stitching method and its capability to generate label reconstructions of cylindrical medication containers.

Chapter 4

Approach 1: Visual Reconstruction of Medication Labels via Image Stitching

This chapter presents Approach 1: Image Stitching Followed by OCR, one of two candidate approaches to cylindrical label reconstruction investigated in this thesis. Approach 1 and the approach presented in Chapter 5 are developed and evaluated as alternatives to the same problem, not as sequential stages of a single pipeline. Their comparative performance is evaluated in Chapter 7, and only one approach is carried forward into the final prototype described in Chapter 8.

Accurately capturing and processing labels on cylindrical objects presents significant challenges due to perspective distortion and the limited visibility of the label from a single viewpoint. In this chapter, we propose a dual-view, ellipse-guided reconstruction pipeline for cylindrical medication labels, designed to recover a more complete label representation from two simultaneous, overlappings views.

The proposed method integrates label segmentation, ellipse-based boundary modeling, mesh-driven cylindrical unwrapping, and dual-view image stitching into a single reconstruction pipeline for cylindrical medication containers. Following segmentation, an ellipse detection model is used to localize the top and bottom boundaries of the

cylindrical label. Subsequently, geometric modeling is applied for distortion correction, transforming the curved label into a flattened visual representation while preserving label details. Finally, the unwrapped views are aligned and merged to produce a single reconstructed label image suitable for downstream Optical Character Recognition (OCR).

We demonstrate the approach on a diverse set of cylindrical bottles with varying types, colors, and sizes. Experimental results demonstrate that the pipeline achieves high accuracy scores for label segmentation and ellipse detection on the datasets, and the reconstructed labels exhibit completeness and visual quality suitable for OCR analysis.

4.1 Introduction

The primary objective of this chapter is to develop and evaluate a geometric reconstruction pipeline that improves label visibility prior to OCR-based analysis in operating theatre and anaesthesia settings. In such cases, a single camera perspective often results in cylindrical distortion and partial occlusion of label information, limiting text extraction by standard OCR systems (Z. Ye, Yi & Tian, 2013; Gromova & Elangovan, 2022).

Due to the inherent challenges of curvature-induced distortion discussed in Sections 2.8 and 3.1.2, this chapter proposes a specialized approach for cylindrical label reconstruction. Specifically, the proposed method operates on a dual camera setup, designed to address the limitations of single-view imaging. The system captures two overlapping segments of the medication container, which are then rectified and merged through an image stitching process. This integration aims to produce more complete label information for downstream OCR analysis.

As illustrated in Fig. 4.1, the workflow sequentially performs several computer vision (CV) tasks:

1. Deep learning (DL)-based segmentation to isolate the label region from the medication container.

2. Ellipse estimation to define the label's top and bottom boundaries to model the curvature. Two ellipse estimation strategies were developed and compared: a geometry-driven contour-based method and a DL-based ellipse detector, in order to examine the trade-off between interpretability, robustness, and reconstruction accuracy.

3. Geometric unwrapping for distortion correction, converting the curved surface into a flattened label representation.

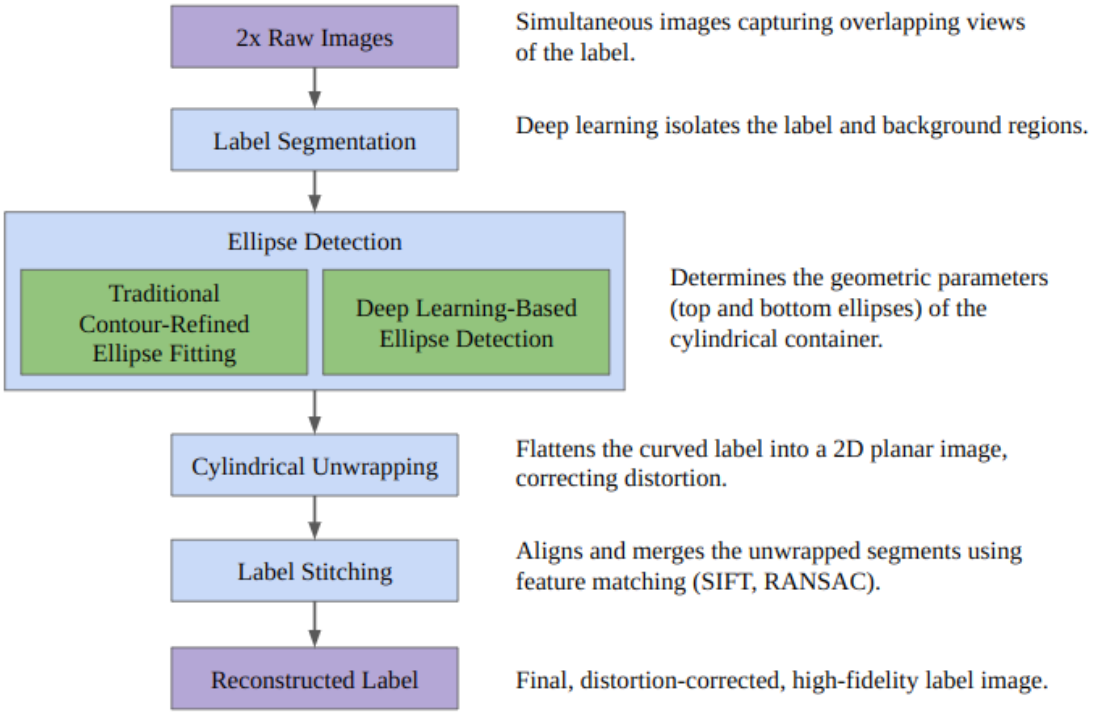


Figure 4.1: Flowchart of the proposed label reconstruction pipeline.

4. Feature-based image stitching to merge the unwrapped labels into a single label reconstruction.

4.2 Related Work

4.2.1 Cylindrical Label Reconstruction Under Partial Visibility and Distortion

The core challenge in medication label reconstruction involves the compound effects of cylindrical projection geometry, partial visibility, and specular reflections. These factors are pronounced in operating theatre environments, where imaging conditions are uncontrolled and time-constrained.

4.2.1.1 Geometry-Based Approaches and Deployment Barriers

Early approaches to cylindrical label reconstruction rely heavily on explicit geometric modeling and multi-camera calibration. Xu *et al.* (Xu, Chen, Xie & Lu, 2017) proposed a four-camera system arranged at 90° intervals, leveraging intrinsic and extrinsic calibration parameters alongside a cylindrical distortion model to reconstruct labels. Similarly, Zhang *et al.* (Z. Zhang, Xu & Chen, 2021) introduced a five-camera setup, where a top-mounted camera estimates the 3D pose of the cylindrical object while four surrounding cameras capture surface details.

These methods demonstrate strong reconstruction accuracy under controlled conditions, but share common assumptions: known cylinder geometry, fixed camera configurations, and precise calibration. This thesis adopts a fixed, dual-camera rig to enable consistent multi-view capture without relying on manual operator positioning.

4.2.1.2 Learning-Based Methods and Practical Considerations

To address the limitations of rigid geometric approaches, learning-based solutions have been explored. Gromova *et al.* (Gromova & Elangovan, 2022) proposed a pipeline using Deep Convolutional Neural Networks (DCNNs) to detect keypoints on cylindrical medication labels, followed by cylindrical dewarping, stitching, and text processing. This approach improves robustness to environmental variation and reduces reliance on explicit calibration.

However, this approach introduces dependencies: reconstruction depends on the accuracy and completeness of detected keypoints. In clinical settings with partially occluded or degraded labels, keypoint detection becomes unreliable. Additionally, the multi-stage architecture incurs significant computational overhead.

Parallel advances in document dewarping demonstrate that end-to-end DL models can recover 3D geometry from single images. While conceptually relevant, these

approaches are primarily designed for planar or mildly curved documents, and their extension to cylindrical geometries with varied label content remains underexplored (Das, Ma, Shu, Samaras & Shilkrot, 2019; Feng, Zhou, Deng, Wang & Li, 2022).

4.2.1.3 Research Gap and Design Rationale

Existing methods broadly fall into two categories: (1) geometry-based approaches that often rely on controlled acquisition setups and manual calibration, and (2) learning-based approaches that provide greater flexibility but can introduce additional computational overhead. In practice, both factors can limit applicability in resource-constrained environments, where computational capacity and acquisition conditions may be restricted.

A practical system therefore needs to balance robustness to material and lighting variation with computational efficiency. In this work, cylindrical geometry is formulated through ellipse-based modeling, and two ellipse detection methods are explored: a contour-based approach and a learning-based approach. This allows a comparison of trade-offs between interpretability and computational cost on one hand, and robustness to noise and variation on the other.

The overall design anchors reconstruction in explicit ellipse parameterization, which provides a structured geometric representation and does not rely on sparse keypoint detection. In addition, classical feature-based stitching is used for alignment, offering a more predictable computational profile compared to fully learning-based alternatives.

Table 4.1: Summary of related works on cylindrical label reconstruction

Study	**Approach**	**Strength**	**Limitation**
(Xu et al., 2017)	Four-camera calibration + geometric distortion model	High accuracy, explicit control	Requires manual calibration, fixed hardware; impractical for hand-held clinical use
(Z. Zhang et al., 2021)	Top-down 3D pose + four surrounding cameras	Robust to label shifts and low-texture	Requires additional hardware; complex setup
(Cai, Kang, Ma, Hong & Qiu, 2022)	Cylindrical back-projection model	Fast; handles variable diameters	Limited to specific viewing angles; narrow applicability
(Gromova & Elangovan, 2022)	DCNN keypoints + cylindrical dewarping + stitching	Robust to varied lighting; minimal hardware	Sensitive to keypoint loss on specular/transparent surfaces; high computational cost
(Das et al., 2019)	Shape network (3D coords) + texture mapping network	Real-time; handles complex geometry	Optimized for planar documents; not designed for cylindrical geometry

(Feng et al., 2022)	Transformer for shape + CNN for textlines (DocGeoNet)	State-of-the-art document OCR accuracy	Requires clear textlines; high computational cost; not designed for cylindrical geometry

4.2.2 Segmentation for Cylindrical Label Isolation

Accurate segmentation is a prerequisite for cylindrical label reconstruction, as it defines the spatial extent of the label and constrains subsequent geometric modeling. Unlike generic medical imaging tasks, segmentation in this setting must handle low-contrast boundaries, specular reflections, partial visibility, and heterogeneous label designs commonly found in pharmaceutical packaging.

Segmentation has been extensively studied across domains. In medical imaging, architectures such as U-Net (Ronneberger, Fischer & Brox, 2015) leverage encoder-decoder structures with skip connections to recover fine grained spatial details, enabling precise delineation of anatomical structures (Shah et al., 2021; Woźniak, Siłka & Wieczorek, 2023; Ranjbarzadeh, Caputo, Tirkolaee, Ghoushchi & Bendechache, 2023; Aamir et al., 2022). Similarly, Mask R-CNN (He, Gkioxari, Dollár & Girshick, 2017) extends the Faster R-CNN framework with a dedicated mask prediction branch and RoIAlign, which ensures accurate pixel-to-pixel alignment through bilinear interpolation. While these approaches achieve high segmentation fidelity, they are computationally demanding.

More recently, YOLO-based segmentation models (Jocher, 2020; Ultralytics, 2024) have introduced a single-stage paradigm optimized for real-time performance. Instead of predicting masks independently per instance, they learn a shared set of prototype masks and generate instance-specific outputs via linear combinations of mask coefficients. This design significantly reduces computational overhead while maintaining competitive accuracy.

YOLOv8's anchor-free design eliminates the need for manual anchor tuning, which is particularly important when label aspect ratios vary due to cylindrical distortion and perspective effects. Its feature fusion strategy (C3 modules and PANet-style aggregation) enhances the representation of curved and partially visible boundaries.

While Mask R-CNN provides higher mask precision through RoIAlign, it incurs higher inference latency and memory usage, critical limitations in resource-constrained hospital environments.

We adopt YOLOv8-s as a practical trade-off: accepting a modest reduction in segmentation precision in exchange for real-time throughput and lower computational cost. Importantly, downstream geometric refinement via ellipse fitting can compensate for segmentation imperfections.

4.2.3 Ellipse Detection Design: Contour-Based vs. Learning-Based Approaches

Ellipse detection plays a central role in cylindrical label reconstruction, as the top and bottom elliptical boundaries directly parameterize the unwrapping transformation. Accurate estimation enables mapping distorted image coordinates back to a consistent 2D representation of the cylindrical surface.

Classical approaches include the Hough Transform (Illingworth & Kittler, 1988), which detects shapes via voting in a discretized parameter space. While robust in theory, its performance depends heavily on accumulator resolution and degrades under noise or fragmented edges, leading to aliasing and false detections. Least-squares fitting (York, 1966; Chernov & Lesort, 2005) provides an analytical solution by minimizing residual errors between observed points and the ellipse model. However, it is highly sensitive to outliers; even small amounts of noise or occlusion can significantly bias the estimated parameters.

Recent DL approaches address these limitations by directly regressing ellipse parameters. Methods such as Ellipse R-CNN (Dong, Roy, Peng & Isler, 2021) extend region-based detection frameworks with dedicated regression branches for predicting ellipse parameters, while anchor-free methods like ElDet (Wang, Lu, Shao, Yuan &

Xia, 2022) model ellipse centers as heatmap peaks and incorporate geometric priors into feature representations. These approaches capture global shape structure, allowing them to remain robust under partial occlusion and incomplete edge information.

In this chapter, we compare classical and DL-based approaches for ellipse detection. Classical contour-based fitting is interpretable and lightweight. Learning-based regression methods are generally more robust to segmentation noise. Our evaluation (Section 4.4.4) highlights this trade-off: traditional fitting performs well under clean conditions but degrades under noisy masks, whereas the DL-based approach achieves higher overall accuracy.

4.2.4 Cylindrical Unwrapping and Classical Feature-Based Stitching

Once geometric boundaries are established, the cylindrical label surface must be transformed into a planar representation. This unwrapping problem has been extensively studied in CV, with traditional approaches relying on explicit geometric transformations such as cylindrical projection, polar unwrapping, and perspective rectification (Hartley & Zisserman, 2003; Szeliski, 2022). These methods typically assume known or estimated camera parameters and object geometry, using detected boundary features (most notably the top and bottom elliptical contours) to parameterize the surface and guide the flattening process (Gromova & Elangovan, 2022; Zankevich, 2021).

In practice, global cylindrical projection is sensitive to errors in geometric estimation. Small inaccuracies in ellipse parameters can propagate across the entire surface, leading to large-scale distortion in the unwrapped result. To mitigate this, we adopt a mesh-based unwrapping strategy, where the surface is divided into local quadrangular patches and each patch is rectified via perspective transformation. This local formulation distributes geometric error across multiple regions, improving robustness to imperfect

ellipse estimation and segmentation noise.

Following unwrapping, reconstructing a complete label requires stitching multiple partial views captured from different viewpoints. Classical image stitching pipelines address this using feature-based matching. Algorithms such as Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) and Speeded-Up Robust Features (SURF) (Bay, Ess, Tuytelaars & Van Gool, 2008) detect invariant keypoints and compute descriptors across images. Correspondences between overlapping regions are then established, and geometric transformations are estimated using RANSAC (Fischler & Bolles, 1981a), enabling robust alignment despite outliers. The aligned images are subsequently blended to produce a unified label reconstruction.

Recent DL approaches aim to improve alignment robustness. Deep homography estimation methods learn direct transformations between image pairs (DeTone, Malisiewicz & Rabinovich, 2016), while unsupervised stitching frameworks jointly optimize alignment and blending (Nie, Lin, Liao, Liu & Zhao, 2021). Learned feature matching models such as SuperPoint (DeTone, Malisiewicz & Rabinovich, 2018) and SuperGlue (Sarlin, DeTone, Malisiewicz & Rabinovich, 2020) further enhance correspondence estimation under challenging conditions, including viewpoint and illumination changes.

While learning-based alignment has seen significant advances, such methods typically demand large-scale training datasets and introduce substantial computational overhead. In contrast, SIFT combined with RANSAC offers a deterministic, well-understood, and computationally efficient solution. It performs reliably under moderate viewpoint variations typical of hand-held yet controlled acquisition in operating theatre settings. Furthermore, this approach remains highly effective on standard hardware without the necessity for GPU acceleration.

In this work, classical feature-based stitching is adopted due to its relative simplicity and predictable computational cost. By addressing geometric distortions in earlier stages, the subsequent stitching process becomes more tractable, allowing conventional feature

matching to perform adequately. This design is motivated by the aim of developing a system that can operate within practical constraints, including environments such as operating theatre settings where computational resources and deployment conditions may be limited.

4.3 Proposed Image Stitching Pipeline

Fig. 4.2 illustrates the proposed pipeline. The workflow consists of three main stages: (1) image acquisition, involving the capture of two overlapping views; (2) label segmentation and geometric modeling, where the label is isolated from the background and top and bottom elliptical contours are detected; and (3) unwarping and stitching, in which these geometric parameters are used to transform the curved labels into planar, distortion-corrected images that are merged into a single composite for high-fidelity analysis.

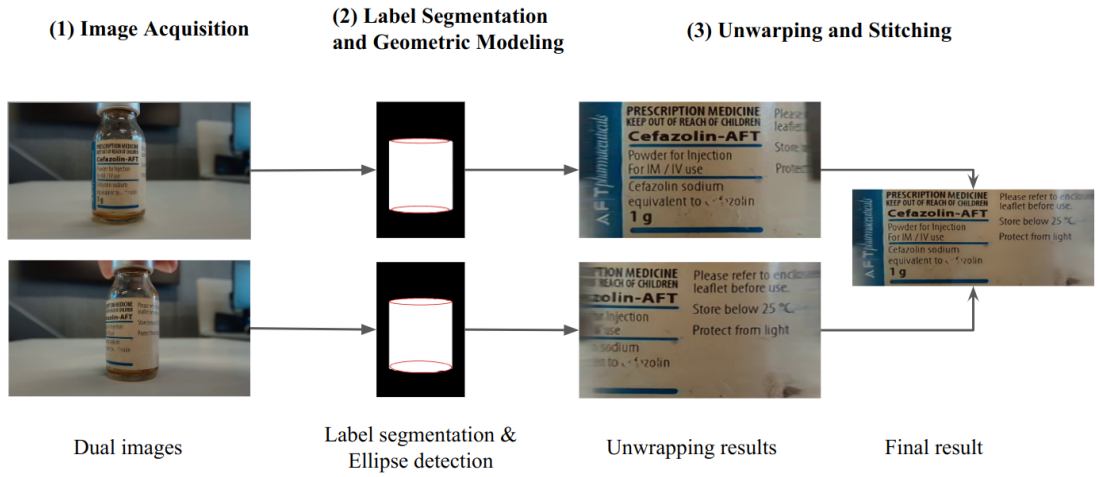


Figure 4.2: Illustration of the reconstruction pipeline for cylindrical medication labels.

4.3.1 Dataset Acquisition

4.3.1.1 Evaluation Proxy and Laboratory Setting

The proposed method is designed for deployment in a dual-camera clinical setup, where two synchronized cameras capture overlapping views of a medication vial from fixed, complementary angles. Such a configuration ensures consistent viewpoint relationships and synchronized acquisition, enabling reliable reconstruction under controlled geometry.

Due to practical constraints in acquiring synchronized dual-camera data in a clinical environment, we instead adopt a single-camera video-based proxy for evaluation. In this setup, each vial is recorded under fixed lighting and camera parameters while being manually rotated to cover a full 360° view. Frames are then extracted at regular intervals to simulate the overlap that would be produced by two cameras observing the object from different viewpoints.

This proxy preserves the essential characteristics required for reconstruction. The extracted frames exhibit controlled and consistent overlap, analogous to that of a fixed dual-camera system, while the underlying cylindrical projection geometry,and therefore the distortion patterns remain unchanged. In fact, the proxy introduces additional variability: manual rotation leads to non-uniform motion and slight inconsistencies in positioning, making the reconstruction task more challenging than the intended deployment scenario.

Nevertheless, the proxy does not fully replicate all aspects of real clinical usage. In particular, it does not capture synchronized multi-view acquisition, hand-held camera jitter, or dynamic lighting variations typical of operating theatre environments. These factors are not modeled in the current evaluation and are reserved for future work involving real-world clinical validation.

4.3.1.2 Image Acquisition

To support model development and evaluation, a dataset was curated comprising 83 cylindrical medication vials and ampoules representing 53 different products commonly used in anaesthesia and operating theatres. The data were provided by Auckland City Hospital (Healthpoint, n.d.) and SaferSleep (SAFERsleep Ltd., n.d.).

To overcome the limitation that a single viewpoint cannot capture the full label of a cylindrical object, we adopted a video-based multi-view acquisition strategy, as shown in Fig. 4.3. For each vial, a short video was recorded under consistent lighting with fixed camera settings to ensure uniform image quality. During recording, the vial was manually rotated to capture a full 360° view of the label. Frames were then extracted at fixed intervals from the video, providing sufficient overlap between adjacent views for reliable reconstruction.

A quality control step was subsequently applied to remove blurry or redundant frames, retaining only images with sharp focus and clear label visibility. This process ensures that the resulting dataset provides dense and consistent coverage of the cylindrical surface, which is essential for accurate unwrapping, distortion correction, and downstream stitching.

The final dataset contains 4,122 images generated, capturing a wide range of real-world variability, including differences in container shape, size, color, and label design. Representative samples from the dataset are shown in Fig. 4.4.

4.3.1.3 Dataset Annotation

Following image acquisition and dataset creation, each extracted frame was manually annotated with polygon masks to isolate the medication label from the background. These annotations serve as ground truth for training and evaluating the segmentation model, enabling accurate localization of the label under varying imaging conditions.

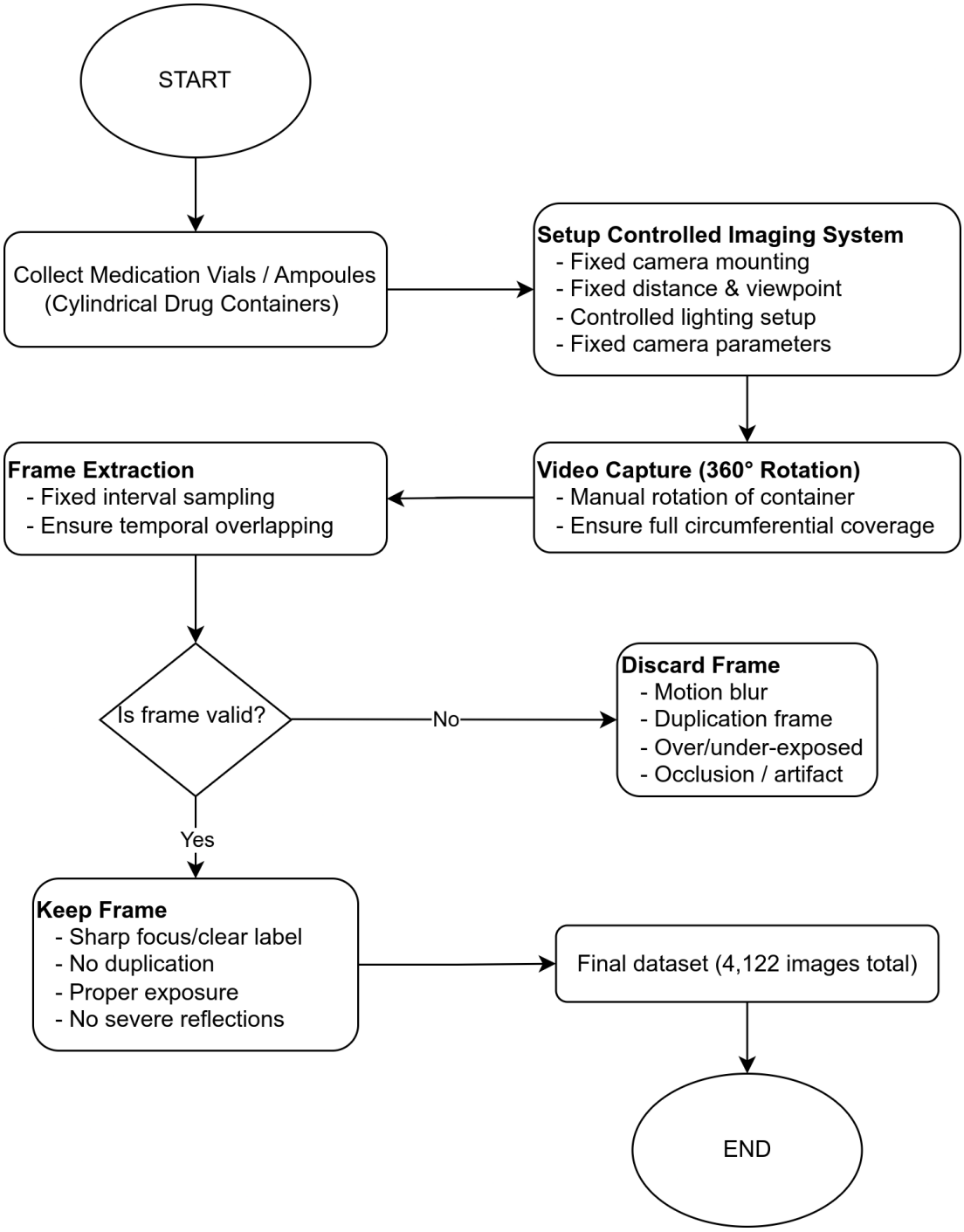


Figure 4.3: The cylindrical container image acquisition workflow.

The annotation process was conducted using the LabelMe tool (Wada, n.d.), which was selected for its flexibility in polygon-based annotation, ease of use, and helps define boundaries precisely. In addition, LabelMe supports different annotation formats that



Figure 4.4: Examples from the multi-view medication containers dataset.

can be readily integrated into common DL pipelines.

To ensure annotation consistency and quality, all images were carefully reviewed during the labeling process, with particular attention to challenging cases such as partial occlusion, specular reflections, and low-contrast label boundaries. Some annotated examples are shown in Fig. 4.5.

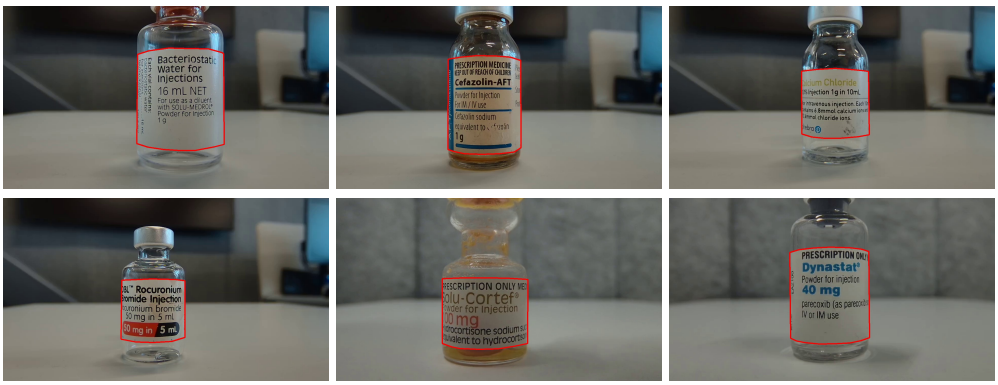


Figure 4.5: Examples of medication bottle with annotations.

4.3.2 Label Segmentation and Geometric Modeling

4.3.2.1 YOLOv8-Based Instance Segmentation

Segmentation is a critical preprocessing step that partitions an image into meaningful regions for subsequent analysis. For cylindrical objects like medication vials, instance

segmentation is especially useful because it precisely separates the label from the background and the container, even when the curved surface distorts the label.

Traditional contour-based methods typically rely on low-level visual cues such as edge detection, with the Canny operator (Canny, 2009) being a foundational technique, or color thresholding and morphological operations (Jähne, 2005) to define object boundaries. While these approaches can be effective under controlled conditions, they exhibit several limitations in real-world scenarios. First, cylindrical geometry introduces non-linear distortion, causing label boundaries to appear curved and inconsistent across viewpoints, which complicates edge continuity and contour extraction. Second, medication labels often contain low-contrast regions, small text, and complex graphical elements, making it difficult to distinguish boundaries using simple intensity or gradient-based methods. Additionally, partial occlusions and complex background may introduce false positives, reducing segmentation reliability.

In contrast, modern DL-based segmentation techniques provide significantly improved robustness to challenging imaging conditions, including variations in illumination, occlusion, and background complexity (Sohail et al., 2022). By learning hierarchical feature representations directly from data, these methods are able to capture both low-level visual cues and high-level semantic information, enabling more reliable identification of label regions even when boundaries are distorted or partially visible.

For this task, we adopt the YOLOv8 segmentation framework due to its ability to balance detection speed, segmentation accuracy, and model efficiency. YOLOv8 leverages an anchor-free architecture and advanced feature aggregation strategies, which enhance its ability to capture complex shapes in real-world images. Upon processing an input image, the model generates a binary segmentation mask, where foreground pixels correspond to the label region and background pixels correspond to non-label areas. This mask serves as the foundational input for subsequent geometric estimation and label unwrapping procedures.

4.3.2.2 Ellipse Estimation Strategies

After segmenting the label regions, the next step is geometric characterization. We aim to estimate the top and bottom boundaries of the cylindrical label, which project as ellipses in the 2D image plane.

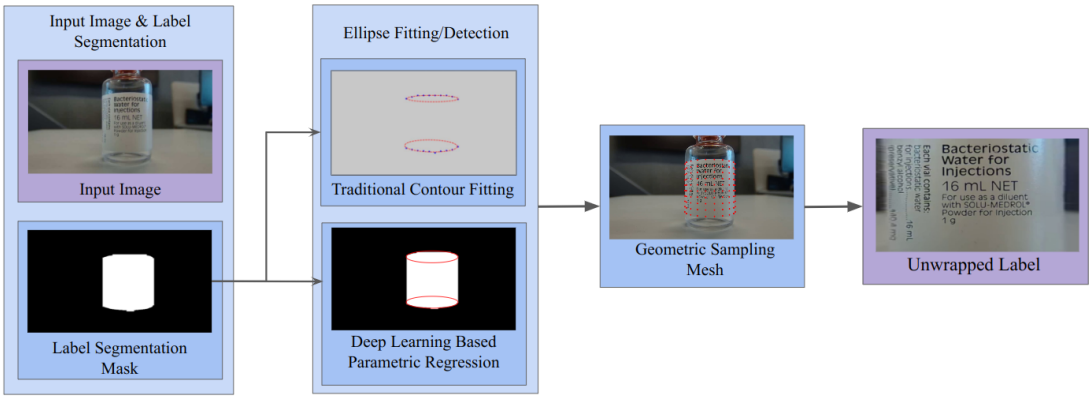


Figure 4.6: Dual path ellipse detection for cylindrical label unwrapping.

We implement two complementary approaches for ellipse estimation, as illustrated in Fig 4.6:

1. **Traditional contour-refined geometric method**: This approach detects connected components in the segmentation mask, extracts the outermost contours, and fits ellipses using least-squares optimization. Contour refinement reduces noise from imperfect segmentation and yields an interpretable geometric model. However, it remains sensitive to segmentation noise and incomplete boundaries.

2. **DL method**: This approach employs a neural network to directly regress ellipse parameters from the binary mask. Although less interpretable, it provides more robust estimation under partial occlusion and lighting variation.

By evaluating these two strategies, we can systematically analyze the trade-offs between the interpretability of classical geometric methods and the robustness of DL-based approaches.

4.3.3 Traditional Contour-Refined Ellipse Fitting

Accurately estimating the top and bottom ellipses of a cylindrical label is essential for correcting perspective distortion and reconstructing a flat, planar label image. We propose a contour-based approach that leverages geometric reasoning to produce interpretable, parameter-based results, as illustrated in Fig. 4.7. The method comprises four sequential stages: (1) contour extraction from the segmentation mask, (2) keypoint localization, (3) horizontal point sampling within each ellipse, and (4) ellipse parameter fitting via regression. This architecture enables explicit control over each stage and facilitates interpretation of the fitted ellipse parameters.

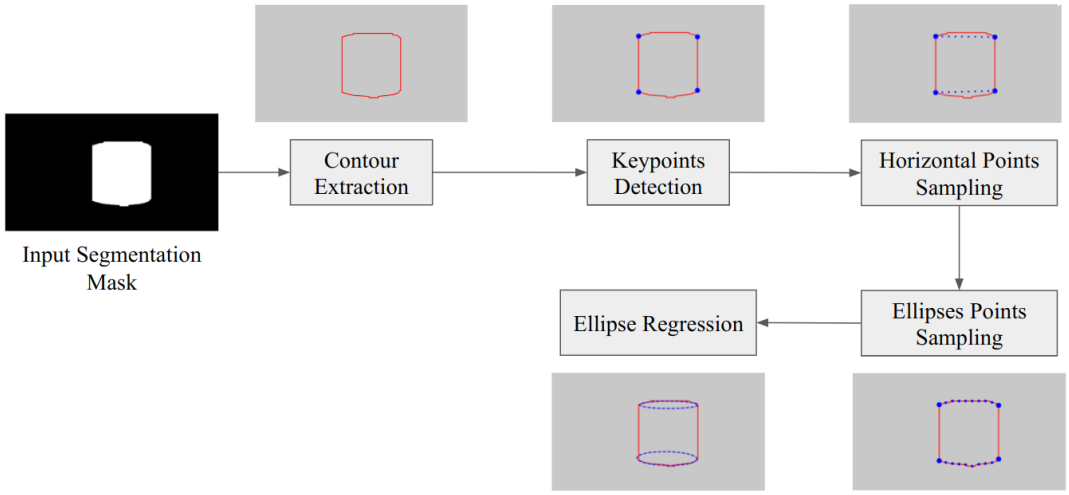
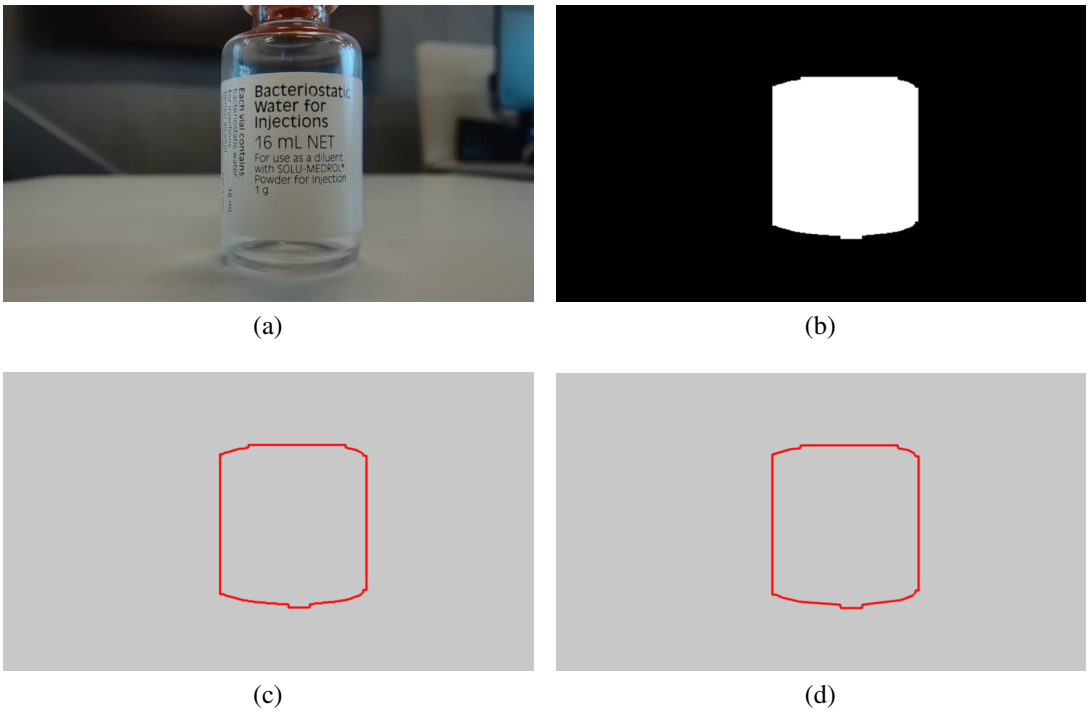


Figure 4.7: The flow chart of the proposed Contour-Based Ellipse Fitting method

4.3.3.1 Contour Extraction

We process the segmented instance mask by extracting its outer topological boundaries using chain-code contour detection. Specifically, to transform this dense pixel-set into a computationally efficient geometric primitive, we apply the Ramer-Douglas-Peucker (RDP) algorithm (Ramer, 1972) via the approxPolyDP implementation in OpenCV. This process recursively simplifies the contour, preserving only the most salient vertices.

As shown in Fig. 4.8, the approximated contour matches the original mask closely. It reduces the number of vertices from 148 to 23 without any noticeable loss in boundary precision. This vertex reduction lowers computational cost and helps ensure that the subsequent ellipse parameter estimation is guided by the vial's overall geometric structure rather than local segmentation noise. This simplification acts as a geometric regularization step, filtering out staircase artifacts and high-frequency irregularities inherent in CNN-based masks.



(a) (b)

(c) (d)

Figure 4.8: Examples are shown as follows: (a) the input drug images and (b) the corresponding binary masks generated by the YOLO segmentation model. (c) displays the approximated contour from the segmented mask with dimensions 148×2, while (d) shows the simplified contour, reduced to 23×2 using the Ramer-Douglas-Peucker algorithm.

In addition, reducing the number of contour vertices minimizes the influence of segmentation noise on subsequent keypoint localization and ellipse regression. This leads to improved computational efficiency while maintaining robustness throughout the fitting pipeline.

4.3.3.2 Keypoint Localization and Horizontal Point Sampling

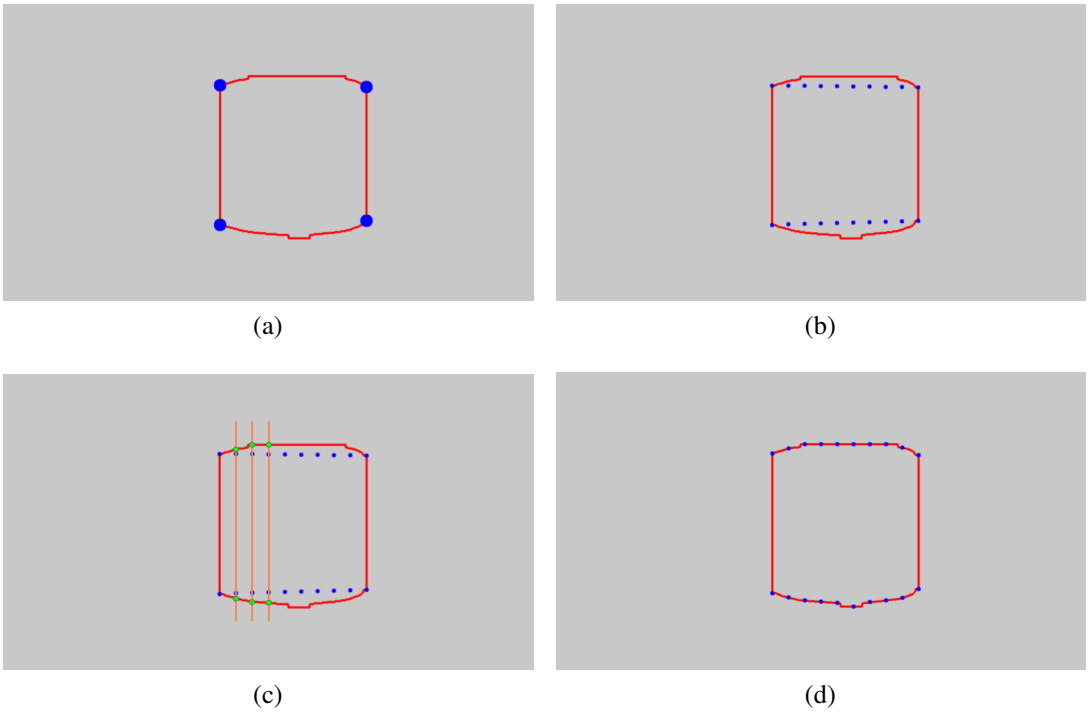


Figure 4.9: Illustration of the sampling process: (a) keypoint localization, (b) horizontal point sampling, (c) vertical line intersection with the contour, and (d) extraction of ellipse points.

To establish the geometric framework for ellipse regression, we implement a structured sampling strategy on the simplified contour. The process begins by localizing four primary keypoints: top-left, top-right, bottom-left, and bottom-right. Specifically, let the simplified contour be defined as:

$$\mathcal{C}' = \{(x_i, y_i)\}_{i=1}^{M}, \tag{4.1}$$

where M denotes the number of contour vertices.

Keypoint localization. Next, we identify the four extremal corners of the projected cylindrical surface. Each point $(x, y) \in \mathcal{C}'$ is evaluated using simple linear combinations of its coordinates:

$$\mathbf{p}_{\text{tl}} = \arg\min_{(x,y)\in\mathcal{C}'} (x + y), \tag{4.2}$$

$$\mathbf{p}_{\text{tr}} = \arg\min_{(x,y)\in\mathcal{C}'} (y + (W - x)), \tag{4.3}$$

$$\mathbf{p}_{\text{bl}} = \arg\min_{(x,y)\in\mathcal{C}'} (x + (H - y)), \tag{4.4}$$

$$\mathbf{p}_{\text{br}} = \arg\min_{(x,y)\in\mathcal{C}'} ((W - x) + (H - y)), \tag{4.5}$$

where W and H denote the image width and height.

This formulation provides an efficient way to estimate the top-left, top-right, bottom-left, and bottom-right points (Fig. 4.9(a)). By relying on global coordinate extrema rather than local curvature, the method is less sensitive to small contour irregularities or segmentation noise, while still capturing the overall geometry of the projected cylindrical region.

Horizontal sampling. To establish a structured representation of the label geometry, we perform uniform sampling along the top and bottom edges defined by the extracted corner keypoints. Let $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^2$ denote the endpoints of a line segment, and let K be the number of samples. We generate K evenly spaced points along the segment using linear interpolation:

$$\mathbf{s}_k = (1 - t_k)\mathbf{p}_1 + t_k\mathbf{p}_2, \quad t_k = \frac{k}{K-1}, \quad k = 0, 1, \ldots, K-1. \tag{4.6}$$

This formulation is applied to the top edge $(\mathbf{p}_{\text{tl}}, \mathbf{p}_{\text{tr}})$ and the bottom edge $(\mathbf{p}_{\text{bl}}, \mathbf{p}_{\text{br}})$, where $\mathbf{p}_{\text{tl}}, \mathbf{p}_{\text{tr}}, \mathbf{p}_{\text{bl}}, \mathbf{p}_{\text{br}}$ denote the top-left, top-right, bottom-left, and bottom-right keypoints, respectively. This produces two sets of anchor points (Fig. 4.9(b)). The uniform parameterization provides consistent spatial coverage across the label width and establishes correspondences between the upper and lower boundaries.

These anchor points are then used as references for subsequent vertical projection and ellipse point extraction.

4.3.3.3 Vertical Line Extension and Ellipse Point Sampling

To recover boundary points for ellipse regression, we refine the initial horizontal correspondences using a directed ray-contour intersection strategy. Given the sampled point pairs $\{\mathbf{s}_k^{\text{top}}\}_{k=0}^{K-1}$ and $\{\mathbf{s}_k^{\text{bot}}\}_{k=0}^{K-1}$, each pair defines a direction vector:

$$\mathbf{d}_k = \mathbf{s}_k^{\text{bot}} - \mathbf{s}_k^{\text{top}}. \tag{4.7}$$

We define two parametric rays along this direction:

$$\ell_k^{\downarrow}(t) = \mathbf{s}_k^{\text{top}} + t\,\mathbf{d}_k, \quad t \in [0, \alpha], \tag{4.8}$$

$$\ell_k^{\uparrow}(t) = \mathbf{s}_k^{\text{bot}} - t\,\mathbf{d}_k, \quad t \in [0, \alpha], \tag{4.9}$$

where $\alpha > 1$ ensures that each ray extends beyond the label boundary.

In this example, the sampled points $\mathbf{s}_k^{\text{top}}$ and $\mathbf{s}_k^{\text{bot}}$ lie within the label region, each ray originates inside the polygon and intersects the contour at a single boundary point along the given direction. We therefore define the refined boundary points as the first intersection between each ray and the contour:

$$\mathbf{e}_k^{\text{bot}} = \ell_k^{\downarrow} \cap \mathcal{C}', \quad \mathbf{e}_k^{\text{top}} = \ell_k^{\uparrow} \cap \mathcal{C}'. \tag{4.10}$$

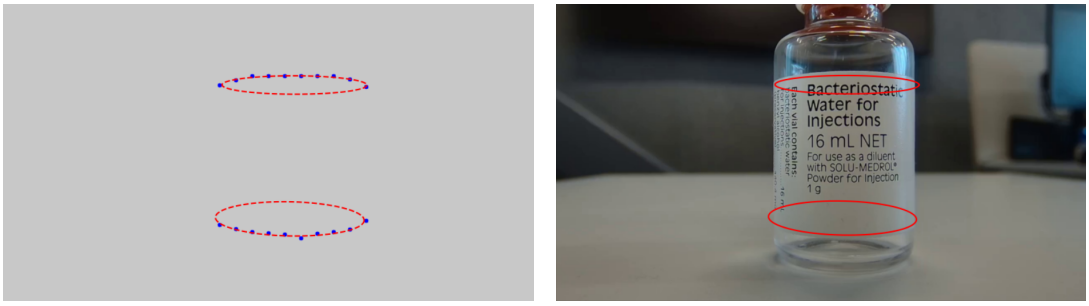
The resulting refined boundary point sets are:

$$\mathcal{E}_{\text{top}} = \{\mathbf{e}_k^{\text{top}}\}_{k=0}^{K-1}, \quad \mathcal{E}_{\text{bot}} = \{\mathbf{e}_k^{\text{bot}}\}_{k=0}^{K-1}. \tag{4.11}$$

As illustrated in Fig. 4.9(c)–(d), this procedure refines the initial horizontal anchors by projecting them onto the label boundary along consistent directions. This helps stabilize the sampling process and reduces the influence of small segmentation irregularities or nearby structures such as caps and specular highlights, which can otherwise affect ellipse fitting.

The formulation does not strictly require all sampled points to lie strictly inside the contour, as boundary points are determined through ray-contour intersection along the prescribed direction.

4.3.3.4 Ellipse Parameter Fitting



(a) (b)

Figure 4.10: Example of ellipse fitting: (a) sampled points and fitted ellipses, and (b) aligned ellipse results on the drug image.

For regressing the ellipse parameters from a set of points, we adopted the numerically stable direct least squares fitting of ellipses algorithm developed by Halíř and Flusser (Halır & Flusser, 1998). This method is robust and can handle cases where earlier methods might fail due to numerical instability or degeneracy. Specifically, this method treats the ellipse as a special case of a general conic section, represented by the equation:

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0, \tag{4.12}$$

where a, b, c, d, e, f are the coefficients. For an ellipse, these coefficients must satisfy

the constraint $b^2 - 4ac < 0$.

