Journal of Engineering and Science in Medical Diagnostics and Therapy. March 04, 2024;
Accepted manuscript posted November 28, 2024. 10.1115/1.4067302

The Role of Artificial Intelligence in Improving Histopathological Diagnosis of Prostate Cancer: A Review

Shereen Afifi[1], M.Hamdy Faragallah[2], Radwa Taha[3], Mirza Baig[4], Ehsan Ullah[5], Hamid GholamHosseini[6] and Sally I. Hassanein [7]

[1]Computer Science and Engineering Department, Faculty of Media Engineering & Technology, German University in Cairo, Egypt

shereen.moataz@guc.edu.eg

[2]Biochemistry Department, Faculty of Science, Helwan University, Cairo, Egypt

20212079@science.helwan.edu.eg

[3]Digital Media and Engineering Department, Faculty of Media Engineering & Technology, German University in Cairo, Egypt

radwa.hussein@student.guc.edu.eg

[4]Senior Solutions Specialist, Digital Front Door, Orion Health, Auckland, New Zealand

mirza.baig@orionhealth.com

[5]Department of Pathology and Laboratory Medicine, Health New Zealand, Te Toka Tumai Auckland, New Zealand

ehsanu@adhb.govt.nz

[6]School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

hgholamh@aut.ac.nz

[7]Biochemistry Department, Faculty of Pharmacy & Biotechnology, German University in Cairo, Egypt

sally.ibrahim@guc.edu.eg

Abstract

This review investigates the effectiveness of exploiting the massive Artificial Intelligence (AI) technology in the diagnosis of prostate cancer histopathological images. It focuses on studying and analyzing the current state and practice for utilizing AI tools, including significant machine learning and deep learning models in the histopathological image analysis process. The PRISMA methodology was adopted for conducting this systematic review to include recent research articles that have been published since 2017. Leveraging novel deep learning models and advanced imaging techniques, AI demonstrates promising capabilities in improving accuracy and efficiency in detecting and classifying prostate cancer. A comprehensive comparison of existing works has been presented with in-depth discussions around current limitations and key challenges while proposing some future advancements. This study aims to pave the way for future research and further integration of AI into the diagnostic processes towards early detection, personalized treatment strategies and enhanced patient outcomes in the context of a prostate cancer diagnosis.

Keywords- Artificial Intelligence; AI; Prostate Cancer; Machine learning; Deep learning; Histopathological images; Digital Pathology.

1. Introduction

Prostate cancer (PCa) is the second most common cancer and the fifth leading cause of death among men [1]. Significant improvements in the diagnosis and treatment of PCa have been achieved by introducing the use of Artificial Intelligence (AI) [2]. AI excels in analyzing and interpreting digital images of tissue samples on histopathology slides, as well as assigning a Gleason Score (GS) which is a key indicator of cancer behavior and guides the treatment plan for PCa [3]. Pathologists manually annotate Whole-Slide Images (WSIs) to train AI models to recognize patterns and features on slides of PCa [4–6]. By integrating AI within traditional histopathology workflows, pathologists can efficiently identify more cases nowadays in real time [7]. That revolutionized the area of digital pathology by enabling AI-powered computational algorithms and Computer-Aided Diagnosis (CAD) systems to enhance diagnostic accuracy while reducing workloads [2].

The massive technology of Machine Learning (ML) and Deep Learning (DL), a subfield of AI, has been widely used in radiology and bioinformatics for boosting diagnosis, prognosis and treatment processes. Modern ML and DL models have been developed and trained using various histopathological images for improving PCa diagnostic accuracy [8–11]. The application of ML and DL models to cancer detection in histology images follows a systematic process involving several essential steps [12,13] Initially, high-resolution histopathological images are acquired and meticulously labelled, with pathologists identifying regions that signify cancerous and healthy tissue. The labelled images undergo preprocessing, which may include resizing, normalization, and augmentation to ensure the data's robustness and to prevent overfitting [9]. The dataset is then partitioned into training, validation, and test subsets. Then, a DL model is trained on the annotated images, enabling it to recognize patterns and features characteristic of cancerous cells. During training, the model iteratively adjusts its parameters to minimize prediction errors, often employing methods such as transfer learning or fine-tuning existing models. After training, the model's performance is rigorously tested on validation and test datasets to gauge its accuracy and ability to generalize across different data samples. Once optimized, the model can be deployed to examine new histology images, autonomously detecting and classifying cancerous regions, thus offering critical support in diagnosis and informing cancer treatment [9]. Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Deep Neural Networks (DNNs) have shown significant results in the diagnosis of PCa and calculating the GS [10,14]. However, standardization of model development and the use of evaluation metrics are crucial for further analysis and comparisons.

Interestingly, a novel AI approach, Large Language and Vision Assistant for BioMedicine (LLaVA-Med) has been recently developed, which exhibits excellent multimodal conversational capability. This approach allows open-ended instructions to be used in biomedical image applications, outperforming previously supervised networks [13]. It has enormous potential to improve the accuracy of diagnosis and reduce the workload on pathologists, leading to ultimate enhancement in the quality of patient care [15].

Consequently, the main objective of this review is to study and analyze the current advancements around histopathology images of prostate cancer using AI, ML, and DL. In addition, this review discusses existing challenges and barriers, as well as highlights future advancements and opportunities for effectively applying AI in the detection and diagnosis of prostate cancer.

2. Methodology

In this review, we adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology for article selection to ensure transparency and replicability of the review process [16]. PRISMA is a widely recognized framework used in systematic reviews to enhance clarity and reduce bias during the selection process. It allows for a structured approach to identifying, screening, and including relevant studies based on predefined criteria.

2.1 Inclusion Criteria

We conducted a comprehensive search of critical databases including PubMed, IEEE Xplore, Scopus, SpringerLink, and Google Scholar. The search was performed using relevant keywords such as "prostate cancer," "AI," "DL," "ML,"

"pathology," and "histopathological images." Journal research articles that are published in the literature from 2017 to 2024 (inclusive) were selected for this review, focusing on the application and impact of AI, ML, and DL in histopathological images of prostate cancer. The proposed keywords were used to cross-search for thousands of related papers in the selected databases to identify initially a total of 503 papers (PubMed (368), IEEE (55), Scopus (15), SpringerLink (7), and Google Scholar (58)).

2.2 Exclusion Criteria

Based on our focus in this review regarding the role of AI in prostate cancer diagnosis using histopathological imaging, we excluded studies that used non-histopathological techniques and other screening modalities, such as MRI and ultrasound. In addition, other articles were excluded that addressed other related topics like biomarkers, genetic detection, surface proteins, or general pathology. We also excluded non-journal papers, such as conference papers.

After removing 67 duplicate records and 389 articles that did not meet our predefined criteria, 47 articles remained for further screening. During the screening , an additional 30 articles were excluded for not meeting the specific histopathology and urology focus of the review. Finally, 17 articles were selected for this review study that met our proposed inclusion and exclusion criteria. All authors reviewed and agreed upon the final selection of articles. Figure 1 visually summarizes this selection process.

