12-1-2025

Precision and Approximation in Digitisation and Transformation of the Individual: Balancing Accuracy and Well-Being in AI-Driven Digital Systems

Yuming Li
Auckland University of Technology, yuming.li@aut.ac.nz

Pooja Sengupta
Information Systems and Operations Management, University of Auckland, pooja.sengupta@auckland.ac.nz

Ruhi Bajaj
The University of Auckland, r.bajaj@auckland.ac.nz

Claris Chung
University of Canterbury, claris.chung@canterbury.ac.nz

David Sundaram
The University of Auckland, d.sundaram@auckland.ac.nz

Precision and Approximation in Digitisation and Transformation of the Individual: Balancing Accuracy and Well-Being in AI-Driven Digital Systems

Full research paper

Yuming Li
Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: yuming.li@aut.ac.nz

Pooja Sengupta
Department of Information Systems and Operations Management
University of Auckland
Auckland, New Zealand
Email: pooja.sengupta@auckland.ac.nz

Ruhi Bajaj
Department of Information Systems and Operations Management
University of Auckland
Auckland, New Zealand
Email: r.bajaj@auckland.ac.nz

Claris Chung
Accounting and Information Systems
University of Canterbury
Christchurch, New Zealand
Email: claris.chung@canterbury.ac.nz

David Sundaram
Department of Information Systems and Operations Management
University of Auckland
Auckland, New Zealand
Email: d.sundaram@auckland.ac.nz

Abstract

Chronic diseases are largely caused by unhealthy lifestyle choices and behaviours. Early diagnosis and transformative management of chronic diseases are vital for the well-being of the global population. Unfortunately, data regarding the lifestyle choices and behaviours of individuals are sparse, fragmented, or nonexistent. These problems motivated the question of whether we can use both rough and precise data on individuals in a complementary fashion to diagnose and manage chronic diseases, ultimately leading to the well-being and transformation of the individual. We develop a holistic Measure, Model, Manage framework and an AI-driven granularity adaptation framework that learns interpretable mappings between rough self-reported lifestyle data and precise clinical indicators. Using both publicly available datasets and AI-generated synthetic datasets, we compare the robustness of models across varying input granularities. We demonstrate that chronic disease risk can be accurately predicted using not only high-precision biometric inputs but also rough, qualitative data.

Keywords: decision-making, digitisation of the individual, chronic disease management, precise and rough data, well-being

1 Introduction

Chronic diseases, such as cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes, are a global health concern and cause 75% of all deaths worldwide (World Health Organisation, 2023). Among these, diabetes alone affects over 10.5% of the global population aged 20–79 (537 million), with an additional 240 million individuals estimated to be living undiagnosed (Magliano et al, 2021). As these conditions are lifelong and largely caused by unhealthy lifestyle choices and behaviours, having quality of life and minimizing the burden on healthcare systems, policymakers, insurers, and other stakeholders are essential (Bloom et al., 2020). Therefore, early diagnosis and transformative lifestyles, particularly in areas such as diet, physical activity, sleep, and stress management, are vital players in preventing disease progression and enhancing long-term well-being.

To support individuals in this transformation, access to proper and relevant lifestyle data is critical. Unfortunately, data regarding the lifestyle choices and behaviours of individuals are sparse, fragmented, or nonexistent. While healthcare providers routinely collect highly precise individual data, most of this data is not relevant for the transformative management of chronic diseases. Clinical indicators such as HbA1c levels, cholesterol, and signs of retinal neuropathy offer an accurate snapshot of disease progression but provide little insight into the underlying lifestyle choices or behaviours of the individual that contribute to disease onset (Britton et al, 2012).

On the other hand, a growing trend in capturing precise individual data using devices such as smart watches, smart phones, and continuous glucose and blood pressure monitoring devices has led to a surge in the availability of fine-grained data on daily activities. Such data are typically extracted from electronic health records (EHRs) or clinical assessments that are conducted by trained professionals (Mosa et al., 2022). However, while frequent, precise data collection may improve accuracy, it often escalates workload and financial costs, and can also heighten patient anxiety (Serrano et al., 2023). Rough data, such as diet and stress, which have a significant influence on the management of chronic disease, particularly diabetes, are often poorly captured.