Next, the design matrix $\mathbf{D}$ is formed using the quadratic and linear terms of the given set of points (x_i, y_i), as follows:

$$\mathbf{D} = \begin{pmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2 y_2 & y_2^2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n & 1 \end{pmatrix} \tag{4.13}$$

The goal is to solve for the vector $\mathbf{p} = [a, b, c, d, e, f]^T$ that minimizes $\|\mathbf{Dp}\|$ subject to the constraint $b^2 - 4ac < 0$. To improve numerical stability and robustness, they proposed partitioning the problem into two distinct components. This partitioning involves separating the design matrix $\mathbf{D}$ into two sub-matrices: $\mathbf{D}_1$, which contains the quadratic terms x^2, xy, y^2, and $\mathbf{D}_2$, which contains the linear terms $x, y, 1$, as follows:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{pmatrix} \tag{4.14}$$

By partitioning $\mathbf{D}$, the problem becomes more manageable and numerically stable. This approach leads to the formation of three key scatter matrices:

$$\mathbf{S}_1 = \mathbf{D}_1^T \mathbf{D}_1 \tag{4.15}$$

$$\mathbf{S}_2 = \mathbf{D}_1^T \mathbf{D}_2 \tag{4.16}$$

$$\mathbf{S}_3 = \mathbf{D}_2^T \mathbf{D}_2 \tag{4.17}$$

The next step involves solving the generalized eigenvalue problem:

$$\mathbf{Mv} = \lambda \mathbf{Cv}, \tag{4.18}$$

where $\mathbf{M} = \mathbf{S}_1 - \mathbf{S}_2\mathbf{S}_3^{-1}\mathbf{S}_2^T$ and $\mathbf{C}$ is a constraint matrix that enforces the ellipse condition $b^2 - 4ac < 0$. This constraint ensures that the solution corresponds to a valid ellipse rather than a degenerate conic section.

To determine the optimal ellipse parameters, the method involves selecting the eigenvector associated with the smallest positive eigenvalue of the generalized eigenvalue problem. This eigenvector provides the coefficients a, b, c, d, e, f, which define the ellipse. These coefficients can then be converted into standard ellipse parameters, such as the center coordinates, the lengths of the semi-major and semi-minor axes, and the orientation angle of the ellipse. The result of this ellipse fitting method is illustrated in Fig. 4.10, where two ellipses are formulated from the sampled points.

This approach fits an ellipse to the provided data points while enforcing the geometric constraints of the ellipse formulation. By leveraging these geometric properties, it produces interpretable ellipse parameters that support the unwrapping of the cylindrical label into a flat representation. Under controlled imaging conditions, the method provides stable results and serves as a baseline for comparison with DL-based approaches.

4.3.4 Deep Learning Based Ellipse Detection

In addition to the traditional contour-based approach, we investigate a DL-based method for detecting the top and bottom ellipses of cylindrical labels. This approach is designed to improve robustness under challenging conditions such as partial occlusion, segmentation noise, and weak or ambiguous boundaries.

4.3.4.1 Synthetic Dataset Creation

To support this, we construct a synthetic dataset of 11,000 binary masks representing simplified 2D projections of cylindrical medication vials. Each sample is generated from a parameter vector $\mathbf{p} = (c_x, c_y, r_x, r_y, \theta)$, which defines the underlying ellipse

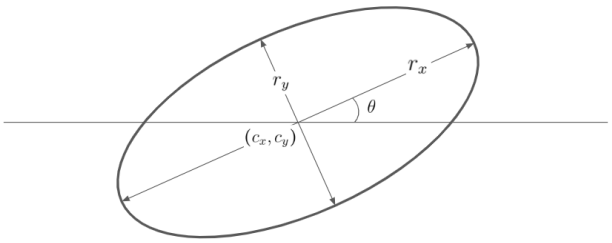


Figure 4.11: Illustration of an ellipse with defined parameters $(c_x, c_y, r_x, r_y, \theta)$

geometry, including center position, semi-axes, and orientation (Fig. 4.11).

In this setting, the model input is the binary segmentation mask, while the target output is the corresponding ellipse parameters $\mathbf{p}$. This formulation enables direct regression of global geometric structure from segmentation output, avoiding the instability of dense contour representations and reducing sensitivity to local noise.

Each binary mask is constructed by rendering two semi-elliptical arcs that approximate the top and bottom boundaries of a cylindrical projection. This representation preserves the essential geometric structure required for unwrapping, where accurate ellipse estimation enables inverse mapping from the cylindrical surface to a flattened label representation.

To enable scalable training without manual annotation, we adopt a procedural generation pipeline that produces diverse yet controlled samples with known ground-truth parameters. To reduce the sim-to-real gap, the generation process incorporates targeted randomization designed to approximate imperfections observed in YOLO-based segmentation outputs:

- Masks are generated on a canvas of size 1280×720 pixels. Each cylindrical label is defined by a pair of ellipses representing the top and bottom boundaries. The horizontal center is sampled as $c_x \in [550, 730]$ to ensure the object remains centered. Vertical positions are sampled as $c_y^{\text{top}} \in [120, 300]$ and $c_y^{\text{bottom}} = c_y^{\text{top}} + \delta_y$, where the height offset $\delta_y \in [300, 400]$. The horizontal radius is sampled as

$r_x \in [200, 300]$, while vertical radii $r_y^{\text{top}} \in [15, 30]$ and r_y^{bottom} are scaled to reflect perspective, with r_y^{bottom} being up to $1.4\times$ larger than r_y^{top}. The orientation θ is sampled within $[-6^\circ, 6^\circ]$, with independent small perturbations applied to the bottom ellipse to account for non-parallel camera alignment.

- The mask is formed by connecting the visible front-facing arcs of the sampled ellipses. By selecting the appropriate semi-elliptical contours and joining them as a single polygon $\mathcal{P}$, we produce a ruled surface that approximates the projected cylindrical label. This formulation captures the dominant geometric structure required for downstream unwrapping while remaining computationally efficient.

- To mimic realistic segmentation imperfections, we apply a two-stage perturbation process. First, random spatial noise $\epsilon \sim \mathcal{N}(0, 2.0)$ is added to each vertex of the polygon $\mathcal{P}$ to introduce jagged edges. Second, to emulate the interpolation effects of CNN upsampling, a Gaussian kernel with $\sigma = 15$ is applied to the binary mask. The result is then re-thresholded at a fixed limit to create organic, slightly rounded boundaries characteristic of deep learning outputs.

- All masks are generated at the native 1280×720 resolution and then will be resized to 256×256 during training process. This prevents resampling artifacts and ensures that the coordinate space remains consistent for subsequent geometric refinement using ellipse fitting.

This synthetic strategy provides precise ground truth for supervised learning and controlled variability that improves robustness to real-world segmentation imperfections, the output cylinder masks and corresponding ground-truth ellipses are illustrated in Fig. 4.12. The dataset is divided into training and validation sets of $10,000$ and $1,000$ samples, respectively.

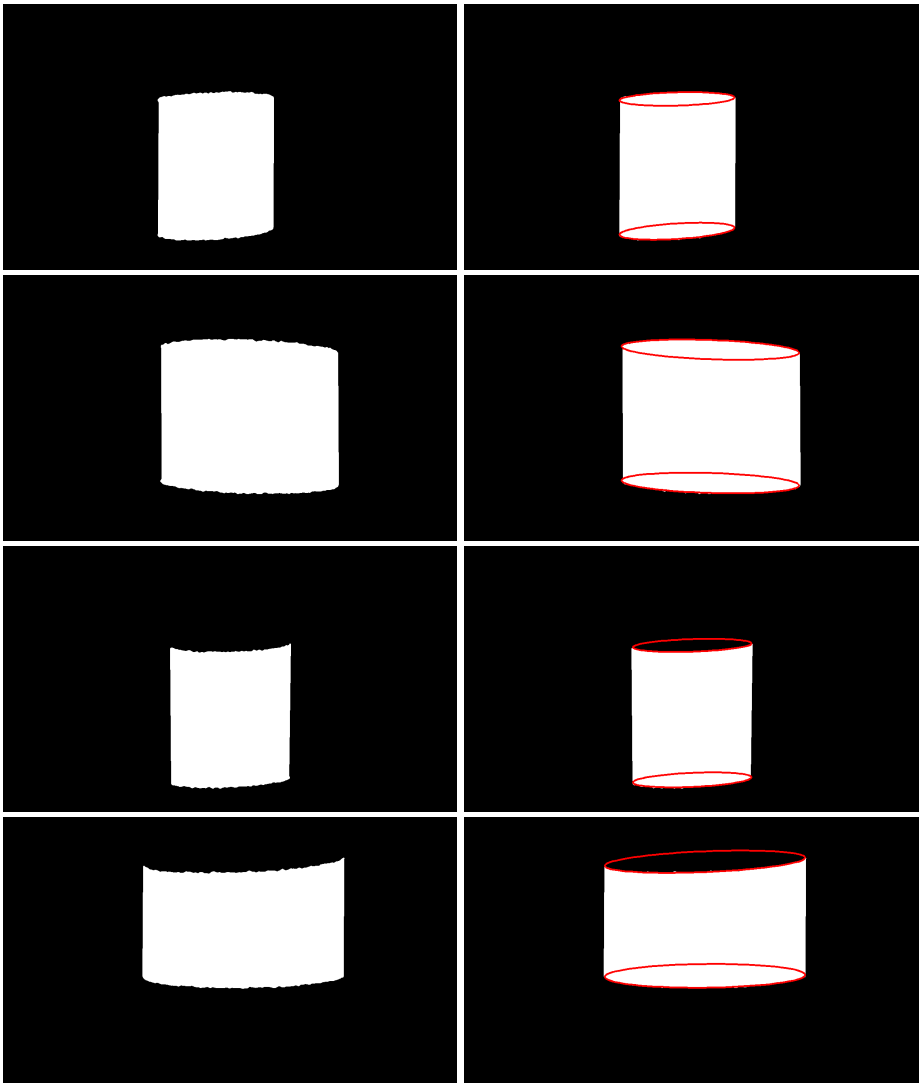


Figure 4.12: Some examples of the generated dataset used in this research. Left: binary cylinder mask; Right: cylinder mask overlaid with the ground truth ellipse parameters used for geometric validation.

4.3.4.2 Network Architecture

Following synthetic dataset generation, we adapt the Fully Convolutional One-Stage (FCOS) framework (Tian, Shen, Chen & He, 2020) for ellipse parameter estimation. FCOS is selected for its anchor-free, per-pixel prediction paradigm, which eliminates the need for manual anchor design and IoU-based heuristics, resulting in a simpler and more stable optimization process. We reformulate the detection head to directly regress

the five ellipse parameters $(c_x, c_y, r_x, r_y, \theta)$ from multi-level feature maps, enabling continuous geometric modeling of the label projection.

The model employs a backbone network with a Feature Pyramid Network (FPN) (T.-Y. Lin et al., 2017) for multi-scale feature fusion, allowing accurate detection across varying label sizes and aspect ratios. This anchor-free formulation provides an efficient, numerically stable, and robust framework for ellipse estimation in diverse cylindrical geometries. Specifically, the FPN outputs a set of feature maps, denoted as P3 through P7, which are used for final prediction. Each level of the FPN is processed by three specialized prediction heads:

- **Classification Head**: Computes the probability that a given spatial location corresponds to an ellipse. This allows the model to distinguish ellipse regions from background pixels effectively, which is critical when the segmented mask may contain noise or irregular shapes.

- **Regression Head**: Directly predicts the five parameters of the ellipse $(c_x, c_y, r_x, r_y, \theta)$ at each location. By performing direct regression rather than relying on anchor boxes, the network can adapt to ellipses of arbitrary size, orientation, and aspect ratio. This improves precision and reduces dependency on manually tuned hyperparameters.

- **Center-ness Head**: Produces a score indicating how close a location is to the center of the predicted ellipse. This helps suppress low-confidence predictions from locations near the edges, improving robustness against partial occlusions, segmentation errors, or irregular label shapes.

Together, these heads enable the model to accurately localize and parameterize ellipses in segmented masks, maintaining high precision even under challenging conditions such as noisy masks, occlusions, or extreme perspective distortion, as illustrated

in Fig. 4.13.

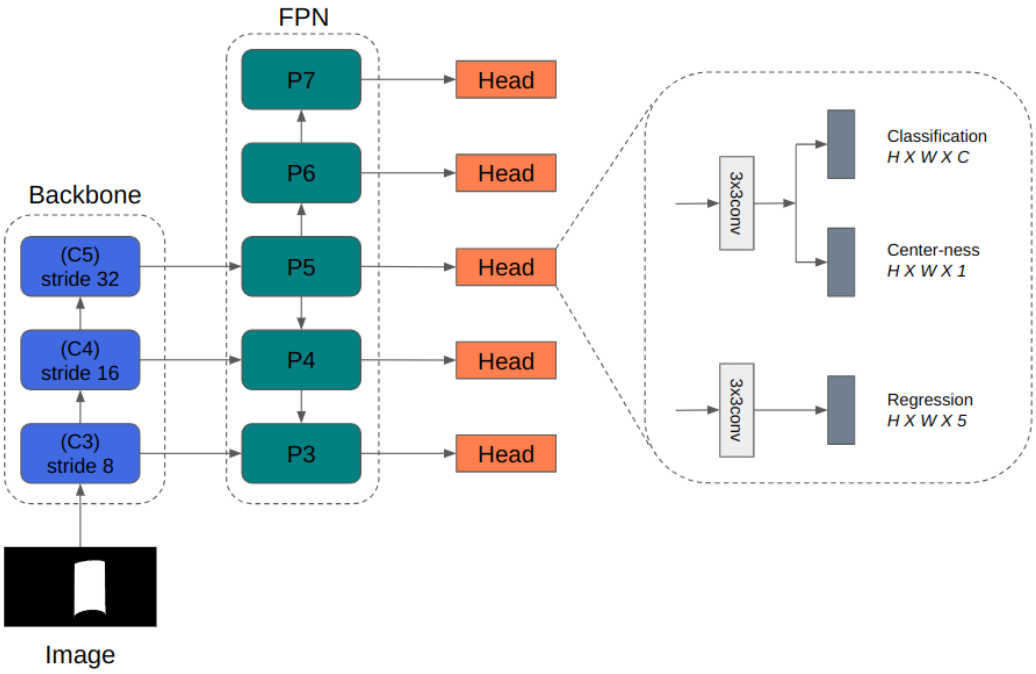


Figure 4.13: The FCOS network architecture. C3, C4, and C5 represent the feature maps from the backbone network, while P3 through P7 denote the feature levels in the FPN used for final predictions.

4.3.4.3 Loss Function

The FCOS-based ellipse detection model is trained using a multi-task loss function that simultaneously optimizes classification, regression, and localization. Each component of the loss is carefully designed to address the unique challenges of detecting ellipses on cylindrical labels.

1. **Classification Loss (Focal Loss):** To handle the inherent class imbalance in ellipse detection where the majority of spatial locations do not correspond to ellipse centers, we adopt Focal Loss (T.-Y. Lin, Goyal, Girshick, He & Dollár, 2018). This loss emphasizes difficult-to-classify positive samples while reducing

the contribution of easy negatives:

$$L_{cls} = -\alpha_t(1 - p_t)^{\gamma} \log(p_t), \tag{4.19}$$

where p_t is the predicted probability, α_t balances positive and negative samples, and $\gamma = 2$ focuses on hard examples. This formulation allows the model to focus on challenging locations, improving detection robustness under cluttered or noisy segmentation masks.

2. **Regression Loss (Gaussian Wasserstein Distance):** Accurate regression of ellipse parameters is critical for precise label reconstruction. We employ Gaussian Wasserstein Distance (GWD) Loss (Yang et al., 2022), which is well-suited for elliptical objects. Unlike standard L1 or L2 losses, GWD is inherently scale-invariant and captures both shape and orientation differences:

$$L_{reg} = \text{GWD}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*), \tag{4.20}$$

where $\hat{\boldsymbol{\theta}} = (\hat{c}_x, \hat{c}_y, \hat{r}_x, \hat{r}_y, \hat{\theta})$ is the predicted ellipse parameter vector and $\boldsymbol{\theta}^*$ is the ground-truth vector. This loss ensures that ellipses with different sizes and rotations are accurately fitted, making the model robust to label variations.

3. **Center-ness Loss (Binary Cross-Entropy):** To improve localization quality, a center-ness score is predicted for each spatial position, encouraging higher confidence for predictions near the center of the ellipse. This component uses binary cross-entropy:

$$L_{cen} = -\big[c^* \log(\hat{c}) + (1 - c^*) \log(1 - \hat{c})\big], \tag{4.21}$$

where c^* is the ground-truth center-ness score and $\hat{c}$ is the predicted score. This

mechanism reduces low-quality detections on the ellipse periphery and enhances the precision of ellipse localization.

The total loss combines these components in a weighted sum over all positive samples (N_{pos}):

$$L = \frac{1}{N_{pos}} \left(\lambda_{cls} \sum_{x,y} L_{cls} + \lambda_{reg} \sum_{x,y} L_{reg} + \lambda_{cen} \sum_{x,y} L_{cen} \right), \tag{4.22}$$

The total loss is defined as a weighted sum over all positive samples, allowing the model to jointly optimize classification, regression, and localization objectives. This encourages the network to learn a balanced representation across tasks, improving the accuracy of ellipse parameter estimation under varying input conditions such as partial occlusion, curvature, and segmentation noise. After training, the network takes a binary segmentation mask as input and predicts the parameters of the top and bottom ellipses, which define the spatial extent of the cylindrical label in the image plane. These predicted parameters are then used to support the subsequent unwrapping and planar reconstruction process.

4.3.5 Cylindrical Label Unwrapping

Once the top and bottom ellipses of the bottle label are detected, the next step is to correct the cylindrical distortion and obtain a flat, undistorted label representation. This unwrapping process enables accurate analysis, OCR, and subsequent stitching of dual label images.

The procedure consists of two main stages: mesh generation for constructing a structured representation of the cylindrical label surface, and geometric transformation for flattening the curved surface into a rectified, unwrapped image.

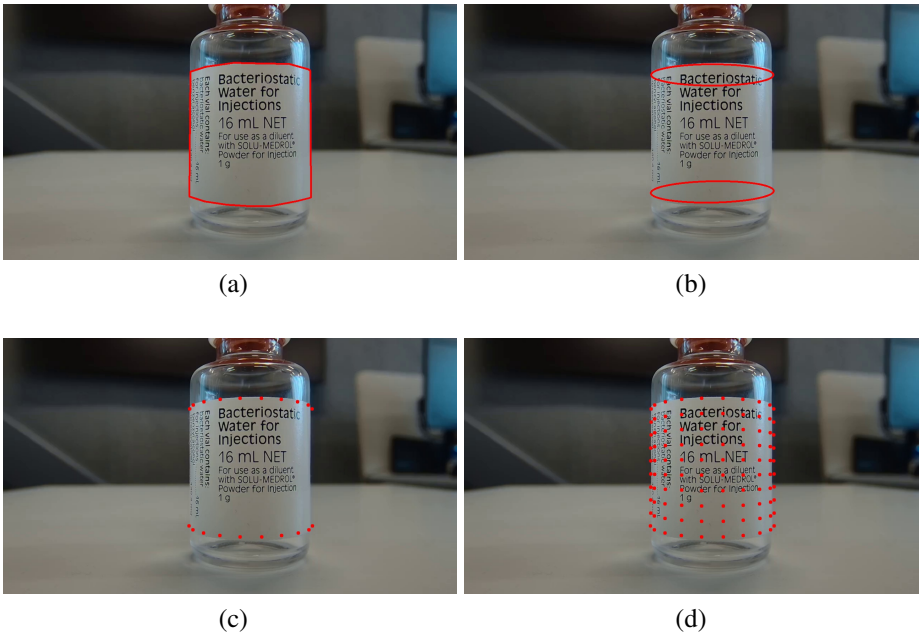


Figure 4.14: The mesh generation process: (a) Input image with segmented mask generated by the YOLO instance segmentation model (b) Top and bottom ellipses of the label detected using the FCOS ellipse detection model (c) Points sampled on top and bottom half ellipses of the bottle label and (d) The final mesh constructed by connecting sampled points.

4.3.5.1 Mesh Generation

To transform the curved label manifold into a rectified planar representation, we construct a discretized geometric mesh that approximates the cylindrical surface of the medication container. This reconstruction process follows a multi-stage pipeline, as illustrated in Fig. 4.14, focusing on identifying the visible surface and establishing a structured coordinate representation.

Selecting Label Halves: Since each detected ellipse represents a full $360°$ boundary, it is divided into two semi-elliptical arcs to separate the visible region from the occluded portion of the cylinder. We compute the IoU between each candidate arc and the initial YOLO segmentation mask, and select the arc pair with the highest overlap. This provides a consistent geometric approximation of the visible label region that is aligned with the observed pixel data.

Point Sampling along Ellipses: Once the correct halves are identified, points are sampled along the top and bottom ellipses using uniform angle intervals. This approach produces denser sampling near the ends of the ellipses where curvature is more pronounced, ensuring a consistent representation of the label surface (Fig. 4.14 (c)). Sampled points (x, y) are calculated as:

$$\begin{aligned} x &= c_x + r_x \cdot \cos(\theta), \\ y &= c_y + r_y \cdot \sin(\theta). \end{aligned} \tag{4.23}$$

where c_x, c_y denote the ellipse center, r_x, r_y the radii, and θ the angle of the sampled point.

Connecting Points and Vertical Sampling: Corresponding points on the upper and lower ellipses are connected to form vertical structures. Uniform sampling along these connections generates a structured quadrilateral mesh, which provides a discrete representation of the label surface (Fig. 4.14(d)). This mesh defines the coordinate mapping used for subsequent planar transformation.

4.3.5.2 Geometric Transformation

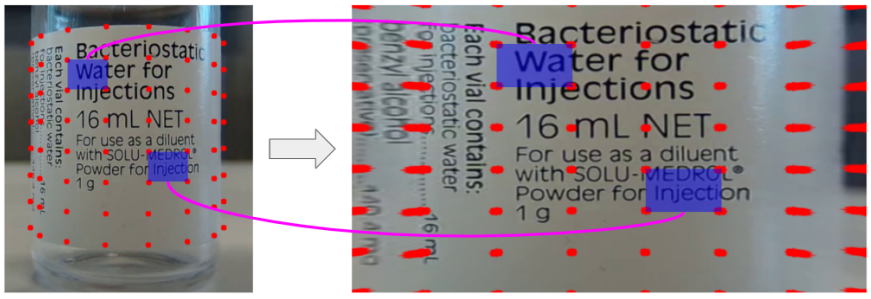


Figure 4.15: Illustration of the cylindrical unwrapping process, while every quadrangle are being applied a perspective transformation to convert the region to a plat plane.

With the mesh constructed, we can now apply geometric transformations to flatten

the cylindrical label, as shown in Fig. 4.15. The process begins with quadrangle formation, where each quadrangle on the cylindrical surface is defined by four neighboring vertices from the sampled mesh points. By iterating through adjacent point pairs along the top and bottom ellipses, we determine all quadrangles to form a continuous mesh that represents the label surface for subsequent geometric transformations.

Specifically, following the mesh construction, a perspective transformation is computed for each individual quadrangle to map its distorted coordinates (x, y) to corrected planar coordinates (x', y'). The transformation matrix $\mathbf{H} \in \mathbb{R}^{3\times3}$ is determined using the four vertices of the quadrangle and is defined as:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \in \mathbb{R}^{3\times3}. \tag{4.24}$$

$$c \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{4.25}$$

where c is a scaling factor and $\mathbf{H}$ encodes rotation, translation, and scaling. This formulation maps each quadrilateral to a planar region using a perspective transformation, providing a local approximation of the cylindrical surface.

The transformation is applied to each quadrilateral in the mesh to produce a flattened representation of the label. This mesh-based approach helps preserve local geometric structure and approximates curvature variations across the surface, which can reduce distortion introduced during unwrapping. As a result, it provides a structured representation that is suitable for downstream OCR and label analysis under the evaluated conditions.

4.3.6 Image Registration and Stitching

After the individual label regions have been detected and geometrically corrected, the next stage focuses on image registration and stitching. This process aligns and merges the two unwrapped label images into a single, continuous view of the label surface. The aim is to reconstruct a more complete label representation from cylindrical containers, providing a consistent visual input for subsequent OCR and text analysis.

4.3.6.1 Feature Detection and Description

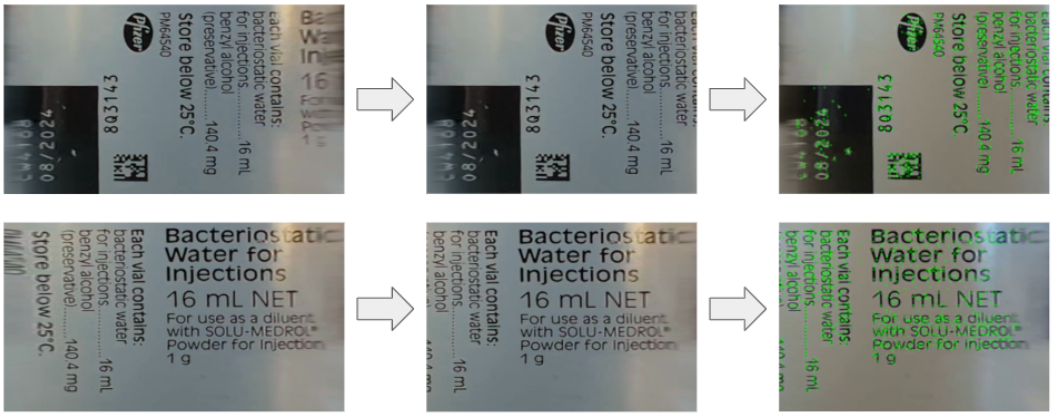


Figure 4.16: Illustration of the keypoint extraction process from unwrapped labels.

To support registration between unwrapped label images, local features are extracted from each image, as shown in Fig. 4.16. After unwrapping, the boundary regions of the label often exhibit stronger geometric distortion due to the stretching effect introduced during the cylindrical-to-planar transformation. In particular, the left and right edges may contain less reliable visual information, as these regions are more sensitive to interpolation and geometric expansion.

To reduce the influence of these unstable regions, both ends of the unwrapped images are cropped prior to feature extraction and registration. In all experiments, a fixed margin (e.g., 5–10% of the image width) is removed from each side to exclude highly distorted boundary areas. This step helps improve the stability of subsequent

feature matching by reducing the number of unreliable keypoints near the edges.

Following this preprocessing step, keypoints are detected using the SIFT algorithm (Lowe, 2004), which is commonly used for its invariance to scale, rotation, and, to some extent, illumination. This property helps maintain feature consistency across different viewing angles and lighting conditions in multi-view captures of medication vials.

After detection, descriptors are computed for each keypoint to generate a 128-dimensional SIFT vector representing local image patterns. These descriptors provide a basis for matching corresponding features across the two unwrapped views, supporting geometric alignment and subsequent image stitching.

4.3.6.2 Feature Matching and Geometric Alignment



Figure 4.17: Demonstration of feature correspondence and geometric alignment between unwrapped labels during the registration and stitching process.

Next, feature matching is performed to identify corresponding points across the unwrapped images. Using a nearest-neighbor search in the descriptor space (implemented in OpenCV), each feature descriptor from one image is compared with those from the

other image to establish candidate correspondences.

Once a sufficient set of correspondences is obtained, the RANSAC algorithm (Fischler & Bolles, 1981b) is used to estimate the homography matrix, which represents the geometric transformation between the two image planes. RANSAC helps reduce the influence of incorrect matches by retaining only geometrically consistent correspondences during estimation.

The resulting homography is then applied to align one unwrapped label image with the reference image, correcting differences in position, orientation, and small perspective variations. This brings corresponding regions into a common coordinate frame, enabling subsequent image merging (Fig. 4.17).

4.4 Experimental Evaluation and Results

The experimental evaluation aims to assess the performance, accuracy, and robustness of the proposed Image Stitching Pipeline (Approach 1). This evaluation provides a baseline for reconstructing medication labels end to end. It measures how well each component helps produce a complete, geometrically consistent label. The experiments are organized in multiple stages to match the pipeline's modular design, covering segmentation, ellipse detection, distortion correction, and image stitching.

4.4.1 Experimental Setup and Dataset

Hardware Configuration All experiments were conducted on a high-performance workstation equipped with an Intel Xeon Silver 4210 CPU, 4 x NVIDIA GeForce RTX 2080 Ti GPU (12GB VRAM), and 100GB DDR4 RAM. This configuration provides sufficient computational resources for both training and inference for DL models such as YOLOv8 and the ellipse regression network, as well as for processing high-resolution multi-view image data.

Multi-view Medication Label Dataset (Segmentation) This dataset comprises 83 cylindrical medication vials and ampoules. A video-based setup was used for acquisition, then frames were sampled at regular intervals to ensure sufficient overlap. After filtering, the final dataset contains 4,122 high-quality images (Section 4.3.1). An 80/20 train/validation split was applied at the drug level to prevent data leakage, ensuring that images of the same physical container do not appear in both sets.

Synthetic Ellipse Projection Dataset (Ellipse Detection) The ellipse regression model is trained on a generated dataset of 11,000 binary masks representing 2D projections of cylindrical labels. Each sample is defined by $(c_x, c_y, r_x, r_y, \theta)$, corresponding to the ellipse center, axes, and orientation. Controlled geometric variations and boundary noise are introduced to simulate real-world segmentation imperfections and improve generalization. The dataset is split into 10,000 training and 1,000 validation samples (Section 4.3.4.1).

Evaluation Metrics Label segmentation and ellipse detection performance are evaluated using mAP, which measure the overlap between predicted masks and ground truth annotations, ensuring sufficient accuracy for downstream geometric modeling and label reconstruction.

4.4.2 Model Training and Hyperparameters

To ensure reproducibility of both the segmentation and geometric regression stages, the training configurations for the YOLOv8 and FCOS-based models are summarized below.

Training Environment Experiments were conducted using four NVIDIA GeForce RTX 2080 Ti GPUs (12GB VRAM each). The models were implemented in PyTorch 2.1 with CUDA 12.1, using the Ultralytics framework for segmentation and MMDetection for FCOS-based ellipse regression.

Optimization Strategy Both models were optimized using Stochastic Gradient Descent (SGD). The initial learning rate was set to 1×10^{-2}, with a 3-epoch warmup followed by linear decay.

Training Hyperparameters The batch size was set to 16 for YOLOv8 and 64 for the FCOS model. Training was performed for up to 100 epochs, with early stopping (patience = 10) to prevent overfitting. Input resolutions were 640×640 for segmentation and 256×256 for ellipse regression. For the FCOS model, the loss weights were set to $\lambda_{cls} = 1$, $\lambda_{reg} = 2$, and $\lambda_{cen} = 1$.

Data Augmentation To improve robustness, standard augmentation techniques were applied, including Mosaic augmentation, random HSV color jittering, and geometric transformations such as translation and scaling (±10%).

4.4.3 Evaluation of Label Segmentation (YOLOv8-s)

The primary objective of the segmentation stage is to generate binary masks of medication labels, which serve as geometric constraints for subsequent ellipse estimation and unwrapping. The YOLOv8-s (Ultralytics, 2024) variant is selected due to its favorable trade-off between inference efficiency and segmentation accuracy, making it suitable for both high-performance and real-time applications.

Table 4.2: Performance of the YOLOv8-s model on the label segmentation task.

Metric	**mAP@0.5**
YOLOv8-s	0.937

The model was evaluated on the validation split (20%) of the multi-view medication dataset. As shown in Table 4.2, the model achieves an mAP@0.5 of 0.937, indicating highly reliable detection of label regions and strong spatial overlap with ground-truth annotations.

We report mean Average Precision (mAP@0.5) as the primary evaluation metric, as the downstream pipeline only requires accurate localization of the label region rather than pixel-perfect boundary delineation. In this setting, precise instance-level segmentation quality at stricter IoU thresholds is less critical, since subsequent geometric refinement is performed during the ellipse fitting stage. Therefore, mAP@0.5 provides a sufficient and task-relevant measure of segmentation performance.

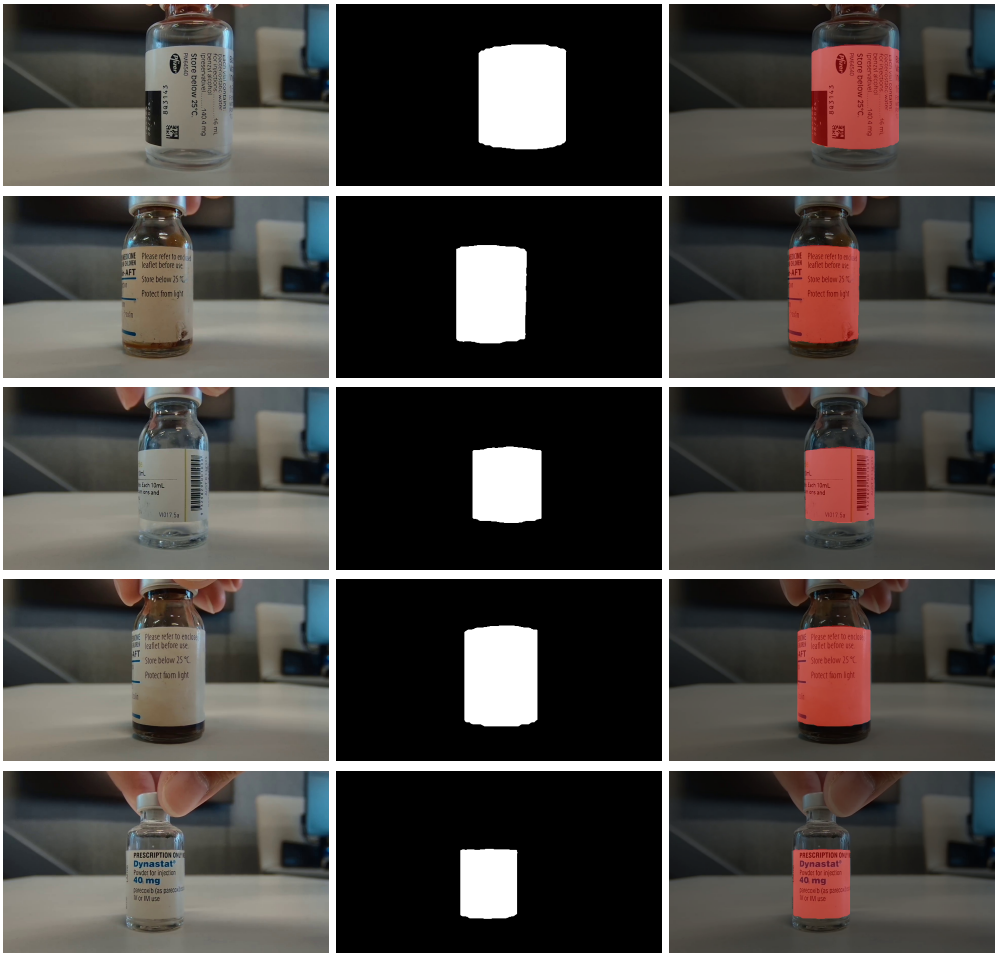


Figure 4.18: Qualitative results of YOLOv8-s label segmentation under varying conditions. Predicted masks are shown as overlays.

In addition, Fig. 4.18 shows segmentation results across different medications and label characteristics. In cases with specular reflections from glass or plastic surfaces, the segmentation generally remains continuous and the label can still be separated from the

background across different viewing angles. However, boundary alignment near curved edges is not always precise. This motivates the use of subsequent ellipse detection to further refine boundary localization and improve geometric consistency.

4.4.4 Ellipse Estimation Results - Traditional vs. Deep Learning Comparison

This section compares the traditional contour-based ellipse fitting approach with the deep learning-based FCOS model. Both methods aim to estimate the geometric parameters of cylindrical medication labels to provide essential inputs for accurate geometric correction and unwrapping. For the FCOS-based method, we employ ResNet-50 (He, Zhang, Ren & Sun, 2015) as the backbone architecture to ensure robust feature extraction for ellipse regression.

Table 4.3: Quantitative comparison of ellipse estimation methods.

Method	**mAP@0.5**
Traditional Ellipse Fitting	0.73
FCOS-based Ellipse Detection	0.91

To evaluate ellipse estimation performance, mAP is adopted as the primary metric. Each predicted ellipse is treated as an object detection result and matched to its corresponding ground-truth ellipse based on the IoU computed between their regions. This evaluation protocol follows standard object detection benchmarks and provides a well-established and robust measure of both spatial localization accuracy and shape agreement.

As shown in Table 4.3, the FCOS-based method outperforms the traditional contour-based approach, achieving an mAP@0.5 of 0.91 compared to 0.73. While both approaches perform reasonably well under controlled conditions, this performance gap

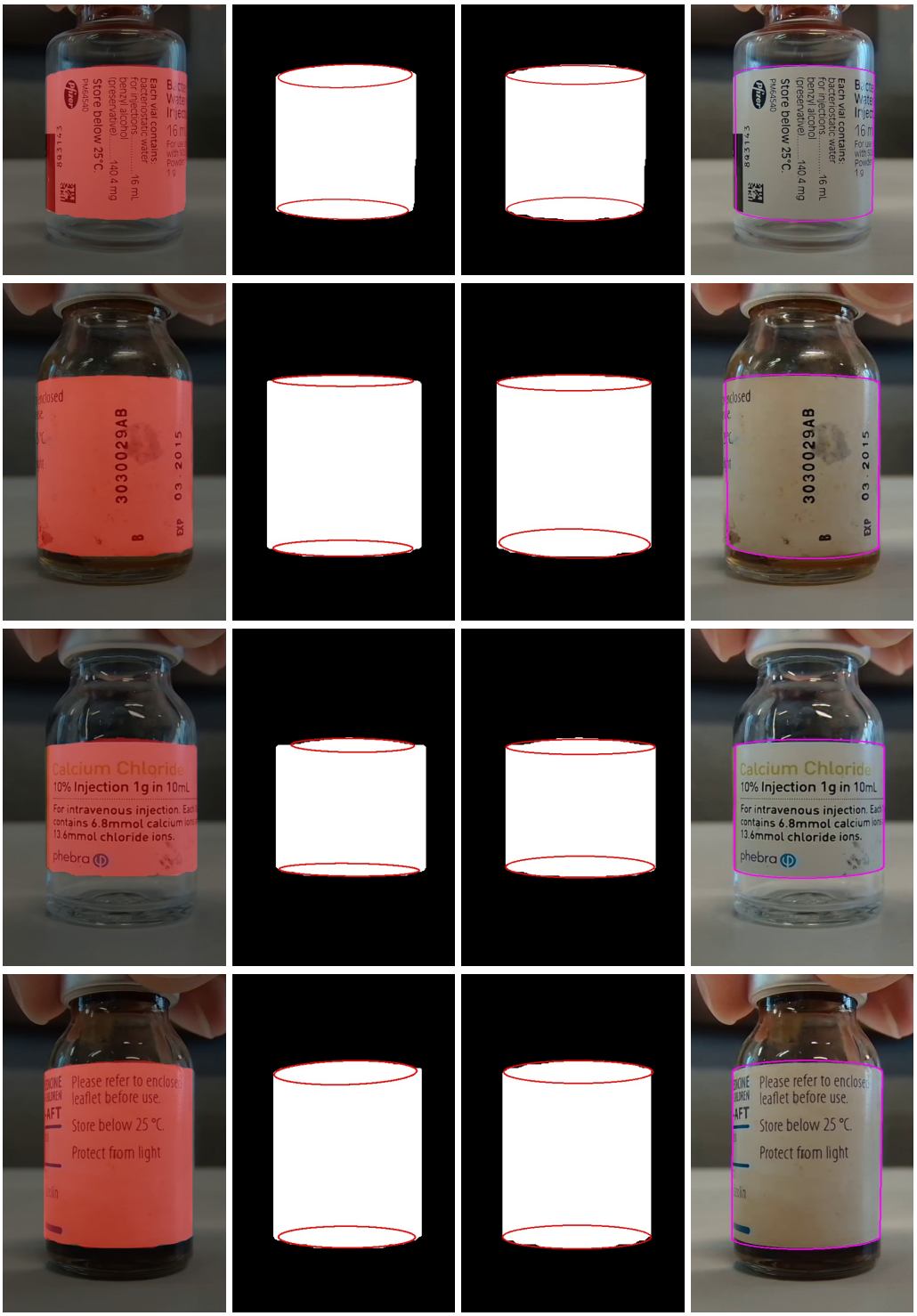


Figure 4.19: From left to right, segmentation mask overlay, traditional contour-based fitting, DL-based regression, and the reconstructed cylindrical representation from the DL ellipse result

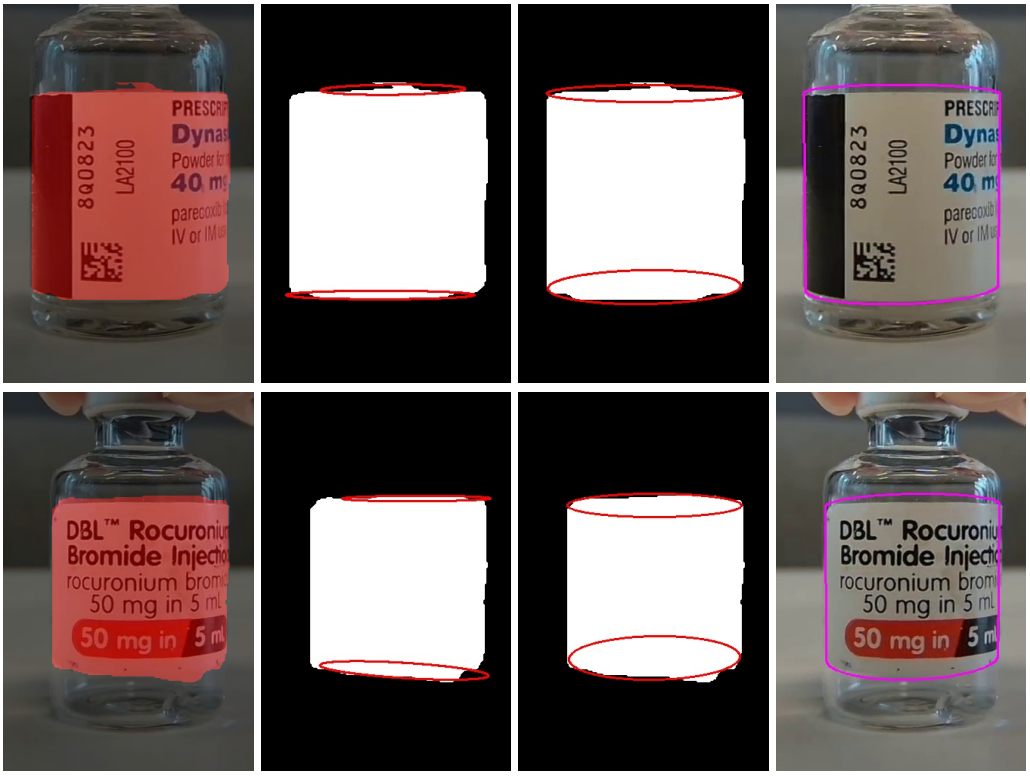


Figure 4.19: From left to right, segmentation mask overlay, traditional contour-based fitting, DL-based regression, and the reconstructed cylindrical representation from the DL ellipse result.

highlights the limitations of direct geometric fitting methods when operating on imperfect or noisy segmentation outputs, whereas learning-based regression provides improved robustness to such variations.

Qualitative results in Fig. 4.19 further illustrate these differences. The traditional contour-based method produces accurate ellipse estimates when segmentation masks are smooth and well-defined. However, its dependence on contour quality makes it sensitive to noise, discontinuities, and perspective distortion. In particular, when the label is viewed from steeper angles, resulting in narrow or flattened ellipses, the method may struggle to capture the underlying geometry, leading to less stable fits.

In contrast, the FCOS-based approach shows more consistent performance across a range of challenging conditions. By learning geometric patterns from data, the

model can estimate ellipse parameters even when the input mask is noisy or lacks clear boundary definition. This tends to produce more stable predictions across varying viewpoints, curvature levels, and label appearances.

The fourth column of Fig. 4.19 presents the reconstructed cylindrical representation of the medication containers, generated by combining two half-ellipses predicted by the model. Despite some minor misalignment, the reconstructed cylinders generally follow the label regions and provide a reasonable approximation of the container geometry.

Overall, both quantitative and qualitative results suggest that the FCOS-based method performs more consistently than the traditional approach under the evaluated conditions. The contour-based method remains effective in controlled settings due to its simplicity and interpretability, while the learning-based approach appears more tolerant to variation in the tested scenarios.

4.4.5 Cylindrical Distortion Correction Results - Unwrapping Evaluation

Following accurate ellipse estimation, the cylindrical distortion correction step was applied to transform the curved label surface into a geometrically rectified, planar representation.

Experimental results of the cylindrical unwrapping process are illustrated in Fig. 4.20. The unwrapped images demonstrate high geometric fidelity and photometric consistency in the central regions of the label, where the cylindrical projection is most stable. Text appears sharp, and label structures are well preserved across a variety of bottle types, orientations, and lighting conditions. Notably, curved text along the original bottle surface is effectively rectified, producing a readable flattened view.

However, quality gradually degrades toward the left and right boundaries of the unwrapped image. These boundary regions come from parts of the cylinder that face



Figure 4.20: Examples of the cylindrical label unwrapping process applied to various medication vials. The results show accurate distortion correction and clear text preservation in central regions.

the camera at a shallow angle. Because these areas are heavily compressed in the input images, the unwrapping step must stretch them much more than the center region. This large stretching often produces visible distortions and lower text clarity. This issue is further amplified by the method's sensitivity to ellipse estimation. Because the mesh is constructed directly from the top and bottom ellipses, even small errors in the ellipse center, orientation, or radii propagate through the mapping and introduce additional distortions, especially near the outer regions of the unwrapped label.

4.4.6 Image Stitching Results - Multi-view Fusion Performance

Figure 4.21 shows the results of stitching two unwrapped medication label images. After segmentation and cylindrical distortion correction, the outer edges of each unwrapped image are cropped to remove regions affected by stretching. Feature correspondences are then extracted using SIFT, and RANSAC is applied to estimate the geometric transformation between the two views, enabling alignment and fusion in the overlapping region. The resulting stitched image preserves most label content in the overlapping area, while minor artifacts may still appear near the cropped boundaries.

The primary focus of this work is a dual-view setup, where only two images are used for reconstruction. To further illustrate the behaviour of the method, multi-image stitching is also demonstrated as an additional experiment. This provides a qualitative assessment of how the approach handles multiple overlapping views and evaluates the consistency of feature matching and geometric alignment beyond the two-view case. As shown in Figure 4.22, the stitched results form a coherent representation of the label surface, preserving both textual and graphical information across the reconstructed region.



(a) (b) (c)

Figure 4.21: Results of stitching two unwrapped images: (a) left image, (b) right image, and (c) the stitched result.