3. Advancements in AI for Prostate Cancer Diagnosis using Histopathological Imaging

Artificial Intelligence has rapidly developed to a point where intelligent systems can now be developed which work very much like human cognition. The use of AI technologies in health care is becoming increasingly prevalent for improving patient care across a range of domains, including the diagnosis of prostate cancer among other complex diseases. Diagnosis or Gleason grading of prostate cancer is usually very time-consuming and prone to errors; thus, it may be subject to observer variability. AI opens up new opportunities to increase diagnostic accuracy for better risk assessment and prognosis. Automation and enhancement of the accuracy of histopathological analysis by AI could change the face of prostate cancer management. Kartasalo et al. showed that AI systems could support pathologists in both Gleason grading and diagnosis of prostate cancer to decrease this variability, since both undertreatment and overtreatment may result from such variability. Once combined with the expertise of pathologists, AI systems may lead to improved consistency in grading with benefits for patient outcomes [17].

Interestingly, AI has shown significant advancements promising to leverage the diagnosis and grading of prostate cancer. This section presents key studies for showing advancements in AI for the diagnosis of prostate cancer using histopathological images, and the selected 17 studies are summarized in Table 1 for further analysis, comparison and discussion.

N.M. Loorutu et al. used AI enhancement and segmentation techniques to classify images with high specificity and sensitivity of 0.9750 for Prostate Cancer. In this study, the best performance was expressed through the InceptionResNetV2 model because of its deep model architecture with residual connections, which helped detect complex patterns in the images. This model uniquely combined the advantages of the two worlds: the feature extraction capability of the Inception module and the efficiency in training from ResNet. Thus, it is an ideal model for histopathological image analysis. Three models, namely SVM, DenseNet121, and InceptionResNetV2, were evaluated in these experiments, using 4000 histopathological images sourced from a dataset called Diagset-A. The performance was higher with the InceptionResNetV2 compared to other solutions being evaluated [4].

In another related advance, Pinckaers et al. proposed an AI-driven approach for detecting prostate cancer in whole-slide images (WSIs). They implemented deep learning, specifically convolutional neural networks, which were trained on slide-level labels extracted from pathology reports. Preparation was done on a dataset with 1243 glass slides and a total of 5759 biopsy sections. The biopsy sections were reviewed by three pathologists to establish a consensus about the grade. Two methods were attempted: a streaming convolutional neural network and a MIL model. The streaming model performed slightly better than the MIL model, achieving an AUC of 0.992 compared to the latter's 0.990. This study exemplified the potential of AI in speeding up model training and enhancing diagnostic accuracy, while several challenges still exist, including poor generalization and missing true positives [6].

SR Duenweg et al. performed two machine learning models: ResNet and ATARI classification of Gleason grading in whole-mount prostate histology. While the ATARI model successfully could predict cancerous versus non-cancerous tissue, the ResNet101 classifier provided more granular predictions of Gleason grade differentiation. This study has proven that deep learning frameworks such as ResNet can enhance the detection capabilities of unique Gleason patterns, thereby enhancing diagnostic accuracy [11].

Among many other approaches, P. Ström et al. developed, based on these, an AI system using DNNs capable of telling benign from cancerous biopsy cores with great accuracy. This automated the quantification of cancer length, something that is quite time-consuming; its κ score was 0.83. However, one of the key limitations of this study was that it lacked pixel-level annotation, hence probably mixing both benign and malignant glands within the annotated regions. The authors stated that refinement techniques in annotation are continuously required to increase precision in diagnosis [18].

P. Raciti et al. investigated the accuracy of pathologists reading digitized WSIs of prostate biopsies with and without AI support. Using the deep learning system Paige Prostate, which was trained on 610 WSIs, they assessed the impact of AI on sensitivity and specificity. PaPr indeed yielded an increase in sensitivity but, under certain circumstances such as rare subtypes of cancer, it did poorly and hence require ongoing refinement and active clinical involvement throughout the diagnostic process [19].

In another study, Mun proposed a deep learning-based automated Gleason grading system that did not require complicated algorithms and massive manual annotations. YAAGGS performs Gleason grade group classification of WSIs in two steps: first, it extracts patch images, and then it classifies the extracted patches through a CNN into one of six classes. This simple process eliminates the need for many expert annotations, thus making the approach very efficient for clinical applications [20].

L. Duran-Lopez et al. proposed another deep learning-based CAD for the classification of malignant versus normal regions in WSIs from the Virgen de Valme Hospital. Very good results were achieved, with an average AUC of 0.996 using three-fold cross-validation. On this dataset, the approach of stain-normalization seemed to have limited utility, though it might be helpful for cross-platform AI systems in different clinical environments [21].

In the presentation of SUK Bukhari et al., the segmentation results of the CNN model in a diagnosis of prostate adenocarcinoma and prostatic hyperplasia were very good. The diagnosis precisions of ResNet-18, ResNet-34, and ResNet-50 models could reach 97.1%, 98%, and 99.5%, respectively. This work underlined the fact that AI-driven systems can reduce human fallibility and offer dependable diagnostic assistance for histopathology assessments [22].

Another promising idea was presented by O. Eminaga et al.: developing a "digital twin" system for managing prostate cancer pathology. This AI-driven digital twin of the physical pathology process can enhance clinical decisions and workflow but requires further guidelines to be laid out in working out this method effectively [23].

Indeed, Kartasalo et al. showed that AI systems could support pathologists in both Gleason grading and diagnosis of prostate cancer to decrease this variability, since both undertreatment and overtreatment may result from such variability. Once combined with the expertise of pathologists, AI systems may lead to improved consistency in grading with benefits for patient outcomes [17].

K. Nagpal et al. proposed a two-stage Deep Learning System that combined regional Gleason pattern classification and classification of the whole-slide Gleason Grade Group. It outperformed board-certified pathologists in the Gleason scoring of prostatectomy slides. Though quite accurate, several challenges regarding digital slide review workflows and integrating AI tools into clinical practice remain to be developed in the future [24].

Deep learning for automatic Gleason grading of prostate biopsies was studied by W. Bulten et al. Their study showed that the performance in multi-observed datasets was comparable to pathologists concerning the detection of prostate cancer. However, the authors declared that reliable assessment of performance could only be provided by expert reference standards and, by extension, needed further validation [25].

This research group also provided the framework for Prostate cANcer graDe Assessment, a global challenge to foster reproducible AI algorithms of the Gleason grading using 10,616 digitized biopsies. In that challenge, participants from 65 countries participated, achieving for most of them pathologist-level performances in blinded tests across different datasets,

representing a quantum leap in AI capabilities [26].