Moreover, diet and stress are inherently difficult to measure with precision. While some emerging apps capture rough attributes related to diet and emotional states, these are rarely integrated with other health data in ways to support holistic chronic disease management. Additionally, the growing emphasis on precise tracking can lead to unintended psychological consequences (Rosman et al, 2020). Individuals may begin to interpret their physical activity data as a measure of personal performance, shifting focus from holistic health and well-being to the narrow goal of hitting daily metrics, potentially undermining motivation and lifestyle change over time.

These practical and research problems motivated us to ask the question whether we can use rough and precise data of individuals in a complementary fashion to diagnose and manage chronic diseases such as diabetes, leading to the ultimate well-being and transformation of the individual. To address these problems, we develop a Measure, Model, Manage framework and an AI-driven granularity adaptation framework that learns interpretable mappings between rough self-reported data and precise clinical indicators. Using both publicly available datasets and AI-generated synthetic datasets, we compare model robustness across input granularities and introduce methods for inverse reasoning, from rough inputs to plausible precise output ranges.

2 Design Science Research Methodology

Considering the research goals, we adopt design science research (Nunamaker et al, 1991; Hevner et al, 2004; Sein et al, 2011) to design our research procedures (Figure 1). Key phases of the methodology relevant to our research were (1) Observe (through Literature Review – section 3) (2) Theorise (through Theory Building – section 4) (3) Design (though Design and Implementation – section 5) (4) Evaluate (through experiments - section 6), and (5) Reflect (section 7). The DSR methodology adapted is an iterative one with the different phases being visited multiple times over the course of this research, leading to refinement of conceptual, procedural, and system artefacts. In terms of the Brendel et al (2022) Design Science Focus Matrix our research focusses on (a) the Designing Quadrant (Theorizing by Exploring) – we design and implement a solution in the diabetes context that could be generalized to chronic diseases and (b) the Solving Quadrant (Instantiating by Applying) – we solve the problem of lack of holism in the measurement (rough and precise data), diagnosis, modelling, prediction, and management of chronic diseases.