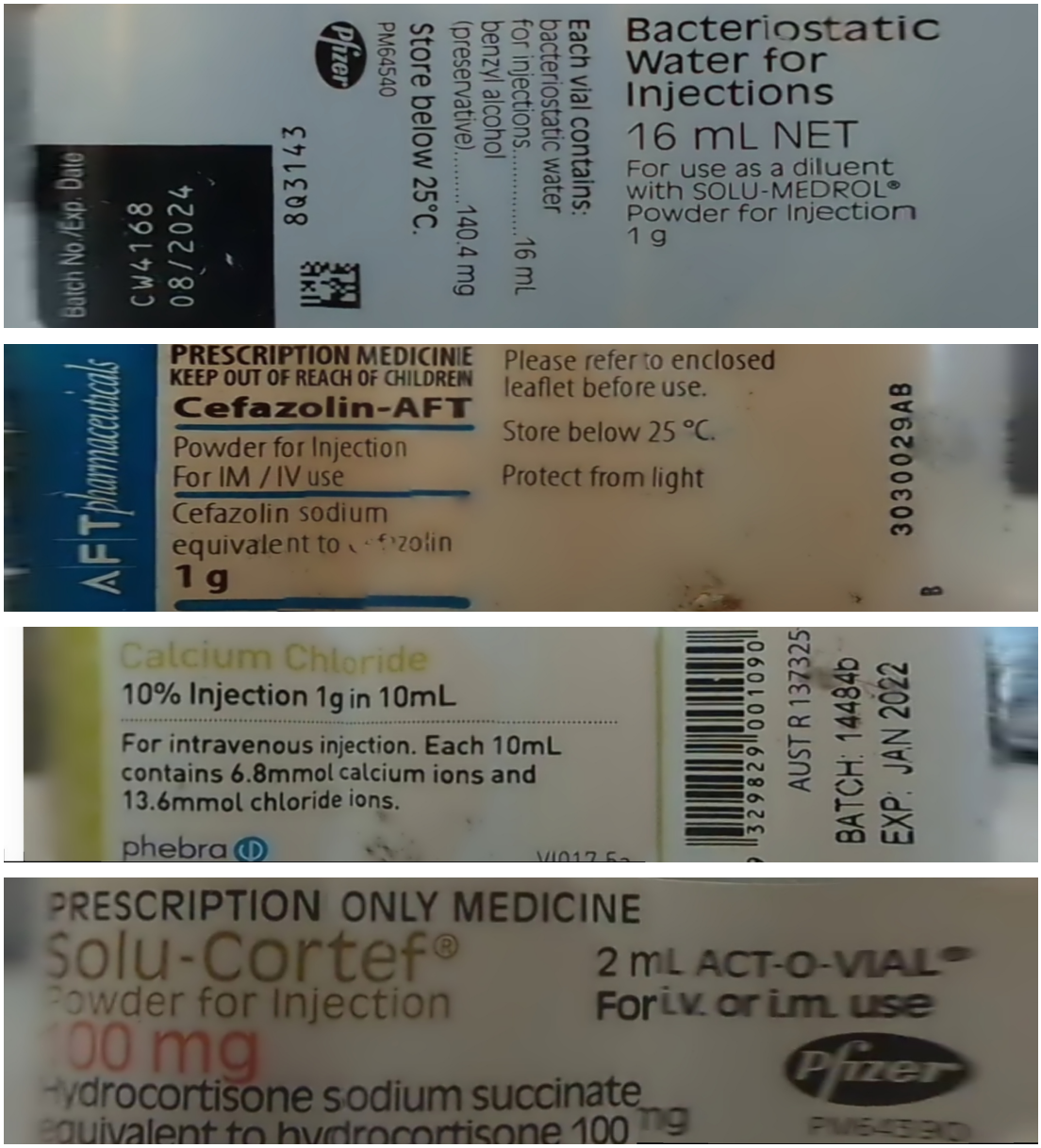


Figure 4.22: Examples of complete stitched labels containing comprehensive drug information (Part 1).



Figure 4.22: Examples of complete stitched labels containing comprehensive drug information (Part 2).

4.4.7 Task-Specific Evaluation: Text Extraction Quantity

While segmentation and ellipse detection are evaluated using standard metrics such as mAP, the unwrapping and stitching stages lack well-defined pixel-level ground truth. We therefore assess their effectiveness using a practical proxy: text extraction quantity, measured as character count from OCR outputs.

The primary limitation of single-view imaging is incomplete surface visibility. Although OCR systems can recognize visible text, they can only process regions captured in the image. For cylindrical vials, a single view captures only a portion of the label due to occlusion and foreshortening. To measure the benefit of dual-view acquisition, we compare text output (character count) from individual views (View A and View B) with the stitched result. OCR is applied to each, and character counts are computed at the word level after case normalization and punctuation removal.

Text quantity is measured as the character count from OCR output:

$$\text{Character count} = N_{\text{OCR output}} \tag{4.26}$$

The quantity gain relative to the best single view is:

$$\text{Quantity gain} = \frac{N_{\text{stitched}} - \max(N_A, N_B)}{\max(N_A, N_B)} \times 100\% \tag{4.27}$$

This conservative baseline ensures stitching is considered beneficial only if it increases text output over the best individual view. For evaluation, 40 image pairs (10 vials × 4 pairs) were constructed from sequential captures with varying viewpoints and partial overlap. Results are aggregated per vial and reported as mean and standard deviation.

As shown in Table 4.4, stitching increases text output by 10–28% across all vials. Larger gains occur when the two views capture distinct label regions, while smaller gains occur when views overlap substantially or label regions consist primarily of non-text elements. This improvement suggests that stitching enables recovery of text invisible in individual views. However, we note that this metric measures the quantity of text extracted by OCR, not its correctness. OCR errors that occur in individual views or potential duplication in stitched output are not separately accounted for. This

Table 4.4: Text extraction quantity gain per container (mean ± standard deviation)

Medication Container	**Mean Gain (%)**	**Std (%)**
Rocuronium Bromide Injection	19.78	24.56
Solu-Cortef 100mg	15.31	12.84
Bridion Sugammadex 200mg	21.40	18.62
Bacteriostatic Water for Injections	9.90	9.37
Vecure 10mg	24.66	18.70
Metoprolol IV Mylan 5mg	28.33	16.19
Cefazolin AFT 1g (old label design)	21.56	22.56
Calcium Chloride	16.99	20.79
Cefazolin AFT 1g (new label design)	26.09	7.75
Prescription Only Medicine	13.58	8.85

evaluation therefore demonstrates that stitching produces increased text output from dual views compared to single-view imaging, without formal verification of output quality or correctness.

4.5 Limitations and Future Work

The label gap remains a significant technical challenge, particularly for transparent or semi-transparent medication containers. When the camera is aligned with the gap between label edges, the absence of strong visual contrast and semantic boundaries often causes the YOLOv8 segmentation model to produce incomplete or fragmented masks. As a result, the subsequent ellipse estimation step fails to recover a coherent and geometrically valid representation of the cylindrical surface.

A related limitation arises when the label does not fully cover the visible portion of the container in the captured image. In such cases, the detected ellipse is constrained by the segmented region and therefore represents only a partial geometry, rather than the full cylindrical shape. This leads to inaccurate ellipse parameters and unreliable geometric reconstruction, as illustrated in Fig. 4.23.

These failure cases highlight an inherent dependency between segmentation quality

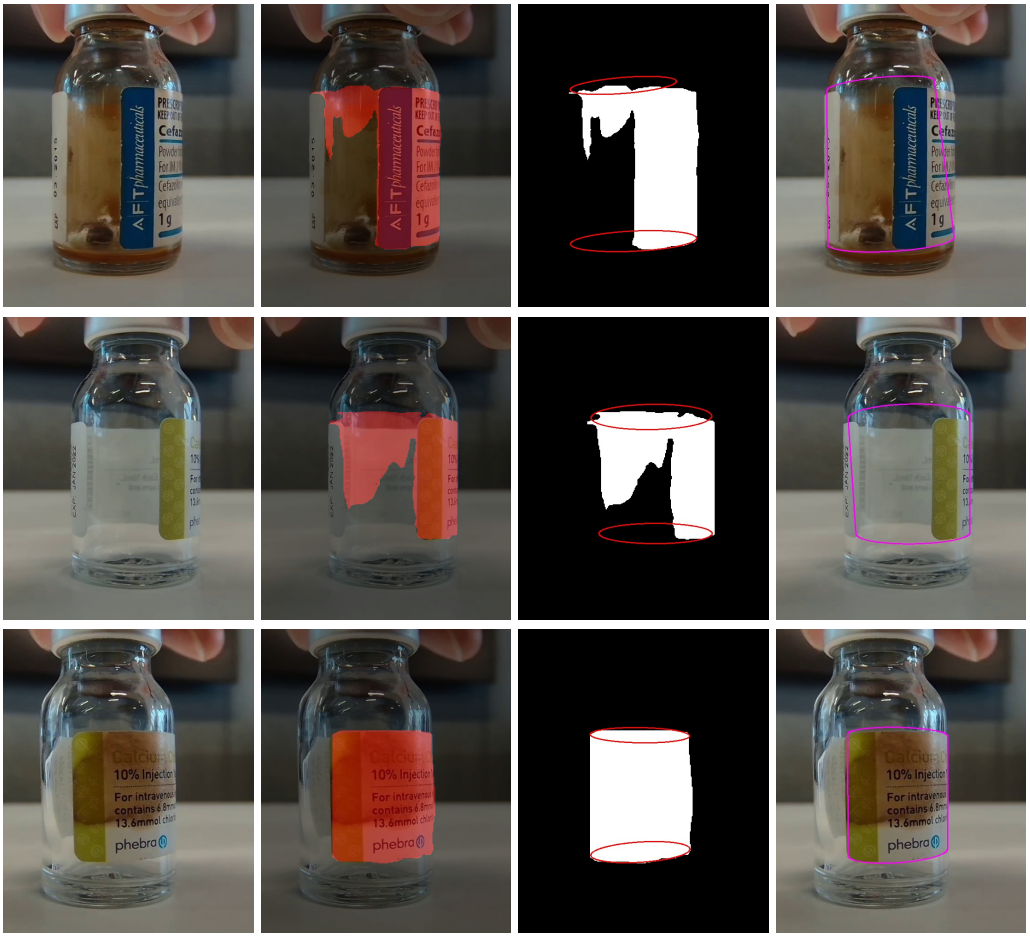


Figure 4.23: The label gap issue occurs when the camera faces the transparent back of the vial, leading to fragmented segmentation and unstable cylindrical reconstruction.

and geometric estimation. When the segmentation output is incomplete or ambiguous, both traditional and learning-based ellipse detection methods are adversely affected.

Future work will focus on improving robustness under these challenging conditions. Potential directions include leveraging multi-view image aggregation to recover full label coverage, and enhancing the segmentation model to better handle low-contrast or transparent surfaces. Additionally, integrating shape completion techniques or enforcing global cylindrical constraints during ellipse estimation could further improve reliability in the presence of partial observations.

4.6 Conclusion

This chapter presented a pipeline for dual-image reconstruction of medication labels on cylindrical containers. The process begins with YOLOv8-based instance segmentation, followed by ellipse estimation using either a traditional contour-based fitting method or an FCOS-based regression approach. The estimated ellipse parameters are then used for cylindrical unwrapping, and the resulting views are finally aligned and stitched to produce a flattened label representation suitable for downstream OCR and medication identification.

Despite the observed results under the evaluated conditions, several limitations remain. In particular, the label gap presents a challenging case, especially for transparent or partially labeled containers. When the camera is aligned with the gap between label edges, the lack of strong visual cues can lead to fragmented or incomplete segmentation. These issues affect both ellipse estimation approaches, resulting in less stable parameter recovery and, in some cases, incomplete or unreliable reconstruction.

More generally, the performance of both ellipse fitting strategies is dependent on segmentation quality. Imperfect masks (such as boundary inaccuracies, missed regions, or sensitivity to reflections) can degrade the quality of the extracted contours, which directly impacts the reliability of either contour-based fitting or FCOS regression. These geometric errors then propagate to the unwrapping stage, where small inconsistencies may accumulate and lead to local distortions in the flattened representation.

Furthermore, variations in imaging conditions such as illumination, specular reflections from glass surfaces, and viewing angles may influence both segmentation and geometric reconstruction stages. Future work will explore the robustness of the proposed pipeline under more varied imaging conditions, including changes in lighting, viewpoint, and surface reflectivity. This could involve constructing a more diverse evaluation dataset or incorporating augmentation strategies that better reflect real-world variability.

In addition, improving the robustness of segmentation and geometric estimation under such conditions may further enhance overall system stability.

Transition from Geometric Image Stitching to Text-Domain Reconstruction

Chapter 4 established the Image Stitching Pipeline (Approach 1) as a geometry-based method for medication label reconstruction, leveraging segmentation, ellipse estimation, and image stitching to recover extended label views. While effective under controlled conditions, this multi-stage pipeline introduces a degree of error propagation across sequential processing stages, where inaccuracies in early components can affect downstream reconstruction quality. In addition, the reliance on multiple geometric processing steps increases system complexity and may limit practical deployment on resource-constrained devices. The final stitching stage can also introduce visible reconstruction artifacts, which may affect the visual consistency of the reconstructed label.

Chapter 5 introduces an alternative methodological direction motivated primarily by the desire to reduce pipeline complexity and mitigate error propagation inherent in geometric reconstruction. Rather than operating on geometrically corrected images, the system transitions to a text stitching approach that reconstructs labels directly from OCR outputs obtained from dual-camera views. This shift simplifies the overall processing pipeline by reducing dependence on explicit geometric modeling, while maintaining the capacity for accurate label reconstruction.

Chapter 5

Approach 2: Textual Reconstruction of Medication Labels via Text Stitching

This chapter presents Approach 2: OCR Followed by Text Stitching, the second of two candidate approaches to cylindrical label reconstruction developed in this thesis (see Chapter 4 for Approach 1). The two approaches are evaluated against one another in Chapter 7.

This chapter presents a text stitching approach for reconstructing medication label information from dual-camera images of cylindrical containers. Motivated by the complexity of the prior image stitching pipeline (Chapter 4), which relied on label segmentation, ellipse fitting, and geometric unwrapping and stitching, we propose a simpler text-domain alternative that operates directly on OCR outputs. By running parallel Optical Character Recognition (OCR) on raw images from two cameras and merging the resulting text segments, the method reduces dependence on computationally intensive image processing and aims to provide a more lightweight pipeline suitable for deployment in resource-constrained and time-sensitive environments, such as operating theatre.

The chapter details the full pipeline, including OCR adaptation for medication

labels and text fragment alignment and merging. We evaluate the approach on a custom dataset with manually annotated ground truth, reporting performance using standard text recognition metrics on synthetic data. Qualitative results on real medication containers show that the method can improve partial text coverage and resolve certain conflicting character predictions; however, persistent OCR errors and alignment failures remain, particularly under challenging visual conditions. Finally, we discuss limitations and the trade-offs of shifting from image-domain reconstruction to text-domain fusion, and outline directions for future improvement.

5.1 Introduction

Accurate medication identification depends on extracting full text from labels. However, cylindrical containers make complete label capture difficult, since a single image often shows only a partial and geometrically distorted view. As described in Sections 2.8 and 3.1.2, this can lead to missing or misread text during OCR, resulting in incomplete or unreliable identification.

In Chapter 4, we presented an image stitching approach for reconstructing medication labels, which relies on a multi-stage geometric pipeline including segmentation, ellipse fitting, and image unwrapping and stitching. While this approach enables label reconstruction under controlled conditions, it introduces a sequence of dependent processing steps, where errors in earlier stages can propagate to subsequent stages. In addition, the reliance on explicit geometric modeling increases system complexity and introduces sensitivity to segmentation and alignment quality. A detailed evaluation of these effects is provided in Chapter 7.

To address this, we propose a shift from the image domain to the text domain. Instead of operating on pixel-level geometry, the proposed approach extracts textual information directly from dual-camera images using OCR and merges the resulting text fragments through linguistic alignment. This reformulation reduces dependence on explicit geometric processing and simplifies the pipeline by reducing the number of sequential image-processing stages. However, this shift also changes the nature of the available information: by operating on OCR outputs, the method discards visual and spatial cues such as character appearance, layout structure, and geometric consistency, which may otherwise assist in resolving ambiguities.

The text-stitching mechanism operates on the OCR outputs of the two camera views independently. Given two text strings S1 and S2, the algorithm first identifies the longest common substring (LCS) between them, assumed to correspond to the region

of the label visible in both views. Text appearing before and after this overlap in each string is then concatenated around the shared region to produce a single merged string representing the fullest available reconstruction of the label text. For example, given S1 = `AMIODARONE 150mg/3ml INJ` (missing text visible only in the second view) and S2 = `150mg/3ml INJ FOR IV USE ONLY` (missing the drug name visible only in the first view), the algorithm identifies the overlapping substring `150mg/3ml INJ` and merges the outputs into `AMIODARONE 150mg/3ml INJ FOR IV USE ONLY`, recovering text neither view alone provided.

This introduces a fundamental trade-off between reduced geometric complexity and loss of visual-spatial information. In particular, once text is extracted, certain classes of OCR errors that are consistent across both views cannot be corrected through fusion alone. We further analyse the implications of this design choice and its impact on performance in Section 5.4 and Chapter 7.

Building on this foundation, this chapter introduces a text-stitching pipeline for reconstructing complete medication label information from dual-camera views. As illustrated in Fig. 5.1, the proposed method consists of three main stages:

1. Parallel extraction of text from dual-camera using a fine-tuned OCR model for medication labels.

2. Text alignment to identify corresponding text lines across images and detect overlapping regions.

3. Text merging to consolidate aligned fragments into a single, continuous label by identifying and removing duplication while preserving unique content.

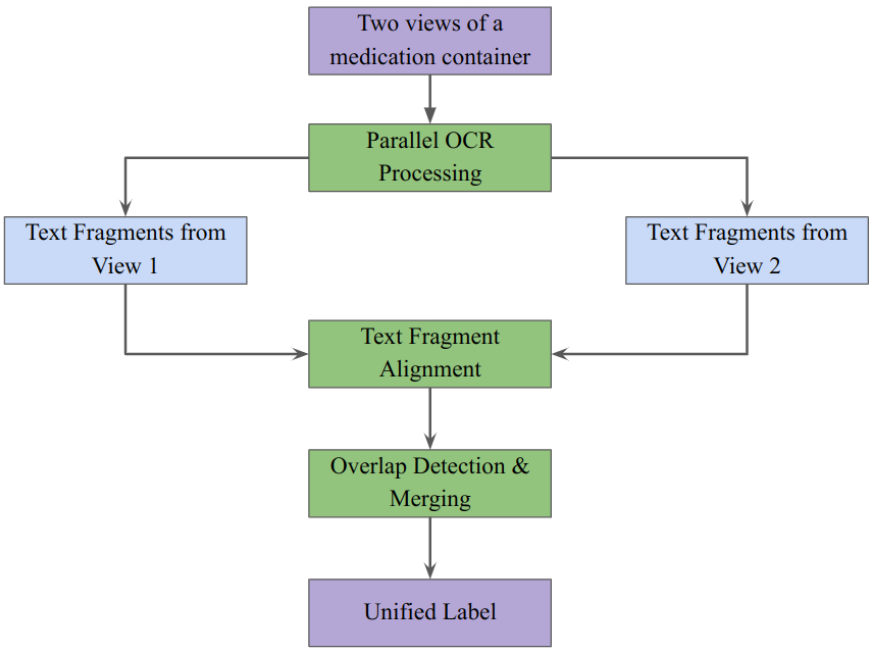


Figure 5.1: Overview of the proposed text-stitching pipeline (Approach 2) for medication label reconstruction from dual-camera images.

5.2 Related Work

5.2.1 OCR in Healthcare and Curved Surfaces

OCR converts images containing printed or handwritten text into machine-readable representations. Modern OCR pipelines typically consist of two stages: text detection, which localizes text regions within an image, and text recognition, which transcribes the detected regions into character sequences. Recent advances in deep learning, including detectors such as EAST (Zhou et al., 2017) and Differentiable Binarization (DB) (Liao, Wan, Yao, Chen & Bai, 2019), as well as recognition models based on CRNNs with CTC decoding (Shi, Bai & Yao, 2016), have improved performance in complex visual conditions. Widely used frameworks such as Tesseract (R. Smith, 2007), Google Vision API (Google, n.d.), and PaddleOCR (Du et al., n.d.) integrate these components to provide generally strong OCR performance in standard document settings.

OCR plays an important role in healthcare, where large volumes of text-based

information must be digitized and processed. However, medical records, prescriptions, and medication labels often exhibit irregular layouts, variable handwriting, and low-quality imaging conditions. To address such challenges, Karthikeyan et al. (Karthikeyan, de Herrera, Doctor & Mirza, 2021) proposed a self-supervised OCR pre-training strategy to improve robustness in skewed or incomplete clinical documents. Batra et al. (Mindee, 2021) evaluated several OCR systems, including Tesseract, EasyOCR, and DocTR, and showed that domain-specific noise and layout variability remain challenging.

While modern OCR systems perform well on flat documents, their performance degrades on curved or non-planar surfaces. Medication labels on cylindrical containers introduce additional challenges, including perspective distortion, text curvature, varying character orientation, shadows, and specular reflections from glossy or transparent materials. Even with advanced detection and recognition modules, systems such as Tesseract, Google Vision API, and PaddleOCR exhibit reduced accuracy under strong geometric deformation or extreme viewpoints (Nakada & Tanaka, 2025).

For medication packaging, these challenges are particularly relevant because labels often contain dense and safety-critical information (e.g., drug names and dosages). Curved surfaces, low contrast, and uneven illumination can lead to incomplete or inaccurate OCR outputs (Saqib et al., 2025). These limitations motivate the use of multi-view acquisition and the integration of complementary OCR outputs, which can improve text coverage while reducing reliance on more complex geometric reconstruction pipelines (Approach 1), and provide a more practical solution in real-world settings.

5.2.2 Domain Adaptation and Fine-Tuning for OCR

Unlike general scene text, pharmaceutical packaging introduces additional constraints, including small font sizes, dense information layouts, and highly specialized medical terminology. These challenges are further compounded by the containers themselves,

such as curved baselines on cylindrical vials and specular reflections from glass or plastic surfaces, which can reduce OCR performance (Baek et al., 2019; Shi et al., 2018). To address the domain gap, researchers increasingly use domain adaptation and parameter-efficient fine-tuning to align general-purpose recognition models with the specific linguistic and geometric requirements of medical labeling (Y. Zhang et al., 2019; Hüttner et al., 2025).

Lightweight text recognition models, such as PaddleOCR (Du et al., n.d.) and CRNN-CTC architectures (Shi et al., 2016), can be effectively fine-tuned on domain-specific datasets to improve recognition accuracy on pharmaceutical labels, which often exhibit small fonts, dense layouts, and specialized terminology. Synthetic data generation has emerged as a powerful strategy to mitigate data scarcity. Early work by Jaderberg et al. (Jaderberg, Simonyan, Vedaldi & Zisserman, 2014) demonstrated that large-scale synthetic text can significantly boost recognition performance. Subsequent approaches, such as SynthText (Gupta, Vedaldi & Zisserman, 2016), further improved realism by rendering text under diverse conditions, including complex lighting, perspective distortion, curvature, and blur.

5.2.3 Text Alignment and Stitching

Merging OCR predictions from dual-camera views is a key step in reconstructing a unified transcription. This process involves two main challenges: aligning partially overlapping text segments and resolving inconsistencies between them.

For alignment, classical methods such as Levenshtein distance (Levenshtein, 1966) measure full-string similarity but are less effective for fragmented OCR outputs, where only partial overlaps are present. In contrast, the Ratcliff/Obershelp algorithm (Ratcliff & Metzener, 1988) identifies longest matching subsequences and recursively aligns

remaining fragments, making it more suitable for substring-level matching under incomplete observations.

In this work, we adopt a deterministic alignment strategy for text stitching. Ratcliff/Obershelp similarity is first used for coarse matching between text sequences, providing a robust measure of overall similarity under fragmented or partially corrupted inputs. This is followed by Longest Common Subsequence (LCS) analysis, which is used to identify overlapping token sequences and determine their optimal ordering for stitching.

This combination provides robustness to incomplete OCR outputs while maintaining computational simplicity, and does not require additional model training. It is therefore consistent with the design goal of reducing pipeline complexity and supporting deployment in resource-constrained environments.

5.3 Proposed Method

This section presents the proposed method for text-based medication label reconstruction. Building on the need to reduce pipeline complexity, our approach shifts focus from combining images to structured textual representations via text stitching. The method is designed to integrate complementary information from dual camera views and produce a merged text description of the medication label. The following subsections detail the core components of the pipeline, including OCR model fine-tuning, cross-view text alignment, and text fragment fusion.



Figure 5.2: Examples of captured drug images. Multiple perspectives are recorded for each container to accommodate text curvature and label distortion.

5.3.1 Fine-tuning PaddleOCR for Medication Label Recognition

5.3.2 Dataset Collection and Annotation

A total of 83 medication containers spanning 53 medication types commonly used in anaesthesia across New Zealand hospitals were collected to construct a domain-specific dataset. These samples were provided by Auckland City Hospital (Healthpoint, n.d.) and SAFERsleep (SAFERsleep Ltd., n.d.), ensuring clinical relevance and coverage of high-risk medications encountered in operating theatre environments (Fig. 5.2).

From these containers, videos were captured and frames were extracted following the acquisition workflow described in Section 4.3.1.2, ensuring consistent capture conditions and sufficient image quality. From this pool, 700 images were sampled for OCR fine-tuning, with 500 used for training and 200 for validation. The dataset

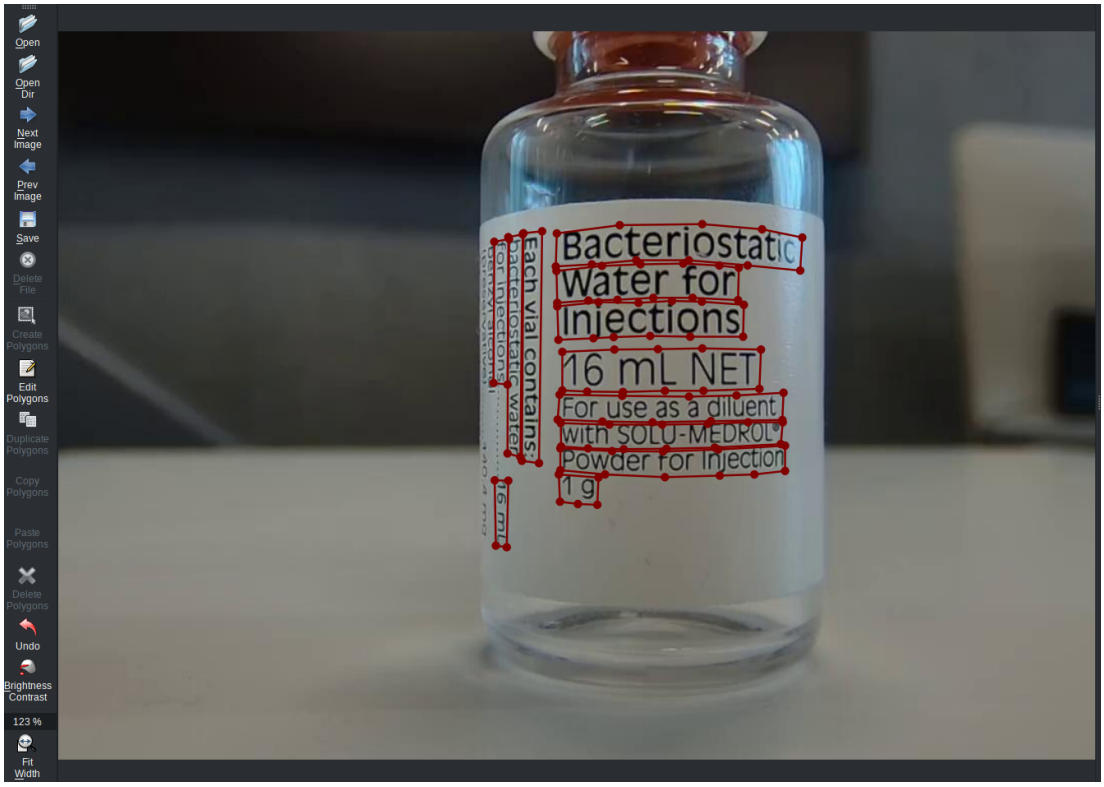


Figure 5.3: Example of text-region annotation using LabelMe.

includes medication containers with varying degrees of curvature and surface reflectivity. Although the dataset size is limited by the availability of physical samples and the cost of manual annotation, it is considered sufficient for domain adaptation when combined with pre-trained models and supplementary public datasets(CTW1500 (Yuliang et al., 2017) and ICDAR 2015 (Karatzas et al., 2015), as described in Section 3.3.4).

Text regions were manually annotated using LabelMe (Wada, n.d.) with polygon-based boundaries to capture curved, rotated, and partially occluded text. This tool was selected for its flexibility in supporting free-form polygon annotations, which is particularly suitable for irregular text shapes on cylindrical surfaces. Annotations were designed to closely follow visible text regions, including non-linear baselines and perspective-induced distortions. Regions were included only when characters were reasonably legible, providing a consistent criterion for ground truth quality (Fig. 5.3).

Training and validation splits were performed at the container level to prevent data

leakage, ensuring that images from the same physical container do not appear in both sets. This setup provides a more realistic evaluation of generalisation across different instances.

5.3.2.1 PaddleOCR Pipeline Architecture

Given the annotated dataset, a robust OCR pipeline is required to detect and recognise text under the challenging conditions introduced by cylindrical containers (Fig. 5.4). In this work, PaddleOCR (Du et al., n.d.) is adopted due to its modular design, availability of pre-trained models, and demonstrated performance in real-world text recognition tasks. Its separation of detection and recognition stages allows targeted adaptation to domain-specific challenges.

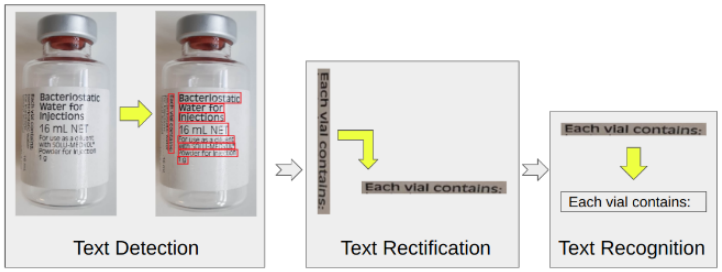


Figure 5.4: Workflow of the PaddleOCR text detection and recognition pipeline.

Text Detection (fine-tuned). The detection stage employs the Differentiable Binarization (DB) algorithm (Liao et al., 2019), which performs text region segmentation in complex scenes. This component is fine-tuned using the medication-specific dataset with polygon-based annotations that delineate curved and irregular text regions. Fine-tuning focuses on improving localization under curvature, distortion, and partial occlusion, which are common in cylindrical packaging. Following detection, the extracted regions are rectified into approximately horizontal representations; however, this text rectification step is not further fine-tuned and is applied using the default PaddleOCR transformation.

Text Recognition (pre-trained, unchanged). The recognition stage uses a CRNN-based architecture (Shi et al., 2016), combining convolutional feature extraction with recurrent sequence modeling. In this work, the recognizer is not further fine-tuned. This design choice reflects a trade-off between domain adaptation and resource constraints. Constructing a sufficiently large and accurately annotated recognition dataset would be labour-intensive, and the available annotations primarily support detection rather than sequence-level transcription.

While detection errors can affect downstream OCR performance, experimental observations suggest that a significant portion of residual errors in the pipeline arises from recognition inaccuracies, particularly under conditions of low contrast, blur, or severe geometric distortion. As such, the decision to retain the pre-trained recognizer priorities simplicity and deployability, while accepting that some recognition errors may remain uncorrected. This limitation is further discussed in Section 5.4.1.2.

Overall, the pipeline adopts a modular strategy: the detection component is adapted to the domain to improve text localisation, while the recognition and rectification components remain general-purpose. This design reduces training complexity while maintaining reasonable performance across irregular, curved, and densely structured medication labels.

5.3.3 Alignment of Text Fragments

Given OCR outputs from two camera views, the pipeline integrates complementary text fragments to reconstruct a more complete label. Each view provides only a partial observation of the cylindrical surface, and extracted fragments may differ in completeness, ordering, and quality. Alignment is therefore required to identify corresponding fragments prior to merging. Specifically, given two sets of OCR text fragments X and Y extracted from dual-camera views, we define:

Algorithm 1 Text Fragment Alignment using Ratcliff/Obershelp with Threshold

Input: Two sets of strings $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$, threshold τ
Output: A set of matched pairs $\{(x_i, y_j)\}$

```
MatchedPairs ← ∅
for each x_i in X do
    best_score ← 0
    best_match ← NULL
    for each y_j in Y do
        similarity_score ← Sim(x_i, y_j)
        if similarity_score > best_score then
            best_score ← similarity_score
            best_match ← y_j
        end if
    end for
    if best_score ≥ τ then
        Add (x_i, best_match) to MatchedPairs
    end if
end for
```

$$X = \{x_1, x_2, \ldots, x_n\}, \tag{5.1}$$

$$Y = \{y_1, y_2, \ldots, y_m\}. \tag{5.2}$$

Here, X and Y denote collections of text fragments from the first and second camera views. Due to viewpoint variations, partial occlusions, and OCR noise, these fragments may differ in completeness, ordering, and quality across views. Without alignment, merging text could introduce duplication, omission, or misplacement of important medication information.

To establish correspondences, we adopt the Ratcliff/Obershelp algorithm (Ratcliff & Metzener, 1988), which measures similarity based on the longest common subsequences between two strings. Unlike edit-distance methods that emphasise global string differences, this approach provides substring-sensitive matching, which is suitable for

fragmented OCR outputs where only partial overlaps are present.

Specifically, for each fragment $x_i \in X$, we compute its similarity with all $y_j \in Y$ and select the one with the highest score exceeding τ, as:

$$Sim(x_i, y_j) = 2 \times \frac{M}{|x_i| + |y_j|}, \tag{5.3}$$

where M is the number of matching characters, and $|x_i|$, $|y_j|$ denote string lengths. The score ranges from 0 (no overlap) to 1 (exact match).

The threshold τ filters low-confidence matches and reduces incorrect alignments. In this work, τ is set empirically to 0.65 based on preliminary validation, balancing the inclusion of valid partial matches against the rejection of spurious ones. Given that each label typically produces only tens of fragments, exhaustive pairwise comparison remains computationally manageable.

Algorithm 1 summarises the alignment procedure. For each fragment in X, the method identifies the most similar fragment in Y and retains the match only if it exceeds the threshold. This deterministic strategy provides a simple and efficient mechanism for aligning fragmented OCR outputs without requiring additional training or complex modelling.

5.3.4 Text Fragment Merging

After aligning OCR fragments from dual-camera views, overlapping content must be detected and consolidated to form a unified transcription. Due to viewpoint differences and OCR noise, fragments may contain repeated, truncated, or partially distorted substrings. A merging strategy is therefore required to reduce redundancy while preserving complementary information across views.

We employ the Longest Common Subsequence (LCS) algorithm (Bergroth, Hakonen & Raita, 2000) to identify overlapping regions between aligned fragments. LCS extracts

the longest sequence of characters shared by both strings in the correct order, without requiring contiguity. This property makes it suitable for OCR outputs where characters may be missing or fragmented due to curvature, occlusion, or detection boundaries.

The choice of LCS reflects a trade-off between robustness and simplicity. Compared to local alignment methods such as Smith-Waterman algorithm (T. F. Smith, Waterman et al., 1981), LCS does not require additional scoring parameters or tuning, which reduces implementation complexity. In contrast, LCS operates directly on character sequences and provides a deterministic mechanism for identifying overlap regions. This aligns with the design goal of maintaining a simple and deployable pipeline without introducing additional model complexity.

Given two aligned fragments s_1 and s_2 of lengths m and n, the LCS is computed using dynamic programming as shown in Algorithm 2. A $(m + 1) \times (n + 1)$ table dp stores the length of the longest common subsequence for all prefixes, and a traceback procedure reconstructs the shared subsequence.

To construct the merged output, the LCS is treated as an overlap anchor. The non-overlapping prefix of s_2 and suffix of s_1 are concatenated around this shared subsequence. In cases where multiple valid overlaps exist, a deterministic tie-breaking rule is applied by selecting the earliest matching position in s_2, ensuring consistent and reproducible merging.

For example, consider two OCR fragments from the same medication label: s_1 = `um Chlo` and s_2 = `Calcium Ch`. The LCS computed is `um Ch`, indicating the overlapping portion. Using this information, the unique prefixes and suffixes of each fragment are preserved while removing duplication, resulting in the merged transcription `Calcium Chlo`, as illustrated in Fig. 5.5.

Another example of a more challenging case illustrates typical OCR noise and conflicting characters:

Algorithm 2 Longest Common Subsequence (LCS)

Input: Two OCR fragments s_1 and s_2 of lengths m and n
Output: LCS length and LCS string
Initialize $dp[0..m][0..n] \leftarrow 0$
for i = 1 to m **do**
    for j = 1 to n **do**
        if $s_1[i-1] = s_2[j-1]$ **then**
            $dp[i][j] \leftarrow dp[i-1][j-1] + 1$
        else
            $dp[i][j] \leftarrow \max(dp[i-1][j], dp[i][j-1])$
        end if
    end for
end for

Traceback:
$i \leftarrow m$, $j \leftarrow n$, $LCS_string \leftarrow$ ""
while $i > 0$ and $j > 0$ **do**
    if $s_1[i-1] = s_2[j-1]$ **then**
        Prepend $s_1[i-1]$ to LCS_string; $i \leftarrow i-1$, $j \leftarrow j-1$
    else if $dp[i-1][j] > dp[i][j-1]$ **then**
        $i \leftarrow i-1$
    else
        $j \leftarrow j-1$
    end if
end while
return $dp[m][n]$, LCS_string

- s_1 = `Cefaz0lin 1g`
- s_2 = `Cefazolin 1 q`

In this example, the LCS captures the shared subsequence `Cefazolin 1`, despite character-level discrepancies (e.g., "0" vs "o", "g" vs "q"). However, because the method operates purely on sequence matching without semantic correction, the merged output may contain duplicated or inconsistent fragments. In this case, the resulting string can take the form:

`Cefazolin 1 Cefazlin 1 1g`

		C	a	l	c	i	u	m	␣	C	h
	0	0	0	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	**1**	1	1	1	1
m	0	0	0	0	0	0	1	**2**	2	2	2
␣	0	0	0	0	0	0	1	2	**3**	3	3
C	0	**1**	1	1	1	1	1	2	3	**4**	4
h	0	1	1	1	1	1	1	2	3	4	**5**
l	0	1	1	**2**	2	2	2	2	3	4	5
o	0	1	1	2	2	2	2	2	3	4	5

Figure 5.5: Example of LCS-based overlap detection between OCR outputs s_1 = `um Chlo` and s_2 = `Calcium Ch`. The LCS identifies the overlapping portion `um Ch`.

This highlights an important limitation: overlap removal is not equivalent to semantic correction. While the method can identify shared structure and improve coverage, it does not resolve conflicting characters or eliminate duplicated tokens. When OCR errors are present in both inputs, they may be preserved or compounded in the merged result.

Overall, this approach is intended as a lightweight and deterministic baseline for multi-view OCR fusion. It improves text coverage by consolidating complementary fragments across views, but does not explicitly address recognition errors. As such, improvements should be interpreted primarily as gains in text coverage rather than intrinsic OCR accuracy. Further limitations of this design are discussed in Section 5.4.3.

5.4 Experimental Results

The evaluation proceeds in three stages: (1) quantitative assessment of the fine-tuned OCR detection model on validation data, (2) controlled evaluation of the stitching algorithm on a synthetic dataset with precise ground truth, and (3) qualitative assessment on real medication containers to characterize practical performance and failure modes.

5.4.1 PaddleOCR Fine-tuning Evaluation

5.4.1.1 Quantitative Results on Validation Set

To improve generalisation beyond the domain-specific dataset, training incorporates two public benchmarks: CTW1500 (Yuliang et al., 2017), which contains curved text samples, and ICDAR 2015 (Karatzas et al., 2015), which includes diverse orientations and complex backgrounds. These datasets are not mixed uniformly with the medication data; instead, a sequential training strategy is adopted.

The model is first pre-trained on the combined public datasets to learn general text detection features across varied layouts and imaging conditions. It is then fine-tuned on the medication-specific dataset using a reduced learning rate, allowing adaptation to domain-specific characteristics such as curvature, specular reflections, and dense label layouts, while preserving previously learned representations. A summary of dataset usage is provided in Table 5.1.

Table 5.1: Datasets used for OCR training and fine-tuning

Dataset	**Train**	**Val**	**Role**
ICDAR 2015 (Karatzas et al., 2015)	1000	500	Pre-training (general scene text)
CTW1500 (Yuliang et al., 2017)	1000	500	Pre-training (curved text)
Medication dataset	500	200	Domain-specific fine-tuning

The medication dataset consists of images captured under controlled conditions, with polygon-based annotations used to train the text detection module. Fine-tuning

focuses exclusively on detection, as the available annotations delineate text regions rather than providing sequence-level transcription required for recognition training.

Table 5.2: Quantitative performance of the pretrained baseline versus fine-tuned PaddleOCR detection model on the 200-image medication label validation set.

Model	**Precision**	**Recall**	**H-mean**
Pretrained baseline	0.7821	0.8114	0.7965
Fine-tuned (ours)	**0.8062**	**0.8300**	**0.8179**
Relative improvement	+3.1%	+2.3%	+2.7%

Table 5.2 shows the performance of the baseline and fine-tuned models on the validation set. The fine-tuned model achieves an H-mean of 0.8179, representing a relative improvement of +2.7% over the pre-trained baseline. Precision improves to 0.8062, indicating fewer false positives, while Recall increases to 0.8300, suggesting better localization of fragmented or distorted text regions. These gains are consistent with the domain-specific challenges of cylindrical medication labels, including curvature and reflective surfaces. However, the improvements are modest, indicating that domain adaptation provides incremental rather than substantial gains in detection performance.

Only the text detection module is fine-tuned, while the recognition module remains unchanged. This decision reflects both practical constraints and a design trade-off. The available dataset does not include sufficiently large or consistently annotated transcription data for reliable recognition fine-tuning, and extending annotation to sequence-level labels would be resource-intensive. While the pre-trained recognizer performs adequately on standard alphanumeric text, qualitative results (Section 5.4.1.2) show that recognition errors persist under challenging conditions such as blur, low contrast, and distortion. As such, this design priorities adaptation of text localization, while acknowledging that recognition remains a limiting factor in overall OCR performance.

Given the relatively small validation set (200 images), we report performance using standard point estimates without statistical confidence intervals or significance testing.

This evaluation scope is consistent with the exploratory nature of the study and provides a practical indication of performance trends under the proposed setting.

5.4.1.2 Qualitative Evaluation and Failure Mode Analysis

Fig. 5.6 presents representative OCR outputs from the validation set. Overall, the system is able to detect and recognise text reliably in regions that are approximately front-facing, including medication names, dosage information, and batch identifiers. In these regions, character sequences are largely preserved, suggesting that recognition quality remains stable when geometric distortion is limited.

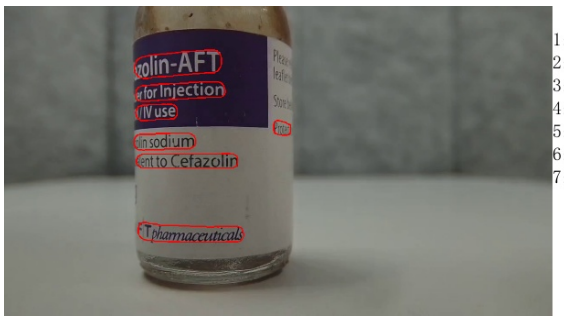
For instance, subfigures (a–d) illustrate these behaviours across different container types. In Fig. 5.6(a), most alphanumeric sequences, such as `433220`, `05-2024`, and `06-2019`, are correctly recognised. However, certain components, such as the missing keyword `EXP` and occasional misread characters (e.g., a character resembling `B`), indicate sensitivity to local distortion and font variation. In more heavily warped regions, incorrect transcriptions such as `iviniocronoy` are observed, reflecting the impact of geometric deformation on recognition.

In contrast, Fig. 5.6(b) shows more pronounced distortion, resulting in fragmented or merged text. Examples include `erforInjection`, `l/IVuse`, and `alent to Cefazolin`. Although some tokens are degraded, key information such as the medication name `zolin-AFT` and its relation to `Cefazolin` remains partially recoverable. These errors are primarily associated with boundary regions, where non-uniform character spacing and curvature affect the OCR output.

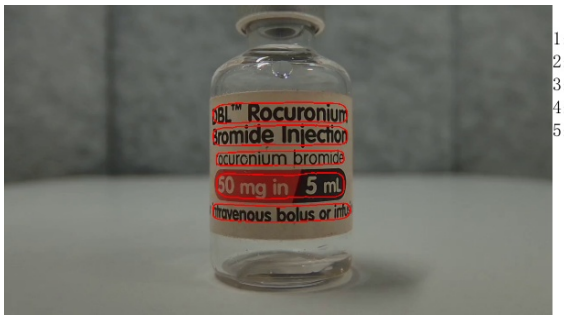
Fig. 5.6(c) demonstrates improved recognition under more favourable geometric conditions. Medication names such as `Rocuronium`, descriptors like `Bromide Injection`, and dosage information (e.g., `50 mg in 5 mL`) are largely preserved, with only minor truncation in phrases such as `ntravenous bolus or infus`. This suggests that recognition performance improves when the visible label region more



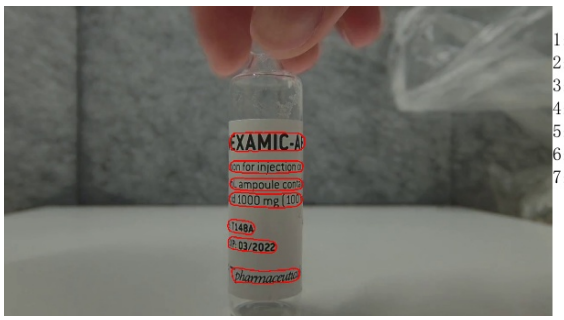
(a)



(b)



(c)



(d)

Figure 5.6: Some OCR results from the validation dataset.

closely approximates a planar surface.

Finally, Fig. 5.6(d) illustrates a limiting case where the visible label region is narrow due to container geometry. This results in incomplete phrases such as `tion for injection or` and `nL ampoule contal`. Even so, certain structured elements, such as the expiration date `XP:03/2022`, remain detectable despite the limited context.

These observations relate to the dual view setting used in this work. Since OCR errors are more frequent near the edges of each view, capturing the container from a different viewpoint can reveal these regions under a more front facing perspective, where distortion is reduced and text becomes more readable.