O. Kott et al. proposed a deep learning algorithm for histopathological diagnosis and Gleason grading of prostate biopsy samples. Tissue was classified at both fine (benign vs Gleason grades 3, 4 and 5) and coarse scales using a deep residual CNN. The AI system identified specific areas which, by its predictions, would have been easily missed by the pathologists and showed it has the potential to serve as a clinically useful decision-support tool [27].

Besides, W Li introduced a novel region-based convolutional neural network (R-CNN) framework combining a grading network head and an epithelial network head for multi-task prediction. Adding an Epithelial Network Head (EHN), adapting the Mask R-CNN to be suitable for the histological image analysis for the Gleason grading task with little additional computational overhead. Then in this study, they developed a two-stage training strategy which enables this model to detect epithelial cells and to predict Gleason grades simultaneously [28].

D. Karimi et al. proposed a method comprising three independent CNNs using varied patch sizes to classify images of prostate cancer. The system used image-space and feature-space augmentation, achieving an accuracy of 92% in distinguishing between cancerous patches and 86% in distinguishing low- and high-grade cancers. This further motivated research into the design of deep learning models to attain expert pathologist-level accuracy [12].

Ambrosini et al. proposed a deep-learning approach for the detection of cribriform growth patterns in prostate biopsies, an architectural feature associated with poor prognosis. Their CNN showed a high sensitivity for the detection of cribriform regions, although a limited number of false positives were reported too. This approach could be an important support for clinical decision-making, offering aggressive cancer feature identification [29].

Last but not least, G. Campanella et al. proposed an integrated deep-learning model of CNNs and RNNs that diagnose prostate cancer in WSIs. This model achieved an AUC of 0.98 on 1,824 slides and for the first time set the breakthrough that may build accurate and reliable decision-support systems into clinical use [30].

The performance evaluation metrics of various AI-based techniques in the detection of prostate cancer have some interesting findings from both the models performing well and the models performing poorly. Duran-Lopez et al. proposed an architecture, called PROMETEO, consisting of four convolution stages and three fully connected layers. The PROMETEO system with a functional CNN obtained the highest AUC score of 0.999, which is remarkable since it has an almost perfect 100 accuracy, sensitivity, and specificity when tested on a dataset comprising 97 whole slide images (WSIs) from Virgen de Valme Hospital [21]. This exceptional achievement might be related to the presence of a specialized deep-learning computational structure optimized for pathology purposes. Likewise, the DDN architecture, which was tested on 1961 historical biopsies comprising STHLM3 data, rendered a promising AUC value of 0.997, stressing its performance in other cross-institute datasets such as the one located in Karolinska University Hospital. [18]. Contrarily, some models such as CNN ResNet-18 employed at Miriam Hospital, produced much lower AUC values (0.83) and accuracy (91.5%) possibly because of a much smaller sample size of only 85 biopsies [27]. It can thus be opined that there are certain bulk datasets or sophisticated models that can enhance the accuracy of the diagnosis.

The next-in-line high-scoring model which has been evaluated successfully is the CNN model at Radboud University Medical Center which reached AUC 0.992, demonstrating the model's capability of conducting biopsy analysis on a massive amount of data as tall as 5759 biopsies [6]. High-performance modelled accuracy (99.5%) and specificity (98%) were also documented by ResNet-50 used at the University of Lahore confirming the efficiency of deep learning frameworks after training on the appropriate data sets [22]. On the other hand, the FSConv+GMP CNN model as applied to the SICAPv2 database achieved one of the poorest AUC values of 0.82 coupled with an accuracy of 67%, which indicates a deficiency in the extraction of relevant features [30].

N Singhal, et al., proposed a Deep Learning system used as an assistive tool to core needle biopsies (CNBs) to detect cancer regions and predict Gleason grades in WSI of the prostate. The system was tested for its capacity to distinguish between benign and malignant biopsies, biopsies with low- and high-grade tumours, and group 2 vs. group 3 biopsies. The system simulates iterative active learning-based data labelling on datasets. This simulation environment was inspired by the Cost-Effective Active Learning paradigm. Using the initial labelled dataset, a Fully Convolutional Network (FCN) semantic segmentation model which is the same design as the segmentation model of epithelium detection to train Gleason grade group identification [31].

Tolkach et al. addressed deep learning for Gleason grading on whole-slide photographs and the diagnosis of prostate cancer. For model development, the convolutional neural network architecture used was NASNetLarge; then, the dataset was assembled, bringing together around 389 slides coming from different patients and institutions. After applying augmentation procedures, the DL model attained more than 98% accuracy in tumour identification with 97.3% accuracy. The Gleason grading system indeed had an excellent performance in prognosis stratification. Nevertheless, several drawbacks are also pointed out by this review, including the concentration on tissue microarray-based data and incompletely digitalized cohorts [32].

J. Silva-Rodríguez, A. Colomer and M.A. Sales et al., the result in the validation sets for their proposed network and the performance of the different models ResNet, VGG19 and FSConv. And in their work, they use the global max-pooling layer to play the role of the global average pooling. All different configurations, fully connected layers with ReLU activation and dropout regularisation (FC), global-average-pooling (GAP) and global-max-pooling (GMP) layers and their combinations are implemented and their performance in this work. FSConv with different top models: fully connected layers (FC), global-max pooling (GMP), global-average pooling (GAP), or a combination of them (GAP+FC or GMP+FC). The results for the best tested fine-tuned architectures, VGG19 and RestNet, using the same top models as FSConv. The optimum hyperparameters were a learning rate of 0.01 for FSConv and 0.0001 for the finned-tunned networks. By the end, FSConv+GMP configuration is the best performing one for the patch-level Gleason grading [33].

4. Analysis and Discussion

The use of ML and DL algorithms in diagnosing and treating prostate cancer has popularized the use of AI tools in histopathology and other fields of diagnostic medicine. Table 1 presents key features and parameters to analyze the reviewed studies/articles, including the reported performance metrics, used datasets and methodologies. Several studies have demonstrated the potential of AI tools and models in diagnosing histopathological images of prostate cancer. However, the success of these AI tools highly depends on the validation of these tools, which requires large datasets of training cases and efficient customization and fine-tuning of the AI models. Also, the pathologist needs considerable time to annotate the cases/images used for the training dataset. In addition, AI has shown the potential to reduce variability, improve diagnostic accuracy and control/automate grading prostate biopsies, towards a massive transformation around digital pathology while reducing the workload of pathologists.

Table 1 also demonstrates how the developed AI models for cancer diagnosis in histopathological images have evolved and improved in evaluation metrics like Area Under the Curve (AUC), sensitivity, specificity, and accuracy. From Table 1, we concluded that accuracy, AUC, sensitivity, and specificity are the common evaluation metrics reported by researchers to evaluate the performance of their developed models. It also includes dataset references as mentioned besides them.