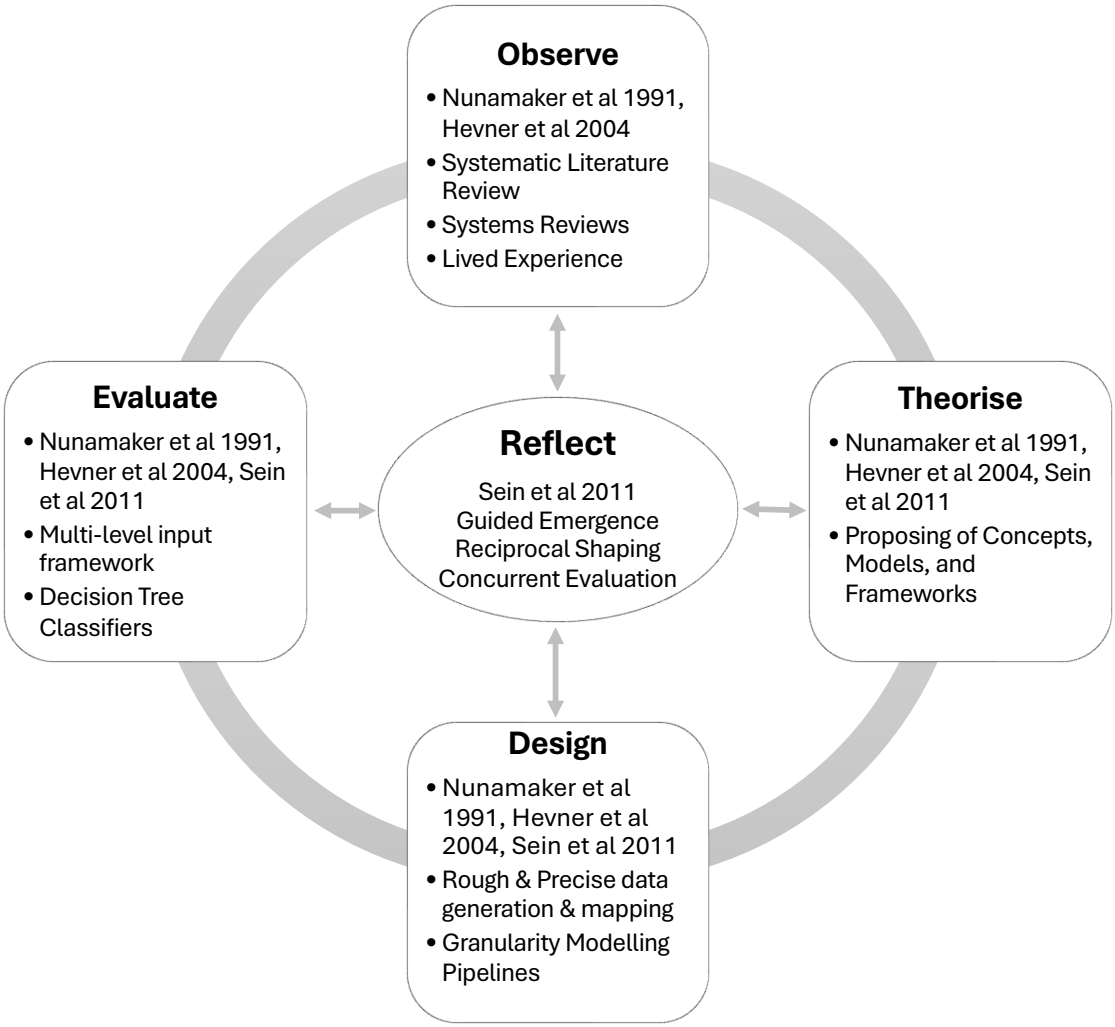

Figure 1: Design Science Research Methodology (Synthesis of Nunamaker et al, 1991; Hevner et al, 2004; and Sein et al, 2011)

3 Literature Review

The first step in the design science research process was observation through a literature review. The methodology of the literature review (section 3.1) and a review of the papers (section 3.2) found therein are discussed below. The practical and research problems (section 3.3) identified through this review are synthesised into issues (section 3.4). These problems and issues motivated our research questions (section 3.5).

3.1 Methodology

To ensure methodological rigour, we conducted an extensive literature review to examine advancements in integrating precise clinical and rough lifestyle measurements for diabetes risk prediction and management, with a focus on AI and machine learning applications. Searches were carried out in Scopus, PubMed, and Web of Science using a structured query, limited to English peer-reviewed articles between 2010 and 2025 across decision sciences, engineering, medical informatics, and computer science. We used Generative AI (GenAI) to enhance data processing and analysis. This approach revealed interdisciplinary insights, practical challenges, and opportunities for applying GenAI to predictive health analytics and improving digital health interventions.

3.2 Literature Review

One of the major global health challenges of the present day is Diabetes mellitus, particularly type 2 diabetes, primarily influenced by diet, physical inactivity, obesity, and smoking. There is an abundance of studies that have applied machine learning models for classifying diabetes risk using structured clinical and lifestyle datasets (Meyer, 2014). The focus of this paper is to improve treatment strategies

for type 2 diabetes patients by predicting and reducing treatment failures using data mining techniques applied to structured electronic health record (EHR) data.

By developing a predictive pipeline incorporating class balancing and multiple algorithms, Mosa et al. (2022) evaluate machine learning models on precise EHR data. Their objective is to predict diabetes-related complications such as eye and kidney diseases. There are similar studies like Özmen and Ozcan (2020), where the authors combine Artificial Neural Networks (ANN) and Classification and Regression Tree (CART), optimized with Genetic Algorithms (GA), for diabetes identification using the Pima Indian Diabetes Dataset (2019). Rawat (2019) applies several machine learning techniques, including AdaBoost, LogicBoost, RobustBoost, Naïve Bayes, and Bagging, for diabetes classification. They conclude that Bagging and AdaBoost provide the best classification accuracy. These studies collectively highlight the growing efficiency of machine learning in early diabetes detection and management. More recently, Shahade and Deshmukh (2025) integrate fuzzy logic with ensemble classifiers to enhance diagnostic precision. This reinforces the strength of structured data-driven models.

However, in real-world scenarios, especially during early or self-reported assessments, precise quantitative data is often unavailable. Variables such as exercise, sleep quality, or calorie intake are frequently reported as rough estimates (e.g., "high sugar consumption," "sedentary," or "occasional exercise"). Most existing studies (Piri et al., 2017) rely on structured inputs such as BMI, HbA1c, or blood pressure, which depend on clinical instruments or continuous monitoring. Although such precise data yield accurate classification models, they are less scalable for population-level or resource-limited settings.

Recent studies have started addressing the uncertainty inherent in rough or qualitative health data. Zhou et al. (2024) propose a deep fuzzy inference system combining Convolutional Neural Networks (CNNs) with fuzzy logic to handle uncertain variables in diabetes diagnosis. Similarly, aligning with rough data representation, Grande-Ramírez et al. (2024) develop an intelligent fuzzy system that evaluates diabetes risk using linguistic variables. Although progress has been made in the exploration of rough, qualitative data, limited attention has been given to predictive models capable of handling incomplete or approximate data inputs on scale.