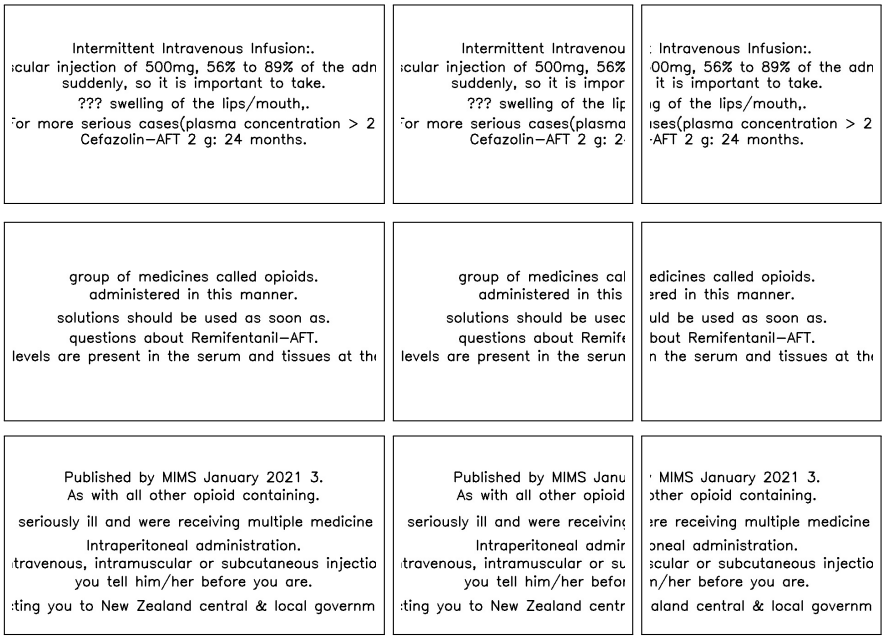
However, this effect depends on the viewpoints. If both images capture similar regions, especially highly curved boundaries, the errors are likely to be similar and cannot be corrected through stitching. This limits the benefit of combining views and highlights the need for sufficient viewpoint variation to improve text coverage.

5.4.2 Synthetic Dataset for Text Stitching Evaluation

To systematically evaluate the text stitching algorithm, we created a synthetic dataset that enables precise control over text content, overlap ratio, and simulated OCR error patterns. While real medication labels provide authentic data, they lack precise ground truth for quantitative reconstruction accuracy.

5.4.2.1 Synthetic Dataset Generation

A synthetic dataset was created to systematically evaluate the robustness of the text stitching algorithm under controlled conditions. Real medication labels provide authentic data, but they do not offer reliable ground truth for evaluating reconstruction accuracy at the text level, nor do they allow controlled variation of key factors such as overlap ratio, fragmentation level, or OCR noise. The synthetic dataset therefore



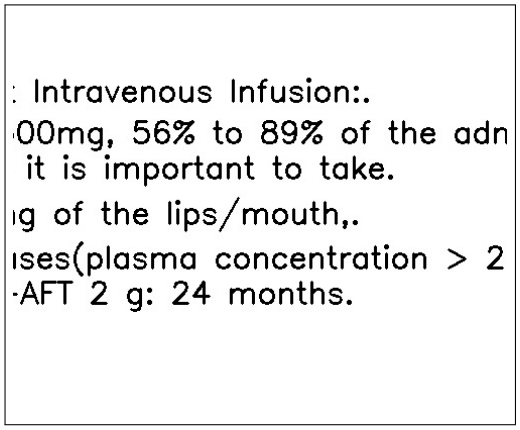
(a) (b) (c)

Figure 5.7: Example images from the synthetic dataset: (a) full text image, (b) left half, and (c) right half.

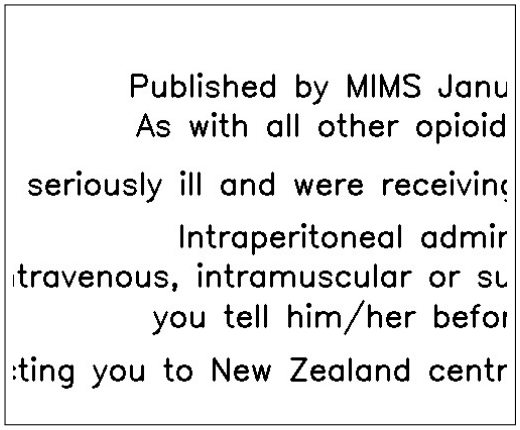
isolates the text fusion problem and enables controlled experimentation on alignment and merging behaviour.

It is important to emphasise that this dataset evaluates the text stitching mechanism in isolation, rather than the full complexity of medication packaging OCR. In particular, it does not model detection errors, complex label layouts, or heterogeneous visual conditions encountered in real pharmaceutical environments.

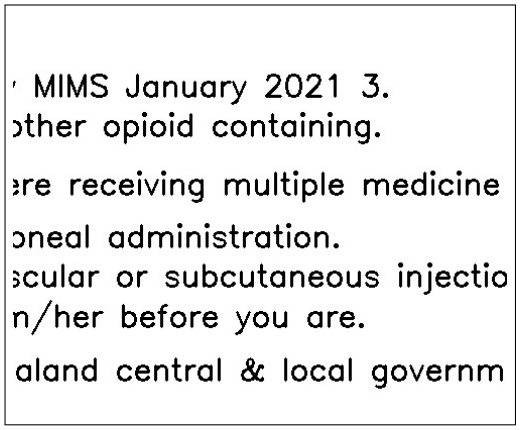
The dataset is generated using a four-stage pipeline that simulates dual-view fragmented text acquisition. First, domain-relevant sentences are collected from medical references, drug monographs, and pharmaceutical safety documentation to ensure realistic vocabulary and syntax. These sentences are then rendered onto a white canvas using OpenCV with a single font type and fixed font size, while random vertical padding is introduced to loosely approximate spacing variations found in real labels. Each rendered text image is then split vertically into two overlapping fragments with controlled overlap ratios ranging from 10 to 50 percent, simulating partial observations of cylindrical



⟶ Intravenous Infusion:
⟶ 00mg, 56% to 89% of the adr
⟶ it is important to take.
⟶ g of the lips/mouth..
⟶ ses(plasma concentration > 2
⟶ AFT 2 g: 24 months



⟶ PubIished by MIMS Janu
⟶ As with all other opioic
⟶ seriously ill and were receivin
⟶ Intraperitoneal admi
⟶ travenous. intramuscular or si
⟶ you tell him/her befo
⟶ ting you to New Zealand centr



⟶ MIMS January 2021 3
⟶ ther opioid containing.
⟶ re receiving multiple medicine
⟶ neal administration.
⟶ cular or subcutaneous injectic
⟶ /her before you are.
⟶ aland central & local governm

Figure 5.8: Images and corresponding OCR results in the synthetic text dataset.

surfaces under dual-camera acquisition. Finally, a fine-tuned PaddleOCR model is applied independently to each fragment, producing realistic noisy OCR outputs that include typical recognition errors such as missing characters, spacing inconsistencies, and fragmented tokens.

The resulting dataset contains 5,000 instances. Each instance consists of a clean ground-truth full text string, two partially overlapping image fragments, and the corresponding OCR outputs from each fragment. This structure enables quantitative evaluation of the proposed stitching algorithm against known ground truth.

While this synthetic dataset provides a controlled evaluation environment, it simplifies several aspects of real-world medication labels. In particular, it assumes a single font family, fixed font size, and simplified layout structure. In contrast, real pharmaceutical packaging contains multiple fonts, varying text sizes, dense multi-line layouts, and complex visual effects such as curvature, glare, and shadows. As a result, the synthetic evaluation primarily assesses the correctness of the string-level alignment and merging strategy under controlled conditions, rather than the full end-to-end robustness of OCR in realistic clinical environments.

5.4.2.2 Quantitative Evaluation

We evaluate the proposed text stitching algorithm on the synthetic dataset using Character Error Rate (CER), Word Error Rate (WER), and overlap recovery metrics (Precision, Recall, and F1-score). CER and WER measure reconstruction accuracy at the character and word levels, respectively, while the overlap recovery metrics quantify the correctness of fragment alignment and merging.

To evaluate alignment performance, we compute Precision, Recall, and F1-score by comparing predicted alignments with ground-truth correspondences. Precision measures the proportion of correct alignments among all predicted alignments, Recall measures the proportion of ground-truth alignments that are successfully recovered, and

F1-score provides a balanced summary of both metrics.

Table 5.3 summarizes the results. The method achieves a CER of 0.027, indicating relatively low character-level deviation from ground truth under synthetic OCR noise conditions. The WER of 0.157 reflects word-level discrepancies arising from a combination of OCR errors and imperfect reconstruction at fragment boundaries. While this indicates that most words are correctly recovered, residual errors remain, particularly in cases where OCR outputs contain missing or partially corrupted tokens.

The overlap recovery metrics show Precision of 0.844 and Recall of 0.852, resulting in an F1-score of 0.848. This suggests that the algorithm performs reliably in identifying and merging corresponding fragments, although some incorrect or missed alignments persist.

These results indicate that the proposed method can effectively reconstruct fragmented text under controlled conditions, with strong performance in overlap detection and moderate robustness to OCR noise. However, it is important to note that WER = 0.157 may still be non-trivial in safety-critical contexts such as medication identification, where even small word-level errors can affect interpretation. A full assessment of downstream identification accuracy is provided in Chapter 7.

Table 5.3: Average performance metrics for the text stitching algorithm on the synthetic dataset.

Metric	**Average Value**
Character Error Rate (CER)	0.027
Word Error Rate (WER)	0.157
Precision	0.844
Recall	0.852
F1-Score	0.848

group of medicines ca + edicines called opioids.
→ **group of medicines called opioids.**

administered in this + red in this manner.
→ **administered in this manner.**

solutions should be usec + Jd be used as soon as.
→ **solutions should be used as soon as.**

guestions about Remif + oout Remifentanil-AFT.
→ **guestions about Remifentanil-AFT.**

evels are present in the serur + n the serum and tissues at th
→ **evels are present in the serum and tissues at th**

PubIished by MIMS Janu - MIMS January 2021 3
→ **PubIished by MIMS January 2021 3**

As with all other opioic - ther opioid containing.
→ **As with all other opioid containing.**

seriously ill and were receivin - re receiving multiple medicine
→ **seriously ill and were receiving multiple medicine**

Intraperitoneal admi - neal administration.
→ **Intraperitoneal administration.**

travenous. intramuscular or si - cular or subcutaneous injectic
→ **travenous. intramuscular or subcutaneous injectic**

Figure 5.9: Examples of text stitching from fragmented OCR outputs. The fragments (in blue) are combined to produce complete sentences (in bold).

5.4.2.3 Qualitative Evaluation

Qualitative examples in Fig. 5.9 illustrate how the algorithm reconstructs text from noisy fragment pairs, even when the OCR output contains partial words or misaligned token boundaries. In many cases, the method identifies the overlap region and avoids duplicating repeated sections, producing a coherent merge.

Fig. 5.10 presents side-by-side comparisons between ground truth text and stitched results. The reconstructed text generally follows the original structure despite minor OCR distortions. For instance, a fragmented pair such as `cular injection of`

```
 ⟶ 1.  Intermittent
Intravenous Infusion:
⟶ 2.  cular injection of
500mg, 56% to 89% of the
adr
⟶ 3.  suddenly, so it is
important to take
⟶ 4.  or more serious
casesiplasmaconcentration>
z
⟶ 5.  cefazolin-AFT 2 a:
24 months
```

```
 ⟶ 1.  group of
medicines called opioids
⟶ 2.  administered in
this manner.
⟶ 3.  solutions should
be used as soon as.
⟶ 4.  auestions about
Remifentanii-AFT
⟶ 5.  evels are present
in the serum and tissues
at th
```

```
 ⟶ 1.  PubIished by MIMS
January 2021
⟶ 2.  As with all other
opioid containing.
⟶ 3.  Seriously ill and
were receiving multiple
medications.
⟶ 4.  Intraperitoneal
administration.
⟶ 5.  Intravenous,
intramuscular, or
subcutaneous injection.
⟶ 6.  Ting you to New
Zealand central & local
government.
```

(a) Processed Result

```
 ⟶ 1.  Intermittent
Intravenous Infusion:
⟶ 2.  cular injection of
500mg, 56% to 89% of the
adr
⟶ 3.  suddenly, so it is
important to take.
⟶ 4.  or more
serious cases(plasma
concentration > 2
⟶ 5.  CefazoIin-AFT 2 g:
24 months.
```

```
 ⟶ 1.  group of
medicines called opioids.
⟶ 2.  administered in
this manner.
⟶ 3.  solutions should
be used as soon as.
⟶ 4.  questions about
Remifentanil-AFT.
⟶ 5.  levels are present
in the serum and tissues
at th
```

```
 ⟶ 1.  Published by MIMS
January 2021
⟶ 2.  As with all other
opioid containing.
⟶ 3.  Seriously ill and
were receiving multiple
medications.
⟶ 4.  Intraperitoneal
administration.
⟶ 5.  Intravenous,
intramuscular, or
subcutaneous injection.
⟶ 6.  Ting you to New
Zealand central & local
government.
```

(b) Ground Truth

Figure 5.10: Comparison between processed results (left) and ground truth (right)

`500mg, 56% to 89% of the adr` can be aligned and merged into a more complete sentence. These examples suggest that the algorithm can recover overlapping content, although it does not explicitly correct recognition errors.

However, the comparative analysis also highlights consistent failure modes within the stitching process. A primary limitation arises from OCR errors, where character substitutions, such as `PubIished` for `Published` (third row), or lexical distortions like `Remifentanii-AFT` (second row), persist across multiple viewpoints. When the same error appears in all input fragments, the merging algorithm has no reliable reference to resolve the correct form, and the error is retained in the output.

Overall, the quantitative and qualitative results indicate that the proposed stitching pipeline can improve text continuity by combining overlapping fragments from multiple views. The method helps recover portions of text that are only partially visible in individual OCR outputs, leading to more complete sequences at the phrase or sentence level. However, the effectiveness of the stitching process depends on the quality of the OCR outputs. When recognition errors are severe, or when similar errors appear across multiple views, the method cannot resolve these inconsistencies and may retain duplicated or corrupted text.

5.4.3 Application to Real Medication Container Images

After validating the stitching pipeline on synthetic data, we applied the method to real dual-camera images of cylindrical medication containers. These experiments evaluate system performance under practical conditions involving different containers, curved surfaces, and partially visible text. For each view, text was first extracted using the fine-tuned PaddleOCR model, and the resulting fragments were then merged into a unified transcription.

5.4.3.1 Quantitative Evaluation on Real Dataset

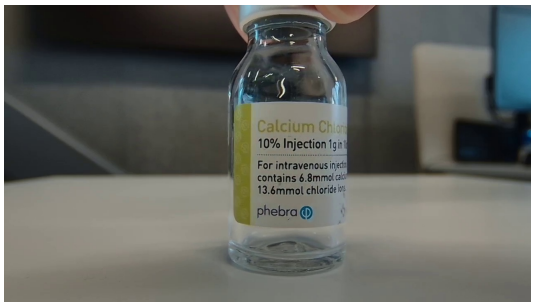
We constructed a dataset of 40 image pairs from 10 medication types (4 pairs per type). Ground truth transcription (GT) was assembled by consolidating readable text across both views of each pair. To ensure reliability, the procedure was conservative: OCR outputs from individual images served as initial candidates, which were manually verified by visual inspection. Text was included in GT only if clearly visible in at least one view, while regions with severe distortion or ambiguity (e.g., label edges) were excluded. This approach acknowledges that real medication labels do not provide pixel-level ground truth due to cylindrical geometry and partial visibility.

We evaluated single-view OCR (View 1 and View 2 separately) and the proposed stitching method against GT using word-level metrics computed after case normalization and punctuation removal. Character Error Rate (CER) and Word Error Rate (WER) measure transcription accuracy, while precision and coverage (recall) measure information recovery. F1-score combines these measures.

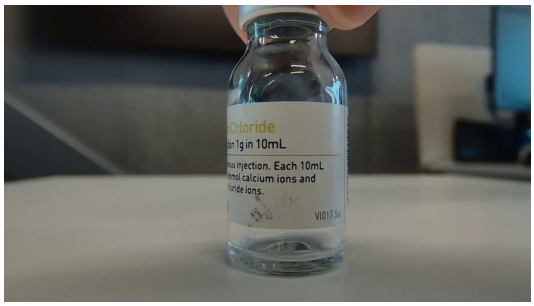
Table 5.4 shows the results of the evaluation. Specifically, single-view OCR achieves high precision (0.81-0.85), meaning that recognized text is accurate. However, coverage remains low (0.53-0.59), indicating that 41-47% of label text is not recovered due to partial visibility on cylindrical surfaces. The stitching method improves coverage to 0.937 (recovery of 93.7% of label text) while maintaining comparable precision (0.877). Consequently, CER decreases to 0.112 and WER to 0.177, indicating more complete label reconstruction. These results show that the primary benefit of dual-view stitching is recovering text invisible in single views, rather than correcting OCR recognition errors.

Table 5.4: Single-view OCR and proposed stitching method on 40 medication label image pairs.

Method	CER ↓	WER ↓	Precision ↑	Coverage ↑	F1 ↑
View 1 OCR	0.531	0.608	0.854	0.586	0.687
View 2 OCR	0.658	0.701	0.812	0.530	0.631
Proposed stitching	**0.112**	**0.177**	**0.877**	**0.937**	**0.903**



(a)



(b)

```
⟶ Calcium
Chlorid
⟶ 10% Injection
1g in 10
⟶ For
intravenous
injection
⟶
contains6.8mmol
calciur
⟶ 13.6mmol
chloride ions
⟶ phebra
```

(c)

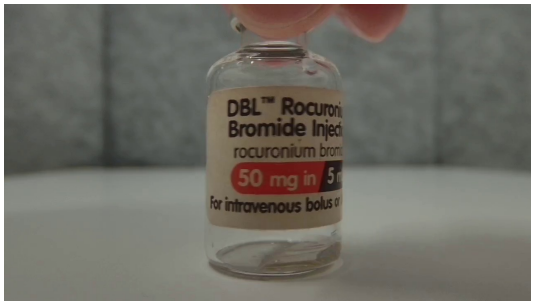
```
⟶ Chloride
⟶ "on 1g in
10mL
⟶ cus
injection.Each
10ml
⟶ umol calcium
ions and
⟶ Morideions.
⟶ V1017.5a
```

(d)

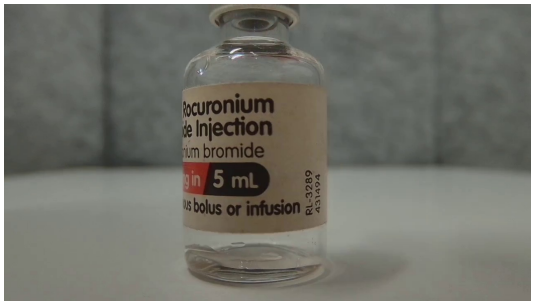
```
⟶ Calcium
Chloride
⟶ 10% Injection
1g in 10mL
⟶ For
intravenous
injections.
⟶
contains6.8mmol
calcium ions and
⟶ 13.6mmol
chloride ion.Each
10ml
⟶ phebra
⟶ V1017.5a
```

(e)

Figure 5.11: Illustration of OCR processing and text merging: (a, b) left and right images, (c, d) corresponding OCR results, and (e) stitched text result.



(a)



(b)

```
⟶ DBLRocuroni
⟶ Bromide Injed
⟶
rocuroniumbrom
⟶ 50mg in 5
⟶ For
intravenous bolus
o
```

(c)

```
⟶ Rocuronium
⟶ de Injection
⟶ ium bromide
⟶ g in 5 mL
⟶ Us bolus or
infusion
```

(d)

```
⟶ DBLRocuronium
⟶ Bromide
Injection
⟶
rocuroniumbromide
⟶ 50mg in 5 mL
⟶ For
intravenous bolus
or infusion
```

(e)

Figure 5.12: Illustration of OCR processing and text merging: (a, b) left and right images, (c, d) corresponding OCR results, and (e) stitched text result.

5.4.3.2 Qualitative Results: Two-View Stitching

We applied the proposed text stitching method to OCR outputs obtained from real dual-camera images of cylindrical medication container labels. For each image, text was extracted independently using the fine-tuned PaddleOCR model, after which the detected fragments were merged to produce a more complete representation of the label content. This dual-view setup improves coverage by capturing complementary regions of the curved surface, mitigating issues caused by perspective distortion, occlusion, and partial visibility.

The qualitative results in Figs. 5.11 and 5.12 show that the proposed method can effectively consolidate fragmented and partially overlapping OCR outputs from two

views, improving overall text completeness.

For example, in Fig. 5.11, the left view contains `10% Injection 1g in 10`, while the right view contains the corrupted fragment `on 1g in 10mL`. After stitching, the merged output becomes `10% Injection 1g in 10mL`, showing that a more complete phrase can be recovered by combining complementary inputs.

A more challenging case in Fig. 5.11 shows the left view correctly recognizes `13.6mmol chloride ions`, while the right view produces the corrupted fragment `Morideions`. The merged result becomes `13.6mmol chloride ion.Each 10ml`, where parts of both inputs are retained but inconsistencies remain. In this case, the output does not fully recover the ground truth `13.6mmol chloride ions.`, but instead reflects a mixture of partial correction and residual OCR errors.

These examples indicate that the stitching process is effective for improving text continuity, but it does not perform semantic correction. When both views contain errors or inconsistent fragments, the merged result may still inherit or combine these inaccuracies rather than fully resolving them.

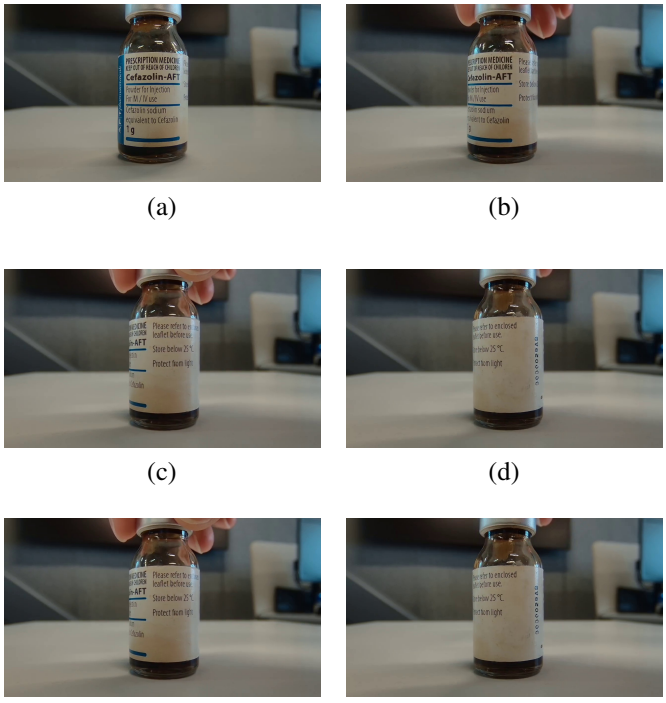
Similarly, in Fig. 5.12, the left and right views capture complementary portions of the label, such as `DBLRocuroni` and `Rocuronium`, which are successfully combined into `DBLRocuronium`. Partial or misrecognized fragments such as `Bromide Injed` and `de Injection` are merged into the correct phrase `Bromide Injection`. These examples highlight the ability of the method to exploit redundancy and complementarity across views to improve overall text reconstruction.

Overall, the results on real dual-camera images show that the proposed stitching method can combine OCR outputs from two viewpoints to improve text completeness compared to single-view recognition. The examples illustrate that fragments which are partially visible or degraded in one view can sometimes be recovered from the other and merged into a more complete sequence. At the same time, the observed failure modes, such as errors that persist across both views, indicate that the effectiveness of

the stitching process depends on the quality and complementarity of the input OCR results.

Taken together with the synthetic experiments, these findings support the effectiveness of the overlap-based merging mechanism under the evaluated conditions and demonstrate its applicability to cylindrical medication containers. However, they do not establish robustness in broader deployment settings, where factors such as severe distortion, low image quality, and limited viewpoint variation may still limit performance.

5.4.3.3 Qualitative Results: Multi-View Stitching



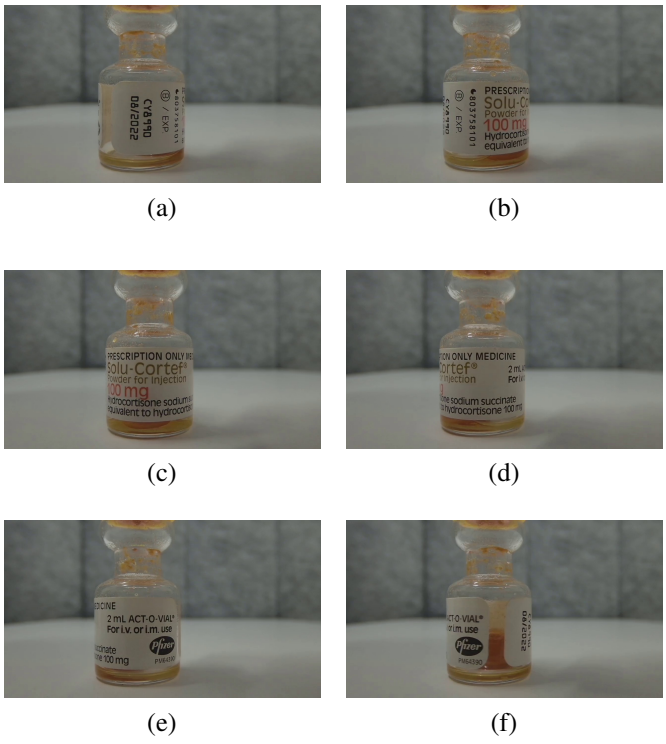
(a) (b) (c) (d) (e) (f)

```
⟶ 03.2015
⟶ PRESCRIPTION
MEDICIN
⟶ Powder for
Injection
⟶ Cefazolin sodium
⟶ equivalent to
Cefazolin
⟶ 1g
⟶ POUT OF REACHOF
CHILDREN
⟶ iM/IVuse
⟶ Please refer
toenctos
⟶ eaflet before
use
⟶ Store below25
⟶ Protect from
light
⟶ 3C30029AB
```

(g)

Figure 5.13: A visual representation of medication label stitching results. (a-f) Images of the drug container captured from different angles, and (g) the corresponding textual information with visual boundaries for clarity.

The proposed text stitching method was further applied to multiple images captured



(a) (b) (c) (d) (e) (f)

```
⟶ PRESCRIPTION
ONLY MEDICINE
⟶ B/EXP
⟶ 08/2022
⟶ CY8990
⟶ 6803758101
⟶ Solu-cortefe
⟶ Powder
forinjection
⟶ iydrocortisone
sodiumst
⟶ equivalent to
hydrocort's
⟶ 2mLA
⟶ 2mLACT-O-VIAL
⟶ For i.v.or
i.m.use
⟶ Phzer
⟶ one100mg
⟶ PM64390
```

(g)

Figure 5.14: A visual representation of medication label stitching results. (a-f) Images of the medication container captured from different angles, and (g) the corresponding textual information with visual boundaries for clarity.

from different angles of the medication. The primary goal in this experiment is to evaluate the effectiveness of the text stitching algorithm in merging text fragments from different perspectives, rather than to produce a fully complete or error-free label reconstruction. By combining text extracted from various images, the method demonstrates its ability to handle partial overlaps, distortions, and inconsistencies in individual OCR outputs. This evaluation focuses on the behavior of the stitching algorithm, while the reconstructed text is used as a qualitative illustration of the aggregation process.

Figs. 5.13 and 5.14 present the results of the multi-view text reconstruction process. In both figures, (a–f) show images of containers captured from multiple angles, which

were processed to extract textual information and stitched together to form a reconstructed representation of the label, as visualized in (g). The reconstructed text includes key information from the labels, such as the medication name, dosage, and instructions, although it may not be complete or fully accurate.

Specifically, in Fig. 5.13, the reconstructed output aggregates text fragments related to Cefazolin-AFT across multiple views, capturing important elements such as dosage form and safety instructions. The multiple viewpoints provide complementary coverage of the label surface. However, OCR errors such as character misinterpretation and missing spaces are present in the reconstructed text (e.g., `Please refer toenctos`), indicating limitations inherited from the recognition stage.

Similarly, Fig. 5.14 illustrates the reconstruction for Solu-Cortef, where text fragments from different views are combined to increase the amount of recovered information, including batch numbers and expiry dates. While the reconstruction improves coverage compared to individual views, OCR-related errors remain, such as `Solu-cortefe` and formatting inconsistencies.

Overall, the results suggest that incorporating multiple viewpoints can improve the coverage of reconstructed text by providing complementary observations of the cylindrical surface. At the same time, the quality of the reconstruction remains dependent on the underlying OCR performance, and errors that are consistently produced across views are not corrected by the stitching process. These experiments therefore illustrate the behavior of the multi-view stitching mechanism in practice, rather than demonstrating complete or fully reliable label reconstruction.

5.5 Conclusion

This chapter introduced a text-domain stitching pipeline as an alternative to the geometry-based image stitching approach presented in the previous chapter. By shifting from

pixel-level manipulation to direct alignment and merging of OCR outputs from dual-camera views, the proposed method simplifies the reconstruction process while avoiding computationally expensive steps such as label segmentation, ellipse fitting, and image unwrapping.

The pipeline consists of three core components: a fine-tuned PaddleOCR detection model adapted to pharmaceutical labels, Ratcliff/Obershelp-based text fragment alignment, and LCS overlap detection for merging. Experimental results indicate that the method can recover reasonably coherent transcriptions from partial, curved, and distorted views. Key information such as medication names, dosages, batch numbers, and expiry dates is often preserved, producing outputs that may be useful for downstream tasks including automated verification and database lookup.

Nevertheless, several limitations remain. First, the performance of the text-stitching pipeline is inherently dependent on the quality of the upstream OCR results. Recognition errors caused by cylindrical distortion, extreme viewing angles, specular reflections, or low image quality are not corrected by the stitching process and may propagate into the final transcription. In addition, the alignment strategy operates at the string level and does not explicitly model spatial layout, which can lead to incorrect ordering or merging when text structure is complex.

Second, the current evaluation is limited in scale and scope. The real-image experiments are based on a relatively small set of medication containers and rely primarily on qualitative assessment. Ground truth annotation is restricted to visually readable regions, meaning that heavily distorted areas are not directly evaluated. As a result, the reported results primarily reflect performance in regions where reliable transcription is feasible, and may not fully capture behavior under more challenging conditions.

Third, while multi-view input improves coverage, it does not guarantee completeness. Text that is consistently misrecognized or poorly captured across all views remains unresolved. The method therefore depends on sufficient viewpoint diversity and image

quality to provide complementary information.

Future work will explore several directions to address these challenges, including context-aware post-processing, integration of language models for semantic error correction and disambiguation, and more structured alignment strategies that incorporate layout or domain-specific constraints. Additional improvements could involve expanding the dataset with more diverse real-world conditions and performing more systematic quantitative evaluation.

In summary, the proposed method demonstrates practical utility as a component within a larger system for medication identification. However, improving robustness and validating performance across broader real-world conditions remain necessary before deployment in safety-critical applications. Future work will focus on addressing OCR propagation, enhancing spatial understanding, and providing more rigorous evaluation across diverse scenarios.

From Reconstructed Text to a Structured Medication Database

Chapter 5 established the Text Stitching Pipeline (Approach 2), which reconstructs label content from cylindrical medication containers using linguistic alignment. Compared to geometry-based image methods, this approach has the potential to reduce computational complexity by operating directly on text fragments. The output is a set of aggregated text fragments representing the label content. However, these results should be treated as intermediate outputs for safety-critical medication identification rather than final records. Both Approach 1 (Image Stitching) and Approach 2 (Text Stitching) provide candidate query strings for a downstream medication database, where the goal is to structure extracted text into standardised fields (e.g., medication name, dosage, manufacturer) and support consistency checking against external authoritative sources.

This chapter connects these outputs to a structured medication representation through LLM-assisted consolidation and integration with external medication knowledge sources. It focuses on organising noisy OCR-derived text into a more structured form, while keeping human oversight as an essential part of the process. In this sense, it bridges the reconstruction stages in Chapters 4 and 5 with downstream representations that may better support medication identification workflows, without assuming full verification of correctness at this stage.

Chapter 6

LLM-Assisted Construction of a Medication Database Using Retrieval-Augmented Generation

The medication database construction process presented in this chapter is an offline, construction-time process carried out prior to deployment. It is independent of the real-time identification pipeline described in Chapters 4, 5, and 8, and is not executed during live medication identification.

This chapter presents an LLM-assisted pipeline for constructing a structured medication database from text extracted from multiple views of medication containers, supporting downstream organization and retrieval of medication information. Text fragments are first obtained from medication images using Optical Character Recognition (OCR), producing noisy and incomplete outputs due to imaging conditions such as blur, occlusion, and cylindrical distortion. These limitations motivate the need for a downstream processing stage that organizes and contextualizes the extracted text.

The proposed pipeline uses a Large Language Model (LLM) to transform unstructured OCR outputs into a structured schema (e.g., JSON fields such as medication name,

dosage, and expiration date). In addition, a Retrieval-Augmented Generation (RAG) mechanism queries an external clinical reference source (e.g., Medsafe datasheets (Medsafe, n.d.)) to provide authoritative contextual information. This retrieved information is used to support consistency checking, ambiguity resolution, and structured representation of medication data. Rather than fully automating medication entry, the system is designed as a human-in-the-loop assistive tool that reduces manual effort while preserving human oversight in decision-making.

This chapter focuses on the role of combining LLM-based structuring with retrieval-based context as an intermediate step between raw OCR output and structured medication records. It does not assume full correctness guarantees, but instead explores how external references and language models can jointly support more consistent organization of extracted medication information.

6.1 Introduction

6.1.1 Background and Motivation

Accurate reading of medication labels is important for patient safety, particularly in time-critical environments such as operating theatres. However, manual transcription is slow and prone to error, making it unsuitable for scalable clinical workflows. Previous chapters addressed the core computer vision (CV) and OCR challenges of extracting text from highly distorted cylindrical medication labels. Using a dual-camera setup, two complementary approaches: image stitching (Chapter 4) and text stitching (Chapter 5) were introduced to reduce geometric distortion and improve multi-view text extraction.

Despite these improvements, the reconstructed text remains an intermediate representation rather than a structured record. Two main limitations remain. First, lack of structure: extracted text is typically free-form, with key fields such as drug name,

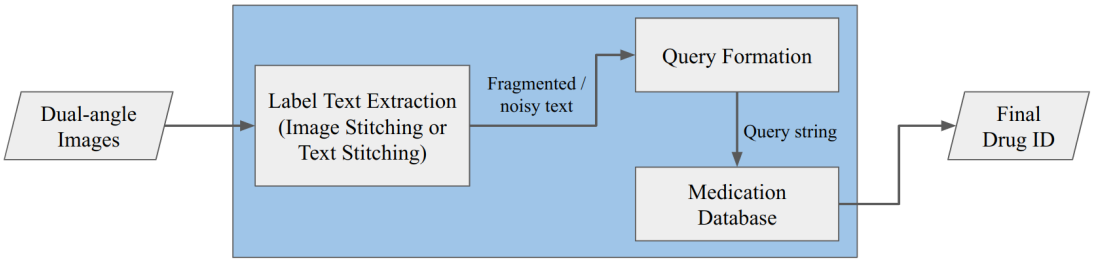


Figure 6.1: Overview of the query-based medication identification pipeline. Text extracted from dual-view images is used solely as a query to retrieve the correct drug identity from a medication database.

dosage, manufacturer, and expiration date not explicitly organized. Second, lack of reliability: OCR outputs may contain errors or ambiguities and are not inherently validated against external references. In addition, limitations in the image acquisition stage may lead to incomplete capture of label content, for example when parts of the label are not visible across the two views or are occluded due to cylindrical geometry. This can result in missing or partial information in downstream OCR outputs.

To address these limitations, this work introduces a database-centric framework in which text detected by Methods 1 and 2 is used primarily as query input. Rather than treating OCR output as structured information, the system uses it as a noisy observation of the underlying label content. Final medication identification and associated fields are retrieved from a curated database, as illustrated in Fig. 6.1.

The database construction process consists of two stages. First, label-derived observations from OCR are organized into a structured format. This step is supported by LLM-based tools, which assist in mapping unstructured text into predefined fields. This is used as a practical approach to handling variability and noise in OCR outputs, rather than relying solely on deterministic parsing rules, which may be brittle under incomplete or ambiguous text.

Second, external knowledge-base enrichment is performed using retrieval from medication reference sources such as Medsafe datasheets (Medsafe, n.d.). This stage

supplements the structured entries with authoritative contextual information, supporting consistency checking and completion of missing fields where applicable. These steps are intended to reduce manual effort and improve consistency, while still relying on external sources and human oversight for validation.

6.1.2 Problem Statement

The dual-camera reconstruction methods introduced in previous chapters are designed for real-time acquisition of medication label in operating theatre. Their primary role is to generate a continuous text fragments suitable for downstream querying. While effective at mitigating distortion and dual-view inconsistencies, the outputs remain unstructured, unverified, and unsuitable for direct clinical use.

6.1.2.1 Limitations of Reconstruction Outputs

Although both approaches consolidate multi-view text, the resulting string inherits noise from image capture and OCR. Method 1 (Image Stitching) corrects geometric distortions but may introduce seams or alignment artifacts, leading to OCR inaccuracies (Section 4.5). Method 2 (Text Stitching) merges OCR fragments but can produce concatenation errors (Section 5.4.3).

In addition, environmental factors such as reflections, uneven lighting, and label damage can reduce OCR quality in both methods. Acquisition constraints, including narrow container geometry, incomplete coverage in the dual-camera setup, or suboptimal positioning where key label regions (e.g., the drug name) are not captured, can also lead to missing or incomplete text prior to OCR. As a result, the reconstructed text is generally human-readable but remains unstructured and may contain errors.

6.1.2.2 Why Reconstruction Outputs Cannot Serve as a Final Database

Even with additional viewpoints from a dual-camera setup, the reconstruction outputs remain intermediate representations and are not directly suitable for use as a medication database. The resulting text is typically unstructured, with key attributes embedded in free-form sequences that are difficult to index or query consistently.

In addition, the reconstructed text may contain OCR-related errors, as observed in earlier evaluations. Incomplete visual coverage during image acquisition can also lead to missing or partial label content, further affecting the completeness of the output.

Overall, reconstruction provides a useful representation of label content, but additional structuring and integration with external references are required before it can support reliable retrieval or practical use.

6.1.2.3 Objective

This chapter investigates whether an LLM- and retrieval-assisted pipeline can improve the structuring of noisy OCR outputs from cylindrical medication labels into consistent structured representations.

The proposed pipeline consists of three stages: (1) LLM-assisted consolidation of fragmented OCR text, (2) extraction into a predefined JSON schema, and (3) optional enrichment using external medication reference sources.

The central hypothesis is that combining LLM-based text consolidation with external reference retrieval can improve the structural quality and completeness of OCR-derived medication information compared to direct rule-based parsing of raw OCR outputs. The effectiveness of the proposed approach is evaluated in terms of structured information quality using two primary metrics:

- **Field Completeness:** the proportion of predefined fields that are successfully populated in the structured output.

- **Critical Field Accuracy:** the correctness of key fields, including medication name, dosage, and expiration date compared to annotated ground truth.

In addition to these quantitative measures, qualitative examples are presented to illustrate how fragmented or noisy OCR outputs are consolidated into more coherent structured representations, as well as to highlight common failure cases.

The focus of this evaluation is on structured representation quality of OCR-derived data rather than clinical decision-making or medical correctness. The study examines whether the proposed pipeline improves the organisation, completeness, and consistency of extracted medication information under a human-in-the-loop setting.

6.2 Related Work: LLMs in Information Extraction

6.2.1 LLMs for Structured Extraction and Schema Mapping

LLMs have been explored for transforming unstructured or semi-structured text into schema-aligned representations. Their ability to model contextual dependencies makes them suitable for consolidating fragmented inputs and mapping them into structured fields such as entity attributes.

In this work, an instruction-tuned LLM accessed via API (e.g., GPT-4.0 mini) is used for structured extraction and schema mapping. The task is formulated as constrained extraction rather than open-ended generation, where adherence to a predefined schema and output consistency are prioritised. This makes compact instruction-following models suitable for producing structured outputs from noisy OCR-derived text.

The input to the model consists of reconstructed text fragments obtained from earlier stages, which may contain fragmentation, duplicated tokens, and inconsistent ordering. The role of the LLM is to organise this input into a structured representation with predefined fields such as medication name, dosage, formulation, and manufacturer.

While this work focuses on API-based models for simplicity, domain-specific adaptation, such as prompt refinement or fine-tuning on pharmaceutical text, may further improve consistency and robustness, and is left for future work.

6.2.2 Retrieval-Augmented Knowledge Integration

Retrieval-augmented approaches enable systems to incorporate external knowledge sources to provide additional context for incomplete or ambiguous inputs (Lewis et al., 2020). In this work, retrieval is used to support the structuring of OCR-derived text rather than to perform automatic verification.

Retrieval-Augmented Generation (RAG) combines a retrieval step with downstream processing, where relevant information is first selected from an external corpus and then used as contextual reference. In this system, medication information is retrieved from Medsafe (New Zealand Medicines and Medical Devices Safety Authority) datasheets (Medsafe, n.d.). Given a structured or partially structured query derived from OCR text, the system retrieves a small set of the most relevant entries (top-k, where k denotes the number of retrieved results) to provide supporting context.

A key challenge is that retrieved entries may correspond to different strengths or formulations of a medication, and partial matches can introduce inconsistencies. In addition, OCR-derived queries may be incomplete or noisy, which can affect retrieval quality. For this reason, retrieved information is not treated as ground truth and is not used for automatic conflict resolution.

Instead, retrieval is used as a supporting mechanism within a human-in-the-loop workflow. The retrieved evidence provides additional context to assist structuring and comparison, while final interpretation and validation remain the responsibility of the user. This design reflects the use of the system as a database construction tool rather than a fully automated decision-making system.

6.3 Proposed Database Construction Pipeline

6.3.1 Overall Pipeline and System Flow

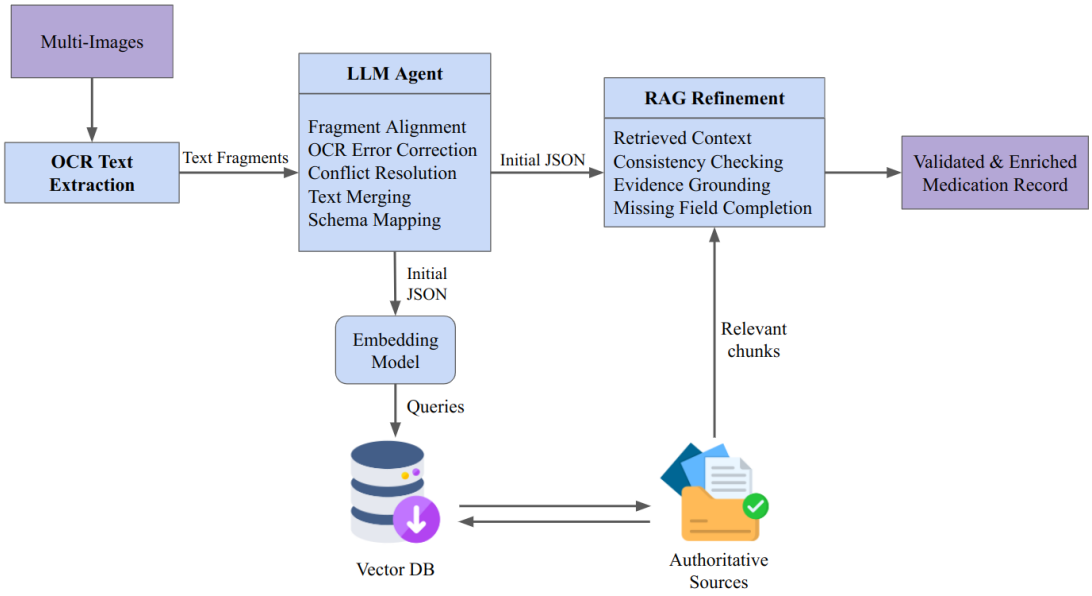


Figure 6.2: Automated Medication Data Extraction and Refinement/Enrichment Architecture.

The proposed system transforms multiple images of medication labels into structured medication records through a three-stage offline workflow, as illustrated in Fig. 6.2. Designed for database construction rather than real-time inference, the pipeline leverages multiple viewing angles to partially mitigate limitations of single-view capture.

First, OCR is applied to all captured medication container images to produce raw text fragments. As discussed in earlier chapters, OCR outputs in this setting can remain highly fragmented and uncertain due to curvature, perspective distortion, lighting variation, and partial visibility of label regions. As a result, the extracted text may contain incomplete or noisy representations of key information such as medication names, dosages, and manufacturer details.

In the second stage, an LLM-based component processes the OCR outputs through a structured workflow. This stage consolidates fragmented text from multiple views

into a unified representation and maps it into a predefined schema. The role of the LLM here is to support structured reformatting and consolidation of noisy inputs rather than to guarantee correction of all errors.

Finally, a RAG component is used to incorporate external medication reference sources. A semantic search is performed over a curated database of medication datasheets, and the retrieved content is used as auxiliary context for checking consistency, and supporting the completion of missing fields where applicable. This step introduces external grounding, although retrieved information may still require interpretation in the presence of ambiguous or incomplete OCR inputs.

Overall, the system combines multi-view OCR extraction, LLM-based consolidation and structuring, and retrieval-based contextual support to progressively improve the usability of extracted medication information for downstream database construction.

6.3.2 LLM Agent for Data Structuring and Normalization

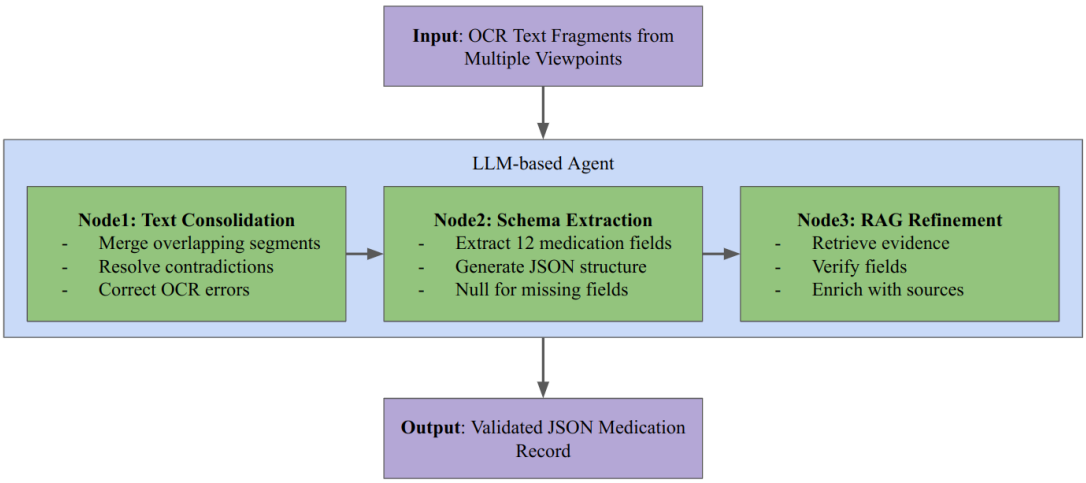


Figure 6.3: LLM agent architecture implemented using LangGraph: (1) Text Consolidation node merges OCR fragments and corrects systematic errors, (2) Schema Extraction node generates structured JSON with 12 medication fields, and (3) RAG Refinement node validates and enriches fields against authoritative sources.