Fig. 2 depicts a visualization chart to compare different accuracy values achieved by some selected models (the best-performed models) reported in Table 1. This figure demonstrates how the accuracy changes with the complexity of the developed models. Similarly, Fig.3 represents different values of the achieved AUC metric of selected studies, as reported in Table 1. Across a range of sample sizes, the existing models have demonstrated reasonable AUC values, showing that advanced machine learning models outperformed traditional ones. Fig. 4, on the other hand, shows the correlation between different studies' sensitivity and specificity values reported in Table 1. The figure underlies the inverse relationship between these metrics: one is high, and the other is low.

From Table 1 and Fig. 2-4, It is clear that several studies achieved good values for the key evaluation metrics at different numbers of samples. Still, we have to consider that none of those studies was trained on a large number of datasets indicating that some of them will suffer from overfitting when tested on other datasets, which is a condition where the DL model memorizes the dataset instead of learning from it that is present in [17] Even with them having achieved the highest sensitivity and specificity almost 100%. This can be overcome by training on a large dataset and applying early stopping techniques. From the table, most of the datasets are private datasets, limiting the comparison between the studies as none of them used the same dataset, highlighting that some of these models may not be able to generalize to other datasets. It can also be seen that CNN is the most investigated DL network in this area proving its reliability. Having CNN as the most used network marks a research gap in whether Encoder-Decoder networks or RNN could have been investigated.

The most noteworthy AUC score of 0.999, accomplished by the PROMETEO framework employing a custom CNN, is

essential for its near-perfect precision, affectability, and specificity, highlighting the vigour of the show when connected to a dataset of 97 entire slide pictures (WSIs) from Virgen de Valme Clinic. [21]. This exceptional result may be credited to the model's custom-fitted plan, sketched out especially for pathology applications. So too, the DDN appeared evaluated utilizing 1961 biopsies from the STHLM3 dataset, finished with an astonishing AUC of 0.997, emphasizing its capability to generalize over distinctive datasets from all-inclusive helpful educate like Karolinska College Clinic. [18]. On the other hand, models like CNN ResNet-18 utilized by Miriam Healing Center appear lower AUC values (0.83) and exactness (91.5%), which may be due to the smaller test estimate of fair 85 biopsies. [27]. This recommends that bigger datasets or more progressed models are pivotal for making strides in symptomatic exactness.

Another basic high-performing show is the CNN appearance utilized at Radboud College Helpful Center, which accomplished an AUC of 0.992, sketching out the model's ampleness in biopsy examination with a broad test of 5759 biopsies. [6]. Besides, models such as ResNet-50 utilized by the College of Lahore have a tall precision (99.5%) and specificity (98%), underscoring the potential of significant learning structures when prepared on carefully curated datasets. [21]. On the other hand, the FSConv+GMP CNN appear associated with the SICAPv2 database recorded one of the most reduced AUC scores at 0.82, close to an exactness of 67%, proposing limitations in counting extraction or dataset appraise (182 WSIs) [33]. These lower values appear to reflect challenges inside the model's generalization capability or the complexity of the dataset utilized.

In terms of affectability, the PROMETEO appeared additionally fulfilled near-perfect affectability at 99.97%, though the PaPr DL appear from MSKCC closely taken after with an affectability of 97.4% [19,21]. Such tall affectability is basic in cancer disclosure since it decreases off-base negatives, ensuring that most cancerous tissues are precisely recognized. On the other hand, the k-nearest-neighbour classifier utilized by NMCSD and TCGA yielded a much lower precision of 70%, which can be due to the less troublesome nature of the calculation compared to more advanced noteworthy learning techniques [24] When huge datasets are accessible for the course of action, models with progressed calculations like CNN, ResNet, and customized critical learning approaches reliably outflank routine techniques, highlighting the significance of dataset degree and complexity in progression aggressive accuracy. To improve tumor differentiating evidence and assess at the master pathologist's level, unique AI models and techniques have been described in the PANDA challenge for Prostate Cancer Grade Appraisal [39].

Indeed, even though AI models have illustrated productivity, actualizing them is still challenging. For case, the execution of these models may be hampered by the whimsies of comments and the dissatisfaction of pathologists to re-evaluate honest-to-goodness positives and veritable negatives. In any case, the change in AI models and their integration interior standard procedures utilized by pathologists has the potential to inside and out lessen work, make strides in accuracy and lessen examining understanding time.

4.1 Challenges and Barriers

Based on our discourse, AI has appeared promising comes about in making strides in the conclusion, guess, and treatment of prostate cancer. Be that as it may, a few restrictions and key challenges related to the integration of AI in this space were recognized and overviewed as follows:

1. Information Predisposition and Data Bias: AI calculations intensely depend on preparing information. On the off chance that the datasets utilized for preparing are one-sided or not agents of different populaces, the AI models may deliver wrong or skewed results.
2. Interpretable Models: AI models regularly work as "dark boxes," making it challenging to decipher their decision-making forms. This need for interpretability can ruin physicians' beliefs and understanding of AI-generated experiences.
3. Need for Standardization: The non-appearance of standardized conventions for information collection, imaging procedures, and symptomatic criteria can lead to inconstancy in AI demonstration execution over diverse healthcare education.
4. Clinical Integration Challenges: The integration of AI into clinical workflows can be challenging, and healthcare experts may confront resistance or scepticism in embracing AI-driven devices.
5. Moral and Security Concerns: AI applications may raise moral and protection concerns, especially concerning quiet information security and educate assent for AI-driven diagnostics.
6. Moderation: Actualizing vigorous information security measures, following security controls, and guaranteeing straightforward communication with patients around the utilization of AI in their healthcare. Building up moral rules for

the sending of AI in prostate cancer care.

Tending to these confinements and challenges requires a collaborative exertion including healthcare experts, analysts, administrative bodies, and innovation designers to guarantee the dependable and successful integration of AI in making strides in prostate cancer care.

In response to these challenges and barriers, emerging AI technologies like Generative AI and Federated Learning show promise in addressing some of the key limitations outlined. While no study directly and efficiently applied such advanced techniques to prostate cancer diagnosis using histopathological images (to the best of our knowledge), its potential should not be overlooked. For handling data bias, Generative AI models, such as Generative Adversarial Networks (GANs), can be employed to synthesize diverse datasets, addressing the issue of biased or unbalanced data. These synthetic datasets can better represent underrepresented populations, thereby improving model generalizability. Yuan Xue et al. [40] demonstrated the use of GANs to generate synthetic histopathological images for a different cancer type, reducing data imbalance and improving model robustness. This approach could be adapted for prostate cancer diagnosis to mitigate similar biases, leading to more scalable and generalized models.

To address concerns around data privacy, Federated Learning (FL) offers a promising solution. FL enables AI models to be trained across decentralized systems, such as hospital networks, without the need to transfer sensitive patient data to a central location. Each institution trains the AI model locally on its dataset, and only model updates (i.e., the learned parameters) are shared with a central server. This approach ensures that patient data remains within the originating institution, significantly reducing the risk of data breaches and preserving patient confidentiality [41].