The interpretability of such systems remains of paramount importance. Several studies like Maass and Storey (2021) and Nodoust and Ghatari (2024) support conceptual modeling in ML to balance interpretability and accuracy. The latter proposes a hybrid ML system that leverages synthetic data for better generalization with rough inputs. Integrating Rough Set Theory (RST) (Pawlak, 2002; Ye et al., 2013) and Concept Hierarchies (Chien et al., 2009) offers a promising approach for handling imprecise or linguistic variables (e.g., "low," "medium," "high" activity levels). Also, RST generates transparent, human-readable "if–then" rules, supporting interpretability crucial for clinical decision-making, for instance: IF lifestyle is Sedentary AND sugar intake is High THEN diabetes risk is High.

The above extant literature motivates the need for exploration of predictive models that incorporate rough, qualitative inputs for diabetes risk classification in combination with precise data. The strength of such models lies in the structured data-based approaches while enabling wider applicability in community health interventions. Given their scalability and interpretability, rough data–based systems could extend diabetes prediction capabilities beyond clinical settings, facilitating inclusive, proactive, and context-aware risk management.

3.3 Practical and Research Problems

Practical Problems: Although precise data has their advantages, they are not free from their disadvantages. As we can see from table 1, precise data might sometimes have the disadvantage of inaccessibility. First, most machine learning models to predict diabetes risk rely on accurate, structured clinical inputs (e.g., BMI, blood pressure, HbA1c, weight). These are often unavailable in low-resource settings or for individuals using self-monitoring tools. In such situations, we must rely on rough data for model building. Second, precise data are collected by professionals, which limits their use in community-based public health settings where only rough, qualitative data is available. Finally, tracking precise, granular data, especially in digital form (e.g., number of steps, calories) to continuously monitor health, may have a counterproductive impact and cause stress and anxiety.

Research Problems: Traditional diabetes risk prediction and diabetes management models predominantly rely on precise, structured clinical data such as blood glucose levels, HbA1c, BMI, and blood pressure. However, in real-world settings, especially in low-resource environments or when individuals self-monitor, they often rely on rough, qualitative inputs (e.g., "high sugar intake", "sedentary lifestyle"). These data sources are more scalable but less granular. There is a need to explore

modeling approaches that can intelligently integrate both rough (qualitative, self-reported) and precise (quantitative, clinical) data for more inclusive and explainable diabetes risk assessment and management. Further, there is a gap in the literature in the methods that can juxtapose such data sources to uphold predictive accuracy while enhancing interpretability, accessibility, and user comfort in tracking health outcomes.

3.4 Issues

Table 1 highlights the strengths and weaknesses of both precise and rough measurements. It is important to strike a balance between their use to improve the prediction of the risk of diabetes, and to improve disease management. Fine-grained tracking can reduce interpretability and overwhelm users, whereas rough data is easier to interpret but harder to model. Although rough is easier to interpret, rough data may introduce ambiguity and variability, requiring robust techniques to manage uncertainty and derive actionable insights. There is also a behavioral trade-off; it is true that rough measurements are easier to keep for individuals who are not professionals, but they may treat meeting precise targets over meaningful health practices or improved lifestyle choices. This may, in turn, lead to reduced long-term adherence and/or increased stress to manage the disease.

Aspect	Precise Measurement Approach	Rough Measurement Approach	Strengths (Precise / Rough)	Weaknesses (Precise / Rough)
Data Type	Numerical, continuous	Categorical, textual, or qualitative	Precise: Enables detailed quantitative analysis. Rough: Captures qualitative nuances and user perception.	Precise: Lacks flexibility for subjective data. Rough: Limited precision and comparability.
Data Collection	Requires clinical tools, lab instruments, and professional monitoring	Self-reported via surveys, interviews, or mobile apps	Precise: High reliability and objectivity. Rough: Low-cost and easy to collect.	Precise: Expensive and resource intensive. Rough: Prone to recall bias and inconsistency.
Accessibility	Restricted to healthcare or research facilities	Accessible across resource-limited and remote populations	Precise: Controlled conditions ensure data integrity. Rough: Broad reach and inclusivity.	Precise: Limited scalability. Rough: Inconsistent data quality.
Handling Missing Data	Requires imputation or case exclusion	Managed through linguistic approximation	Precise: Supports statistical corrections. Rough: Naturally tolerant of gaps.	Precise: Sensitive to missing values. Rough: Oversimplifies variability.