Processing medication-related text extracted from cylindrical containers introduces

several challenges. Curved-surface distortion, partial occlusion, motion blur, and uneven lighting can lead to OCR outputs that are fragmented, duplicated, or inconsistently segmented, as discussed in Sections 2.8 and 3.1.2. As a result, raw OCR output often does not directly form a coherent representation suitable for structured storage.

To address these limitations, an LLM-based agent is used as a post-OCR processing module. The agent is implemented as a graph-based workflow using LangGraph, which allows explicit definition of processing stages and controlled execution flow, as illustrated in Fig. 6.3. The pipeline consists of three stages: text consolidation, schema extraction, and retrieval-based contextual checking. The design provides modularity and traceability compared to single-step prompt-based processing.

6.3.2.1 Text Consolidation and Error Correction

```
 You are an AI specialized in merging text fragments from
medication bottle images.
Given a list of text fragments captured from different
angles, merge them into a coherent, complete label.
Remove duplicated substrings, resolve overlaps, and
correct clear OCR errors.
Return only the merged text with no additional
commentary.
```

Figure 6.4: Prompt used for medication text merging.

The first stage focuses on consolidating OCR fragments obtained from multiple viewpoints of the same label. Because each image captures only a partial view of the cylindrical surface, the resulting text may contain overlaps, missing segments, or duplicated substrings.

Rather than relying on heuristic string matching alone, the LLM is used to perform rule-guided consolidation based on explicit prompting constraints (see Fig. 6.4). The goal of this stage is to produce a single coherent text representation by merging overlapping segments and reducing duplicated content. Any correction of OCR artifacts is

restricted to cases where the intended text can be inferred from consistent contextual evidence across fragments. The model is instructed to avoid introducing new information beyond what is present in the input fragments. Furthermore, all experiments were conducted using a fixed model version (GPT-4.0 mini) and prompt configuration to further reduce variability.

The consolidation stage aims to improve the coherence of fragmented OCR outputs by merging overlapping segments and reducing duplication under explicit prompting constraints. The model is instructed to preserve medically relevant terminology and to avoid introducing information not supported by the input fragments. The resulting output is a single consolidated text representation intended to improve readability and support downstream structured extraction, although it does not guarantee recovery of missing content when label regions are not captured in any input view.

6.3.2.2 Schema Mapping and JSON Generation

Table 6.1: Standardized schema fields for medication database records.

Field	Purpose
`drug_name`	Primary medication identification
`active_ingredients`	Ingredient verification, interaction checking
`dosage_information`	Dose confirmation, concentration validation
`expiration_date`	Safety, shelf-life management
`manufacture_details`	Traceability, batch recall support
`diluent_name`	Dilution compatibility, preparation guidance
`indications`	Clinical use verification
`side_effects`	Risk awareness, adverse effect monitoring
`storage_instructions`	Handling compliance, proper storage
`administration_route`	Safe delivery method confirmation
`precautions`	Contraindication, allergy screening
`regulatory_information`	Regulatory standards, approvals

After text consolidation, the next step is schema mapping, where the unified label

```
 You are an AI system for structuring medication label
text.
Given a consolidated text string, extract the following
fields when explicitly present:
- drug_name (string)
- active_ingredients (list of strings)
- dosage_information (string; include units such as mg,
mL, IU)
- expiration_date (string)
- manufacture_details (string)
- diluent_name (string)
- indications (string)
- side_effects (string)
- storage_instructions (string)
- administration_route (string)
- precautions (string)
- regulatory_information (string)

Do not infer missing information unless explicitly
supported by the input text.
Return a JSON object and use null for missing fields.
```

Figure 6.5: Prompt used for structured medication information extraction.

text is transformed into a structured representation for database storage and retrieval (Table 6.1). The selected schema reflects fields commonly present on medication packaging and those relevant to operating-theatre preparation workflows, such as identification (drug name), preparation (dosage, diluent, route), and safety-related information (expiry, precautions, storage).

The extraction prompt (Fig. 6.5) enforces a constrained mapping from text to schema fields. The model is explicitly instructed to extract only information supported by the input text and to avoid inferring missing values. This is important because medication labels are often incomplete or inconsistently structured after OCR, and unsupported inference may introduce unsafe or incorrect assumptions.

The output is formatted as a structured JSON object, with missing fields represented as `null`. This design enables consistent downstream processing without assuming

completeness of the extracted information. It also allows later integration with external reference sources, where label-derived fields and externally sourced information can be clearly distinguished.

6.3.3 Knowledge Integration: Retrieval and Enrichment

Despite text consolidation and schema-based extraction, certain medication fields may remain incomplete, ambiguous, or uncertain due to OCR noise, curved-surface distortions, or limited label visibility. To address these limitations, a RAG component is incorporated as an additional processing stage within the agent pipeline. This component retrieves relevant information from external medication reference sources and provides contextual support for interpreting and completing extracted fields where applicable.

Rather than acting as a definitive verification step, the retrieval component serves as an auxiliary mechanism for comparison and enrichment. Retrieved information may help identify inconsistencies, highlight potential discrepancies, or supplement missing fields, but it does not guarantee correctness. Final interpretation remains dependent on the quality of both the OCR-derived input and the retrieved evidence, and may require human review in ambiguous cases.

6.3.3.1 Knowledge Base Construction

The knowledge base is constructed from publicly available medication datasheets published by Medsafe (Medsafe, n.d.), the New Zealand Medicines and Medical Devices Safety Authority. These documents contain clinical, regulatory, and product-specific information for medications approved for use in New Zealand.

The datasheets are provided in PDF format and exhibit variability in layout, structure, and length. They typically include information such as indications, dosage instructions,

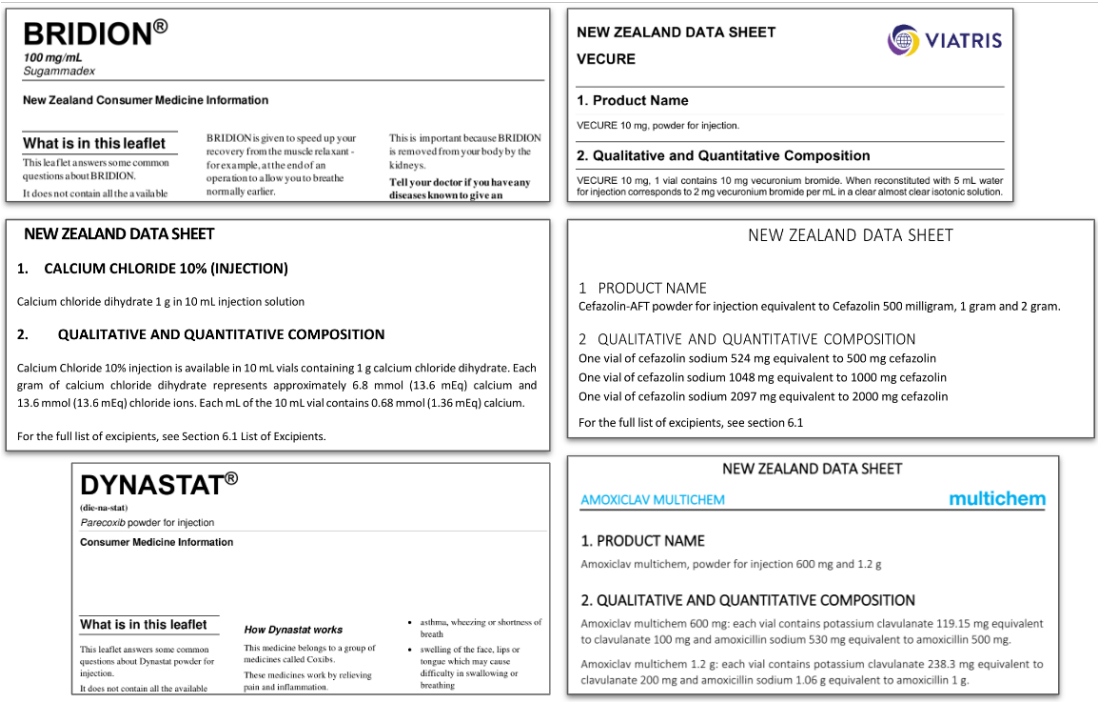


Figure 6.6: Examples of Medsafe medication datasheets used as external reference sources in this study.

contraindications, warnings, interactions, and adverse effects. However, these documents may also contain multiple formulations, strengths, or product variants within a single file, which can introduce ambiguity when mapping retrieved information to a specific medication instance.

For use in retrieval, the PDFs are processed through a text extraction and chunking pipeline. Text is extracted using the `pdfplumber` library, with basic preservation of document structure such as headers and section boundaries. The extracted text is then segmented into overlapping chunks (512 tokens with 128-token overlap) to support semantic retrieval while maintaining local context. Tables and structured sections are retained as linearised text where possible, and selected metadata (e.g., section headers) is preserved to provide additional context during retrieval.

These processed documents form the basis of the retrieval corpus used in the RAG component. Retrieved content is treated as supplementary reference material rather

than a definitive source of truth, particularly in cases where multiple product variants or ambiguous descriptions are present.

6.3.3.2 Two-Stage Retrieval Process

Stage 1: Global Retrieval via Semantic Similarity. To support interpretation of extracted medication details, the system retrieves relevant context from the knowledge base using semantic similarity. Retrieval is performed using dense vector embeddings generated from the consolidated text output, enabling comparison with chunked medication datasheets beyond simple keyword matching.

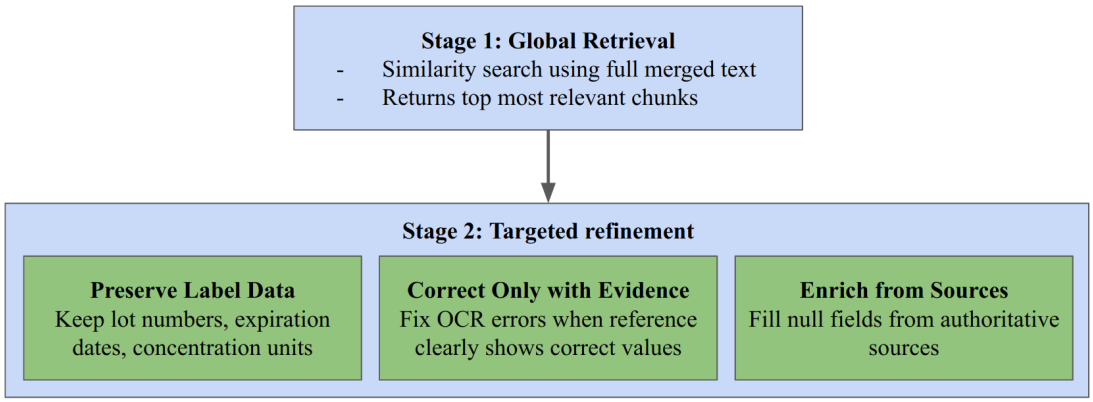


Figure 6.7: Two-stage retrieval process consisting of global search followed by targeted refinement. The framework emphasises preserving label-derived information, limiting corrections to cases supported by consistent reference evidence, and using retrieved content as supplementary context for incomplete fields.

The retrieval pipeline operates in two steps, as shown in Fig. 6.7. In the first stage, a global similarity search is performed using selected structured fields and consolidated text as the query. The system retrieves the top-k most similar knowledge base chunks based on cosine similarity, which serve as candidate reference contexts for subsequent processing. This step does not assume a single correct match, but instead provides a set of potentially relevant documents that may correspond to different formulations or related products.

Dense vector embeddings are generated using the `all-MiniLM-L12-v2` model, which provides a balance between semantic representation quality and computational efficiency for short medical phrases. The embeddings are indexed using a FAISS-based vector database with a flat L2 index to support efficient similarity search over the document corpus. For retrieval, the system returns the top $k = 5$ most similar chunks based on cosine similarity, with a minimum similarity threshold of 0.5 applied to filter weak matches.

Stage 2: Targeted Refinement In the second stage, additional queries are issued using selected extracted attributes (e.g., drug name, active ingredients, formulation) to retrieve more specific context such as dosage descriptions or regulatory sections. The goal is to provide relevant reference material that can assist in refining the initially extracted fields. Retrieved content may correspond to different strengths or product variants, and therefore requires careful interpretation.

Once relevant contexts are retrieved, the LLM performs a refinement step in which the initial JSON is updated using the provided evidence. The process is guided by the prompt (Fig. 6.8), which encourages the model to preserve existing values when appropriate, update fields when clearer evidence is available, and fill missing attributes where possible. In practice, this step can improve the completeness and coherence of the structured output, particularly when OCR-derived text is fragmented or partially missing.

However, this refinement does not enforce strict consistency checks or guarantee correct matching to a specific product variant. Retrieved documents may contain overlapping or ambiguous information, and the model may not always distinguish between similar formulations or strengths. As such, the output should be interpreted as a refined but not fully verified representation, and ambiguous cases may require human review.

A limitation of the current implementation is that field-level provenance is not

```
 You are an AI tasked with refining and enriching
medication information extracted from OCR data.  You will
receive:
- A JSON object containing the initial extraction (may be
incomplete or noisy).
- Retrieved PDF chunks from a vector database (trusted
medication references).

Your task is to refine the JSON by:
- Using `initial_json` as the base.
- Comparing each field with `retrieved_chunks`.
- Preserving original values unless the retrieved context
provides a clear and specific correction.
- Never replacing dosage or expiration values unless
contradicted by authoritative evidence.
- Filling null fields with concise, relevant information
extracted from retrieved chunks.
- Paying special attention to warnings for side effects
and precautions.

Return only the refined JSON object with no extra
commentary.
```

Figure 6.8: Prompt for medication information refinement with RAG.

explicitly encoded in the final JSON output. Although intermediate outputs from each stage (OCR consolidation, LLM structuring, and retrieval-augmented refinement) are available within the processing pipeline, the final representation does not retain explicit metadata linking each field to its origin or modification history. As a result, provenance can be reconstructed through comparison of intermediate states in the execution graph, but it is not directly embedded in the final structured output. Addressing this by incorporating explicit field-level provenance tracking within the JSON schema is left for future work.

6.4 Experimental Results

6.4.1 Overview

The proposed system transforms multi-view medication container images into structured medication records through OCR, LLM-based consolidation, and retrieval-based contextual support. The primary goal is to reduce manual data entry effort in clinical and pharmacy settings. Rather than conducting a large-scale quantitative evaluation, this chapter presents qualitative case studies to illustrate system behavior across representative scenarios, highlight practical limitations, and discuss potential implications for workflow efficiency.

The evaluation focuses on practical usability: whether the system can produce structured, database-ready records that reduce the extent of manual review required. A record is considered usable for downstream processing if key fields (medication name, active ingredients, dosage, expiration date) are present with plausible values, the JSON output is syntactically valid, and no obvious internal inconsistencies are observed. This definition is used for analysis purposes only and does not imply clinical validation.

6.4.2 Experimental Setup

The experimental setup integrates OCR, LLM-based structuring, and embedding-based retrieval into an end-to-end pipeline for medication label interpretation. Each component is selected to provide a practical trade-off between computational cost, deployment constraints, and the level of structured output required for downstream processing, given the hardware described in Section 4.4.1.

For medication text extraction, the fine-tuned PaddleOCR model introduced in Chapter 5 is employed. Medication vials present challenges such as small fonts, irregular spacing, and geometric distortion from cylindrical surfaces. The fine-tuned model is

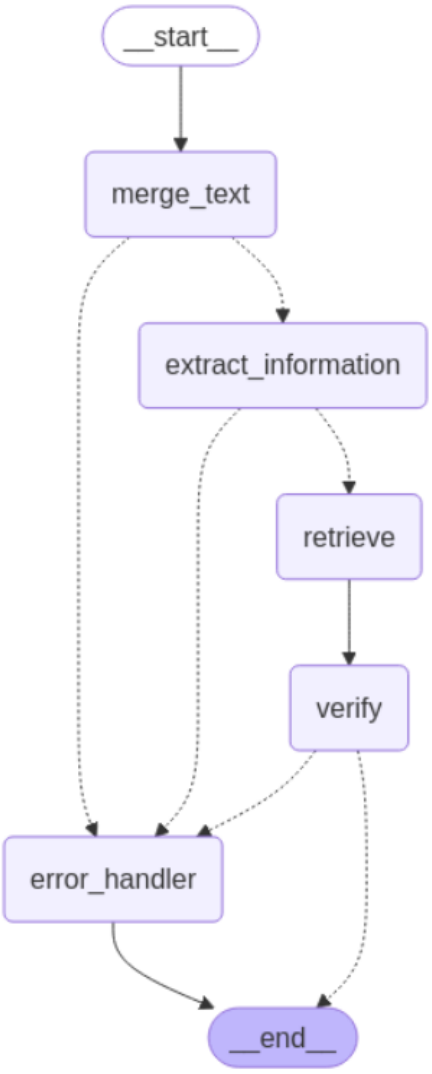


Figure 6.9: Workflow graph of the LLM agent implemented using LangChain and LangGraph.

used because it provides more consistent recognition on this domain-specific data compared to generic OCR models, while maintaining processing speed suitable for batch image processing. This stage produces noisy but usable text fragments for downstream consolidation.

For LLM inference, an instruction-tuned language model (GPT-4o mini, accessed via API) is used for structured extraction and schema mapping. This model is selected as a balance between output quality, latency, and ease of integration within the pipeline. While larger proprietary models (e.g., full GPT-4-class systems) may offer stronger reasoning capability, they introduce higher cost and similar API dependency constraints. In contrast, smaller open-weight models may be more suitable for local deployment but can be less consistent in following strict structured output formats.

The chosen model provides sufficient instruction-following ability for schema-constrained generation, which is the primary requirement of this task. Its role is limited to consolidating OCR-derived text and mapping it into a predefined schema, rather than performing open-ended reasoning or clinical decision-making. This design ensures predictable outputs while maintaining compatibility with downstream database storage and retrieval.

For the retrieval component, embeddings are generated using the `all-MiniLM-L12-v2` model (Reimers & Gurevych, 2019). This model is selected due to its efficiency and acceptable performance on short text queries, such as medication names and dosage phrases. Domain-specific embedding models were considered; however, many are optimized for longer clinical narratives and introduce additional computational overhead without clear benefit for short, label-derived text. The chosen model provides a practical balance between embedding quality and retrieval speed for this task.

The LLM agent is implemented using LangChain and LangGraph, which provide an explicit node-based control flow for structuring the processing pipeline. As illustrated in Fig. 6.9, the agent comprises four primary processing nodes aligned with the design

described in Section 6.3. The behavior of these nodes is guided by task-specific prompts introduced in earlier sections and shown in the corresponding figures.

1. a *merge_text* node that consolidates OCR fragments from multiple viewpoints (see prompt in Fig. 6.4),

2. an *extraction_information* node that maps the consolidated text into a structured JSON representation of medication records (see prompt in Fig. 6.5),

3. *retrieve* and *refine* nodes that incorporate retrieval-based contextual information into the extracted fields. The behavior of these nodes is guided by task-specific prompts defined in Fig. 6.8 and earlier sections,

4. an *error_handler* node that manages parsing failures or malformed outputs, allowing partially structured results to be returned rather than terminating execution.

All prompts used within the LLM agent are kept fixed throughout the experiments reported in this thesis. The same prompt definitions are applied across all test cases for text consolidation, schema extraction, and retrieval-based refinement. This ensures that the reported results reflect the behavior of the underlying model and overall pipeline design under a consistent prompting setup.

6.4.3 Quantitative Evaluation

We evaluate the medication label extraction system using 15 real medication containers representing different pharmaceutical products commonly used in operathing theatre, provided by Auckland City Hospital (Healthpoint, n.d.) and SAFERsleep (SAFERsleep Ltd., n.d.). Each container is captured from 8-10 viewpoints under realistic conditions, including variations in occlusion, curvature, and lighting.

Each label is first processed using OCR on the raw images. The resulting text is directly used as input to the LLM-based extraction and RAG refinement pipeline, without additional text stitching or post-OCR reconstruction.

Extracted information is organized into two categories based on clinical importance. Critical fields include drug name, dosage information, and expiration date, which are essential for safe medication identification. Non-critical fields include supporting information such as active ingredients, manufacturer details, storage instructions, administration route, and regulatory or precautionary information.

Based on this schema, we evaluate the system using two metrics. Field completeness measures the proportion of non-null extracted fields across both categories, reflecting how fully the structured representation is populated from OCR input. Critical field accuracy measures whether the essential fields (drug name, dosage, and expiration date) are correctly extracted with respect to the label content, as these directly affect medication identification reliability.

Limitation: The evaluation is conducted on a small dataset of 15 containers, which limits statistical generalization. In addition, the correctness of fields in the non-critical category is not formally verified against an external reference. This may introduce ambiguity for attributes that have multiple valid textual representations or are less clearly defined on the label.

6.4.3.1 Field Completeness

Table 6.2 reports field-level completion performance over 15 medication containers. Overall, the system achieves an average completeness of 80.6%, indicating moderate variability across field types and label conditions.

A clear pattern emerges across clinically critical fields. Core identification attributes such as drug name, active ingredients, administration route, manufacture details, and expiration date achieve 100% completeness, indicating that these fields are consistently

present and reliably extracted across all samples in the dataset.

Dosage information also shows high coverage at 93.3%, reflecting strong robustness for primary clinical parameters. In contrast, secondary fields exhibit lower and more variable completeness. Regulatory information and indications both achieve 73.3%, while storage instructions decrease to 66.7%. More context-dependent fields such as precautions and side effects further drop to 60.0% and 53.3%, respectively. The lowest completeness is observed for diluent name at 46.7%, which is often inconsistently presented or absent depending on the medication packaging format.

Overall, the results indicate that the system performs reliably for structured core attributes that are consistently visible on medication labels, while completeness decreases for supplementary fields that are less standardized or not always present in the input data.

Table 6.2: Field-level completion rates across 15 medication containers with summary statistics.

Field	**Completion Rate (%)**
drug_name	100.0
active_ingredients	100.0
administration_route	100.0
manufacture_details	100.0
expiration_date	100.0
dosage_information	93.3
regulatory_information	73.3
indications	73.3
storage_instructions	66.7
precautions	60.0
side_effects	53.3
diluent_name	46.7
Mean	**80.6**

6.4.3.2 Critical Field Accuracy

Critical fields (drug name, dosage information, and expiration date) are essential for safe medication identification and are evaluated against manually inspected ground truth derived from the original label images. The results are summarized in Table 6.3.

Overall, drug names are correctly extracted in 14 out of 15 cases (93.3%), dosage information in 13 out of 15 cases (86.7%), and expiration dates in all cases (100.0%). These results indicate strong performance for structured and visually prominent attributes, particularly expiration dates and medication identity.

Errors are mainly concentrated in dosage information, which accounts for most of the mismatches. These errors typically arise from formatting variability such as unit representation, spacing, or line break inconsistencies, rather than complete failure of detection. In contrast, drug names and expiration dates are more consistently formatted on labels, leading to higher reliability.

Performance differences across fields reflect variations in how information is presented on medication packaging. While drug names and expiration dates follow relatively standardized formats, dosage information is more heterogeneous, making it more prone to extraction inconsistencies.

Although additional fields such as active ingredients are extracted, they are not included in this evaluation due to the absence of detailed ground-truth annotation. Therefore, the analysis focuses on safety-critical attributes as defined in Table 6.3.

6.4.4 Case Studies: System Behavior and Outcomes

To better understand system behavior beyond aggregate metrics, we present a set of representative case studies illustrating both strengths and limitations of the proposed pipeline.

Field	Correct / Total	Accuracy (%)
Drug name	14 / 15	93.3
Dosage information	13 / 15	86.7
Expiration date	15 / 15	100.0
Mean Field Accuracy	-	93.3
Fully Correct Medication	11 / 15	73.3

Table 6.3: Critical field accuracy evaluated on 15 medication labels using manually annotated ground truth.

6.4.4.1 Case 1: Successful reconstruction under multi-view input.

The example in Figure 6.10 shows a scenario where fragmented OCR-derived text is consolidated into a single representation that preserves key semantic elements of the medication label. Despite noise and repetition in the input, clinically relevant information such as drug name, dosage strength, storage conditions, and regulatory identifiers is retained in the merged output.

For instance, the drug name is correctly consolidated as `Fresofol 1% MCT/LCT`, while dosage information is consistently extracted as `500 mg/50 mL`. Regulatory and identification details such as `AUST R 193609` and batch-related metadata are also preserved. In addition, key safety and handling instructions, including `Store below 25°C. Do not freeze.` and `For intravenous use only`, remain intact in the consolidated text despite fragmentation in the raw input.

The subsequent structured extraction maps this consolidated text into a consistent JSON schema. Core fields, including drug name, dosage information, and regulatory reference, are correctly populated. Additional fields such as storage instructions, administration route, and precautions are also recovered where explicitly present in the input text, while less explicit attributes (e.g., side effects) remain absent.

Overall, this example illustrates that text consolidation improves the stability of downstream structured extraction by reducing fragmentation in OCR inputs, leading to

```
"merged_text":
"PRESCRIPTION ONLY MEDICINE  \nKEEP OUT OF THE REACH OF
   CHILDREN  \n500 mg/50 mL Fresofol\u00ae 1% MCT/LCT (
   propofol 500 mg 50 mL injection)  \nFor intravenous use
   only.  \n50 mL in each 50 mL vial.  \nContains excipients:
   soya oil 2.5 g, medium chain triglycerides, glycerol 1.125
   g, egg lecithin 600 mg, oleic acid and sodium hydroxide for
    pH adjustment, Water for Injections to 50 mL.  \nFor
   single use only. Contains no antimicrobial agent.  \nShake
   well before use. Conditions: Prior to use, the rubber
   membrane should be cleaned with an alcohol swab. Use only
   if the container is undamaged.  \nDirections: Please refer
   to the enclosed insert. Store below 25\u00b0C. Do not
   freeze.  \nFresenius Kabi Australia, Level 2, 2 Woodland
   Way, Mount Kuring-gai NSW 2080.  \nManufactured by
   Fresenius Kabi Austria GmbH, Austria.  \nFresenius Kabi AB,
    Sweden.  \nBatch no.: 16PB6592  \nExp. 01 2022  \nAUST R
   193609"
```



```
"refined_json": {
        "drug_name": "Fresofol 1% MCT/LCT",
        "active_ingredients": ["propofol 500 mg"],
        "dosage_information": "500 mg/50 mL",
        "expiration_date": "01 2022",
        "manufacture_details": "Fresenius Kabi Austria GmbH,
           Austria. Fresenius Kabi AB, Sweden.",
        "diluent_name": "Water for Injections",
        "indications": "For intravenous use only.",
        "side_effects": null,
        "storage_instructions": "Store below 25\u00b0C. Do not
            freeze.",
        "administration_route": "Intravenous",
        "precautions": "Shake well before use. Prior to use,
           the rubber membrane should be cleaned with an
           alcohol swab. Use only if the container is
           undamaged.",
        "regulatory_information": "AUST R 193609"
}
```


Figure 6.10: Example of multi-view OCR aggregation and structured extraction under noisy and fragmented input conditions.

more complete and coherent representations under challenging acquisition conditions.

6.4.4.2 Case 2: Recovery of missing context via RAG.

From the example shown in Figure 6.11, the initial extraction reliably captures core attributes such as `drug_name`, `dosage_information`, and `active_ingredients`, but leaves most contextual fields empty. For example, fields such as `indications`, `administration_route`, and `storage_instructions` are null in the initial output.

After retrieval-augmented refinement, these missing fields are populated using external reference information. For instance, `indications` is filled with clinical usage categories, and `administration_route` is expanded to include multiple routes such as intravenous and intramuscular injection. Safety-related fields such as `precautions` and `storage_instructions` are also completed based on retrieved documentation.

Overall, this case shows that RAG improves the coverage of contextual and descriptive fields. By enriching initially sparse outputs with relevant external information, the approach can better support downstream clinical documentation workflows that require more complete medication context.

The evaluation focuses on completeness with respect to retrieved references. However, the correctness of the added information has not been independently validated against expert annotations, and would require expert review for clinical deployment.

6.4.4.3 Case 3: Incorrect entity reconstruction under ambiguous OCR input.

This case (Fig. 6.12) illustrates a failure mode arising from incomplete and fragmented OCR outputs, where key identifiers are only partially captured and cannot be fully resolved during downstream processing. In the merged text, the drug name appears as a truncated form (`METOPRO`), while related contextual information such as the

```
 "initial_json": {
        "drug_name": "solu-Cortef",
        "active_ingredients": [
        "Hydrocortisone sodium succinate"
        ],
        "dosage_information": "100 mg",
        "expiration_date": "08/2022",
        "manufacture_details": "Pfizer",
        "diluent_name": null,
        "indications": null,
        "side_effects": null,
        "storage_instructions": null,
        "administration_route": null,
        "precautions": null,
        "regulatory_information": "PRESCRIPTION ONLY MEDICINE"
}
```

```
"refined_json": {
        "drug_name": "SOLU-CORTEF",
        "active_ingredients": [
        "Hydrocortisone sodium succinate"
        ],
        "dosage_information": "100 mg",
        "expiration_date": "08/2022",
        "manufacture_details": "Pfizer",
        "diluent_name": "Water for Injections",
        "indications": "Endocrine Disorders, Rheumatic
           Disorders",
        "side_effects": "Increased susceptibility to
           infections, peptic ulceration.",
        "storage_instructions": "Store at or below 25\u00b0C,
           protect from light.",
        "administration_route": "Intravenous injection,
           intravenous infusion, intramuscular injection.",
        "precautions": "Monitor for infections; gradual
           withdrawal after long-term therapy.",
        "regulatory_information": "PRESCRIPTION ONLY MEDICINE"
}
```


Figure 6.11: Example of retrieval-augmented enrichment for Solu-Cortef. The initial LLM extraction contains limited clinical context, while the RAG-enhanced output incorporates additional prescribing information such as indications, precautions, and administration details.

manufacturer (`Mylan`) and active ingredient (`metoprolol tartrate`) is present in separate segments.

```
"merged_text":
"M1550  \n1121  \nPRESCRIPTION ONLY  \nMETOPRO  \nMylan  \n5
   mg/5 mL  \nSolution  \nfor injection  \nmetoprolol tartrate
     \nI.V. use only  \nAUST R 204665  \nCode No.: KR/DRUGS/
   KTK28  \n1029099  \nLOT: 70085  \nEXP: 05/20.  \nLOT:
   7008550  \nEXP: 05/2021",
```

```
"initial_json": {
        "drug_name": "METOPRO",
        "active_ingredients": ["metoprolol tartrate"],
        "dosage_information": "5 mg/5 mL",
        "expiration_date": "05/2021",
        "manufacture_details": "Mylan",
        "diluent_name": null,
        "indications": "I.V. use only",
        "side_effects": null,
        "storage_instructions": null,
        "administration_route": "I.V.",
        "precautions": null,
        "regulatory_information": "AUST R 204665"
},
```

Figure 6.12: Example of incorrect entity reconstruction under ambiguous OCR input, where fragmented text leads to partial or inaccurate drug name normalization.

In this example, the structured extraction reflects limitations introduced at the OCR stage rather than errors in the downstream reasoning or extraction process. Although supporting contextual cues such as the manufacturer and active ingredient are available, they are not sufficient to reliably infer the full canonical drug name. As a result, the extracted entity remains partially incomplete, propagating upstream ambiguity into the final structured output.

This case highlights a limitation of the pipeline in scenarios where critical identifiers are partially missing or heavily fragmented in the OCR input. In such cases, subsequent merging and structured extraction are constrained by the available evidence, even when complementary contextual information is present.

6.4.4.4 Case 4: Retrieval-induced enrichment with misaligned context.

As shown in Figure 6.13, the initial structured output correctly captures the core attributes directly supported by the label, including drug name, active ingredient, dosage, and administration route. These fields remain unchanged after the enrichment stage, indicating stable extraction of primary identifiers.

The enriched output introduces additional fields such as `diluent name` and `storage instructions`, which are derived from retrieved external references. In some cases, these additions may reflect general product-level information, and their inclusion depends on the quality of retrieval and query relevance.

This example highlights a characteristic behaviour of the RAG stage: it can improve field completeness by incorporating external context, but retrieval may also fail to return fully aligned evidence, leading to incomplete or loosely matched enrichment. For this reason, enrichment is treated as a supportive step for expanding contextual coverage rather than a strict verification mechanism.

Importantly, core fields such as drug name, dosage, and expiration date remain unchanged, as they are directly grounded in the OCR-derived extraction. Retrieval effects are therefore mainly observed in auxiliary fields, which are more variable across sources. Future improvements will focus on improving query specificity and retrieval filtering to reduce cases of weak or mismatched context.

6.5 Web Application Development and Clinical Workflow Integration

The proposed system includes a web-based interface to support the management, review, and storage of medication information extracted from container images. The application provides a centralized platform for uploading images, monitoring processing, and

```
"initial_json": {
        "drug_name": "Dynastat",
        "active_ingredients": [
        "parecoxib (as parecoxib sodium)"
        ],
        "dosage_information": "40 mg",
        "expiration_date": "07/2022",
        "manufacture_details": "Pfizer",
        "diluent_name": null,
        "indications": null,
        "side_effects": null,
        "storage_instructions": null,
        "administration_route": "IV or IM use",
        "precautions": null,
        "regulatory_information": null
}
```

```
"refined_json": {
        "drug_name": "Dynastat",
        "active_ingredients": [
        "parecoxib (as parecoxib sodium)"
        ],
        "dosage_information": "40 mg",
        "expiration_date": "07/2022",
        "manufacture_details": "Pfizer",
        "diluent_name": "bacteriostatic Water for Injections
           with benzyl alcohol",
        "indications": null,
        "side_effects": null,
        "storage_instructions": "Store unreconstituted product
            below 25\u00b0C. Protect from light.",
        "administration_route": "IV or IM use",
        "precautions": null,
        "regulatory_information": null
}
```


Figure 6.13: Example of retrieval-augmented enrichment where additional fields are introduced based on external context. While core attributes remain unchanged, new fields such as diluent and storage instructions are added without explicit support from the original OCR-derived content.

inspecting structured outputs generated by the OCR and LLM pipeline. Rather than replacing clinical workflows, the interface is designed to support candidate identification and structured record generation, while enabling human verification prior to database entry.

6.5.1 Clinical Workflow and System Integration

The system is designed for integration into pharmacy and operating theatre workflows. The recommended workflow comprises four stages that balance automation with clinical safety oversight:

Stage 1: Image Capture Clinical staff capture 8-10 images of a medication vial/ampoules from different angles. Images are acquired using a standard camera or mobile device, requiring minimal training. Multi-angle capture ensures that OCR can extract text from portions of the label that may be obscured or distorted in any single image. This stage is quick and can be performed by pharmacy technicians or nursing staff without specialized equipment.

Stage 2: Automated Processing The captured images are uploaded to the web application backend, which automatically processes them through the complete pipeline: OCR extraction, LLM consolidation, schema extraction, and RAG-based refinement/enrichment.

Stage 3: Pharmacist Review and Verification After automated processing, the web application displays the extracted JSON record. The pharmacist reviews the structured data, comparing key fields (medication name, dosage, expiration date, manufacturer) with the physical label and vial. This stage ensures clinical safety by requiring expert judgment before data entry, while the automated system handles the most labor-intensive aspects of manual transcription.

Stage 4: Database Entry and Clinical Retrieval Verified and approved records

are committed to the medication database (MongoDB). The database is queryable by clinical staff for rapid medication identification and information retrieval, supporting tasks such as medication verification.

6.5.2 Workflow Overview

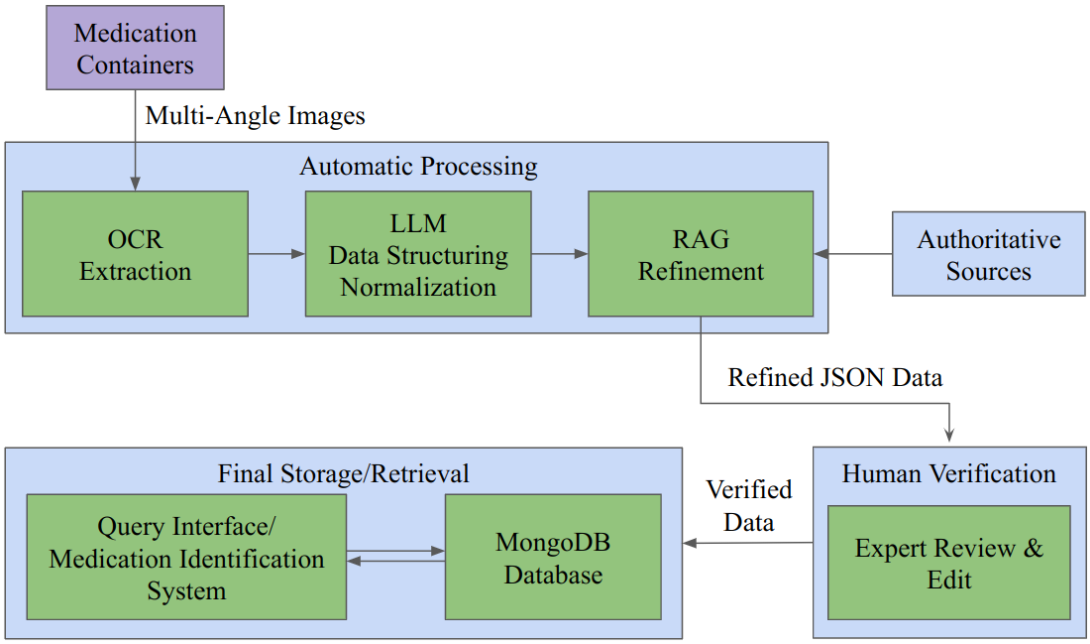


Figure 6.14: Schematic overview of the medication database pipeline, flow from multi-angle image acquisition, through automated processing utilizing OCR, LLM, and RAG refinement, to human expert validation and final storage in MongoDB for clinical retrieval.

Figure 6.14 illustrates the overall pipeline. Multi-view images are processed using OCR and LLM-based structuring to generate a candidate record, which is then reviewed before storage. The inclusion of a manual verification step reflects the current limitations of automated extraction and ensures that the stored data remains consistent with the source label.

6.5.2.1 Medication Image Upload and Automatic Processing

As shown in Fig. 6.15, users can upload multiple images of a medication container. These images are processed by the backend pipeline, which includes OCR-based text extraction, LLM-driven text consolidation, and RAG-based refinement. The output is a structured JSON representation of the medication information, forming the initial candidate data for pharmacist review. The interface provides real-time progress feedback during processing, allowing users to monitor completion and identify any system errors that prevent data generation.

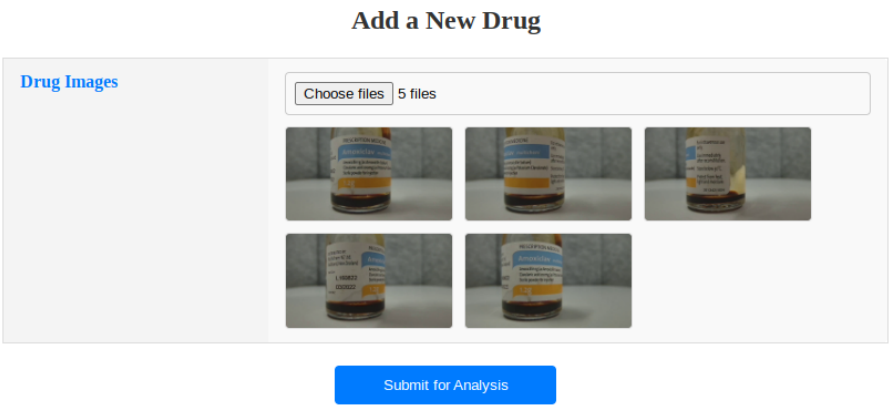


Figure 6.15: Users can upload multiple images of drug containers for processing. The backend pipeline processes these images using OCR, LLM, and RAG to generate structured drug information.

6.5.2.2 Manual Verification and Editing

After automatic extraction and refinement, the web application allows users to manually inspect, verify, and edit the extracted medication data, as illustrated in Fig. 6.16. The interface displays the extracted JSON fields alongside the original vial images, enabling direct comparison. Pharmacists can modify any field with a single click (e.g., correcting an OCR error, entering missing information visible on the label, or confirming values from the physical vial). This verification step is essential for correcting any discrepancies

or ambiguities in OCR or LLM outputs, maintaining high data integrity while leveraging the efficiency of automated extraction.

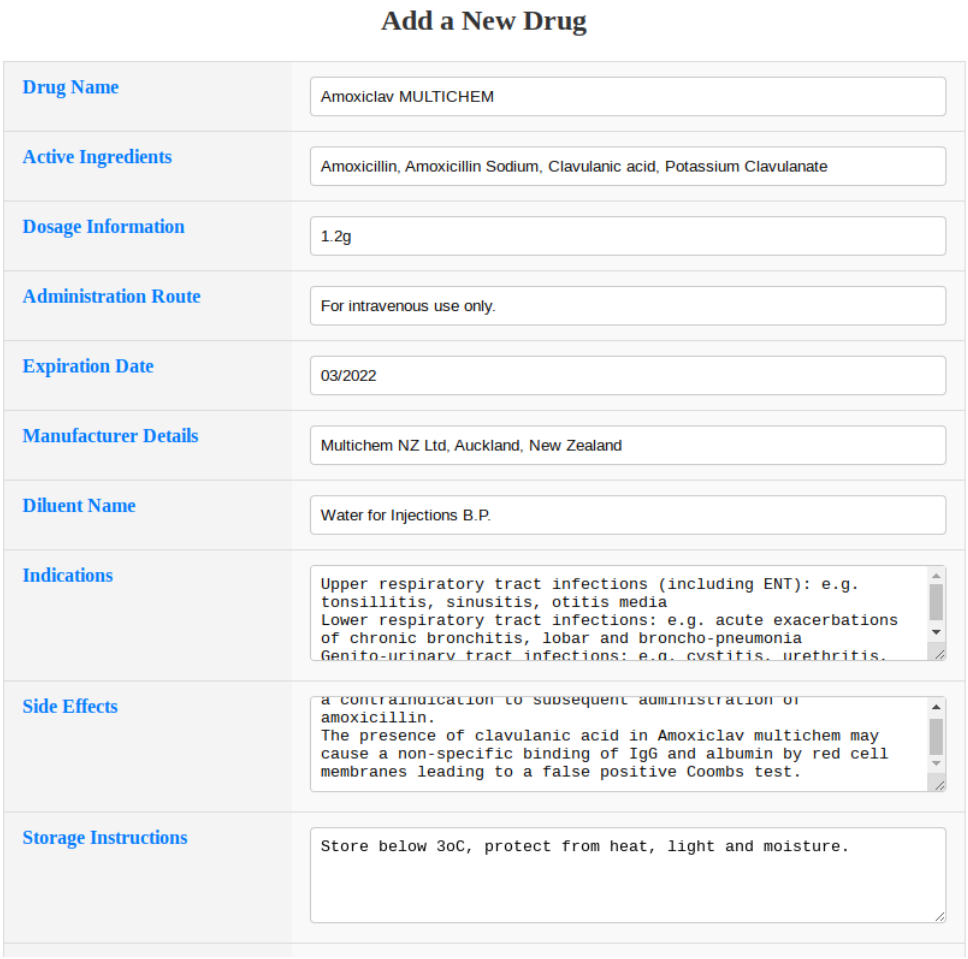


Figure 6.16: Users can manually verify and edit extracted drug data before storing it in the database. This step ensures data quality and accuracy.

6.5.2.3 Database Storage and Retrieval

Verified records are stored in MongoDB as structured JSON documents. The interface (Figure 6.17) supports basic search and retrieval, allowing users to access stored medication information when needed. This functionality is intended to assist with documentation and lookup, rather than to replace existing clinical verification processes.



Drug Management System

Drug List | Add a New Drug

Drug List

Drug Name	Active Ingredients	Dosage Information	Administration Route	Expiration Date	Manufacture Details	Diluent Name	Side Effects	Storage Instructions	Precautions	Regulatory Information	Actions
Protamine Sulfate 10mg/ml Solution for Injection	Protamine Sulfate	For slow I.V. injection over a period of about 10 minutes. No more than 50mg of protamine sulfate should be given in any one dose.	I.V.	OCT23	BNPU40307	Not specified in the provided data	In gross excess, protamine itself acts as an anticoagulant.	Not specified in the provided data	N MEDICINE, RESCRIPTIO	Distributed by: Artex Ltd., N.Z, CP Pharmaceuticals Ltd., U.K.	View Details Edit Delete
Cefazolin-AFT powder for injection equivalent to Cefazolin	Cefazolin sodium	1g	For IM/IV use	EXP03.2015	AFTpharmaceuticals	Sterile Water for Injection	Not specified	Store below 25°C, Protect from light	Please refer to enclosed leaflet before use	Not specified	View Details Edit Delete
Amoxiclav Multichem	Amoxicillin, Amoxicillin Sodium, Clavulanic acid, Potassium Clavulanate	1.2g	For intravenous use only.	03/2022	DISTRIBUTED BY: PrescripTO Multichem NZ Ltd Auckland, New Zealand	Water for Injections B.P.	Amoxiclav multichem should be used with caution in patients with evidence of hepatic dysfunction. Amoxiclav multichem should be given with caution to patients who have experienced symptoms of allergy associated with a cephalosporin or penicillin. Massive doses of Amoxiclav multichem can cause hypokalaemia and sometimes hypernatraemia. Use of a potassium-sparing diuretic may be helpful. In patients undergoing high-dose treatment for more than 5 days, electrolyte balance, blood counts and renal functions should be monitored. Hepatic events have been reported predominantly in males and elderly patients and may be associated with prolonged treatment. These events have been very rarely reported in children. Signs and symptoms usually occur during or shortly after treatment but in some cases may not become apparent until several weeks after treatment has ceased. These are usually reversible. Hepatic events may be severe and in extremely rare circumstances, deaths have been reported. These have almost always occurred in patients with serious underlying disease or taking concomitant medications known to have the potential for hepatic effects. In patients with renal impairment, dosage should be adjusted according to the degree of impairment (see section 4.2).	Store below 30°C, protect from heat, light and moisture.	Use immediately after reconstitution.	261342/1604 (or) 261342	View Details Edit Delete

Figure 6.17: The database view displays all stored drugs, supporting quick search, query, and retrieval of structured drug information.