The integration of AI tools into existing healthcare infrastructures is critical for ensuring the seamless adoption of advanced diagnostic technologies, especially in the context of prostate cancer diagnosis. Efforts to establish industry-wide interoperability standards, such as FHIR (Fast Healthcare Interoperability Resources), are enabling AI tools to connect seamlessly with Electronic Health Record (EHR) systems and other healthcare technologies. These standards ensure that data, including histopathological images and patient records, can be efficiently shared and processed across diverse platforms while maintaining compatibility and security [42]. Moreover, APIs and middleware solutions are increasingly facilitating the integration of AI systems with diverse healthcare platforms, allowing AI models to retrieve and analyze prostate cancer-related data from various systems [43]. This streamlined integration reduces the need for manual data transfer and allows clinicians to make more informed decisions faster, without disrupting their routine.

In prostate cancer care, such integration can enhance the accuracy of diagnosis and improve patient outcomes by leveraging AI-driven insights in real time, all while ensuring that data privacy and interoperability are maintained.

4.2 Future Advancements

The longer-term of AI in moving forward prostate cancer holds promising conceivable outcomes, with progressing headways that can revolutionize diagnostics, treatment arranging, and quiet results. A few key patterns and potential advancements in this field are as follows:

- Improved diagnostics will proceed to advance, advertising more exact and effective demonstrative capabilities. This incorporates the early location of prostate cancer through forward investigation of imaging information, such as MRI and ultrasound.
- Personalized treatment plans utilizing AI-driven models will play a pivotal part in fitting treatment plans based on a person's persistent characteristics, hereditary profiles, and reaction designs. This personalized approach can optimize restorative intercessions and minimize side impacts.
- Prescient analytics for forecasting utilizing AI models will end up more modern in foreseeing persistent results, repeat dangers, and in general survival rates. This data can help clinicians in making more educated choices concerning post-treatment care and reconnaissance.
- Computerized pathology examination utilizing AI calculations will proceed to refine the examination of pathology slides, helping pathologists recognize and review prostate cancer with higher accuracy. This will speed up the demonstrative preparation and decrease changeability in elucidations.
- Persistent strengthening and instruction by giving them with open data about their condition, treatment alternatives, and potential results. This may upgrade shared decision-making between patients and healthcare suppliers.

As AI advances advance, their part in prostate cancer administration will doubtlessly grow, contributing to the more exact analysis, personalized medicines, and forward by and large quiet care. Be that as it may, it is significant to address moral, regulatory, and usage challenges to saddle the total potential of AI in prostate cancer healthcare.
AI brings about a meaningful alteration within the conclusion of prostate cancer and may be a herald for the advanced selection of AI in histopathological conclusions because of the promising results in terms of exactness and productivity. In any case, certain confinements, such as information inclination, interpretability, and moral contemplations, must be tended to guarantee that AI is mindfully and viably coordinated into demonstrative workflows and clinical hone. Encouraging inquiry and improvement endeavours and collaborative work will offer assistance completely unleash the potential of AI for way better results in prostate cancer care, and other zones of demonstrative and clinical pharmaceutical.

References

[1] Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med 2020;288:62–81. https://doi.org/10.1111/joim.13030.

[2] Chu TN, Wong EY, Ma R, Yang CH, Dalieh IS, Hung AJ. Exploring the Use of Artificial Intelligence in the Management of Prostate Cancer. Curr Urol Rep 2023;24:231–40. https://doi.org/10.1007/s11934-023-01149-6.

[3] Frewing A, Gibson AB, Robertson R, Urie PM, Corte D Della. Don't Fear the Artificial Intelligence: A Systematic Review of Machine Learning for Prostate Cancer Detection in Pathology. Arch Pathol Lab Med 2023. https://doi.org/10.5858/arpa.2022-0460-ra.

[4] Loorutu NM, Yazid H, Ab Rahman KS. Prostate Cancer Classification Based on Histopathological Images. International Journal on Robotics, Automation and Sciences 2023;5:43–53. https://doi.org/10.33093/ijoras.2023.5.2.5.

[5] Busby D, Grauer R, Pandav K, Khosla A, Jain P, Menon M, et al. Applications of artificial intelligence in prostate cancer histopathology. Urologic Oncology: Seminars and Original Investigations 2023. https://doi.org/10.1016/j.urolonc.2022.12.002.

[6] Pinckaers H, Bulten W, Van Der Laak J, Litjens G. Detection of Prostate Cancer in Whole-Slide Images through End-to-End Training with Image-Level Labels. IEEE Trans Med Imaging 2021;40. https://doi.org/10.1109/TMI.2021.3066295.

[7] Ayyad SM, Shehata M, Shalaby A, Abou El-Ghar M, Ghazal M, El-Melegy M, et al. Role of ai and histopathological images in detecting prostate cancer: A survey. Sensors 2021;21. https://doi.org/10.3390/s21082586.

[8] Castaldo R, Cavaliere C, Soricelli A, Salvatore M, Pecchia L, Franzese M. Radiomic and genomic machine learning method performance for prostate cancer diagnosis: Systematic literature review. J Med Internet Res 2021;23. https://doi.org/10.2196/22394.

[9] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson J V., Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 2021;13. https://doi.org/10.1186/s13073-021-00968-x.

[10] Hegazi MAAA, Taverna G, Grizzi F. Is Artificial Intelligence the Key to Revolutionizing Benign Prostatic Hyperplasia Diagnosis and Management? Arch Esp Urol 2023;76:643–56. https://doi.org/10.56434/j.arch.esp.urol.20237609.79.

[11] Duenweg SR, Brehler M, Bobholz SA, Lowman AK, Winiarz A, Kyereme F, et al. Comparison of a machine and deep learning model for automated tumor annotation on digitized whole slide prostate cancer histology. PLoS One 2023;18. https://doi.org/10.1371/journal.pone.0278084.

[12] Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. Deep Learning-Based Gleason Grading of Prostate Cancer from Histopathology Images - Role of Multiscale Decision Aggregation and Data Augmentation. IEEE J Biomed Health Inform 2020;24:1413–26. https://doi.org/10.1109/JBHI.2019.2944643.

[13] Li C, Wong C, Zhang S, Usuyama N, Liu H, Yang J, et al. LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day 2023.

[14] Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. Deep Learning-Based Gleason Grading of Prostate Cancer from Histopathology Images - Role of Multiscale Decision Aggregation and Data Augmentation. IEEE J Biomed Health Inform 2020;24:1413–26. https://doi.org/10.1109/JBHI.2019.2944643.

[15] Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nat Rev Urol 2019;16. https://doi.org/10.1038/s41585-019-0193-3.

[16] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ 2021;372. https://doi.org/10.1136/bmj.n71.

[17] Kartasalo K, Bulten W, Delahunt B, Chen PHC, Pinckaers H, Olsson H, et al. Artificial Intelligence for Diagnosis and Gleason Grading of Prostate Cancer in Biopsies—Current Status and Next Steps. Eur Urol Focus 2021;7:687–91. https://doi.org/10.1016/j.euf.2021.07.002.