Table 1. Comparison Between Rough and Precise Measurements for Diabetes Risk Prediction and Management

3.5 Research Questions

The above practical and research problems and issues motivate us to ask the following research questions:

> *RQ1: Can diabetes risk be accurately predicted using a hybrid model that incorporates both rough (qualitative) and precise (quantitative) measurements?*
>
> *RQ2: How does the predictive performance (e.g., accuracy, sensitivity, specificity) vary across only rough data, only precise data, and a combination of both?*
>
> *RQ3: What interpretable rules or decision support can be generated from models that combine rough and precise data to guide individual and public health management systems?*

As an extension of the above research questions, this study also seeks to examine how the integration of rough and precise data influences users' psychological outcomes within AI-driven decision support systems. Such outcomes are in the form of stress, adherence, and perceived control. In addition, it aims to identify the types of data most crucial for effective diagnosis, management, and potential reversal of diabetes. These extensions further extend the investigation into how hybrid data integration not only

supports medical decision-making but also shapes users' behavioural and emotional engagement with digital health technologies, such as healthcare apps.

4 Theory Building

Observation through the review of the literature motivates us to propose that rough and precise data are both required to help us (1) diagnose chronic diseases such as diabetes, (2) manage chronic diseases, and (3) ultimately transform individuals and potentially reverse chronic diseases. Furthermore, we propose that chronic disease prediction and management models integrate and interweave rough and precise data in a way that provides a holistic view of the individual that includes data about their disease as well as their behaviours. To support these we need to have (a) mechanisms, devices, and apps that help us to *measure* rough as well as precise date (b) *models* that help us appropriately integrate and transform rough and precise data into a wholistic view of the individual to various decision makers such as the individual themself and the health practitioner and (c) models, devices, visualizations, games, gamification and apps that helps persuade the individual to *manage* and reverse their chronic disease effectively. We propose the Measure, Model, and Manage framework as illustrated in Figure 2. The data and models described above not only help us in *describing* the condition of the individual, but also to *predict* the future condition of the individual based on current lifestyle practices and behaviours, as well as potentially different practices and behaviours. Such data and models are also valuable in persuading as well as prescribing courses of action that could result in the transformation of the individual.