6.5.3 Front-end Design

The web application's user interface is designed to be intuitive and responsive. Key components include:

Image Upload Interface Users can upload images of medication labels for processing, with clear visual feedback during each processing stage (OCR extraction $\rightarrow$ LLM consolidation $\rightarrow$ Schema extraction $\rightarrow$ RAG refinement).

Medication List and Search A medication list displays all medications stored in the database, allowing clinical staff to browse, search, and filter by attributes such as name, dosage, manufacturer, or storage requirements. The search interface supports partial name matching and autocomplete, reducing typing errors and enabling rapid access to information in time-critical scenarios (e.g., verifying a medication during anesthesia induction).

Detailed Medication View and Editing Selecting a medication opens a detailed view showing all extracted attributes: medication name, active ingredients, dosage, expiration date, manufacturer, indications, side effects, storage instructions, administration route, precautions, and regulatory information. Clinical staff can review critical safety

information (precautions, side effects, drug interactions) directly from this interface.

6.5.4 Backend and Database Architecture

The backend architecture manages medication data and provides seamless communication between the frontend and the database through RESTful APIs. The system is designed for reliability, scalability, and integration with existing clinical information systems.

Database Design Medication records are stored in MongoDB as JSON documents, allowing flexible and scalable storage of diverse attributes, including name, dosage, ingredients, side effects, regulatory information, and clinical use metadata. The schema is consistent with the structured format defined in Table 6.1, enabling direct ingestion of LLM-extracted data without transformation.

RESTful API Endpoints The backend exposes several RESTful API endpoints to support typical operations:

- **GET /drugs:** Retrieves a list of all stored medications, with optional filtering by name, dosage, or manufacturer.
- **GET /drugs/{id}:** Retrieves detailed information for a specific medication.
- **POST /process:** Accepts uploaded images and initiates the automated processing pipeline (OCR + LLM + RAG).
- **POST /drugs:** Adds new medication records to the database (after pharmacist verification).
- **PUT /drugs/{id}:** Updates existing medication records with pharmacist edits.
- **DELETE /drugs/{id}:** Removes specific medication records.
- **GET /search?q={query}:** Full-text search across all medication fields.

6.5.5 Implementation Details

The backend system is implemented using modern web and AI technologies to ensure robust, efficient, and scalable operation, using the technology stack:

- **Python:** Serves as the primary language for backend logic, including OCR processing, LLM-based text consolidation, and API management.
- **LangChain:** LangChain provides orchestration of workflow management and integration of the RAG pipeline components (retrieval, LLM reasoning, refinement).
- **FastAPI:** Provides fast, asynchronous backend services and RESTful APIs, connecting the frontend and the database efficiently. FastAPI's async/await support enables concurrent processing of multiple medication images without blocking.
- **MongoDB:** A schema-less NoSQL database that stores structured medication data in flexible JSON format. MongoDB's scalability and flexible schema support future extensions to the medication record structure.
- **React.js and Axios:** React.js builds a dynamic and responsive frontend, while Axios manages API requests for real-time CRUD operations and interface updates. React's component-based architecture enables modular design of the upload, review, and search interfaces.

The system can be deployed on cloud platforms (AWS, Azure, Google Cloud) or on-premises depending on institutional requirements and data privacy regulations. The containerized architecture (using Docker) enables rapid deployment across multiple clinical sites with consistent behavior. The modular design allows horizontal scaling:

additional API servers and LLM inference nodes can be added to handle increasing medication processing volume without architectural changes.

6.6 Conclusion

This chapter presented a system for assisting medication label data entry from container images in clinical settings. The approach combines OCR-based text extraction, LLM-based structuring, and retrieval-based enrichment to produce structured representations of medication labels. The primary objective is to reduce the effort associated with manual transcription while maintaining a human-in-the-loop workflow that supports verification.

Through qualitative case studies and limited quantitative evaluation, the system demonstrates the ability to recover key safety-critical fields, such as drug name, dosage information, and expiration date from multi-view inputs, even in the presence of OCR noise and partial occlusion. In addition, the structured outputs provide a consistent representation of medication data that can support downstream storage and retrieval. These findings suggest that the approach can assist with candidate identification and preliminary record generation, particularly in scenarios where manual transcription would otherwise be required.

The system is implemented within a web-based interface that integrates image capture, automated processing, manual review, and database storage. This workflow reflects a design choice to prioritize safety and traceability: extracted outputs are treated as intermediate results requiring human verification, rather than as authoritative clinical records. In this context, the system shifts the role of clinical staff from manual data entry to validation and correction, while retaining responsibility for final decision-making.

Overall, this work demonstrates the feasibility of combining OCR, language models, and retrieval mechanisms to support structured data generation from medication

labels. However, the current results should be interpreted as an initial step toward semi-automated workflows, rather than evidence of readiness for autonomous clinical deployment.

6.7 Limitations

Several limitations should be considered when interpreting the results of this study.

First, the evaluation is based on a relatively small dataset and relies primarily on case-based analysis, supplemented by limited quantitative metrics. The study does not include large-scale benchmarking or statistical validation across a diverse range of medications and label formats. As such, the reported performance should be viewed as indicative rather than conclusive.

Second, system performance is dependent on OCR quality and image acquisition conditions. Variations in lighting, blur, label curvature, and viewing angles can affect text extraction, which propagates to downstream structuring. While multi-view capture improves coverage, it does not fully resolve ambiguity or recognition errors in all cases.

Third, the retrieval-augmented component introduces a trade-off between completeness and traceability. While RAG can enrich missing fields and provide useful contextual information, the retrieved content is not guaranteed to be fully aligned with the specific input label. As a result, outputs may appear more complete, particularly for non-critical fields, but still require human review to confirm correctness before use.

Finally, the system is evaluated as a prototype with a standalone interface and database. Integration with existing hospital information systems, as well as compliance with clinical safety standards and regulatory requirements, has not been addressed. In addition, no formal user studies have been conducted to assess usability, efficiency gains, or impact on clinical workflows.

6.8 Future Directions

Future work will focus on improving the reliability of extraction and the alignment between retrieved information and source evidence.

On the perception side, OCR robustness can be improved through preprocessing techniques such as denoising, contrast enhancement, and geometric correction. Multimodal approaches that incorporate both visual layout and text may further improve performance on curved or partially occluded medication labels.

On the reasoning side, performance may be enhanced using instruction-tuned language models adapted to biomedical or pharmaceutical domains. This includes both API-based models and locally deployed open-weight alternatives, depending on constraints such as privacy, latency, and offline deployment. Fine-tuning and prompt optimization may further improve consistency in structured extraction, particularly for drug names and dosage fields.

For retrieval, future work should explore improved query design, relevance filtering, and ranking strategies to reduce misaligned or loosely related context. Adding confidence estimation or evidence tracking may also improve interpretability of enriched outputs.

Finally, more systematic evaluation in clinical settings is required. This includes expanding the dataset, defining formal annotation protocols, and conducting user studies to assess usability and workflow impact.

Overall, future work aims to strengthen the system as an assistive tool for structured medication data extraction, with clear separation between automated suggestions and clinically validated information.

From Pipeline to Experimental Analysis

Having established the full processing pipeline from dual-camera acquisition to reconstructed text and verified database creation, we now turn to a detailed experimental evaluation. This chapter examines how each component performs in practice, focusing on the accuracy, robustness, and speed of the two reconstruction methods, as well as their impact on medication identification via database retrieval. The results presented here provide empirical evidence for the system's technical performance under controlled conditions.

Chapter 7

Experimental Results

This chapter presents a comparative experimental evaluation of two pipelines for medication identification on cylindrical containers: an image stitching-based pipeline followed by Optical Character Recognition (OCR) (Approach 1) and an OCR-based text stitching pipeline (Approach 2). Experiments were conducted on a controlled dataset to examine the accuracy, robustness, and efficiency of both approaches under challenges such as label distortion, varying orientations, and multi-angle visual inputs. The analysis includes quantitative metrics such as top-k identification accuracy, score distribution observations, and processing time measurements, complemented by qualitative visual comparisons. The comparison provides an initial indication of the relative strengths and limitations of each method, offering insights into their potential applicability in real-world operating theatre environments. In addition, a brief error analysis highlights representative failure cases to illustrate system limitations and inform future improvements. Overall, the findings suggest the technical feasibility of the proposed end-to-end medication identification pipeline, particularly in anaesthesia and operating theatre settings.

7.1 Introduction

Accurate medication identification is essential for minimizing medication errors in operating theatre environments, particularly where cylindrical containers introduce curvature, perspective distortions, and incomplete label visibility. To address these challenges, the preceding chapters introduced two complementary label reconstruction pipelines: Image Stitching (Chapter 4) and Text Stitching (Chapter 5) with the Large Language Model (LLM)/Retrieval-Augmented Generation (RAG)-constructed verified medication database (Chapter 6), each designed to leverage dual-view image data while mitigating different sources of visual ambiguity. This chapter presents a comparative results analysis of the two label-reconstruction approaches developed in Chapters 4 and 5, evaluated under identical dataset and hardware conditions. As the two approaches differ in multiple respects beyond the domain in which reconstruction occurs, this comparison is presented as an evaluation of two complete candidate systems rather than a controlled single-variable test, specifically:

- **Approach 1**: Image Stitching reconstructs an unwrapped label by segmenting, aligning, and stitching images captured from dual viewpoints prior to performing OCR. This approach provides a consolidated visual representation that reduces curvature effects and enables OCR to operate on a more complete label surface. Its performance, however, depends on the accuracy of geometric alignment and the ability to minimize stitching artifacts.

- **Approach 2**: Text Stitching applies OCR directly to each individual view and subsequently merges the extracted text fragments via overlap analysis and sequence alignment. By operating in the text domain, this method avoids image-level geometric constraints and can be more efficient, though it must contend with inconsistent OCR outputs, partial words, and perspective-dependent variations.

Both methods ultimately produce a structured query string that is matched against the medication database. This ensures that performance metrics reflect not only OCR or stitching quality but the reliability of the full identification system operating under real-world constraints.

This chapter presents an empirical evaluation of the end-to-end pipeline, comparing accuracy, robustness, runtime efficiency, and failure modes across the two methods. The results provide insight into the operational trade-offs of each approach and their suitability for deployment in clinical environments requiring high confidence and safety.

7.2 Experimental Setup

7.2.1 Dataset Description

The experimental dataset consists of images of cylindrical medication containers captured from multiple angles using a dual-camera setup. This configuration ensures systematic visual coverage of each drug label. The dataset contains 53 distinct medication products, with each product represented by 20 pairs of images (one from the left and one from the right perspective), resulting in a total of 1,060 samples. Each image pair is annotated with a corresponding DrugId, serving as ground truth for evaluation.

All medication containers in the dataset were provided by Auckland City Hospital (Healthpoint, n.d.) and SAFERsleep (SAFERsleep Ltd., n.d.) and are commonly used in operating theatres, particularly for anesthesia. The dataset includes a variety of labels with different designs, text sizes, and orientations, reflecting real-world clinical conditions. Examples include Amoxiclav, Calcium Chloride, and Cefazolin Aft, as shown in Fig. 7.1. The samples represent naturalistic variation in lighting, container orientation, and camera positioning, though controlled under standardized laboratory capture conditions.

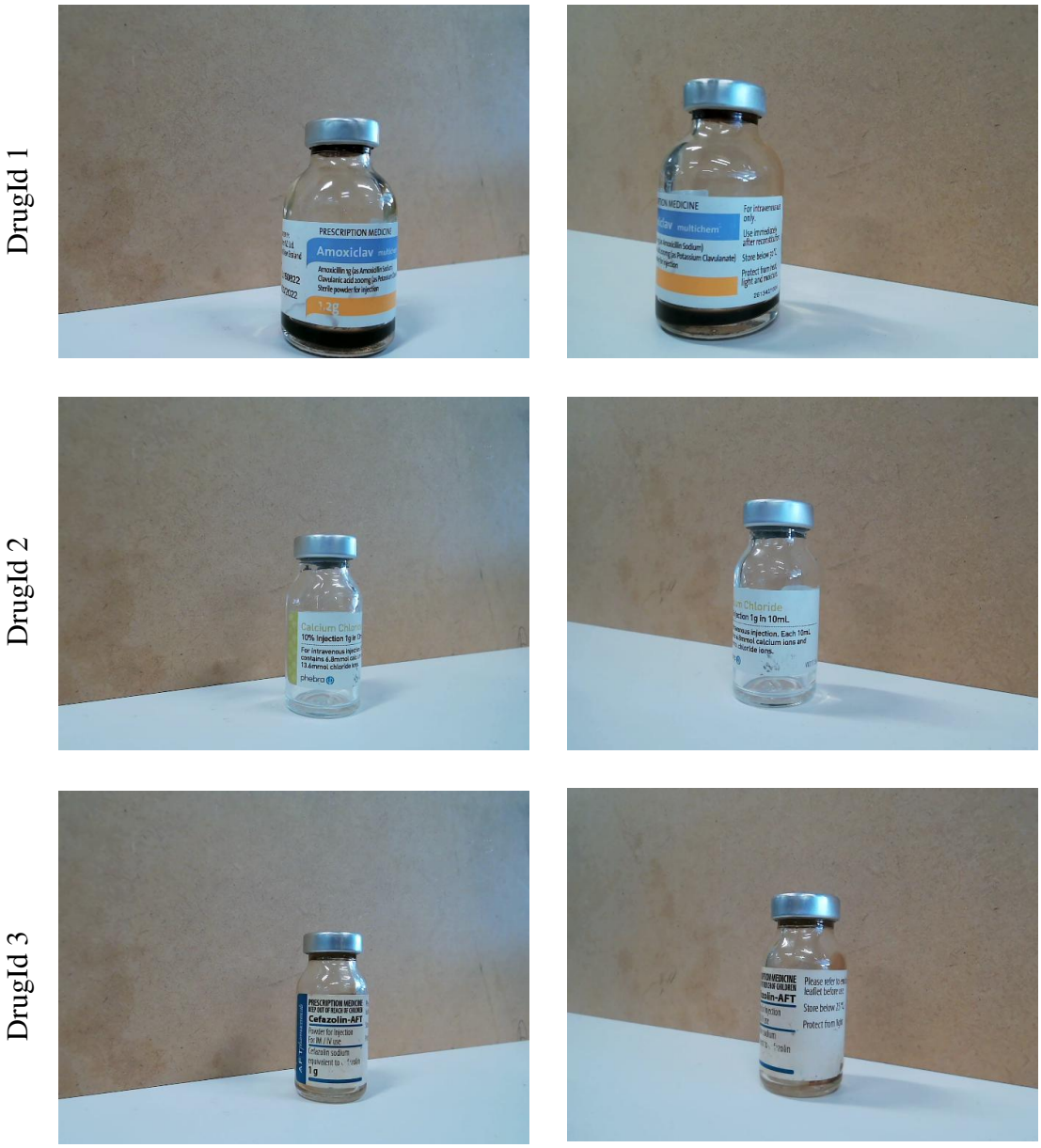


Figure 7.1: Sample images from the experimental dataset. Each medication ID is represented by two images captured from different angles (left and right views) using a dual-camera setup.

All 53 medication classes are equally represented with 20 samples each, ensuring no class imbalance that could bias evaluation metrics. The dual-camera capture protocol provides two complementary viewpoints per medication, simulating practical conditions involving partial visibility and perspective variation. All physical medication containers used in this evaluation are the same objects as those used in earlier data collection. However, the evaluation images are independently acquired and are not reused from any training, tuning, or validation pipelines.

In particular, although the same containers are observed, the evaluation dataset is constructed using a separate dual-camera acquisition protocol rather than video frame sampling or previously collected images. This produces a distinct set of image pairs with different viewpoints and capture conditions. This setup ensures no overlap in image data between training-related stages and evaluation, allowing results to reflect generalization to unseen observations of the same physical items, as described in Section 3.3.4.3 and Section 3.3.4.4.

7.2.2 Experimental Setup

To evaluate the proposed methods for medication identification, a shared experimental setup was established to ensure consistency across both approaches. The experiments were conducted using two computing platforms to reflect both high-performance and resource-constrained settings (Section 3.3.2). This setup enables a practical comparison of accuracy, processing speed, and scalability under different hardware conditions, providing indicative insights for potential deployment in anaesthesia environments.

It should be noted that the evaluation reported in this chapter was conducted under controlled laboratory conditions, using images captured in a consistent setting rather than the variable lighting, occlusion, and handling conditions of an actual operating theatre. The reported accuracy and latency figures should therefore be interpreted as

indicative of system performance under the tested conditions, rather than as a guarantee of real-world clinical performance. This limitation is discussed further in Section 9.3.

7.2.2.1 Hardware

Machine 1: High-Performance System This system is used for computationally demanding operations, including image stitching, OCR processing, and deep learning inference. It serves as a reference for performance under favourable hardware conditions.

- CPU: Intel(R) Xeon(R) Silver 4210 @ 2.20GHz
- GPU: NVIDIA GeForce RTX 2080 Ti (12GB)
- RAM: 100GB DDR4
- OS: Ubuntu 22.04 LTS

Machine 2: Resource-Constrained Laptop A standard laptop is used to approximate a resource-limited setting, reflecting scenarios where specialised hardware may not be available. While not a dedicated edge device, it provides a reasonable proxy for evaluating runtime behaviour and efficiency under constrained computational resources.

- CPU: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz
- GPU: TigerLake-LP GT2 [Iris Xe Graphics]
- RAM: 8GB
- OS: Ubuntu 22.04 LTS

Using these two platforms allows for a comparative assessment across differing hardware capabilities. In particular, differences in memory capacity and GPU architecture provide a practical indication of how performance may vary under constrained conditions.

7.2.2.2 Evaluation Metrics

The performance of the methods was evaluated using a set of metrics selected to reflect both identification accuracy and practical system behaviour.

Top-1 Accuracy: Measures the percentage of test samples where the correct medication is ranked as the top prediction. This metric provides a direct indication of identification correctness under the assumption of a single final decision.

$$\text{Top-1 Accuracy} = \frac{\text{Correct Top-1 Predictions}}{\text{Total Test Samples}} \times 100$$

Top-5 Accuracy: Measures the percentage of test samples where the correct medication appears within the top five predictions. This metric captures partial correctness and is useful in cases where multiple candidates may be considered before final confirmation.

$$\text{Top-5 Accuracy} = \frac{\text{Number of Samples with Correct Label in Top-5}}{\text{Total Test Samples}} \times 100\%$$

Top-1 and Top-5 accuracy are reported together to provide complementary views of system performance, reflecting both strict identification and tolerance to ambiguity in the prediction space.

Processing Time: Measures the total time required from input image to final prediction. This includes all major stages of the pipeline, such as label segmentation, image unwrapping and stitching (for Approach 1), OCR processing, text stitching (for Approach 2), and final matching or identification. This metric is used to assess computational efficiency across different hardware platforms and to provide an indication of runtime feasibility.

These metrics are chosen to provide a focused, quantitative comparison between the two approaches in a controlled experimental setting. While they offer insight into

identification performance and computational behaviour, they do not constitute a full evaluation of integration into clinical workflows. In particular, factors such as user interaction, decision latency in practice, and downstream clinical impact are not assessed in this study.

7.2.3 Method 1: Image Stitching Setup

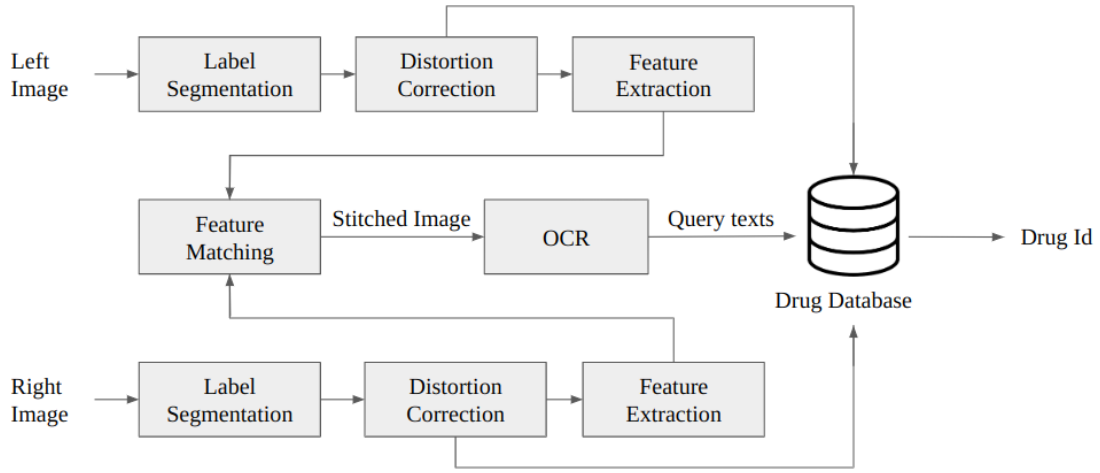


Figure 7.2: Overview of Method 1: Image stitching followed by OCR for drug label identification. The pipeline emphasizes geometric reconstruction prior to text extraction, enabling a consolidated visual representation of the medication label.

Method 1 applies an image stitching approach to generate a continuous representation of cylindrical medication labels, with the aim of supporting OCR performance. Figure 7.2 illustrates the end-to-end pipeline, which consists of the following steps:

1. **Image Capture**: Images are captured using a dual-camera setup to provide broader coverage of the cylindrical label under varying viewpoints.

2. **Label Segmentation**: A segmentation model isolates the medication label from the background, reducing interference from surrounding objects.

3. **Distortion Correction**: Ellipse detection and geometric transformations are applied to unwrap the cylindrical label into a flat 2D image. This step assumes an

approximately cylindrical surface, which represents an inherent limitation of the approach, as deviations from this assumption may introduce reconstruction errors (Section 4.5).

4. **Image Stitching**: Key features are extracted and matched across multiple images to produce a more complete label representation. Feature matching is based on SIFT descriptors with RANSAC-based geometric verification.

5. **OCR Text Extraction**: OCR is applied to the stitched image to produce a consolidated text output.

6. **DrugId Query**: The extracted text is used to query a structured drug database to retrieve relevant medication information. The database follows the schema introduced in Chapter 6 and is stored using SQLite in this experimental setup.

All core components, including segmentation, ellipse detection, OCR, and retrieval, follow the configurations described in Chapter 4. This method focuses on improving label continuity prior to text extraction. However, as the pipeline involves multiple stages, errors may propagate through subsequent steps, which can influence overall system performance.

7.2.4 Method 2: Text Stitching Setup

Method 2 adopts an alternative strategy to Method 1 by applying OCR directly to individual images and then merging the extracted text fragments to reconstruct the medication label. The workflow, illustrated in Figure 7.3, consists of the following steps:

1. **Image Capture:** As in Method 1, left and right images of the medication labels are captured from two different viewpoints using a dual-camera setup.

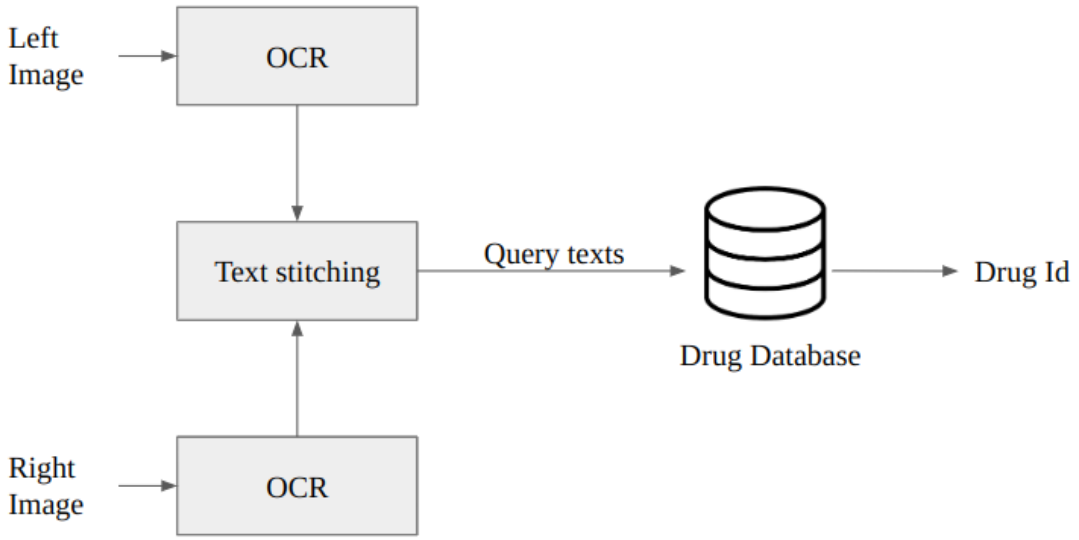


Figure 7.3: Overview of Method 2: OCR followed by text stitching. By operating in the text domain, this pipeline avoids geometric reconstruction, reducing computational cost and sensitivity to misalignment.

2. **OCR on Individual Images:** Each image is processed independently using OCR to extract text fragments from the medication label. This approach avoids geometric correction and reduces sensitivity to misalignment and perspective distortion, as recognition is performed locally within each view.

3. **Text Stitching:** Extracted text fragments are aligned and merged to form a single coherent representation of the medication label. The stitching procedure is described in detail in Section 5.3.

4. **DrugId Query:** The reconstructed text is used to query the same structured drug database as in Method 1. For both methods, matching is restricted to information present on the medication label only, ensuring a consistent and fair comparison between approaches.

Method 2 is designed to complement Method 1 by shifting the reconstruction process from the image domain to the text domain. This introduces a trade-off: while it removes dependency on geometric modelling and reduces sensitivity to distortion, it becomes

more reliant on OCR accuracy and the quality of text alignment during stitching. This trade-off motivates the comparative evaluation of both approaches.

In terms of efficiency, Method 2 is expected to reduce computational overhead by avoiding geometric processing steps. This suggests potential benefits in runtime performance on resource-constrained hardware. However, these aspects are further examined empirically in the results section.

7.3 Quantitative Results

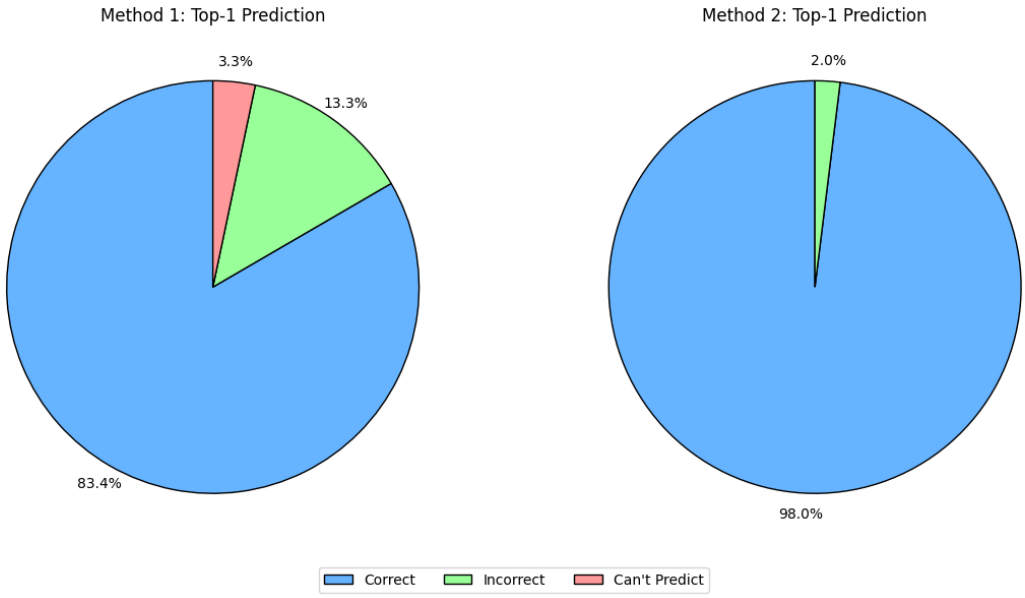
7.3.1 Top-1 and Top-5 Accuracy

Table 7.1: Top-1 and Top-5 accuracy comparison between Method 1 (Image Stitching) and Method 2 (Text Stitching) on 1,060 test samples.

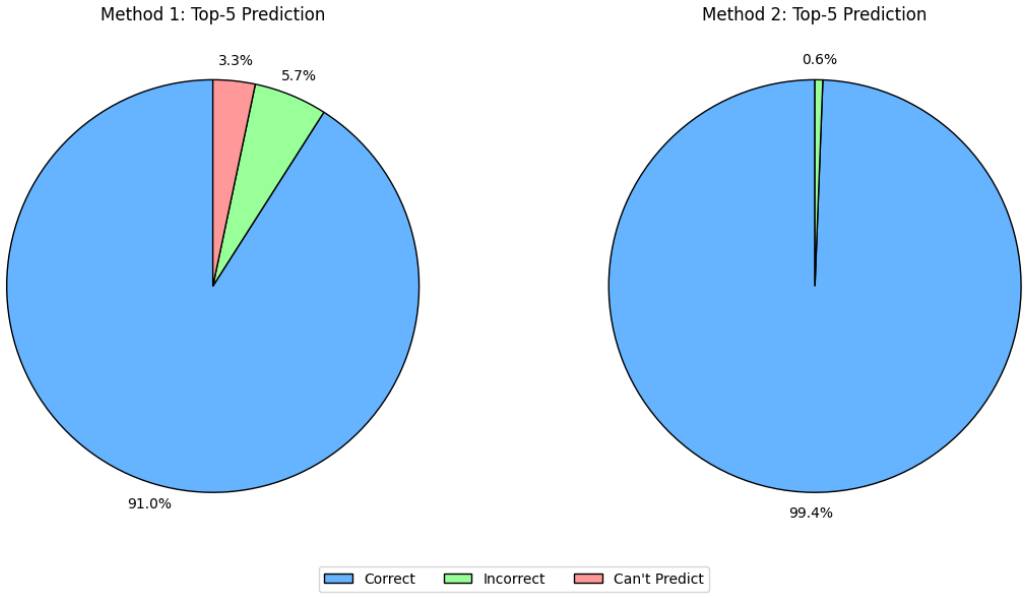
Metric	Method 1	Method 2	Difference
Top-1 Accuracy	83.40% (884/1060)	98.02% (1039/1060)	+14.62 pp
Top-5 Accuracy	91.04% (965/1060)	99.43% (1054/1060)	+8.39 pp

Table 7.1 summarizes the Top-1 and Top-5 accuracy achieved by the two proposed methods over the full test set. Method 2 (Text Stitching) achieves a Top-1 accuracy of 98.02% (1,039 correct predictions from 1,060 samples) compared to 83.40% for Method 1 (884 correct predictions from 1,060 samples). This represents an absolute improvement of 14.62 percentage points. A similar trend is observed for Top-5 accuracy, where Method 2 reaches 99.43%, exceeding Method 1 by 8.39 percentage points.

The accuracy metrics reported above are calculated relative to the total sample count (1,060 samples), including cases where predictions are unavailable. Figure 7.4 provides a complete breakdown of prediction outcomes. Method 1 exhibits “unable-to-predict” cases in 35 out of 1,060 samples (3.3%). These failures occur when image stitching does not produce usable output, preventing downstream OCR processing. In these cases, no prediction is generated. In contrast, Method 2 produces predictions for all 1,060



(a) Top-1 Prediction



(b) Top-5 Prediction

Figure 7.4: Distribution of prediction outcomes for Top-1 and Top-5 evaluation. Categories include correct predictions, incorrect predictions due to database mismatch, and unable-to-predict cases caused by upstream pipeline failures (e.g., unsuccessful image stitching).

samples (100%), demonstrating greater robustness across varying viewpoints and label conditions. When Method 1 failures are excluded and accuracy is recalculated on the

1,025 samples with predictions, Method 1 achieves 86.14% Top-1 accuracy (883/1,025). This distinction clarifies that Method 1's lower absolute accuracy reflects both lower per-sample performance and loss of entire samples to system failure.

7.3.2 Top-1 Similarity Scores and Confidence Analysis

Table 7.2: Statistical summary of Top-1 similarity scores for both methods. Scores range from 0 to 1, with 1.0 indicating perfect match confidence.

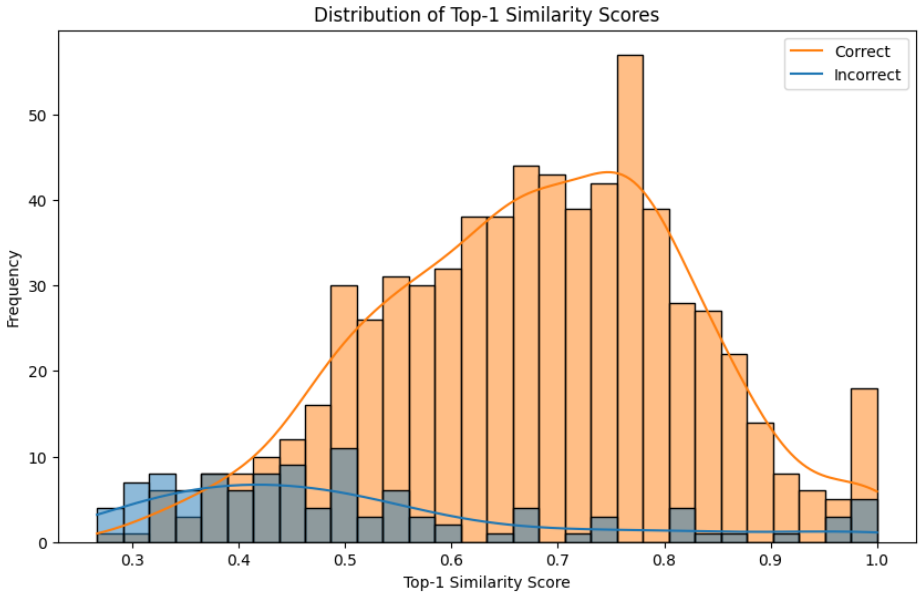
Statistic	Method 1	Method 2	Difference
Mean	0.8622	0.9804	+0.1182
Std. Dev.	0.3447	0.1385	-0.2062

Figure 7.5 illustrates the distribution of Top-1 similarity scores for both methods. In general, both approaches produce high similarity scores for correct predictions, indicating reasonable overall performance. However, differences are observed in the spread and consistency of scores.

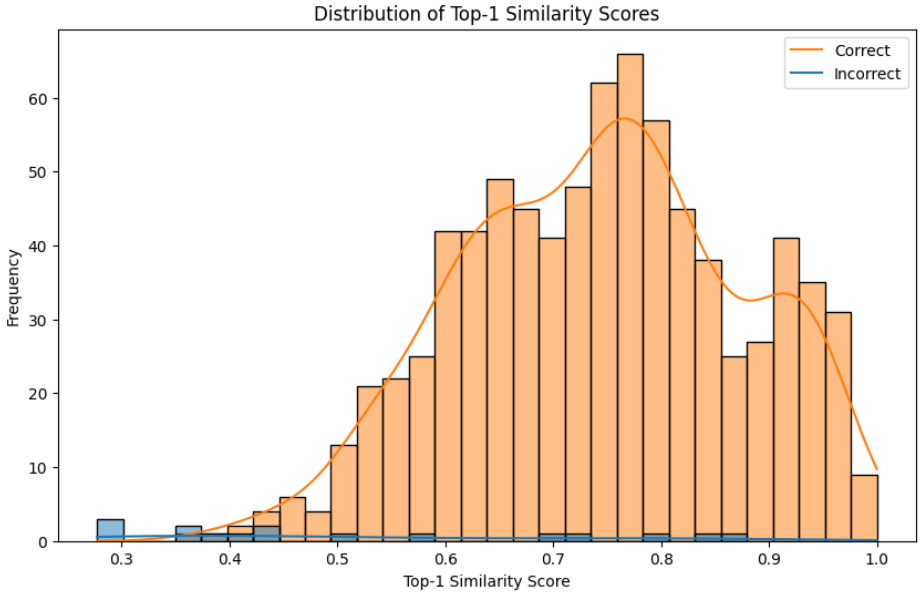
Method 1 shows a broader distribution, with incorrect predictions appearing across a wider range of similarity values (approximately 0.3-0.5). This suggests that similarity scores are less consistently aligned with prediction correctness. In contrast, Method 2 exhibits fewer incorrect predictions and a more concentrated distribution. On this dataset, Method 2 shows a strong empirical trend toward higher and tighter similarity scores.

These observations are supported by the summary statistics in Table 7.2. Method 2 achieves a higher mean similarity score (0.9804 vs. 0.8622) and lower variability (standard deviation 0.1385 vs. 0.3447). The larger spread in Method 1 indicates less consistent scoring across samples, whereas Method 2 produces more stable similarity values.

Figure 7.6 further highlights this difference, where Method 2 shows a higher median and a tighter interquartile range. Both methods contain low-score outliers, which likely



(a) Method 1



(b) Method 2

Figure 7.5: Distribution of Top-1 similarity scores for Method 1 (Image Stitching) and Method 2 (Text Stitching). Method 2 produces substantially higher scores for correct matches and exhibits tighter clustering.

correspond to more challenging cases such as strong OCR noise or ambiguous label content.

Overall, Method 2 produces more consistent similarity scores, which may better

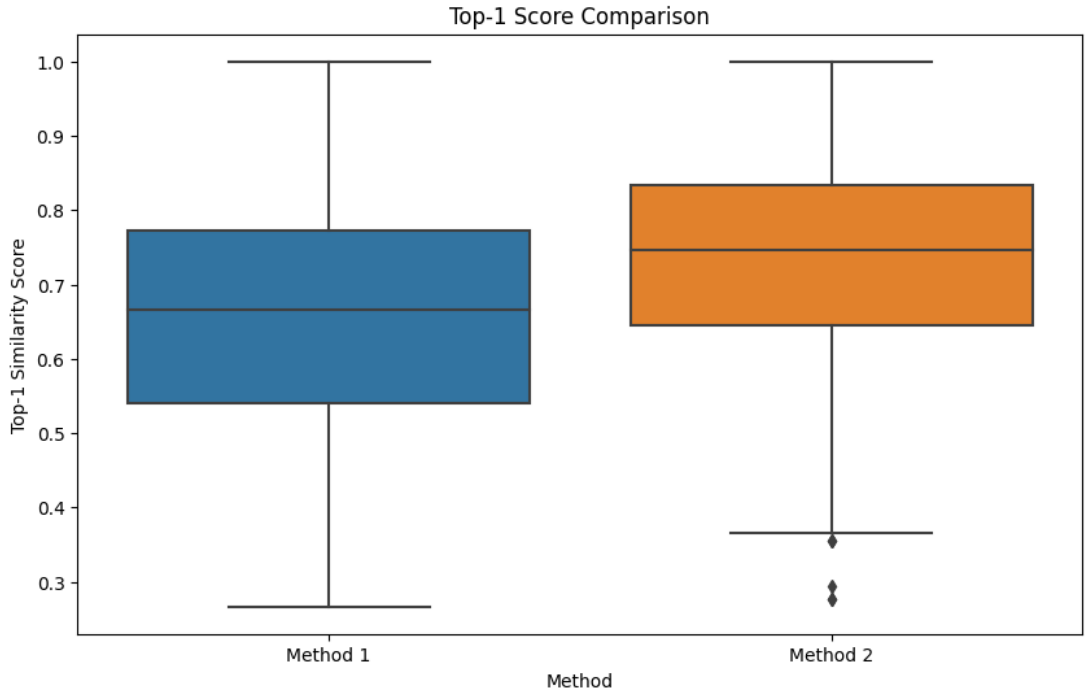


Figure 7.6: Comparison of Top-1 similarity score distributions. Method 2 exhibits higher mean, tighter interquartile range, and greater separation between correct and incorrect predictions.

support ranking-based review. In this work, the system returns the highest-similarity candidate along with a ranked list of alternatives, rather than relying on a fixed decision threshold. This design avoids the need for explicit score calibration while still providing a notion of confidence through relative ranking.

7.3.3 Speed and Computational Efficiency

Table 7.3 reports the execution time breakdown for each processing stage on Machine 1 (high-performance system). All timings were measured end-to-end within a single process, including all data loading and preprocessing overheads.

Specifically, method 1 incurs substantial computational overhead due to its image-centric processing pipeline, which includes label segmentation (21.90 ms for both views), ellipse detection (105.78 ms), label unwrapping (19.18 ms), and feature-based

Table 7.3: Execution time breakdown for Method 1 and Method 2 on Machine 1 (high-performance system). Times are reported in milliseconds, averaged across all 1,060 test samples. The "(2 ×)" notation indicates processing of both camera views.

Processing Step	**Method 1 (ms)**	**Method 2 (ms)**
Label Segmentation	(2 ×) 10.95	N/A
Ellipse Detection	(2 ×) 52.89	N/A
Label Unwrapping	(2 ×) 9.59	N/A
Image Stitching	714.21	N/A
Text Stitching	N/A	3.00
OCR	60.03	(2 ×) 60.03
Database Query	20.05	20.05
Total Time	**941.15**	**143.11**

image stitching (714.21 ms). Image stitching is the dominant bottleneck, accounting for 714.21 ms and representing approximately 75.9% of total pipeline runtime. This is expected, as stitching relies on SIFT descriptor extraction, feature matching, and RANSAC-based verification. The cumulative time for geometric preprocessing adds further overhead.

In contrast, Method 2 achieves significantly lower latency with a total processing time of 143.11 ms, representing a speedup factor of 6.58× compared to Method 1. By eliminating image-space geometric reconstruction, Method 2 replaces the most expensive stage (image stitching at 714.21 ms) with lightweight text stitching (3.00 ms). Although OCR is executed independently on both images (120.06 ms combined vs. 60.03 ms for Method 1), the overall savings from eliminating geometric operations is substantial. This architectural difference represents the key explanation for the runtime gap between methods: text-domain reconstruction avoids the computational cost of geometry-based image operations.

Table 7.4 details the runtime performance of Method 2 on Machine 2, a resource-constrained system representative of portable deployment scenarios. Method 1 could not be evaluated on this hardware due to resource constraints: the combined requirements

Table 7.4: Execution time breakdown for Method 2 on Machine 2 (portable edge-device system). Times are reported in seconds, averaged across all 1,060 test samples.

Processing Step	**Time (s)**
OCR	(2 ×) 0.434
Text Stitching	0.004
Database Query	0.248
Total Time	**1.12**

of GPU memory (for neural network-based segmentation and feature detection) and system RAM exceeded the available 8GB, preventing successful execution of the deep learning components required by Method 1.

Method 2's performance on Machine 2 is characterized by longer absolute latency but remains computationally feasible. The total pipeline completes in 1.12 seconds. OCR remains the dominant component, reflecting the computational cost of text recognition on resource-limited hardware. Database query time increases from 20 ms on Machine 1 to 248 ms on Machine 2, reflecting memory bandwidth limitations. Text stitching remains negligible (4 ms) even on constrained hardware.

These timing results demonstrate that Method 2 is computationally feasible on both high-performance and portable hardware platforms, whereas Method 1 is restricted to adequately resourced systems.

7.4 Qualitative Results and Failure Mode Analysis

7.4.1 Method 1: Image Stitching Failure Characteristics

Method 1 leverages image stitching to reconstruct a continuous representation of the medication label prior to OCR. When alignment is successful, this approach preserves spatial continuity and can improve readability.

Two types of failure are observed under the acquisition conditions considered in

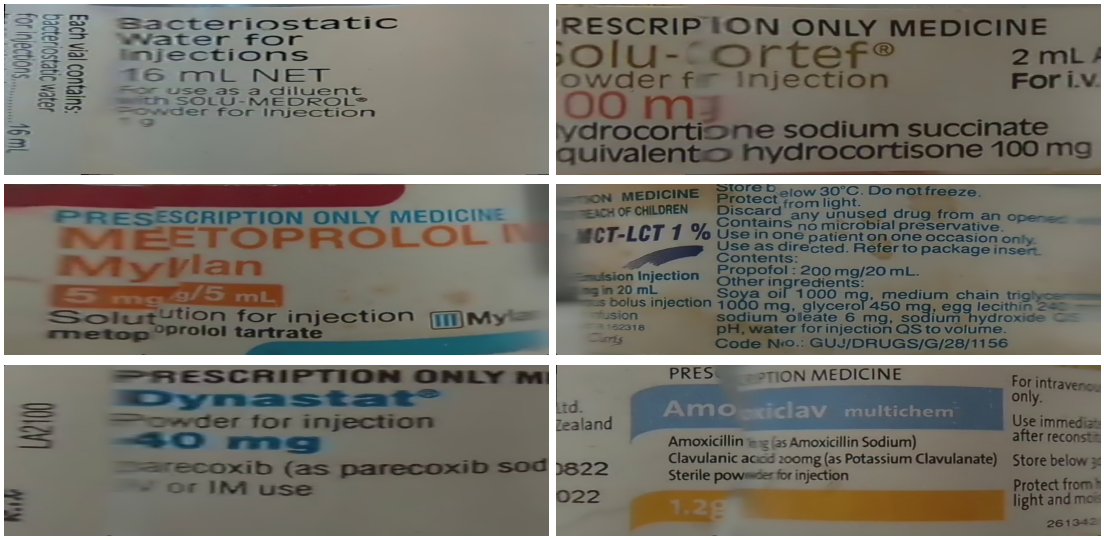


Figure 7.7: Representative failure cases of Method 1, illustrating artifacts caused by image stitching and geometric unwrapping errors..

this study. First, in 35 cases, stitching fails completely and no reconstructed image is produced. In these cases, OCR cannot be applied, and the pipeline produces no output for downstream querying.

Second, even when stitching succeeds, the reconstructed images may contain geometric distortions, misalignment, or discontinuities. As shown in Fig. 7.7, these artifacts can deform or fragment text regions, which in turn reduces OCR quality and may lead to incorrect or incomplete predictions.

These issues are primarily observed under challenging conditions such as limited overlap between views, strong curvature, reflections, and non-ideal viewpoints, where geometric alignment becomes less stable. Within this setting, the method shows sensitivity to reconstruction quality: failures can either prevent the pipeline from producing any output or degrade downstream performance.

Overall, these results highlight a limitation of Method 1 under the tested acquisition variability. While effective when alignment is reliable, its dependence on geometric reconstruction can reduce robustness in less controlled capture conditions.