[18] Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol 2020;21. https://doi.org/10.1016/S1470-2045(19)30738-7.

[19] Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, et al. Clinical Validation of Artificial Intelligence-Augmented Pathology Diagnosis Demonstrates Significant Gains in Diagnostic Accuracy in Prostate Cancer Detection. Arch Pathol Lab Med 2023;147. https://doi.org/10.5858/ARPA.2022-0066-OA.

[20] Mun Y, Paik I, Shin SJ, Kwak TY, Chang H. Yet Another Automated Gleason Grading System (YAAGGS) by weakly supervised deep learning. NPJ Digit Med 2021;4. https://doi.org/10.1038/s41746-021-00469-6.

[21] Duran-Lopez L, Dominguez-Morales JP, Conde-Martin AF, Vicente-Diaz S, Linares-Barranco A. PROMETEO: A CNN-Based Computer-Aided Diagnosis System for WSI Prostate Cancer Detection. IEEE Access 2020;8. https://doi.org/10.1109/ACCESS.2020.3008868.

[22] Bukhari SUK. The assessment of deep learning computer vision algorithms for the diagnosis of prostatic adenocarcinoma. Annals of Clinical and Analytical Medicine 2021;12. https://doi.org/10.4328/acam.20322.

[23] Eminaga O, Abbas M, Kunder C, Tolkach Y, Han R, Brooks JD, et al. Critical Evaluation of Artificial Intelligence as Digital Twin of Pathologist for Prostate Cancer Pathology. n.d.

[24] Nagpal K, Foote D, Liu Y, Chen PHC, Wulczyn E, Tan F, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. NPJ Digit Med 2019;2. https://doi.org/10.1038/s41746-019-0112-2.

[25] Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol 2020;21. https://doi.org/10.1016/S1470-2045(19)30739-9.

[26] Bulten W, Kartasalo K, Chen PHC, Ström P, Pinckaers H, Nagpal K, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat Med 2022;28:154–63. https://doi.org/10.1038/s41591-021-01620-2.

[27] Kott O, Linsley D, Amin A, Karagounis A, Jeffers C, Golijanin D, et al. Development of a Deep Learning Algorithm for the Histopathologic Diagnosis and Gleason Grading of Prostate Cancer Biopsies: A Pilot Study. Eur Urol Focus 2021;7. https://doi.org/10.1016/j.euf.2019.11.003.

[28] Li W, Li J, Sarma K V., Ho KC, Shen S, Knudsen BS, et al. Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images. IEEE Trans Med Imaging 2019;38. https://doi.org/10.1109/TMI.2018.2875868.

[29] Ambrosini P, Hollemans E, Kweldam CF, Leenders GJLH van, Stallinga S, Vos F. Automated detection of cribriform growth patterns in prostate histology images. Sci Rep 2020;10. https://doi.org/10.1038/s41598-020-71942-7.

[30] Campanella G, Silva VWK, Fuchs TJ. Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology 2018.

[31] Singhal N, Soni S, Bonthu S, Chattopadhyay N, Samanta P, Joshi U, et al. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. Sci Rep 2022;12. https://doi.org/10.1038/s41598-022-07217-0.

[32] Tolkach Y, Dohmgörgen T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. Nat Mach Intell 2020;2. https://doi.org/10.1038/s42256-020-0200-7.

[33] Silva-Rodríguez J, Colomer A, Sales MA, Molina R, Naranjo V. Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. Comput Methods Programs Biomed 2020;195. https://doi.org/10.1016/j.cmpb.2020.105637.

[34] Koziarski M, Cyganek B, Niedziela P, Olborski B, Antosz Z, Żydak M, et al. DiagSet: a dataset for prostate cancer histopathological image classification 2021.

[35] Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 2016;6. https://doi.org/10.1038/srep26286.

[36] Egevad L, Delahunt B, Berney DM, Bostwick DG, Cheville J, Comperat E, et al. Utility of Pathology Imagebase for standardisation of prostate cancer grading. Histopathology 2018;73. https://doi.org/10.1111/his.13471.

[37] Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013;45:1113–20. https://doi.org/10.1038/ng.2764.

[38] Zhong Q, Guo T, Rechsteiner M, Rüschoff JH, Rupp N, Fankhauser C, et al. A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients. Sci Data 2017;4. https://doi.org/10.1038/sdata.2017.14.

[39] Bulten W, Kartasalo K, Chen PHC, Ström P, Pinckaers H, Nagpal K, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat Med 2022;28:154–63. https://doi.org/10.1038/s41591-021-01620-2.

[40] Xue Y, Ye J, Zhou Q, Long LR, Antani S, Xue Z, et al. Selective synthetic augmentation with HistoGAN for improved histopathology image classification. Med Image Anal 2021;67. https://doi.org/10.1016/j.media.2020.101816.

[41] KPMG LLP. Federated Learning: Strategies for Improving Communication Efficiency. Iclr 2018;2008.

[42] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association 2016;23. https://doi.org/10.1093/jamia/ocv189.

[43] Bukowski M, Farkas R, Beyan O, Moll L, Hahn H, Kiessling F, et al. Implementation of eHealth and AI integrated diagnostics with multidisciplinary digitized data: are we ready from an international perspective? Eur Radiol 2020;30. https://doi.org/10.1007/s00330-020-06874-x.

Table 1. Summary of the reviewed articles.

Ref.	year	Method	Dataset	Number of samples	Performance metrics			
					AUC	Accuracy	Specificity	Sensitivity
[4]	*2023*	*InceptionResNet V2, DenseNet121, SVM*	*Diagset-A [34]*	4000 images	0.9750 0.9675 0.9448	97.25% 96% 85.50%	97.50 % 96.50 % 86.50 %	97.50 % 97 % 84.50 %
[6]	2021	CNN (ResNet- 34), MIL	Radboud University Medical Center [6,35]	5759 biopsies *(1243 glass slides)*	0.992 0.990	-	-	-
[11]	*2023*	*ResNet (101) , ATARI*	Private study [11]	*330 slides(47 patients)*	-	83 % 89%	-	-
[18]	2020	DDN	STHLM3 participants, International, Karolinska University Hospital [18,36]	*1961 (biopsies)*	0.997	-	-	-