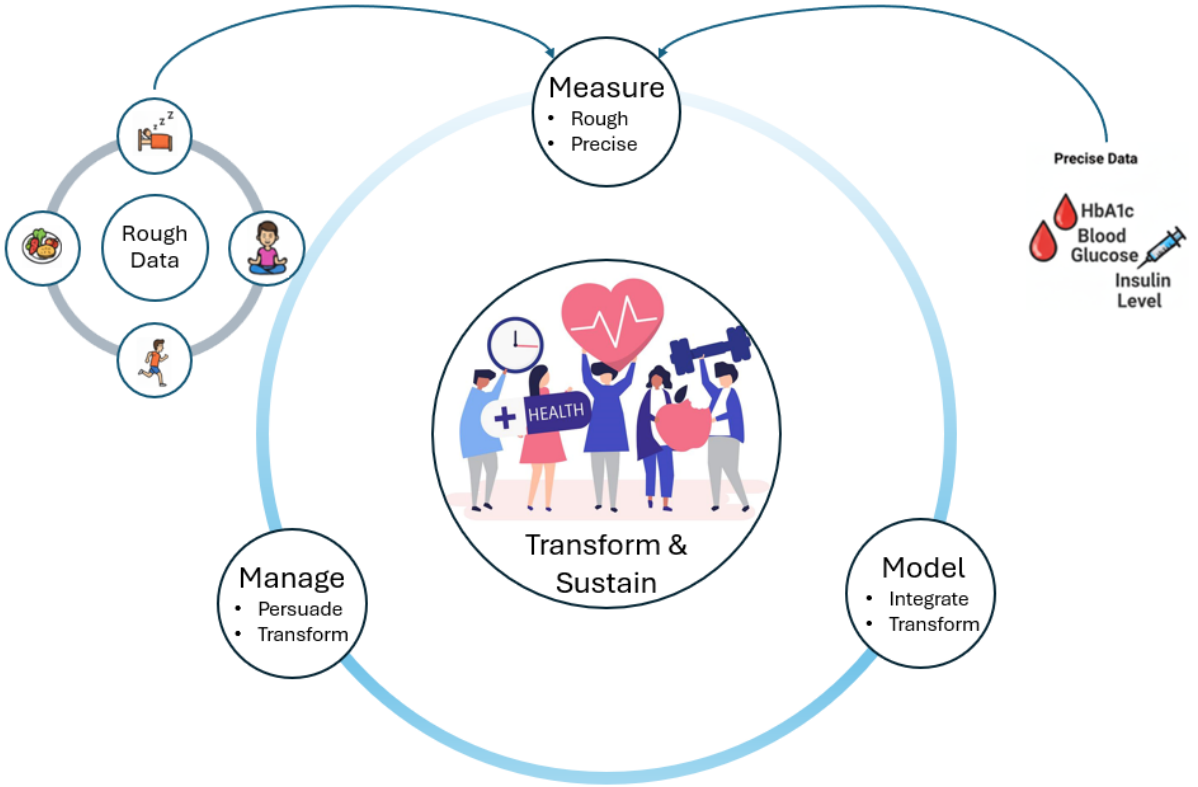

Figure 2: Measure, Model, and Manage for Individual Digitisation and Transformation

For example, in the context of diabetes management, precise measurements are essential for clinical assessment and treatment planning, including blood glucose levels, HbA1c, and insulin levels. The rough measurements are lifestyle data that indicate daily health, which influences metabolic outcomes. Tracking this rough data is particularly vital as diabetes can often be improved by improving lifestyle behaviours (Abdel-Rahman et al., 2022; Svedbo Engström et al., 2016). Especially sleep, exercise, emotional stress, and diet are the most recommended lifestyle areas for the self-management of the disease. By transforming lifestyle behaviours, particularly in these areas, disease can be managed or even reversed, allowing individuals to live sustainably. Each component reflects a foundational aspect of daily life that can offer valuable insights into an individual's overall well-being. This includes assessing whether adequate sleep was obtained, whether some physical activity was undertaken, whether efforts were made to reduce emotional stress from sources like work or finances, and whether dietary intake

was reasonably balanced in terms of macronutrients. Therefore, we introduce the SEEDs, an acronym that stands for Sleep, Exercise, Emotional stress, Diet, and Sustainability, as important lifestyle aspects to manage for advocating a complementary approach that integrates these qualitative aspects into the broader care model. The SEEDs provides a broad approach for managing diabetes, as it focuses on the key lifestyle factors that can influence the disease. As such, the added focus on sustainability highlights the importance of maintaining these behaviours consistently over time. While not as precise as clinical data, these indicators provide critical context and support a more holistic and actionable approach to managing diabetes effectively.

5 Design

To investigate the effects of measurement granularity on predictive performance and interpretability, we systematically transformed the original features from the PIMA Indians Diabetes dataset (PIMA Indians Diabetes Dataset, 2019) into four increasingly abstracted levels of categorical inputs (L0–L3), while preserving the original continuous values at L4. Notably, the PIMA dataset is composed of continuous numerical variables (e.g., glucose concentration, BMI, blood pressure), and we designated this raw, high-fidelity representation as Level 4 (L4), the most precise input format in this case. We then utilised Generative AI to define transformation criteria and generate corresponding rough data (L0–L3) using inverse reasoning. Examples of these transformed feature values at each level are illustrated in Tables 2 to 5.

L0 represents the most abstract, simulated qualitative input, focusing on behavioural and lifestyle proxies (e.g., sleep, diet, stress, and exercise) (Table 2). Each L0 feature was derived using composite logic based on clinical thresholds. For instance, Sleep_L0 was labelled as Poor if either the participant's BMI exceeded 30 or diastolic blood pressure surpassed 90 mmHg—both indicators of physiological strain; otherwise, it was labelled Good (note: only diastolic blood pressure is available in the PIMA Indians Diabetes dataset). Similarly, Diet_L0 was labelled as High Sugar if glucose or BMI levels exceeded metabolic risk thresholds, while Emotion_L0 and Exercise_L0 were derived from glucose, age, and diabetes pedigree indicators.