7.4.2 Method 2: Text Stitching Robustness and Localized Failures

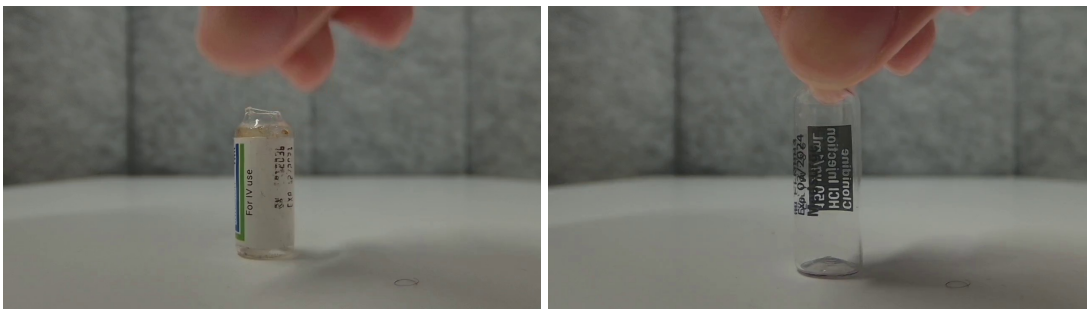


Figure 7.8: Representative failure cases of Method 2, illustrating performance degradation under poor label quality and overlapping text. Left: Metaraminol ampoule with faded, low-contrast printing; Right: Clonidine ampoule with densely packed and overlapping text on a curved surface, leading to ambiguous OCR outputs.

Method 2 performs OCR independently on each captured view and reconstructs the full label by stitching the extracted text fragments. By operating directly in the textual domain, this approach avoids geometric unwrapping and explicit image alignment, resulting in lower computational cost and improved robustness to cylindrical distortion. As demonstrated in the quantitative results, this design leads to more consistent prediction performance under the evaluated conditions.

Figure 7.8 illustrates representative scenarios where Method 2 performance degrades, highlighting limitations under challenging visual conditions. Errors primarily arise from poor label quality and complex text layouts that reduce OCR reliability. In cases of missing or faded text, OCR outputs become fragmented, and although multi-view stitching can partially recover missing information, reconstruction remains limited when all views are similarly degraded. Transparent ampoules with text printed directly on curved glass surfaces also present challenges, as the lack of a distinct label region reduces text visibility and hinders accurate extraction.

A key distinction between the two methods lies in their failure modes (Table 7.5). Method 1 exhibits abrupt failure behaviour: when intermediate steps such as image stitching break down, the pipeline may fail to produce any valid output. In contrast,

Table 7.5: Comparative failure mode characteristics of Method 1 and Method 2. Method 1 exhibits catastrophic failures (unable-to-predict), while Method 2 shows graceful degradation.

Failure Type	**Method 1**	**Method 2**
Complete Failure	Yes (35 cases, 3.3%)	No (0 cases)
Failure Characteristics	Catastrophic; no output	Graceful; low-confidence output
Primary Cause	Geometric misalignment	OCR errors on individual views
Clinical Implication	System unavailable	Reduced confidence but usable

Method 2 tends to degrade more gradually. Imperfect OCR outputs can still be partially consolidated, producing a ranked list of candidate predictions rather than a single failure outcome.

However, the practical usefulness of these lower-confidence outputs is not formally evaluated in this study. While ranking provides a relative measure of similarity, it does not guarantee correctness, and some outputs may remain incomplete or misleading. As such, these results should be interpreted as providing partial decision support rather than definitive identification, and human verification remains necessary, particularly in safety-critical settings.

Overall, these observations highlight a limitation of Method 2 under the tested acquisition variability: although it is more tolerant to upstream failures, its performance is ultimately bounded by OCR quality, and degraded inputs may still lead to uncertain or ambiguous predictions.

7.5 Comparative Analysis and Discussion

Table 7.6 provides a comprehensive side-by-side comparison of the two methods across multiple performance and deployment dimensions. The results reveal consistent advantages for Method 2 across accuracy, reliability, and computational efficiency

Table 7.6: Comprehensive comparison of Method 1 and Method 2. Method 2 demonstrates clear advantages in accuracy, robustness, speed, and clinical deployment feasibility, making it the preferred approach for operating theatre environments.

Aspect	**Method 1 (Image Stitching)**	**Method 2 (Text Stitching)**
Top-1 Accuracy	83.4%	98.02%
Complete Failures	3.3% unable-to-predict	0% unable-to-predict
Prediction Consistency	High variability (SD: 0.3447)	High consistency (SD: 0.1385)
Processing Speed (M1)	941.15 ms	143.11 ms (6.58× faster)
Edge Device Viability	Not feasible (GPU/memory constraints)	Fully feasible (1.12 s on Machine 2)
OCR Dependency	Moderate (geometric processing provides recovery)	High (errors propagate from both views)
Failure Mode	Catastrophic (no output)	Graceful (low-confidence output)
Visual Interpretability	High (reconstructed label visible)	Low (text-only output)
Computational Complexity	High (deep models, feature extraction, RANSAC)	Low (text algorithms only)
Overall Assessment	**Better for offline analysis with abundant resources**	**Better for scenarios prioritizing reliability and speed**

metrics.

7.5.1 Accuracy and Reliability

In terms of identification accuracy, Method 2 achieves higher performance across both Top-1 and Top-5 metrics. The 14.62 percentage point difference in Top-1 accuracy (98.02% vs. 83.40%) is substantial and represents a meaningful gap in experimental performance. This improvement is further strengthened by Method 2's lower variability in confidence scores (0.1385 vs. 0.3447 standard deviation), indicating more consistent prediction behavior across diverse inputs.

Critically, Method 1's 3.3% complete failure rate (35 samples) represents a system reliability limitation: when predictions are unavailable, the system provides no

assistance. Method 2's 100% prediction rate eliminates this availability gap.

7.5.2 Computational Efficiency and Deployment Scalability

Method 2 provides a 6.58× speedup on high-performance hardware (143.11 ms vs. 941.15 ms) and remains executable on resource-constrained hardware (1.12 seconds on Machine 2). In contrast, Method 1 cannot be executed on Machine 2 due to GPU and memory constraints, limiting its deployment to adequately resourced systems. This difference has practical implications for deployment flexibility.

7.5.3 Trade-offs in Method Selection

Method 2 demonstrates better performance for applications emphasizing reliability, speed, and computational efficiency within the constraints of the experimental conditions tested. Its consistent output availability, 14.62 percentage point accuracy advantage, and 6.58× speedup represent measurable technical advantages under these conditions.

Method 1 may retain value in scenarios where visual quality and interpretability are primary requirements (e.g., label auditing for regulatory compliance or offline analysis with abundant computational resources). However, within the experimental setup evaluated, the technical trade-offs favor Method 2 for speed and reliability-critical applications.

7.5.4 Limitations of the Current Evaluation

The evaluation was conducted on a controlled dataset of 1,060 images covering 53 medication products, collected under standardized laboratory conditions with a fixed dual-camera setup and consistent lighting. Because of this, the results may not fully

reflect performance under more variable real-world conditions, such as changes in illumination, container types, medications, or imaging systems.

In addition, the experiments were limited to a closed set of 53 known medications. The behaviour of the system on previously unseen medications or out-of-distribution inputs was not evaluated. Although Method 2 can still produce ranked predictions when inputs are degraded, the reliability of low-confidence outputs was not explicitly analysed in terms of their suitability for downstream use.

7.6 Conclusions

This chapter compared two approaches for automated medication identification on cylindrical containers. Method 1 uses image stitching followed by OCR, relying on geometric reconstruction before text extraction. This can produce visually coherent results under well-controlled conditions, but it is more sensitive to misalignment and requires higher computational cost. Method 2 instead performs OCR on individual views and reconstructs the result in the text domain, avoiding explicit geometric processing.

Under the evaluated conditions, Method 2 achieved higher accuracy (98.02% vs. 83.40% Top-1), produced predictions for all test samples (vs. 96.7% for Method 1), showed more stable confidence scores (Std. Dev. 0.1385 vs. 0.3447), and was significantly faster (143.11 ms vs. 941.15 ms on high-performance hardware, with feasibility on lower-resource systems).

A key difference between the two methods is how they fail. Method 1 may fail when geometric reconstruction is unstable, resulting in missing outputs. Method 2, on the other hand, tends to degrade more gradually and can still produce ranked predictions even when the input quality is poor. In this setting, this leads to more consistently available outputs.

While Method 1 may still be useful when visual reconstruction is required, Method 2

offers a more robust and efficient alternative under the tested conditions, particularly in terms of speed and output stability.

It is important to note that these results are based on a controlled dataset of 1,060 images covering 53 medication products under standardized capture conditions. The findings therefore reflect performance in this specific experimental setup and may not directly generalize to more variable environments, different imaging systems, or unseen medications without further validation.

Overall, the results suggest that operating in the text domain can be an effective and practical strategy for this task under the evaluated conditions, while also highlighting the dependence of system performance on OCR quality.

From Method Evaluation to System Implementation

While the previous chapters evaluated the proposed medication identification methods at an algorithmic level, translating research findings into a functional system requires addressing practical implementation challenges. This chapter presents a prototype system that combines the text stitching approach with hardware, software, and interface components designed for anaesthesia and operating theatre environments. The prototype serves as a proof-of-concept (PoC) demonstration that the methods evaluated in Chapter 7 can be implemented as a working system suitable for further development and evaluation in realistic clinical settings.

Chapter 8

Prototype Development for Medication Identification System

Section 2.7 characterised computer-vision-based medication identification as high-cost and poorly scalable relative to barcode and RFID alternatives, due to typical reliance on high-performance computing and cloud infrastructure. The prototype described in this chapter directly addresses this limitation: it runs entirely offline on a standard laptop with two consumer USB cameras, without GPU acceleration or network dependency, while meeting the accuracy threshold agreed with the industrial partner.

This chapter describes the development of a proof-of-concept medication identification system integrating dual-camera image capture, optical character recognition (OCR), text stitching, and database-based text matching for medication identification on cylindrical containers. The system is designed to address practical challenges inherent to clinical workflows, including variable label orientations, time-constrained decision-making, offline operation requirements (due to hospital network policies), and computational constraints of portable devices.

8.1 Introduction

The transition from algorithmic evaluation to system implementation requires careful consideration of operational requirements and constraints. The experimental results in Chapter 7 demonstrated that text stitching (Approach 2) achieves 98.02% Top-1 accuracy with processing times of 143 ms on high-performance hardware and 1.12 seconds on resource-constrained devices. However, translating these results into a functional system requires addressing implementation concerns beyond algorithmic performance, including integration with clinical workflows, offline operation, and reliability under real-world conditions.

Operating theatre environments present challenges that differ from controlled laboratory evaluation. The system must operate offline to comply with hospital network security policies, integrate with existing medication handling procedures, and provide outputs that clinicians can interpret and use to inform decisions. While this proof-of-concept (PoC) aims to demonstrate technical feasibility, it is not intended to address all practical challenges of a fully deployable medication recognition system. Developing a complete solution that satisfies all clinical, regulatory, and operational requirements is beyond the scope of this study. Instead, this work focuses on evaluating the feasibility of integrating the proposed methods into a functional prototype under simplified assumptions.

To reduce dependency on network connectivity and mitigate associated risks, the system is designed with an emphasis on local processing wherever possible. Core components, including image acquisition, OCR, and text stitching, are executed on-device, while external dependencies are minimized in the prototype design. This reflects a deployment scenario where reliable operation is required even in constrained or offline environments.

The prototype development process prioritizes practical implementation of the

algorithms from earlier chapters while maintaining awareness of deployment constraints. Rather than pursuing an exhaustive system design, the focus is on demonstrating how the text stitching approach can be integrated with supporting components (hardware, user interface, database access) in a functional prototype. The design choices presented here reflect considerations of hardware efficiency, interface clarity, and operational simplicity, acknowledging that clinical deployment would require additional regulatory, safety, and validation work beyond the scope of this research.

The system is designed with anaesthesia and operating theatre workflows in mind, where medication identification must occur on a timescale of seconds and errors have serious consequences. Accordingly, the prototype provides ranked candidate medications with associated similarity scores rather than a single deterministic output, allowing clinicians to interpret results in context and make informed decisions. The components of the system is presented in Table 8.1.

Table 8.1: System Component Roles

Component	**Role**
Dual-camera acquisition	Real-time prototype
OCR (fine-tuned PaddleOCR)	Real-time prototype
Text stitching	Real-time prototype
Database matching (Levenshtein)	Real-time prototype
LLM-assisted structuring	Offline / construction-time only
RAG enrichment	Offline / construction-time only
Web-based database interface	Offline / construction-time development tool

8.2 Design Rationale: Approach Selection

The prototype development began by selecting between the two identification approaches evaluated experimentally in Chapter 7. While Approach 1 (image stitching

followed by OCR) was conceptually appealing due to its potential to provide visually continuous label representations, experimental results revealed practical limitations making it less suitable for a real-time clinical prototype.

Latency Constraints: Approach 1 exhibits significant latency, with a total pipeline time of 941.15 ms, dominated by the image stitching stage (714.21 ms). Although this is still under one second, the processing delay may disrupt time-sensitive clinical workflows. Approach 2 achieves a 6.58× speedup, with total latency of 143.11 ms and text stitching contributing only 3.00 ms, making it more suitable for rapid medication identification.

System Reliability: Approach 1 produces no output in 3.3% of test cases, representing a complete system failure that provides clinicians with no information (neither prediction nor confidence indication). Approach 2 produces outputs for all samples, with failures manifesting as low confidence predictions rather than system breakdown. This graceful degradation is more clinically interpretable.

Complexity and Maintainability: Approach 1 relies on a multi-stage geometric pipeline (segmentation, ellipse fitting, geometric transformation, feature matching), introducing multiple points of potential failure and requiring careful tuning and maintenance. Approach 2 follows a simpler pipeline of independent OCR, text merging, and database matching, reducing implementation complexity and improving robustness.

Computational Efficiency: Approach 1 cannot execute on resource-constrained devices due to computational requirements. Approach 2 completes in 1.12 seconds on such devices, enabling deployment on portable hardware without specialized infrastructure.

Based on these considerations, Approach 2 (text stitching) was selected as the foundation for the prototype, as it better aligns with the practical demands of clinical implementation.

8.3 Technology Stack and Component Selection

The selection of software tools and frameworks influences prototype performance, maintainability, and implementation efficiency. The following choices were made with consideration for real-time operation, ease of integration, and proven applicability to computer vision tasks.

8.3.1 Optical Character Recognition: PaddleOCR

PaddleOCR was selected as the OCR engine because it offers a practical balance of performance and ease of integration for medication label recognition. The framework is designed for real-time text detection and recognition and provides Python bindings that integrate with the rest of the prototype. The framework is actively maintained and widely used, supporting long-term viability.

8.3.2 Programming Language: Python

Python was chosen as the primary implementation language due to its extensive ecosystem of libraries for computer vision (OpenCV), user interface development (PyQt6), database interaction (SQLite), and numerical computation (NumPy). Python's syntax enables rapid development and iteration, which is practical for research prototypes. The language also provides cross-platform compatibility important for diverse hospital IT environments.

8.3.3 User Interface: PyQt6

PyQt6 was selected for user interface development because it provides cross-platform compatibility and supports real-time display of dual-camera feeds and identification results. The framework provides responsive event handling and efficient rendering

suitable for displaying video at acceptable frame rates even on resource-constrained hardware.

The framework also enables development of configuration dialogs allowing users to adjust system parameters (OCR model selection, confidence thresholds, frame-skipping rates) without code modification. This is practical for a prototype that may require adjustment in different evaluation settings.

8.3.4 Image Processing: OpenCV

OpenCV was chosen for image processing because it is the standard library for real-time computer vision applications, providing highly optimized implementations of image acquisition and preprocessing routines. The library's multi-camera support is relevant to the prototype's dual-camera architecture. OpenCV is extensively documented and widely deployed, providing confidence in its reliability.

8.3.5 Local Data Storage: SQLite

SQLite enables fully offline operation on resource-constrained devices by providing an embedded, serverless database stored as a local file. This avoids the need for network connectivity or external infrastructure, which is important for hospital environments with strict IT constraints.

In this system, the SQLite database is constructed by storing structured medication text extracted from container labels. A separate MongoDB-based framework (described in Chapter 6) supports scalable database construction using LLM-based processing and is intended to enable semi-automated curation and expansion of medication records. However, the runtime system uses SQLite for practical deployment. Curated records from the MongoDB framework can be exported and transferred into the SQLite database

for field deployment, allowing the system to benefit from advanced database construction methods while maintaining a lightweight, self-contained runtime environment.

8.3.6 Dual-Camera Hardware Setup



Figure 8.1: Creality Nebula Camera selected for the dual-camera setup. The camera features a short focal length enabling close-up capture of medication labels and manual focus adjustment for precise label imaging.

The prototype employs a dual-camera configuration positioning two cameras to capture medication labels from different angles. The Creality Nebula Camera (Fig. 8.1) was selected based on specifications suitable for medication label imaging, including short focal length for detailed close-up shots and manual focus adjustment critical for handling vials at varying distances.

The physical prototype was designed in Tinkercad to provide a stable mounting platform. The design includes a bottle holder preventing container movement during image capture, two adjustable camera stands for multi-angle positioning, and fine-grained distance adjustment, as shown in Figures 8.2 and 8.3. This mechanical design

enables straightforward positioning of medications for image capture without extensive setup.

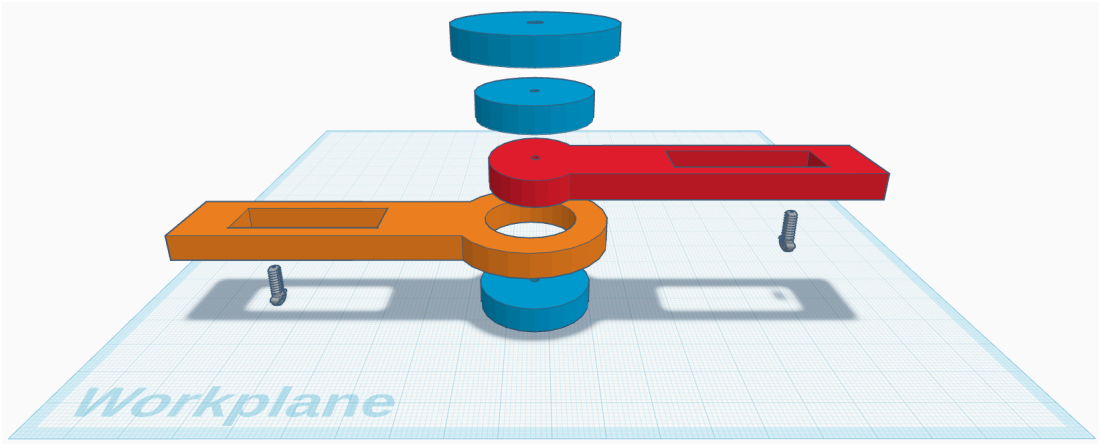


Figure 8.2: 3D design of the prototype hardware, created in Tinkercad, showing the arrangement of the bottle holder, dual adjustable camera stands, and camera mounts for multi-angle label capture.

The dual-angle imaging strategy addresses a fundamental challenge: cylindrical vial geometry creates variable perspective distortion depending on viewing angle. Capturing images from two distinct angles (typically separated by 30-45 degrees) provides different perspectives on label text, with different portions clearly visible in each view. The text stitching algorithm leverages these complementary views to reconstruct a more complete representation of label text than would be possible from a single viewpoint.

8.3.7 Software Architecture and Processing Pipeline

The software architecture follows a modular design separating concerns and enabling independent testing of components. The system consists of four main functional modules.

Image Acquisition Module: Manages simultaneous capture from both cameras using OpenCV, handling frame synchronization, image pre-processing (resizing, color normalization), and buffering. The module abstracts hardware differences, allowing

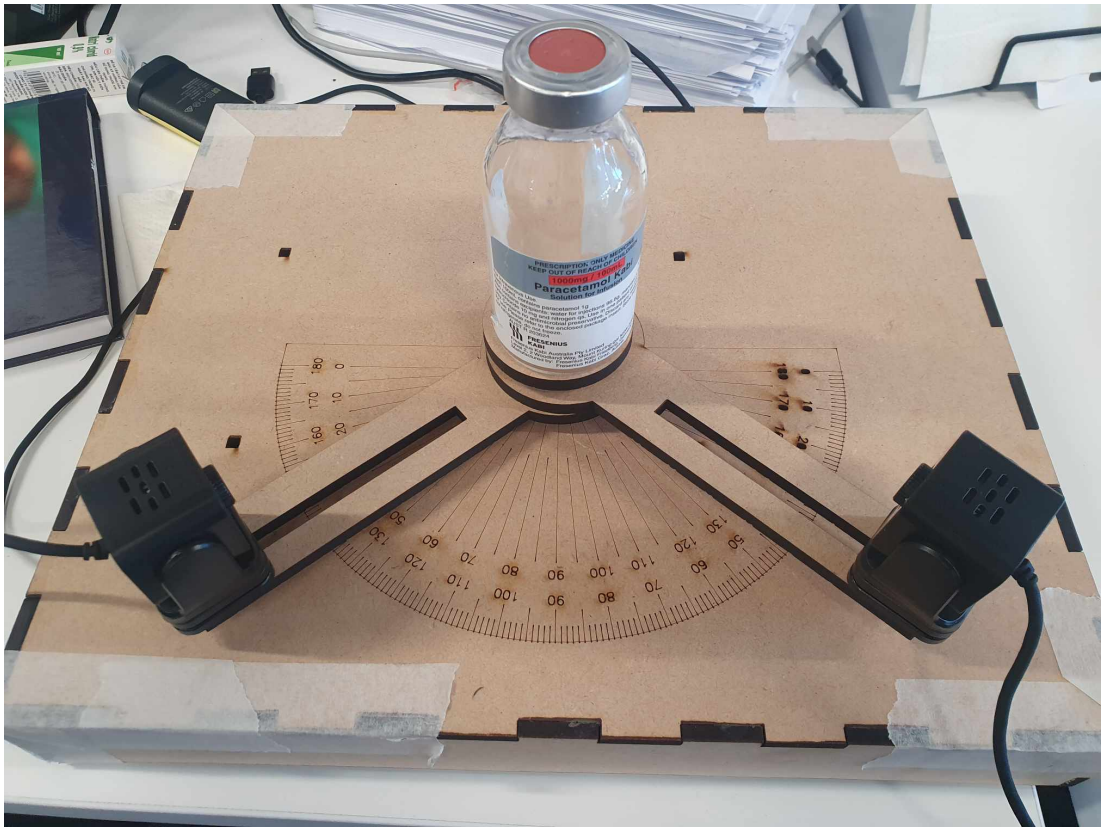


Figure 8.3: Physical prototype constructed using a laser cutter. The hardware provides stable positioning for medication containers and flexible camera positioning to capture labels from complementary viewing angles.

operation with different camera models.

OCR Processing Module: Applies PaddleOCR independently to each dual-camera image. Each image is processed separately, preserving local text context and avoiding error accumulation that can occur with pixel-level image stitching. The OCR engine outputs text sequences with bounding box information indicating character locations.

Text Stitching and Database Matching Module: Implements the text stitching algorithm (detailed in Chapter 5) to merge OCR outputs from both cameras into a single medication label representation. The algorithm identifies overlapping text regions between camera views, resolves duplicates, and produces an ordered text sequence.

The merged text is queried against the medication database using Levenshtein distance (edit distance). This metric is robust to minor OCR errors, with single character

misrecognitions contributing a distance of 1. Distances are normalized by string length to produce similarity scores in the range [0, 1], avoiding bias toward longer medication names. Medications are ranked by similarity score, and the top-5 candidates are presented with associated scores.

Database matching is performed using Levenshtein edit distance between the stitched OCR text and each candidate medication label in the reference database, with candidates ranked by descending similarity score. The main computational bottlenecks in the pipeline are OCR inference, which represents the most expensive step, particularly on resource-constrained hardware. Additionally, for Approach 1, bottlenecks occur during the sequential geometric processing stages, including segmentation, ellipse fitting, and unwrapping, as each of these phases adds latency and introduces a potential point of failure.

The confidence threshold used for match acceptance is configurable via the Config Window interface. Based on the mean similarity score of 0.9804 reported for correct matches in Section 7.3.2, a default threshold of 0.90-0.95 is recommended as a starting point, set below the mean to accommodate normal variation while still rejecting low-confidence matches. A full calibration, for example, selecting the threshold that best separates the similarity-score distributions of correct and incorrect matches was not performed in this work and is proposed as a direct extension in Section 9.4.5.

User Interface Module: Renders medication identification results in a format suitable for clinician review. The interface displays live video feeds from both cameras, the top-ranked medication candidate with its confidence score, and the merged text extracted from the label. Optional features include manual verification or override of identification, confidence threshold adjustment, and OCR model selection.

8.3.7.1 Class and Sequence Diagrams

The modular architecture is formalized in object-oriented design through a set of interconnected classes, each responsible for a specific functional area.

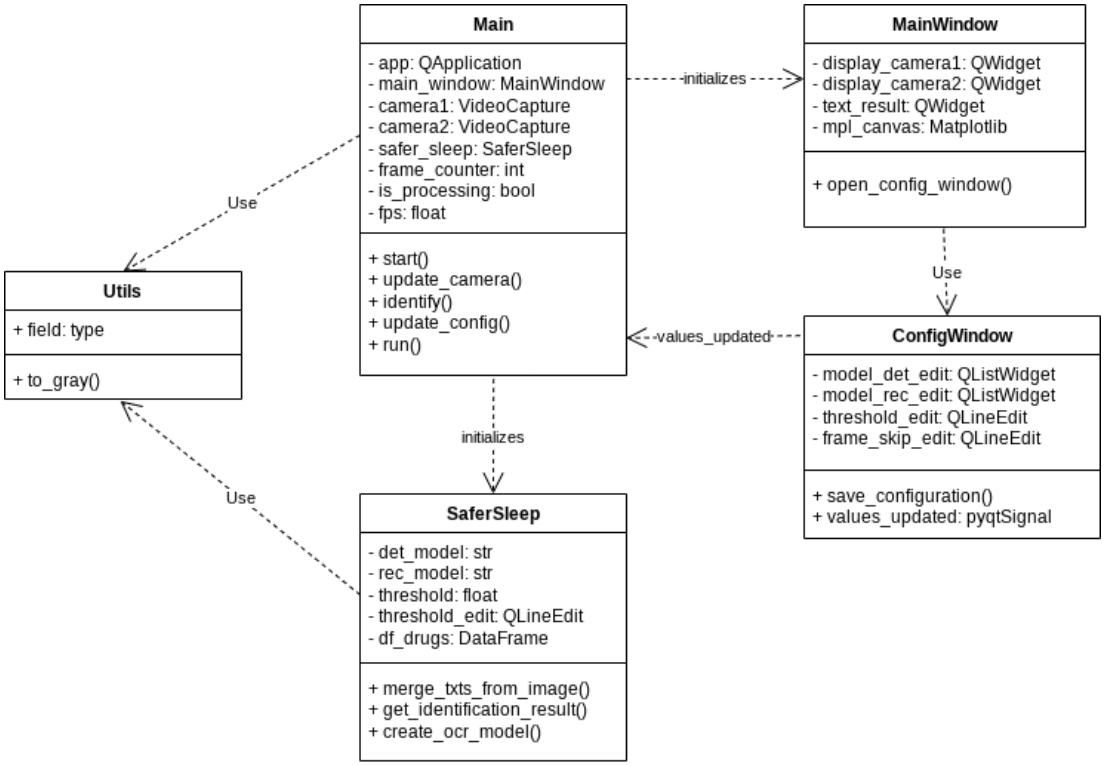


Figure 8.4: Class diagram showing the modular architecture of the prototype. Arrows indicate dependencies between classes.

Figure 8.4 shows the system architecture, organized into five core classes for real-time, reliable medication identification.

Main Class: Acts as the central controller, managing the application lifecycle, dual-camera input, and coordination between components. It continuously acquires frames, sends them to `SaferSleep` for processing, and forwards results to MainWindow.

SaferSleep Class: Implements the core identification pipeline. It performs OCR on multi-view images, merges text, and matches results against a medication database. It manages OCR models, confidence thresholds, and database access.

MainWindow Class: Handles the user interface, displaying live camera feeds and

identification results. It updates outputs in real time and ensures smooth rendering.

ConfigWindow Class: Provides a UI for adjusting system parameters such as OCR models, confidence thresholds, and frame-skipping rates. It supports saving/loading settings and validating inputs.

Utils Class: Contains reusable helper functions for image preprocessing and text normalization, supporting OCR and display operations.

The system follows a centralized dependency structure in which the Main class orchestrates all components. `SaferSleep` (identification logic), `MainWindow` (user interface), and `ConfigWindow` (configuration) operate independently and communicate exclusively through `Main`, eliminating circular dependencies, while `Utils` provides shared helper functions across modules. This modular design enhances testability, maintainability, and extensibility, enabling independent updates to the OCR pipeline, UI, and configuration without impacting the overall system.

Figure 8.5 illustrates the runtime workflow of the system during a medication identification task. The process consists of five stages: (1) initialization, where `Main` launches the application and instantiates `SaferSleep` and `MainWindow`; (2) continuous frame capture from dual cameras; (3) periodic identification, where `Main` sends frames to `SaferSleep` for OCR, text stitching, and database matching to produce ranked candidates with confidence scores; (4) result handling, where outputs are returned to `Main`; and (5) real-time display, where `MainWindow` updates the UI with camera feeds and identification results.

8.4 User Interface Design

The user interface displays dual live camera feeds side-by-side, allowing operators to see what the system is analyzing and diagnose issues such as poor angles or occlusions. Identification results are displayed prominently with the top candidate and confidence

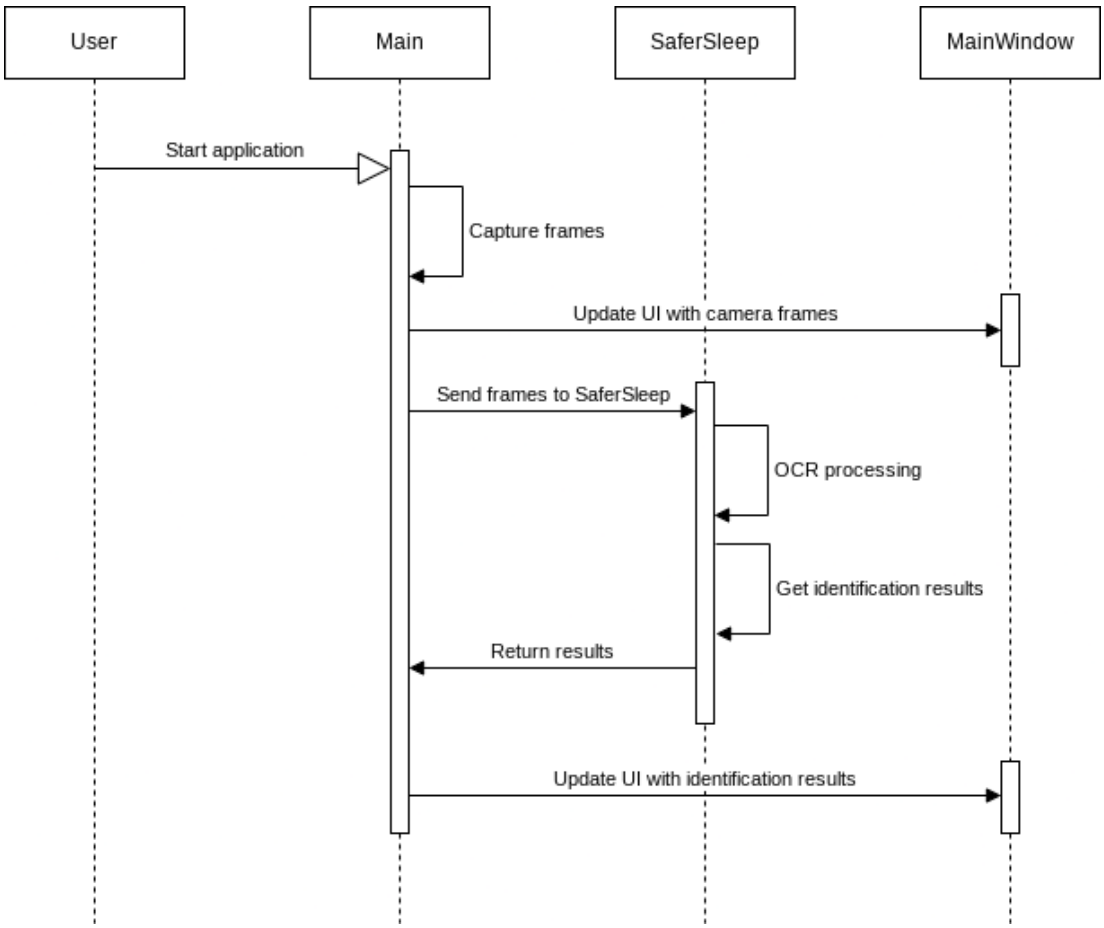


Figure 8.5: Sequence diagram illustrating the interaction between system components during a typical medication identification workflow. The diagram shows the progression from user interaction through image capture, OCR processing, database matching, and result presentation.

score, while additional candidates provide context. The merged label text is also shown, enabling operators to verify extracted information and identify potential OCR errors, as illustrated in Fig. 8.6.

A configuration window (Fig. 8.7) enables adjustment of system parameters without code modification. These parameters include OCR model selection (enabling rapid switching between model versions), confidence threshold controlling result acceptance, and frame-skipping rate controlling identification frequency (lower rates for rapid iteration, higher rates for power efficiency).

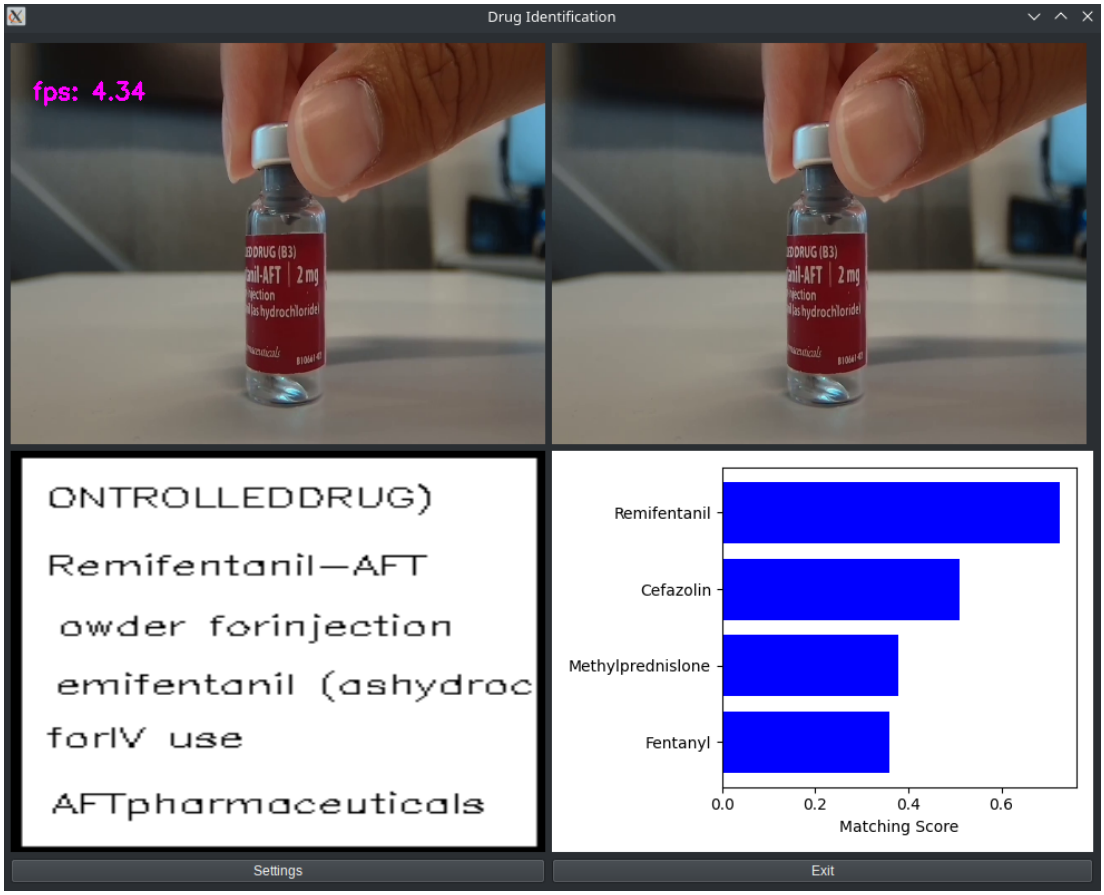


Figure 8.6: Main application window showing dual live camera feeds (left and right), the top-ranked medication identification result with confidence score (center panel), and merged label text (bottom panel). The interface is designed for rapid comprehension in high-pressure clinical environments.

8.5 Key Functional Features and Implementation Details

8.5.1 Dual-Angle Coverage and Text Merging

The core capability of this prototype is its use of dual-camera coverage to address challenges posed by cylindrical label geometry. Medication vials present an imaging challenge: the label wraps around a cylinder, so complete label capture in a single frontal view is difficult without distortion. Dual-camera imaging captures complementary views

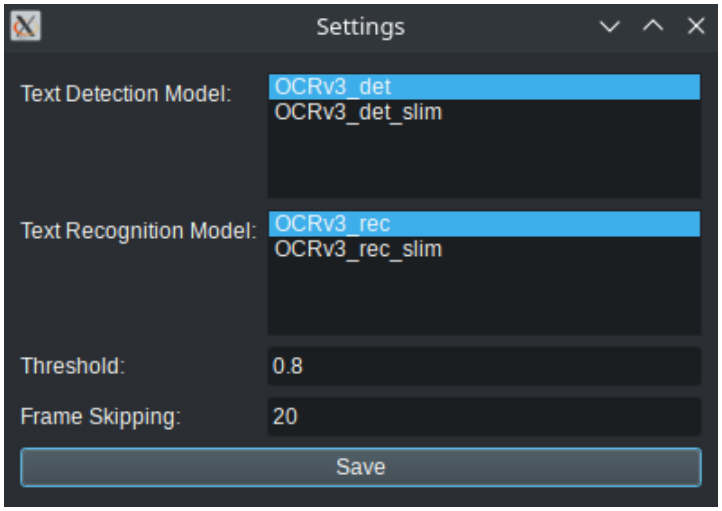


Figure 8.7: Configuration interface enabling users to adjust system parameters including OCR model selection, confidence threshold for automatic acceptance of results, and frame-skipping rate controlling the frequency of identification attempts.

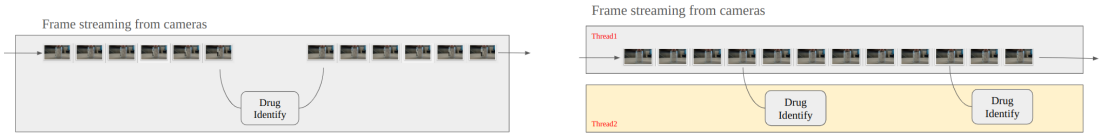
with different portions of text being sharp and clearly visible in each.

The text stitching algorithm processes OCR outputs from both cameras to produce a merged text sequence representing the complete medication label. The algorithm identifies overlapping text regions, resolves duplicates, and produces an ordered sequence. This text-based merging approach is more robust than image-level stitching because it operates on recognized text rather than pixel-level features, avoiding distortions and artifacts introduced by geometric image transformations, as demonstrated in Chapter 7.

8.5.2 Concurrent Processing via Multi-Threading

Operating theatre workflows require timely feedback from medication identification systems. The prototype achieves responsive operation through multi-threaded design, where image acquisition and identification processing execute concurrently, as illustrated in Fig. 8.8.

One thread continuously acquires frames from both cameras, ensuring fresh image data is available. A second thread performs OCR and database matching on buffered



(a) Single-Threaded Processing (b) Multi-Threaded Processing

Figure 8.8: Comparison of single-threaded and multi-threaded processing architectures. In single-threaded execution (left), image acquisition is blocked during OCR, causing frame drops and lag. In multi-threaded execution (right), acquisition and processing run concurrently, improving responsiveness and stability.

frames without blocking acquisition. This decoupling avoids blocking during computationally expensive OCR and maintains acceptable frame display rates.

8.5.3 Configurable Processing Rate Through Frame Skipping

Continuous identification attempts incur high computational cost and can deplete battery capacity on portable devices or cause thermal throttling. The prototype implements configurable frame skipping, where identification is attempted on every n-th frame rather than every frame (Fig. 8.7). Users can adjust the skip rate based on their operational requirements and device capability.

8.5.4 Medication Matching via String Similarity

The system matches merged label text to the database using Levenshtein distance, which measures the minimum number of character edits required to transform one string into another. This metric is robust to OCR errors, with small misrecognitions producing minor differences. Distances are normalized by string length to produce similarity scores in the range [0, 1], avoiding bias toward longer medication names. All database entries are ranked by similarity, with the top-5 candidates displayed.

8.5.5 Model Flexibility

The prototype supports selectable OCR models, enabling deployment adaptation based on available hardware. Lightweight models can be used on resource-constrained devices, while larger models can be deployed on high-performance hardware. New or fine-tuned models can be integrated and selected through the configuration interface without code changes.

8.6 Summary and Future Directions

This chapter has described the design and implementation of a proof-of-concept medication identification system demonstrating technical feasibility of the text stitching approach from Chapter 7 in a functional prototype. The system integrates dual camera image acquisition, PaddleOCR-based text extraction, text stitching for multi-view label reconstruction, and SQLite-based database matching into a coherent processing pipeline.

The technology stack was selected to balance performance, development efficiency, and practical deployability. The modular architecture enables independent development and refinement of components. The user interface provides straightforward display of system outputs for operator review.

Within this framework, the prototype addresses several practical implementation concerns such as offline operation, computational efficiency on resource-constrained devices, and concurrent processing to maintain responsive operation. These design choices allow the system to implement the text stitching approach under realistic operational constraints, though the prototype is a research implementation rather than a production-ready clinical system.

Further work includes evaluating the system on a broader range of medication labels,

testing performance under realistic lighting and positioning conditions, and conducting usability studies with clinical users. Formal safety and regulatory assessment would also be required before any clinical deployment. In addition, improvements such as better OCR models trained on domain-specific data, expansion of the medication database, and optimization for specific deployment environments would further enhance system performance.

Overall, the prototype demonstrates that the proposed text stitching approach can be implemented as a functional system, providing a foundation for further development and evaluation toward potential future clinical application.

Chapter 9

Conclusions and Future Work

This concluding chapter synthesizes the key findings and contributions of this thesis, which investigated the development of an automated medication identification system aimed at enhancing patient safety in anaesthesia and operating theatre environments. The research built upon a comprehensive review of existing identification methods, including barcode/QR codes, RFID/NFC technologies, and computer vision (CV) techniques to propose and evaluate two novel approaches for cylindrical medication label reconstruction. Through extensive experimentation, prototype development, and critical analysis, the thesis demonstrates that CV combined with optical character recognition (OCR) can provide accurate medication identification from dual-camera images of medication labels. This chapter summarizes the research contributions, reflects on limitations, and outlines directions for future work necessary before clinical deployment.

9.1 Summary of Research Contributions

This thesis presented a comprehensive, end-to-end framework for automated medication identification in clinical environments, addressing both the technical challenges of accurate text recognition from curved labels and the practical requirements of operating theatre integration. The research spanned five major areas, each building on previous findings to create a coherent system for medication identification.

9.1.1 Image-Based Label Reconstruction

The first major contribution was the development of an image-based framework for reconstructing medication labels from cylindrical containers through geometric analysis and dual-view stitching. This work, presented in Chapter 4, addressed the fundamental challenge that cylindrical vial geometry creates perspective distortion, curved surfaces, and incomplete views of label text.

The framework integrated multiple complementary techniques. Label segmentation isolated medication labels from background clutter, providing clean input for subsequent geometric processing. Two alternative methods for ellipse fitting were proposed: a traditional contour-refined approach providing precise geometric extraction based on image contours, and a deep learning-based method offering automated boundary estimation robust to noise and partial occlusions. Following ellipse estimation, cylindrical label unwrapping and geometric transformation mapped the curved label surface onto a planar domain suitable for standard image processing. Image stitching then fused multiple unwrapped views into a unified, flattened representation.

Experimental evaluation in Chapter 7 highlights several limitations of the image-based pipeline under the tested conditions. The multi-stage geometric process is sensitive to errors in intermediate steps such as segmentation, ellipse fitting, and stitching, which in some cases lead to the inability to produce a usable output (3.3% of test cases).

The approach also introduces higher computational cost (941.15 ms total latency), which may limit its suitability for real-time use on resource-constrained devices.

9.1.2 Text-Based Label Reconstruction

Building on the limitations of image-based approaches, the second major contribution was the development of a text-based framework for label reconstruction, presented in Chapter 5. This approach shifted focus from pixel-level geometric transformation to character-level text fusion, operating directly in the semantic domain where medication identification ultimately occurs. The thesis contributes a dual-view OCR text-stitching framework that avoids full geometric label reconstruction and demonstrates improved robustness and efficiency for cylindrical medication label identification under the tested conditions.

The framework employed fine-tuned PaddleOCR to extract text directly from images without prior geometric unwrapping. Fine-tuning focused on improving text localisation for the characteristics of medication labels: curved text, variable font sizes, and degraded print quality, with text recognition remaining general-purpose. Rather than attempting perfect geometric alignment, the system leveraged the insight that imperfect OCR outputs from complementary viewing angles contain sufficient redundancy for text stitching to recover label text missed in a single view, rather than to correct recognition errors common to both views.

Text stitching employed longest common substring analysis and sequence alignment to merge OCR outputs from dual-camera views. The algorithm identified overlapping text regions, resolved duplicates, and produced an ordered sequence representing complete medication label information. This text-level fusion proved more robust than image-level stitching, handling imperfect OCR outputs and variable imaging conditions.

Experimental results demonstrated the practical advantages of this approach: text

stitching achieved 98.02% Top-1 medication identification accuracy with zero complete system failures (100% prediction rate), compared to image stitching's 83.40% accuracy and 3.3% complete failure rate. Processing time was also substantially reduced, with text stitching completing in 143 ms on high-performance hardware and 1.12 seconds on resource-constrained edge devices.

9.1.3 Medication Database Construction Using AI-Based Processing

The third component of this work was the construction of a medication database using Large Language Model (LLM)-assisted processing, described in Chapter 6. This reflects the observation that medication identification depends not only on text extraction, but also on the quality and consistency of the reference data used for matching. This component supports system implementation and demonstrates a feasible workflow for database construction, but it is not claimed as a fully validated database automation method.

The database construction process incorporated LLMs to support data structuring and normalization, consolidating information from multiple heterogeneous sources into a more standardized schema. Retrieval-Augmented Generation (RAG) was used to refine and enrich the extracted information by referencing external knowledge sources.

A web-based interface was developed to support database interaction, including querying and manual inspection of entries. While this interface is not part of the prototype (which operates offline), it illustrates a possible workflow for reviewing and maintaining the database. The effectiveness of this process for ensuring data accuracy or supporting large-scale updates was not evaluated in this work.

9.1.4 Experimental Evaluation Under Realistic Conditions

The fourth contribution was an experimental comparison of image-based and text-based approaches under conditions designed to reflect operating theatre challenges, presented in Chapter 7. The evaluation was conducted using the same dataset and hardware setup for both methods to ensure a fair comparison.

The dataset consists of 1,060 real medication label images covering 53 medications with balanced representation. Two hardware platforms (a high-performance workstation and a resource-constrained device) were used to assess performance under different computational conditions. Evaluation focused on standard metrics such as accuracy and processing time, as well as practical considerations such as robustness and runtime behaviour.