[19]	*2022*	*PaPr DL*	(MSKCC) [19]	*610 (biopsies)*	0.99	-	94.8%	97.4%
[20]	2021	(YAAGGS)	(HUMC) and (KUGH) [12,20]	7600 WSIs	0.983	94.7%	96 %	93.6%
[21]	2020	PROMETEO (Custom CNN)	Pathological Anatomy Unit of Virgen de Valme Hospital in Seville Spain [21]	*97 WSIs*	0.999	99.98 %	100 %	99.97 %
[22]	*2021*	*ResNet-50* *ResNet-34* *ResNet-18*	University of Lahore - Islamabad Campus [22]	802 images	-	99.5% 98 % 97.1 %	-	-
[24]	2019	k-nearest-neighbor classified	NMCSD, TCGA, Marin Medical Laboratories [24,37]	*912 slides*	0.96	70 %	-	-
[25]	*2020*	*DLS*	Radboud University Medical Center, Department of Pathology and Molecular Pathology, University Hospital Zurich [25,38]	*1243 slides (5759 biopsies)*	0.990	95.3%	95.2%	99%
[27]	2022	(CNN; ResNet-18)	Miriam Hospital institutional pathology [27]	*85 biopsies*	0.83	91.5 %	90 %	93 %
[28]	2019	Path R-CNN	Pathology Department at Cedars-Sinai Medical Center [28]	*513 images*	0.998	99.07%	-	-
[29]	2020	CNN	Department of Pathology of Erasmus University Medical Center Rotterdam [29]	*128 (biopsies)*	0.80	-	-	90 %
[30]	2022	VGG11-BN under the MIL setup	*MIL setup Dataset from* the original pathology reports in the (LIS) at MSKCC [30]	*1824 slides*	0.98	-	-	-
[31]	2022	FCN	*MPUH Radboud Karolinska[31]*	*6670 WSIs*	0.997	89.4%	-	-
[32]	*2020*	*NASNetLarge*	GDC Data Portal of the National Cancer Institute, TCGA [32]	389 slides	-	97.3%	-	-
[33]	2020	FSConv+GMP (CNN)	*SICAPv2 database [33]*	*182 WSIs*	0.82	0.67	-	-

AUC= Area Under Curve, DDN= Deep Drive Network, DL= Deep Learning, CNN= Convolutional Neural Network, AI = Artificial Intelligence, SVM = Support Vector Machine, PaPr DL =Paige Prostate Deep Learning, FCN=Fully Convolutional neural network, NA = Not Available,
WSI = Whole Slide Image, DLS= Deep Learning System

MSKCC= Memorial Sloan Kettering Cancer Center
YAAGGS= Yet Another Automated Gleason Grading System, MPUH= Muljibhai Patel Urological Hospital

HUMC= Hanyang University Medical Center and KUGH=Korea University Guro Hospital, LIS= Laboratory Information System

NMCSD=Naval Medical Center San Diego , TCGA =The Cancer Genome Atlas,
MIL = Multiple-Instance Learning

Journal of Engineering and Science in Medical Diagnostics and Therapy. March 04, 2024;
Accepted manuscript posted November 28, 2024. 10.1115/1.4067302

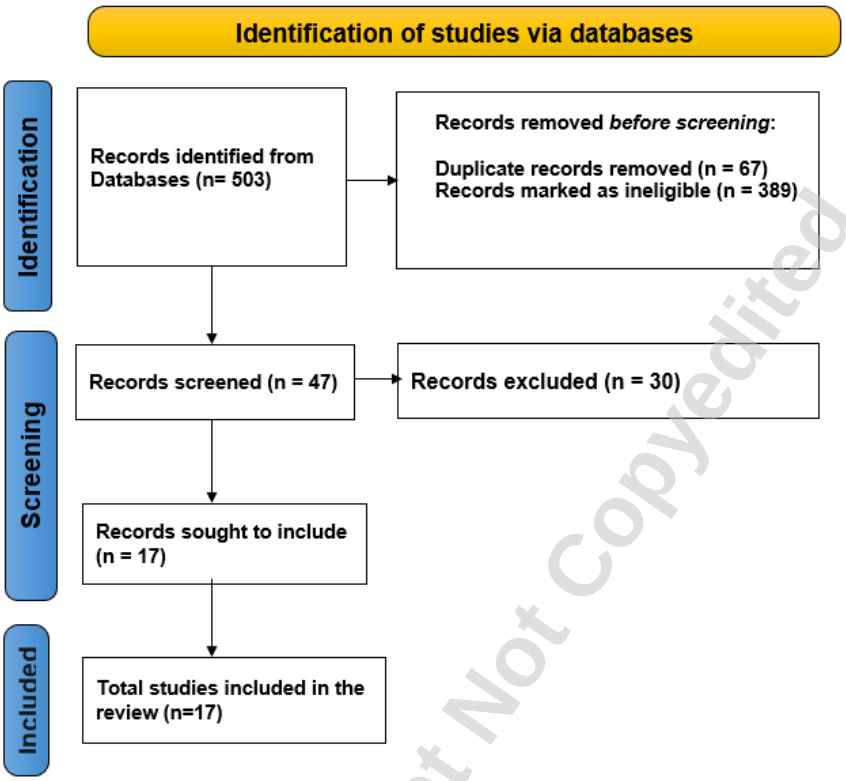

Fig. 1. The PRISMA chart for the selected studies.

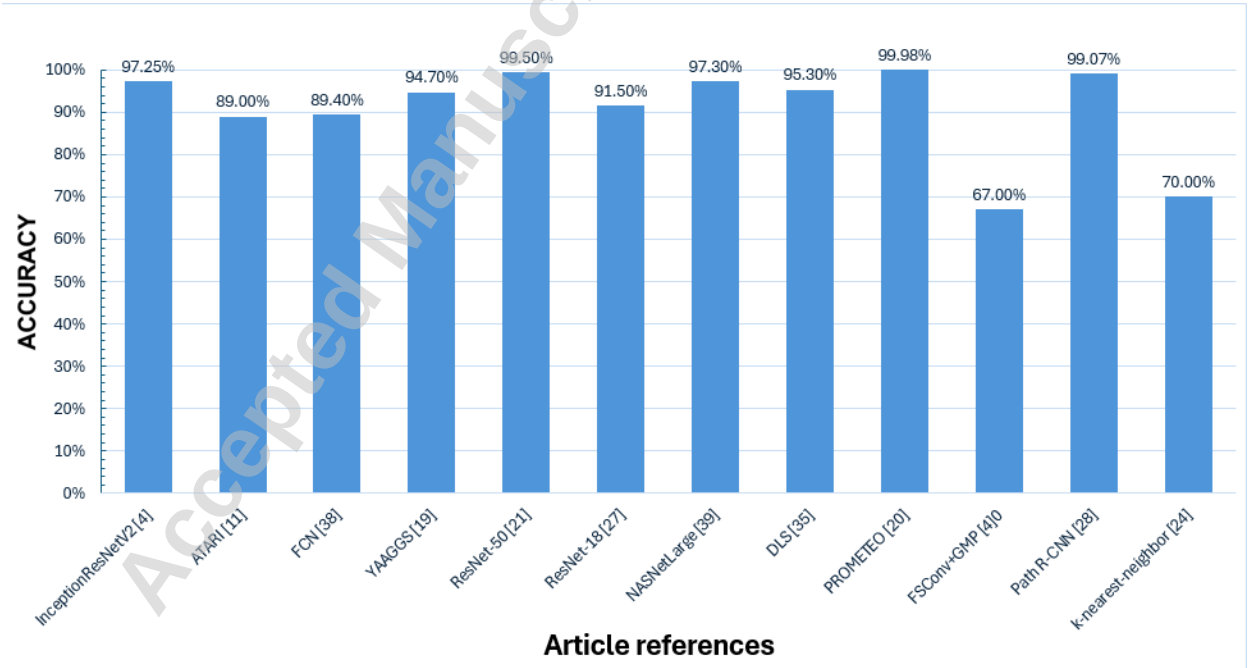

Fig. 2. The representation of Accuracy values of selected studies in Table 1.