	Feature	**Mapping Rule**	**Categories**
L0	Sleep_L0	BMI ≥ 30 or BP > 90 → Poor, else Good	Good, Poor
	Diet_L0	Glucose ≥ 140 or BMI ≥ 30 → High Sugar, else Balanced	Balanced, High Sugar
	Emotion_L0	DPF > 1.0 or Glucose > 160 → Stressed, else Stable	Stable, Stressed
	Exercise_L0	Age ≥ 50 or BMI ≥ 30 → Inactive, else Active	Active, Inactive

Table 2. Mapping Rule from Rough to Precise - Focusing on behavioural and lifestyle proxies

L1 introduced a low-resolution biomedical interpretation, dichotomising core metabolic and demographic features using established clinical cutoffs (e.g., Glucose ≤140 mg/dL as Normal, otherwise High) (Table 3). This level maintains closer alignment with biomedical assessment norms while facilitating categorical model interpretation.

	Feature	**Mapping Rule**	**Categories**
L1	Pregnancies_L1	≤ 3 → Low, else High	Low, High
	Glucose_L1	≤ 140 → Normal, else High	Normal, High
	BloodPressure_L1	60–80 → Normal, else High	Normal, High
	SkinThickness_L1	≤ 35 → Normal, else Thick	Normal, Thick
	Insulin_L1	< 200 → Normal, else High	Normal, High
	BMI_L1	< 25 → Normal, else High	Normal, High
	DPF_L1	≤ 0.5 → Low, else High	Low, High
	Age_L1	< 50 → Young, else Senior	Young, Senior

Table 3. Mapping Rule from Rough to Precise - Low-resolution Biomedical Interpretation

L2 increased the resolution by subdividing the same features into three ordinal categories (Table 4), capturing gradations (e.g., Low, Normal, High or Young, Middle-aged, Senior), based on guidelines from endocrinology and public health literature. L3 further refined select numerical variables into four to five brackets, approximating the type of numeric yet discretised inputs a digital health platform may offer via drop-down interfaces (Table 5). These partitions reflect medically meaningful ranges (e.g., Glucose_L3 includes ranges such as <70, 70–89, 90–140, and >140 mg/dL). L4 retained the original continuous numerical values without any transformation, serving as the high-precision baseline for comparison (Table 6).

We employed a consistent modelling pipeline across all levels of granularity. A Random Forest-based classifier was selected due to its robustness to overfitting, ability to handle high-dimensional and mixed-type input features, and its inherent support for feature importance analysis. For each granularity level (L0–L4), the dataset was vectorised using one-hot encoding and split into training and test sets (80:20 ratio). We used accuracy, macro precision, macro recall, and macro F1 score as our evaluation matrix.

	Feature	**Mapping Rule**	**Categories**
L2	Pregnancies_L2	0 → 0, 1–3 → 1–3, else 4+	0, 1–3, 4+
	Glucose_L2	< 90 → Low, ≤140 → Normal, >140 → High	Low, Normal, High
	BloodPressure_L2	< 60 → Low, ≤80 → Normal, >80 → High	Low, Normal, High
	SkinThickness_L2	< 20 → <20, ≤35 → 20–35, >35 → >35	<20, 20–35, >35
	Insulin_L2	<100 → <100, ≤200 → 100–200, >200 → >200	<100, 100–200, >200
	BMI_L2	<18.5 → Underweight, <30 → Normal, ≥30 → Obese	Underweight, Normal, Obese
	DPF_L2	<0.5 → <0.5, ≤1.0 → 0.5–1.0, >1.0 → >1.0	<0.5, 0.5–1.0, >1.0
	Age_L2	<30 → Young, ≤50 → Middle-aged, >50 → Senior	Young, Middle-aged, Senior

Table 4. Mapping Rule from Rough to Precise - Medium-resolution Capturing Gradations

	Feature	**Mapping Rule**	**Categories**
L3	Glucose_L3	<70 → <70, <90 → 70–89, ≤140 → 90–140, >140 → >140	<70, 70–89, 90–140, >140
	BloodPressure_L3	<50 → <50, <70 → 50–69, <90 → 70–89, ≥90 → 90+	<50, 50–69, 70–89, 90+
	BMI_L3	<18.5 → Underweight, <25 → Normal, <30 → Overweight, ≥30 → Obese	Underweight, Normal, Overweight, Obese
	Age_L3	<20 → <20, <30 → 20–29, <40 → 30–39, <50 → 40–49, ≥50 → 50+	<20, 20–29, 30–39, 40–49, 50+

Table 5. Mapping Rule from Rough to Precise - Refined and Discretised such as in Digital Health Platforms

	Feature	**Mapping Rule**	**Categories**
L4	All Features	Original Value in dataset	Precise, Continuous values

Table 6. Mapping Rule from Rough to Precise – Original Continuous Numerical Values – High Precision Baseline