Overall, the results indicate that text-based stitching performs more consistently in terms of accuracy and efficiency across the tested conditions, particularly on the resource-constrained device. Image-based stitching provides more interpretable visual outputs but is more sensitive to geometric reconstruction failures. These findings informed the design choices used in the subsequent prototype development.

9.1.5 Prototype System Design and Implementation

The fifth major contribution was the design and implementation of a working prototype system integrating all previous components into a cohesive application for medication identification, presented in Chapter 8. The prototype demonstrated that the text stitching approach could be implemented as a functional system.

The system integrated dual-camera hardware for multi-angle label coverage, fine-tuned PaddleOCR for text extraction, text stitching for multi-view fusion, and SQLite-based medication database matching. Software architecture employed modular design enabling independent component refinement. Performance optimizations including multi-threading and configurable frame skipping enabled operation on resource-constrained hardware. The user interface displayed dual live camera feeds, top candidate medications with confidence scores, and merged label text enabling operator verification.

Critical design choices reflected lessons learned throughout the research. Approach 2 (text stitching) was selected over Approach 1 (image stitching) based on experimental evidence. Offline operation was chosen to accommodate hospital security policies while enabling decentralized deployment. Confidence-aware output (ranked candidates with scores) was selected to support operator interpretation and informed decision-making.

As documented in Chapter 8, the prototype is a research implementation and proof-of-concept. It demonstrates technical feasibility but is not presented as a production-ready clinical system. Clinical deployment would require additional validation, regulatory assessment, and integration work.

9.2 Key Research Findings and Their Implications

Beyond the main contributions, this work highlights several practical findings relevant to medication identification systems in healthcare settings.

A central observation is the trade-off between image-based and text-based approaches. Image-based methods retain richer visual and spatial information, which can be useful for inspection and interpretability, but they introduce higher computational cost and are more sensitive to geometric reconstruction errors. In contrast, text-based methods are more efficient and robust to geometric distortion since they operate directly

on extracted OCR text, but they lose visual context that may assist human verification. In time-critical environments such as operating theatres, the text-based approach provides a more practical balance between accuracy, speed, and robustness under the tested conditions.

Another important finding is the benefit of dual-view imaging. Compared to single-view capture, multiple viewpoints help mitigate common issues such as occlusion, perspective distortion, and incomplete label visibility. This leads to more complete OCR inputs and improves the reliability of downstream extraction, particularly for cylindrical or curved medication containers where a single viewpoint is often insufficient.

Finally, deployment constraints play a significant role in system design. Hospital environments often have strict limitations on network connectivity and external data access, which makes reliance on cloud-based processing impractical in many cases. This motivates an offline-first design, where all core processing is performed locally. Such constraints do not only restrict system design but also help define practical architectures that are compatible with real clinical infrastructure.

9.3 Limitations

The evaluation dataset includes 53 medications from Auckland City Hospital (Healthpoint, n.d.) and SaferSleep (SAFERsleep Ltd., n.d.). While these cover commonly used anaesthetic and emergency drugs, the dataset is still limited in terms of manufacturer diversity, packaging styles, and label design variability. In particular, medications with alternative layouts or non-standard label formats are underrepresented. As a result, performance on broader hospital inventories or different clinical settings may vary and would require further validation.

A second limitation relates to OCR dependency. The system relies on the quality of OCR outputs as the primary input for downstream processing. Therefore, label

conditions such as unusual fonts, low contrast text, non-Latin characters, fading, or physical damage can directly affect recognition quality. In such cases, errors at the OCR stage propagate through the pipeline and reduce the reliability of extracted structured information.

In addition, the dual-camera acquisition setup improves overall coverage but does not guarantee complete visibility of all label content. Highly curved regions, occluded areas, or text located on the reverse side of cylindrical containers may still be missed. This can lead to incomplete inputs even before text consolidation or structured extraction is applied.

Another limitation concerns deployment on edge devices. While the system achieves acceptable performance under controlled conditions, real-world latency may vary depending on hardware specifications, available memory, concurrent processes, and thermal constraints. The reported 1.12 second processing time reflects a stable experimental setup and should be interpreted as an indicative rather than guaranteed runtime.

Finally, an important limitation is the absence of prospective clinical evaluation. The system has been tested under controlled experimental conditions but not in real operating theatre environments. Factors such as workflow variations, user interaction, integration with existing hospital systems, and potential failure modes are not fully captured in this setting. As a result, conclusions regarding usability and safety are based on experimental observations and remain preliminary.

9.4 Future Research Directions

This thesis identifies several promising directions for advancing automated medication identification and extending the research toward clinical deployment.

9.4.1 Expanding Dataset Diversity

The immediate priority should be expanding the medication dataset to include medications from diverse manufacturers and label design conventions. Collaborating with different hospitals and healthcare systems would reveal how system performance generalizes when applied to medications not represented in the current dataset.

Incorporating newly introduced medications as they come to market would maintain system relevance in dynamic clinical environments where medication inventories continuously evolve.

Expand the evaluation dataset to include a broader range of medications, manufacturers, and label designs.

9.4.2 Advanced OCR Enhancement and Domain Adaptation

While baseline PaddleOCR achieves relatively high word-level accuracy on medication labels, there remains scope for improvement. Future work could explore several directions. Language model-based context restoration may help correct implausible character sequences based on surrounding text. Confidence-weighted OCR could incorporate uncertainty estimates at the character level, allowing the stitching process to prioritise higher-confidence predictions across views. Synthetic data generation may be used to expand training data by creating realistic label variations in font, scale, orientation, and degradation. In addition, self-supervised or semi-supervised approaches could leverage unlabeled label images to improve generalisation.

9.4.3 Multi-Camera and Sensing Approaches

The current dual-camera setup may miss information on highly curved or wrapped labels. Future systems could employ three or more cameras positioned around the cylinder to improve label coverage. Alternatively, robotic systems could rotate the

medication vial under fixed cameras, capturing the complete cylindrical surface. Such approaches would increase hardware complexity but potentially improve coverage.

Integrating additional sensing modalities beyond visual imaging-such as infrared or depth-based sensing could provide complementary information about label geometry and robustness to variable lighting conditions.

Also, evaluate three- or four-camera configurations to improve coverage of highly curved or fully wrapped labels.

9.4.4 Clinical Validation and Real-World Deployment Studies

The most critical next step is prospective clinical validation in actual operating theatre environments. Such studies would:

Assess real-world performance under actual clinical conditions, not laboratory conditions. Identify unexpected failure modes and edge cases not encountered in controlled evaluation. Evaluate clinician acceptance, workflow integration, and usability. Measure actual impact on medication errors and patient safety outcomes. Gather feedback from anaesthetists and perioperative nurses on interface design and workflow integration.

In addition, test the system under more realistic lighting and occlusion conditions representative of operating theatre environments. Conduct clinical and user validation studies with anaesthetists and perioperative staff in real operating-theatre workflows

9.4.5 Advanced Matching and Knowledge-Based Reasoning

Current medication matching is based on Levenshtein distance, which ranks candidates using surface-level text similarity. More advanced approaches could incorporate semantic similarity to account for relationships between medications with similar pharmacological properties. Another possible direction is the use of knowledge graphs to

represent structured relationships such as drug classes, contraindications, and interactions.

These extensions would move beyond exact or near-text matching toward a more semantic understanding of medication identity. This could improve robustness to OCR noise, particularly when errors affect non-critical parts of the label, while still preserving correct identification of clinically relevant medications.

9.5 Final Remarks

This thesis has documented a comprehensive exploration of automated medication identification spanning image-based and text-based label reconstruction methods, AI-based database construction, experimental evaluation, and prototype system development. The research demonstrates that integrating dual-view imaging, fine-tuned OCR, text stitching, and database matching can provide accurate medication identification from optical images. Overall, this thesis contributes to the emerging field of AI-assisted medication safety by demonstrating that local, camera-based, OCR-driven identification can be technically feasible without relying on barcode infrastructure or cloud-based processing.

The research journey involved pursuing promising technical approaches (image stitching), conducting rigorous evaluation revealing their limitations, and pivoting to alternative approaches (text stitching) better suited to practical requirements. This iterative, evidence-based process produced a system better aligned with deployment constraints.

The achieved Top-1 identification accuracy of 98.02% meets the >97% acceptance threshold agreed with the industrial partner, SAFERsleep, at the outset of this research. The prototype system demonstrates that text stitching can be implemented as a functional system. However, feasibility in controlled settings does not guarantee success

in clinical deployment. As documented in Chapter 8, the prototype is a research implementation intended to demonstrate technical feasibility. Clinical deployment would require additional validation, regulatory assessment, and integration work beyond the scope of this research.

The research journey from initial motivation (medication errors represent preventable harm) through extensive research (multiple technical approaches, comprehensive evaluation) to a working prototype demonstrates the potential of systematic research to address important clinical problems. This thesis documents both technical contributions and the research process, including constraints and lessons learned, providing a foundation for future researchers building on this work toward clinical application.

Future work will focus on the directions outlined above, with particular emphasis on clinical validation and real-world assessment before any clinical deployment. The technical feasibility demonstrated in this thesis provides grounds for optimism that automated medication identification can eventually contribute to safer medication administration in clinical practice, but substantial additional work is necessary to translate laboratory findings into clinical reality.

References

Aamir, M., Rahman, Z., Dayo, Z. A., Abro, W. A., Uddin, M. I., Khan, I., ... others (2022). A deep learning approach for brain tumor classification using mri images. Computers and Electrical Engineering, 101, 108105.

Aghavali, R., Roberts, E. G., Kurokawa, Y. K., Mak, E., Chan, M. Y., Wong, A. O., ... Costa, K. D. (2024). Enhanced drug classification using machine learning with multiplexed cardiac contractility assays. Pharmacological Research, 209, 107459.

Al-Ahmadi, R. F., Al-Juffali, L., Al-Shanawani, S. & Ali, S. (2020). Categorizing and understanding medication errors in hospital pharmacy in relation to human factors. Saudi Pharmaceutical Journal, 28(12), 1674-1685. Retrieved from https://www.sciencedirect.com/science/article/pii/S1319016420302498 doi: https://doi.org/10.1016/j.jsps.2020.10.014

Albarrak, A. I., Al Rashidi, E. A., Fatani, R. K., Al Ageel, S. I. & Mohammed, R. (2014). Assessment of legibility and completeness of handwritten and electronic prescriptions. Saudi Pharmaceutical Journal, 22(6), 522–527.

Ameer, A., Dhillon, S., Peters, M. J. & Ghaleb, M. (2015). Systematic literature review of hospital medication administration errors in children. Integrated Pharmacy Research and Practice, 153–165.

An, M. & Jiang, J. (2025). Comprehensive evaluation of flumazenil adverse reactions: Insights from faers data and signal detection algorithms. Medicine, 104(10), e41721.

Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., ... Lee, H. (2019). What is wrong with scene text recognition model comparisons? dataset and model analysis. In Proceedings of the ieee/cvf international conference on computer vision (pp. 4715–4723).

Barakat, S. & Franklin, B. D. (2020). An evaluation of the impact of barcode patient and medication scanning on nursing workflow at a uk teaching hospital. Pharmacy, 8(3), 148.

Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., ... others (1995, Jul). Incidence of adverse drug events and potential adverse drug events. implications for prevention. ade prevention study group. JAMA, 274(1), 29–34.

Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L. (2008). Speeded-up robust features (surf). Computer vision and image understanding, 110(3), 346–359.

Bergroth, L., Hakonen, H. & Raita, T. (2000). A survey of longest common subsequence algorithms. Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000, 39-48. Retrieved from https://api.semanticscholar.org/CorpusID:10375334

Biotium. (2023). Product authenticity verification. Retrieved from https://biotium.com/product-authenticity-verification/ (Image source: Biotium.)

Bouzillé, G., Morival, C., Westerlynck, R., Lemordant, P., Chazard, E., Lecorre, P., ... Cuggia, M. (2019). An automated detection system of drug-drug interactions from electronic patient records using big data analytics. In Medinfo 2019: Health and wellbeing e-networks for all (pp. 45–49). IOS Press.

Bowdle, T., Jelacic, S., Nair, B., Togashi, K., Caine, K., Bussey, L., ... others (2018). Facilitated self-reported anaesthetic medication errors before and after implementation of a safety bundle and barcode-based safety system. British Journal of Anaesthesia, 121(6), 1338–1345.

Bowdle, T. A., Jelacic, S., Nair, B., Zucker, F., Bussey, L. S., Togashi, K., ... Lang, J. (2019). Electronic audit and feedback with positive rewards improve anesthesia provider compliance with a barcode-based drug safety system. Anesthesia & Analgesia, 129(2), 418–425.

Bryan, R., Aronson, J. K., Williams, A. & Jordan, S. (2021). The problem of look-alike, sound-alike name errors: drivers and solutions. British journal of clinical pharmacology, 87(2), 386–394.

Burkoski, V., Yoon, J., Solomon, S., Hall, T. N., Karas, A. B., Jarrett, S. R. & Collins, B. E. (2019). Closed-loop medication system: leveraging technology to elevate safety. Nurs Leadersh, 32, 16–28.

Cai, X., Kang, Y., Ma, H., Hong, N. & Qiu, G. (2022). Research on pre-processing method based on cylindrical image restoration of steel pipe. In Proceedings of the 2022 IEEE International Conference on Signal and Image Processing (CCIS (pp. 173–177). doi: 10.1109/CCISP.2022.9974159

Canny, J. (2009). A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence(6), 679–698.

Carver, K. H., Burgess, L. H., Cooper, M., Elders, T. & Kramer, J. (2018). Use of clinical decision support to identify iv-to-oral conversion opportunities and cost savings. American Journal of Health-System Pharmacy, 75(23_Supplement_4), S82–S86.

Chernov, N. & Lesort, C. (2005). Least squares fitting of circles. Journal of Mathematical Imaging and Vision, 23(3), 239–252.

Chiu, Y.-J. (2024). Automated medication verification system (amvs): System based on edge detection and cnn classification drug on embedded systems. Heliyon, 10(9).

Dang, B., Zhao, W., Li, Y., Ma, D., Yu, Q. & Zhu, E. Y. (2024). Real-time pill identification for the visually impaired using deep learning. In 2024 6th international conference on communications, information system and computer engineering (cisce) (pp. 552–555).

Das, S., Ma, K., Shu, Z., Samaras, D. & Shilkrot, R. (2019). Dewarpnet: Single-image document unwarping with stacked 3d and 2d regression networks. In Proceedings of the ieee/cvf international conference on computer vision (pp. 131–140).

Decia, I., Farías, A., Szerman, D., Grundel, L., Briatore, D., Piñeyrúa, M., ... Simini, F. (2017). Camacua: low cost real time risk alert and location system for healthcare environments. In Vii latin american congress on biomedical engineering claib 2016, bucaramanga, santander, colombia, october 26th-28th, 2016 (pp. 90–93).

DeTone, D., Malisiewicz, T. & Rabinovich, A. (2016). Deep image homography estimation. arXiv preprint arXiv:1606.03798.

DeTone, D., Malisiewicz, T. & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In Proceedings of the ieee conference on computer vision and pattern recognition workshops (pp. 224–236).

Dhayanithi, A., Dalali, S., Madhavi, B., Surla, G., Marapelli, B. & Arigela, A. K. (2024). Revolutionizing healthcare: A smart pill detection system empowered by deep learning for intelligent drug recognition. In 2024 international conference on trends in quantum computing and emerging business technologies (pp. 1–6).

Dong, W., Roy, P., Peng, C. & Isler, V. (2021). Ellipse r-cnn: Learning to infer elliptical object from clustering and occlusion. IEEE Transactions on Image Processing, 30, 2193–2206.

Du, Y., Li, C., Guo, R., Cui, C., Liu, W., Zhou, J., ... Ma, Y. (n.d.). PP-OCRv2: Bag of Tricks for Ultra Lightweight OCR System.

Dunn, L. & Anderson, J. (2019). Barcode medication administration implementation in the operating room. American Journal of Health-System Pharmacy, 76(10), 636–637.

Dutt, M. I. & Saadeh, W. (2023). Monitoring level of hypnosis using stationary wavelet transform and singular value decomposition entropy with feedforward neural network. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 1963–1973.

Elliott, R. A., Camacho, E., Jankovic, D., Sculpher, M. J. & Faria, R. (2021, February). Economic analysis of the prevalence and clinical and economic burden of medication error in England. BMJ Quality & Safety, 30(2), 96–105. Retrieved 2023-12-17, from https://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2019-010206 doi: 10.1136/bmjqs-2019-010206

Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303–338.

Feng, H., Zhou, W., Deng, J., Wang, Y. & Li, H. (2022). Geometric representation learning for document image rectification. In European conference on computer vision (pp. 475–492).

Finkenzeller, K. (2010). Rfid handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication. John wiley & sons.

Fischler, M. A. & Bolles, R. C. (1981a). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.

Communications of the ACM, 24(6), 381–395.

Fischler, M. A. & Bolles, R. C. (1981b, jun). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6), 381–395. Retrieved from https://doi.org/10.1145/358669.358692 doi: 10.1145/358669.358692

Fox, B. I., Pedersen, C. A. & Gumpper, K. F. (2015). Ashp national survey on informatics: assessment of the adoption and use of pharmacy informatics in us hospitals—2013. American Journal of Health-System Pharmacy, 72(8), 636–655.

Fram, B. R., Rivlin, M. & Beredjiklian, P. K. (2020). On emerging technology: what to know when your patient has a microchip in his hand. The Journal of hand surgery, 45(7), 645–649.

Gala, D. V., Gandhi, V. B., Gandhi, V. A. & Sawant, V. (2021). Drug classification using machine learning and interpretability. In 2021 smart technologies, communication and robotics (stcr) (pp. 1–8).

Gariel, C., Cogniat, B., Desgranges, F.-P., Chassard, D. & Bouvet, L. (2018). Incidence, characteristics, and predictive factors for medication errors in paediatric anaesthesia: a prospective incident monitoring study. British Journal of Anaesthesia, 120(3), 563-570. Retrieved from https://www.sciencedirect.com/science/article/pii/S0007091217542014 doi: https://doi.org/10.1016/j.bja.2017.12.014

Geoffrey, B., Sanker, A., Madaj, R., Tresanco, M. S. V., Upadhyay, M. & Gracia, J. (2022). A program to automate the discovery of drugs for west nile and dengue virus—programmatic screening of over a billion compounds on pubchem, generation of drug leads and automated in silico modelling. Journal of Biomolecular Structure and Dynamics, 40(10), 4293–4300.

Google. (n.d.). Google cloud vision api. Retrieved from https://cloud.google.com/vision (Computer software)

Grailey, K., Brazier, A., Franklin, B. D., McCrudden, C., Crespo, R. F., Brown, H., ... others (2024). Raising the barcode: improving medication safety behaviours through a behavioural science-informed feedback intervention. a quality improvement project and difference-in-difference analysis. BMJ Quality & Safety, 33(10), 682–690.

Gromova, K. & Elangovan, V. (2022). Automatic extraction of medication information from cylindrically distorted pill bottle labels. Machine Learning and Knowledge Extraction, 4(4), 852-864.

Gupta, A., Vedaldi, A. & Zisserman, A. (2016). Synthetic data for text localisation in natural images. In Proceedings of the ieee conference on computer vision and pattern recognition (pp. 2315–2324).

Hachesu, P. R., Zyaei, L. & Hassankhani, H. (2016). Recommendations for using barcode in hospital process. Acta Informatica Medica, 24, 206 - 210.

Haddara, M. & Staaby, A. (2018). RFID Applications and Adoptions in Healthcare: A Review on Patient Safety. Procedia Computer Science, 138, 80–88. Retrieved 2023-03-23, from https://linkinghub.elsevier.com/retrieve/

pii/S1877050918316430 doi: 10.1016/j.procs.2018.10.012

Halır, R. & Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In Proc. 6th international conference in central europe on computer graphics and visualization. wscg (Vol. 98, pp. 125–132).

Hartley, R. & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge university press.

He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017). Mask r-cnn. In Proceedings of the ieee international conference on computer vision (pp. 2961–2969).

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep residual learning for image recognition. Retrieved from https://arxiv.org/abs/1512.03385

Healthpoint. (n.d.). Auckland city hospital. Retrieved from https://www.healthpoint.co.nz/auckland-city-hospital/

Heo, J., Kang, Y., Lee, S., Jeong, D.-H. & Kim, K.-M. (2023). An accurate deep learning–based system for automatic pill identification: Model development and validation. Journal of medical Internet research, 25, e41043.

Ho, J. & Burger, D. (2020). Improving medication safety practice at a community hospital: A focus on bar code medication administration scanning and pain reassessment. BMJ Open Quality, 9(3), e000987.

Huang, Y.-H. & Gramopadhye, A. K. (2016). Recommendations for health information technology implementation in rural hospitals. International Journal of Health Care Quality Assurance, 29(4), 454–474. doi: 10.1108/IJHCQA-04-2015-0051

Hüttner, L., Mayr, M., Gorges, T., Wu, F., Seuret, M., Maier, A. & Christlein, V. (2025). Low-rank adaptation vs. fine-tuning for handwritten text recognition. In Proceedings of the winter conference on applications of computer vision (pp. 1323–1332).

Illingworth, J. & Kittler, J. (1988). A survey of the hough transform. Computer vision, graphics, and image processing, 44(1), 87–116.

Insights, I. (2023). Tracking critical medication with RAIN RFID. Retrieved from https://www.impinj.com/library/partner-solutions/intelliguard-track-critical-healthcare-inventory-with-anesthesia-station/ (Image source: IMPINJ Insights.)

Jaderberg, M., Simonyan, K., Vedaldi, A. & Zisserman, A. (2014). Synthetic data and artificial neural networks for natural scene text recognition. arXiv preprint arXiv:1406.2227.

Jähne, B. (2005). Digital image processing. Springer.

Jasim, M. J. M., Kak, S. F., Ageed, Z. S. & Zeebaree, S. R. (2023). Spotted hyena optimizer driven deep learning-based drug-drug interaction prediction in big data environment. Computer Systems Science & Engineering, 46(3).

Jelacic, S., Bowdle, A., Nair, B. G., Kusulos, D., Bower, L. & Togashi, K. (2015). A system for anesthesia drug administration using barcode technology: the codonics safe label system and smart anesthesia manager™. Anesthesia & Analgesia, 121(2), 410–421.

Jessurun, J. G., Hunfeld, N. G. M., Van Rosmalen, J., Van Dijk, M. & Van Den Bemt, P. M. L. A. (2021). Effect of automated unit dose dispensing with barcode

scanning on medication administration errors: an uncontrolled before-and-after study. International Journal for Quality in Health Care, 33(4), mzab142.

Jin, H., Teng, Z., He, Y., Chen, Q., Wang, R. & Hu, Y. (2022, July). Medicine bottle recognition based on machine vision and deep learning in intravenous medicine dispensing robot. In 2022 IEEE International Conference on Real-time Computing and Robotics (RCAR) (pp. 272–277). Guiyang, China: IEEE. Retrieved 2023-05-23, from https://ieeexplore.ieee.org/document/9872280/ doi: 10.1109/RCAR54675.2022.9872280

Jocher, G. (2020, October). ultralytics/yolov5: v3.1 - bug fixes and performance improvements. Zenodo. Retrieved from https://github.com/ultralytics/yolov5 doi: 10.5281/zenodo.4154370

Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., ... Valveny, E. (2015). ICDAR 2015 competition on Robust Reading. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (p. 1156-1160). doi: 10.1109/ICDAR.2015.7333942

Karthikeyan, S., de Herrera, A. G. S., Doctor, F. & Mirza, A. (2021). An ocr post-correction approach using deep learning for processing medical reports. IEEE Transactions on Circuits and Systems for Video Technology, 32(5), 2574–2581.

Kelly, K., Harrington, L., Matos, P., Turner, B. & Johnson, C. (2016). Creating a culture of safety around bar-code medication administration: an evidence-based evaluation framework. JONA: The Journal of Nursing Administration, 46(1), 30–37.

Khan, S., Khan, S. & Kothandan, H. (2016). Simulator evaluation of a prototype device to reduce medication errors in anaesthesia. Anaesthesia, 71(10), 1186–1190.

Kittisorayut, S., Nonsiri, S., Kamin, P. & Makdee, S. (2024). A machine learning algorithms for drug identification: Enhancing medication safety in thailand. In 2024 21st international joint conference on computer science and software engineering (jcsse) (pp. 47–52).

Koppel, R., Wetterneck, T., Telles, J. L. & Karsh, B.-T. (2008). Workarounds to barcode medication administration systems: their occurrences, causes, and threats to patient safety. Journal of the American Medical Informatics Association, 15(4), 408–423.

Laohakijvitoon, P. & Nilwong, S. (2024). Vision-based drug classification system using transfer learning approach. In 2024 9th international conference on business and industrial research (icbir) (pp. 0400–0404).

Lee, E. K., Tran, D. D., Keung, W., Chan, P., Wong, G., Chan, C. W., ... Khine, M. (2017). Machine learning of human pluripotent stem cell-derived engineered cardiac tissue contractility for automated drug classification. Stem Cell Reports, 9(5), 1560–1572.

Lee, H. & Youm, S. (2021). Development of a wearable camera and ai algorithm for medication behavior recognition. Sensors, 21(11), 3594.

Leitch, S., Dovey, S. M., Cunningham, W. K., Smith, A. J., Zeng, J., Reith, D. M., ... Lillis, S. (2021, August). Medication-related harm in New Zealand general practice: a retrospective records review. British Journal of General Practice, 71(709),

e626–e633. Retrieved 2023-05-29, from http://bjgp.org/lookup/doi/10.3399/BJGP.2020.1126 doi: 10.3399/BJGP.2020.1126

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. Proceedings of the Soviet physics doklady.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... others (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459–9474.

Li, L., Pan, X., Chen, W., Wei, M., Feng, Y., Yin, L., ... Yang, H. (2020). Multi-manufacturer drug identification based on near infrared spectroscopy and deep transfer learning. Journal of Innovative Optical Health Sciences, 13(04), 2050016.

Liao, M., Wan, Z., Yao, C., Chen, K. & Bai, X. (2019). Real-time scene text detection with differentiable binarization.

Lin, C.-H., Tsai, F.-Y., Tsai, W.-L., Wen, H.-W. & Hu, M.-L. (2012, December). The feasibility of QR-code prescription in Taiwan: QR-code prescription. Journal of Clinical Pharmacy and Therapeutics, 37(6), 643–646. Retrieved 2023-03-23, from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2710.2012.01358.x doi: 10.1111/j.1365-2710.2012.01358.x

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the ieee conference on computer vision and pattern recognition (pp. 2117–2125).

Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. (2018). Focal loss for dense object detection. Retrieved from https://arxiv.org/abs/1708.02002

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740–755).

Lin, Y.-S., Tsai, C.-C., Chang, K.-M., Shih, P.-C. & Cheng, C.-L. (2020). A development of medication assist device based on multi-object recognition. In 2020 ieee region 10 conference (tencon) (pp. 224–228).

Litman, R. (2018, 03). How to prevent medication errors in the operating room? take away the human factor. British Journal of Anaesthesia, 120. doi: 10.1016/j.bja.2018.01.005

Lowe, D. G. (2004, nov). Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60(2), 91–110. Retrieved from https://doi.org/10.1023/B:VISI.0000029664.99615.94 doi: 10.1023/B:VISI.0000029664.99615.94

Macias, M. (2018). Impact of a barcode medication administration system on patient safety. Number 1/January 2018, 45(1), E1–E13.

Macriga, G. A., Samyukthaa, V. & Yogavarshini, S. (2023). Hospital management system. In 2023 intelligent computing and control for engineering and business systems (iccebs) (pp. 1–6).

Martinez Perez, M., Dafonte, C. & Gómez, Á. (2018). Traceability in patient healthcare through the integration of rfid technology in an icu in a hospital. Sensors, 18(5),

1627.

Mathioudakis, N. N., Abusamaan, M. S., Shakarchi, A. F., Sokolinsky, S., Fayzullin, S., McGready, J., ... Golden, S. H. (2021). Development and validation of a machine learning model to predict near-term risk of iatrogenic hypoglycemia in hospitalized patients. JAMA Network Open, 4(1), e2030913–e2030913.

Maydana, T., Giraldo, L., González, Z., Schachner, B., Mayan, J., Luna, D. & Benítez, S. (2017). Barcode medication administration in icu: Learning from our nurses. In Medinfo 2017: Precision healthcare through informatics (pp. 1095–1098). IOS Press.

Medsafe. (n.d.). New Zealand Medicines and Medical Devices Safety Authority. Retrieved from `https://www.medsafe.govt.nz/`

Merry, A. & Wahr, J. (2021).
In Medication safety during anesthesia and the perioperative period (p. i–ii). Cambridge University Press.

Mindee. (2021). doctr: Document text recognition. `https://github.com/mindee/doctr`. GitHub.

Mohd Yusof, N., Muda, A. K., Pratama, S. F. & Carbo-Dorca, R. (2022). Amphetamine-type stimulants (ats) drug classification using shallow one-dimensional convolutional neural network. Molecular diversity, 26(3), 1609–1619.

Mulac, A., Mathiesen, L., Taxis, K. & Granås, A. G. (2021). Barcode medication administration technology use in hospital practice: a mixed-methods observational study of policy deviations. BMJ quality & safety, 30(12), 1021–1030.

My, L. N. T., Le, V.-T., Vo, T. & Hoang, V. T. (2025). A comprehensive review of pill image recognition. Computers, Materials & Continua, 82(3).

Naik, U., Bhuatara, S. & Chougala, B. (2019). An experimental framework for automated bed localization and drug identification using zigbee signal strength and mobile robot. In 2019 ieee international conference on distributed computing, vlsi, electrical circuits and robotics (discover) (pp. 1–6).

Nakada, H. & Tanaka, Y. (2025). Robustness of structured data extraction from perspectively distorted documents. arXiv preprint arXiv:2511.17607.

Nanji, K. C., Patel, A., Shaikh, S., Seger, D. L. & Bates, D. W. (2016). Evaluation of perioperative medication errors and adverse drug events. Survey of Anesthesiology, 60(6), 259–260.

Nguyen, A. D., Nguyen, T. D., Pham, H. H., Nguyen, T. H. & Nguyen, P. L. (2022). Image-based contextual pill recognition with medical knowledge graph assistance. In Asian conference on intelligent information and database systems (pp. 354–369).

Nguyen, A. D., Pham, H. H., Trung, H. T., Nguyen, Q. V. H., Truong, T. N. & Nguyen, P. L. (2023). High accurate and explainable multi-pill detection framework with graph neural network-assisted multimodal data fusion. Plos one, 18(9), e0291865.

Nie, L., Lin, C., Liao, K., Liu, S. & Zhao, Y. (2021). Unsupervised deep image stitching: Reconstructing stitched features to images. IEEE Transactions on Image Processing, 30, 6184–6197.

Ou, Y.-Y., Tsai, A.-C., Zhou, X.-P. & Wang, J.-F. (2020). Automatic drug pills detection based on enhanced feature pyramid network and convolution neural networks. IET Computer Vision, 14(1), 9–17.

Owens, K., Palmore, M., Penoyer, D. & Viers, P. (2020). The effect of implementing bar-code medication administration in an emergency department on medication administration errors and nursing satisfaction. Journal of emergency nursing, 46(6), 884–891.

O'Connell, E., Pegler, J., Lehane, E., Livingstone, V., McCarthy, N., Sahm, L. J., ... Corrigan, M. (2016). Near field communications technology and the potential to reduce medication errors through multidisciplinary application. Mhealth, 2, 29.

Palenychka, R., Lakhssassi, A. & Palenychka, M. (2018). Verification of medication dispensing using the attentive computer vision approach. In 2018 ieee international symposium on circuits and systems (iscas) (pp. 1–4).

Parthasarathi, A., Puvvada, R., Patel, H., Bhandari, P., Nagpal, S. & Puvvada, R. K. (2021). Evaluation of medication errors in a tertiary care hospital of a low-to middle-income country. Cureus, 13(7).

Patel, S. & Loveridge, R. (2015). Obstetric neuraxial drug administration errors: a quantitative and qualitative analytical review. Anesthesia & Analgesia, 121(6), 1570–1577.

Pedersen, C. A., Schneider, P. J. & Scheckelhoff, D. J. (2016). Ashp national survey of pharmacy practice in hospital settings: monitoring and patient education—2015. American journal of health-system pharmacy, 73(17), 1307–1330.

Pilania, U., Kumar, M. & Singh, S. (2023). Leveraging convolutional neural networks for improved drug identification. In 2023 6th international conference on contemporary computing and informatics (ic3i) (Vol. 6, pp. 671–676).

Poon, E. G., Keohane, C. A., Yoon, C. S., Ditmore, M., Bane, A., Levtzion-Korach, O., ... Gandhi, T. K. (2010). Effect of bar-code technology on the safety of medication administration. New England Journal of Medicine, 362(18), 1698-1707. Retrieved from https://doi.org/10.1056/NEJMsa0907115 (PMID: 20445181) doi: 10.1056/NEJMsa0907115

Profetto, L., Gherardelli, M. & Iadanza, E. (2022). Radio frequency identification (rfid) in health care: where are we? a scoping review. Health and technology, 12(5), 879–891.

Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. Computer graphics and image processing, 1(3), 244–256.

Ranjbarzadeh, R., Caputo, A., Tirkolaee, E. B., Ghoushchi, S. J. & Bendechache, M. (2023). Brain tumor segmentation of mri images: A comprehensive review on the application of artificial intelligence tools. Computers in biology and medicine, 152, 106405.

Ratcliff, J. W. & Metzener, D. (1988). Pattern-matching: The gestalt approach. Dr. Dobb's Journal, 46–72.

Reimers, N. & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Resnik, P. & Lin, J. (2010). Evaluation of nlp systems. The handbook of computational

linguistics and natural language processing, 271–295.

Rice, V., Glass, N., Ogle, K. & Parsian, N. (2014). Exploring physical health perceptions, fatigue and stress among health care professionals. Journal of multidisciplinary healthcare, 155–161.

Rodriguez-Gonzalez, C. G., Martin-Barbero, M. L., Herranz-Alonso, A., Durango-Limarquez, M. I., Hernandez-Sampelayo, P., Sanjurjo-Saez, M. & iPharma. (2015). Use of failure mode, effect and criticality analysis to improve safety in the medication administration process. Journal of Evaluation in Clinical Practice, 21(4), 549–559.

Rodziewicz, T. L., Houseman, B., Vaqar, S. & Hipskind, J. E. (2024, 12 Feb). Medical error reduction and prevention. StatPearls [Internet]. (PMID: 29763131; Bookshelf ID: NBK539956)

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18 (pp. 234–241).

SAFERsleep Ltd. (n.d.). SAFERsleep. Retrieved from `https://safersleep.com/`

Sakowski, J. A. & Ketchel, A. (2013). The cost of implementing inpatient bar code medication administration. The American journal of managed care, 19(2), e38–45.

Saleem, M. (2023). Barcode medication administration technology to prevent medication errors. Journal of the College of Physicians and Surgeons–pakistan: JCPSP, 33(1), 111–112.

Saqib, S. M., Muhammad, O., Mazhar, T., Iqbal, M., Guizani, S. & Hamam, H. (2025). Medicine image classification using deep learning: highlighting the mednet-mobil hybrid model. Discover Computing, 28(1), 103.

Sarlin, P.-E., DeTone, D., Malisiewicz, T. & Rabinovich, A. (2020). Superglue: Learning feature matching with graph neural networks. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 4938–4947).

Shah, V., Keniya, R., Shridharani, A., Punjabi, M., Shah, J. & Mehendale, N. (2021). Diagnosis of covid-19 using ct scan images and deep learning techniques. Emergency radiology, 28, 497–505.

Shi, B., Bai, X. & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 39(11), 2298–2304.

Shi, B., Yang, M., Wang, X., Lyu, P., Yao, C. & Bai, X. (2018). Aster: An attentional scene text recognizer with flexible rectification. IEEE transactions on pattern analysis and machine intelligence, 41(9), 2035–2048.

Silva, M. C., e Silva, A. C., Orlando, M. T. & Bezzon, V. D. (2024). Towards a machine-learning-based application for amorphous drug recognition. IEEE Latin America Transactions, 22(9), 755–760.

Smith, R. (2007). An overview of the tesseract ocr engine. In Ninth international

conference on document analysis and recognition (icdar 2007) (Vol. 2, pp. 629–633).

Smith, T. F., Waterman, M. S. et al. (1981). Identification of common molecular subsequences. Journal of molecular biology, 147(1), 195–197.

Sohail, A., Nawaz, N. A., Shah, A. A., Rasheed, S., Ilyas, S. & Ehsan, M. K. (2022). A systematic literature review on machine learning and deep learning methods for semantic segmentation. IEEE Access, 10, 134557–134570.

Spina, A. C., Yang, C. D., Jain, A., Ha, C., Chen, L. E., Yee, P. & Lin, K. Y. (2025). Deep learning–driven glaucoma medication bottle recognition: A multilingual clinical validation study in patients with impaired vision. Ophthalmology Science, 5(4), 100758.

Srikamdee, S., Suksawatchon, U., Suksawatchon, J. & Werapan, W. (2022). Clinicya: An application for pill identification using deep learning and k-means clustering. In 2022 26th international computer science and engineering conference (icsec) (pp. 117–122).

Strykowski, J., Hadsall, R., Sawchyn, B., VanSickle, S. & Niznick, D. (2013, January). Bar-code-assisted medication administration: A method for predicting repackaging resource needs. American Journal of Health-System Pharmacy, 70(2), 154–162. Retrieved 2023-03-23, from `https://academic.oup.com/ajhp/article/70/2/154/5112843` doi: 10.2146/ajhp120200

Suksawatchon, U., Srikamdee, S., Suksawatchon, J. & Werapan, W. (2022). Shape recognition using unconstrained pill images based on deep convolution network. In 2022 6th international conference on information technology (incit) (pp. 309–313).

Suzuki, R., Imai, T., Sakai, T., Tanabe, K. & Ohtsu, F. (2021). Medication errors in the operating room: An analysis of contributing factors and related drugs in case reports from a japanese medication error database. Journal of Patient Safety, 18, e496 - e502.

Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature.

Tani, A., Norberto, J., da Silva, M. & Durand-Petiteville, A. (2024). Automated computation of cutting paths for unit-dose repackaging of pharmaceutical blister packs. In 2024 ieee 20th international conference on automation science and engineering (case) (pp. 1678–1683).

Tariq RA, S. A. S. Y., Vashisht R. (2022). Medication dispensing errors and prevention. StatPearls Publishing.

Tavakoli, M. J., Fazl, F., Sedighi, M., Naseri, K., Ghavami, M. & Taghipour-Gorjikolaie, M. (2025). Enhancing pharmacy warehouse management with faster r-cnn for accurate and reliable pharmaceutical product identification and counting. International Journal of Intelligent Systems, 2025(1), 8883735.

Ti, Y.-W., Chen, S.-K. & Wu, W.-C. (2020). A new visual cryptography-based qr code system for medication administration. Mobile Information Systems, 2020(1), 8885242.

Tian, Z., Shen, C., Chen, H. & He, T. (2020). Fcos: A simple and strong anchor-free object detector. IEEE transactions on pattern analysis and machine intelligence,

44(4), 1922–1933.

Ting, H.-W., Chung, S.-L., Chen, C.-F., Chiu, H.-Y. & Hsieh, Y.-W. (2020). A drug identification model developed using deep learning technologies: experience of a medical center in taiwan. BMC health services research, 20, 1–9.

Tran, T. T., Richardson, A. J., Chen, V. M. & Lin, K. Y. (2022, March). Fast and Accurate Ophthalmic Medication Bottle Identification Using Deep Learning on a Smartphone Device. Ophthalmology Glaucoma, 5(2), 188–194. Retrieved 2023-05-23, from https://linkinghub.elsevier.com/retrieve/pii/S2589419621001848 doi: 10.1016/j.ogla.2021.08.001

Truitt, E., Thompson, R., Blazey-Martin, D., Nisai, D. & Salem, D. (2016). Effect of the implementation of barcode technology and an electronic medication administration record on adverse drug events. Hospital pharmacy, 51(6), 474–483.

Tseng, M.-H. & Wu, H.-C. (2014, February). A cloud medication safety support system using QR code and Web services for elderly outpatients. Technology and Health Care, 22(1), 99–113. Retrieved 2023-03-23, from https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/THC-140778 doi: 10.3233/THC-140778

TTelectronics. (2023). RFID:the technology making industries smarter. Retrieved from https://www.ttelectronics.com/blog/rfid-technology/ (Image source: TTelectronics.)

Ultralytics. (2024). Yolov8: A novel object detection algorithm with enhanced performance and robustness. International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Retrieved from https://docs.ultralytics.com/models/yolov8/ doi: 10.1109/adics58448.2024.10533619

Upton, J., Olsson-Brown, A., Marshall, E. & Sacco, J. (2017). Using qr codes to enable quick access to information in acute cancer care. British Journal of Nursing, 26(10), S4–S12.

Uy, R. C. Y., Kury, F. P. & Fontelo, P. A. (2015). The state and trends of barcode, rfid, biometric and pharmacy automation technologies in us hospitals. In Amia annual symposium proceedings (Vol. 2015, p. 1242).

Uzun, V. & Bilgin, S. (2016, December). Evaluation and implementation of QR Code Identity Tag system for Healthcare in Turkey. SpringerPlus, 5(1), 1454. Retrieved 2023-03-23, from http://springerplus.springeropen.com/articles/10.1186/s40064-016-3020-9 doi: 10.1186/s40064-016-3020-9

van der Veen, W., Taxis, K. & van den Bemt, P. M. (2023). Safe administration of medicines in hospitals.

Vencer, L. V. T., Marqueses, L. G. M., Espineli, J. D., Caballero, A. R. & Caballero, J. M. (2024). Capsure: A novel application of cnns, openal's language model, and google text-to-speech for reliable drug recognition in the philippines. In 2024 9th international conference on business and industrial research (icbir) (pp. 0673–0678).

Verma, G., Chauhan, R. & Kumar, G. R. (2024). Automated pharmaceutical drug and vitamin classification using the efficientnetb3 model in deep learning. In 2024 international conference on information science and communications technologies (icisct) (pp. 458–463).

Vu, H.-D., Pham, V. H. & Le, Q. D. (2025). Drug classification system based on drug composition and usage instructions. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 12(1).

Wada, K. (n.d.). Labelme: Image Polygonal Annotation with Python. Retrieved from https://github.com/wkentaro/labelme doi: 10.5281/zenodo.5711226

Wang, T., Lu, C., Shao, M., Yuan, X. & Xia, S. (2022). Eldet: An anchor-free general ellipse object detector. In Proceedings of the asian conference on computer vision (pp. 2580–2595).

Warnakulasuriya, D., Dewangi, T., Sewwandi, N., Rathnayake, M., Kodagoda, N. & Suriyawansha, K. (2022). Mobile medical assistant system for laboratory report analysis and medical drug identification. In 2022 4th international conference on advancements in computing (icac) (pp. 102–107).

Wheeler, S. J. & Wheeler, D. W. (2005). Medication errors in anaesthesia and critical care. Anaesthesia, 60(3), 257-273. Retrieved from https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2044.2004.04062.x doi: https://doi.org/10.1111/j.1365-2044.2004.04062.x

Wilson, N., Jehn, M., Kisana, H., Reimer, D., Meister, D., Valentine, K., ... Clarke, H. (2020). Nurses' perceptions of implant barcode scanning in surgical services. CIN: Computers, Informatics, Nursing, 38(3), 131–138.

Woźniak, M., Siłka, J. & Wieczorek, M. (2023). Deep neural network correlation learning mechanism for ct brain tumor detection. Neural Computing and Applications, 35(20), 14611–14626.

Xia, J. (2024). Construction of an intelligent identification model for drugs in near infrared spectroscopy and research on drog classification based on improved deep algorithm. Scalable Computing: Practice and Experience, 25(3), 1361–1372.

Xie, N., Kalia, K., Strudwick, G. & Lau, F. (2019). Understanding mental health nurses' perceptions of barcode medication administration: a qualitative descriptive study. Issues in Mental Health Nursing, 40(4), 326–334.

Xu, J., Chen, C., Xie, H. & Lu, F. (2017). Cylindrical product label image stitching method. In Proceedings of the 2017 2nd ieee international conference on computational intelligence and applications (iccia) (pp. 331–335). doi: 10.1109/CIAPP.2017.8167215

Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X. & Tian, Q. (2022). Rethinking rotated object detection with gaussian wasserstein distance loss. Retrieved from https://arxiv.org/abs/2101.11952

Ye, J. (2023). Patient safety of perioperative medication through the lens of digital health and artificial intelligence. JMIR Perioperative Medicine, 6, e34453.

Ye, Z., Yi, C. & Tian, Y. (2013). Reading labels of cylinder objects for blind persons.

In 2013 ieee international conference on multimedia and expo (icme) (pp. 1–6).

Yoo, S., Hwang, H. & Jheon, S. (2016). Hospital information systems: experience at the fully digitized seoul national university bundang hospital. Journal of thoracic disease, 8(Suppl 8), S637.

York, D. (1966). Least-squares fitting of a straight line. Canadian Journal of Physics, 44(5), 1079–1086.

Yuliang, L., Lianwen, J., Shuaitao, Z. & Sheng, Z. (2017). Detecting curve text in the wild: New dataset and new solution.

Zadvinskis, I. M., Smith, J. G. & Yen, P.-Y. (2018). Nurses' experience with health information technology: Longitudinal qualitative study. JMIR medical informatics, 6(2), e8734.

Zankevich, A. (2021). How to unwrap wine labels programmatically. (Available online: `https://medium.com/hackernoon/how-to-unwrap-wine-labels-programmatically-31c8c62b30ce` (accessed on 15 June 2021))

Zhang, Y., Nie, S., Liu, W., Xu, X., Zhang, D. & Shen, H. T. (2019). Sequence-to-sequence domain adaptation network for robust text image recognition. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 2740–2749).

Zhang, Z., Xu, M. & Chen, S. (2021). Research on cylindrical labels unfolding algorithm based on machine vision. In Proceedings of the 2021 ieee 5th advanced information technology, electronic and automation control conference (iaeac) (pp. 2568–2572). doi: 10.1109/IAEAC51226.2021.9390987

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W. & Liang, J. (2017). East: an efficient and accurate scene text detector. In Proceedings of the ieee conference on computer vision and pattern recognition (pp. 5551–5560).

Zhu, D., Huang, Z., Yung, K. & Ip, A. W. (2024). Drug recognition detection based on deep learning and improved yolov8. Journal of Organizational and End User Computing (JOEUC), 36(1), 1–21.

Zhu, Z., Wu, R., Luo, M., Zeng, L., Zhang, D., Hu, N., ... Li, Y. (2024). Two-dimensional deep learning frameworks for drug-induced cardiotoxicity detection. ACS sensors, 9(6), 3316–3326.