Journal of Engineering and Science in Medical Diagnostics and Therapy. March 04, 2024;
Accepted manuscript posted November 28, 2024. 10.1115/1.4067302

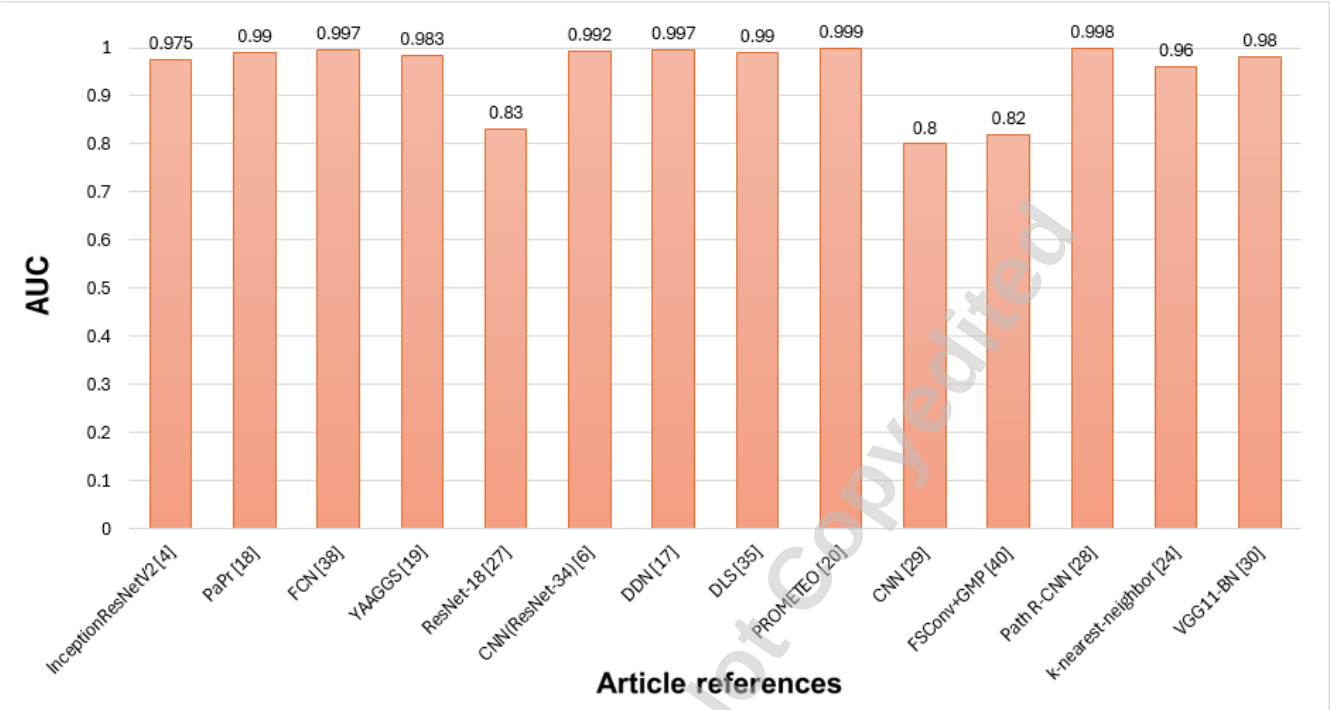

Fig. 3. The representation of AUC values of selected studies in Table.

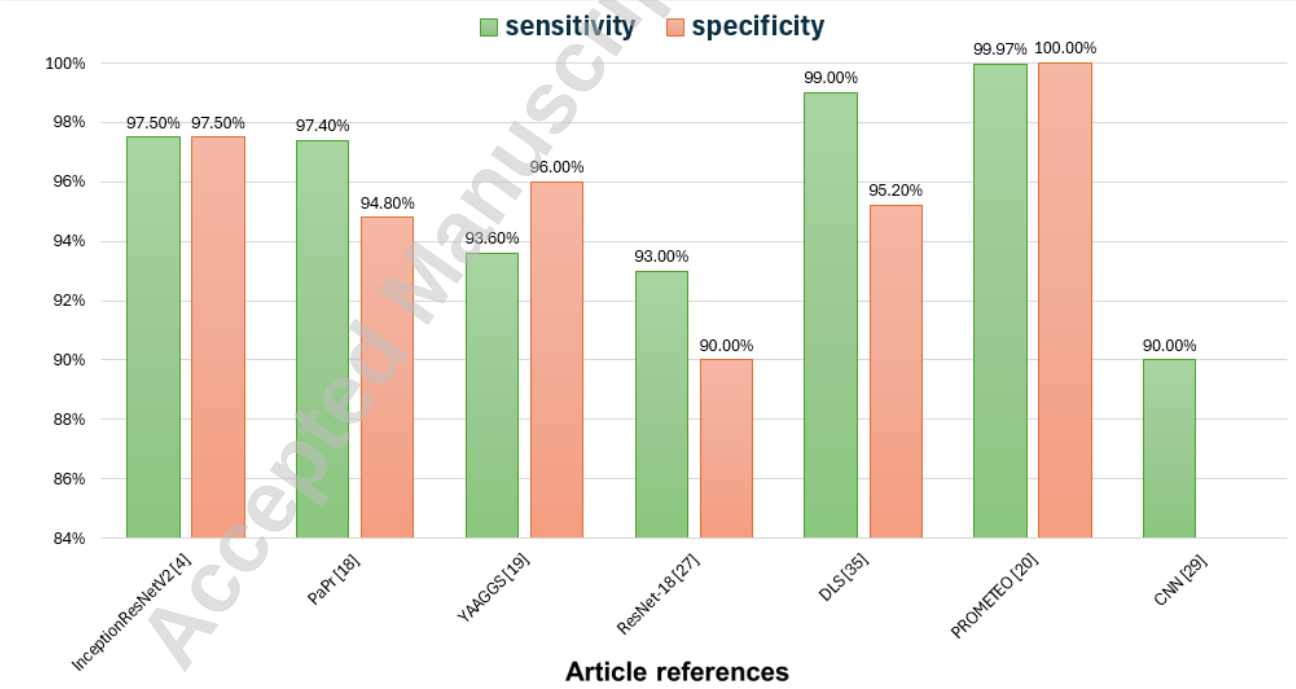

Fig. 4. The representation of sensitivity and specificity values of selected studies in Table 1.