6 Evaluate

To evaluate the influence of input granularity on predictive performance, we trained decision tree classifiers using various combinations of features at different levels of granularity, ranging from single-level inputs (L0 through L4) to multi-level combinations (pairs, triplets, quadruplets, and the full set of all five levels). Each model was trained on 80% of the data and evaluated on the remaining 20%, with performance assessed using accuracy, macro-averaged precision, recall, and F1-score. Importantly, our goal is not to optimize model performance per se, but to evaluate whether a multi-level input framework

can support robust and interpretable prediction under varying input conditions. In this context, L0 features are not intended as construct-valid measures, but rather as rough-level approximations of lifestyle factors. Their comparable performance supports the idea that even approximate or heuristically defined features can contribute meaningfully to health prediction tasks, particularly in settings where precise data may be unavailable.

Combination	Acc.	Pre.	Rec.	F1
L0	0.734	0.748	0.652	0.657
L1	0.714	0.688	0.665	0.671
L2	0.714	0.691	0.697	0.694
L3	0.747	0.726	0.734	0.729
L4	0.747	0.725	0.730	0.727

Table 7. Test 1 – Prediction with 1 level granularity

Combination	Acc.	Pre.	Rec.	F1
L0+L1	0.708	0.680	0.676	0.579
L0+L2	0.727	0.707	0.715	0.638
L0+L3	0.740	0.718	0.721	0.643
L0+L4	0.747	0.724	0.726	0.649
L1+L2	0.714	0.693	0.701	0.621
L1+L3	0.688	0.662	0.665	0.571
L1+L4	0.747	0.724	0.718	0.636
L2+L3	0.734	0.710	0.712	0.631
L2+L4	0.747	0.724	0.726	0.649
L3+L4	0.779	0.761	0.772	0.707

Table 8. Test 2 – Prediction with 2 level Granularity Combination

Surprisingly, **fine-grained data does not always guarantee superior performance**. The results across all configurations were largely comparable. Models trained on coarse-grained inputs (e.g., L0 or L1) achieved performance metrics within a narrow margin of those trained on more detailed inputs (e.g., L3 or L4). Likewise, combining multiple levels of precision, whether two (e.g., L1+L4), three (e.g., L0+L1+L2), or all five levels (L0–L4), did not yield significant gains in predictive accuracy or macro-level F1 scores. Across all models, macro Accuracy plateaued around 0.70–0.78. This suggests a performance ceiling, possibly reflecting limitations of the underlying dataset or binary label definition. It also implies diminishing returns when stacking more levels of granularity after a certain point.

Combination	Acc.	Pre.	Rec.	F1
L0+L1+L2	0.727	0.705	0.711	0.708
L0+L1+L3	0.708	0.684	0.688	0.686
L0+L1+L4	0.747	0.724	0.726	0.725
L0+L2+L3	0.779	0.760	0.760	0.760
L0+L2+L4	0.760	0.738	0.732	0.735
L0+L3+L4	0.766	0.746	0.749	0.747
L1+L2+L3	0.740	0.717	0.717	0.717
L1+L2+L4	0.740	0.717	0.717	0.717
L1+L3+L4	0.740	0.718	0.721	0.719
L2+L3+L4	0.766	0.746	0.749	0.747

Table 9. Test 3 – Prediction with 3 level Granularity Combination

Combination	Acc.	Pre.	Rec.	F1
L1+L2+L3+L4	0.753	0.731	0.731	0.731
L0+L2+L3+L4	0.753	0.732	0.735	0.733
L0+L1+L3+L4	0.740	0.719	0.725	0.721
L0+L1+L2+L4	0.727	0.704	0.707	0.705
L0+L1+L2+L3	0.721	0.698	0.702	0.699

Table 10. Test 4 – Prediction with 4 Granularity Combination

Combination	Acc.	Pre.	Rec.	F1
L0 + L1 + L2 + L3 + L4	0.734	0.713	0.720	0.716

Table 11. Test 5 – Prediction with 5 Granularity Combination

7 Reflect

Reflection was something that happened during every phase of the DSR as well as at the end every phase of DSR. This reflection allowed the guided emergence of new knowledge leading to the creation of the various artefacts of this research (a) identification of problems and issues (b) emergence of research questions (c) proposal of the Measure, Model, Manage framework and (d) the implementation of the AI-driven granularity adaptation framework. The results of evaluating these frameworks reveal that diabetes risk can be accurately predicted across all levels of data granularity (L0–L4), with model performance remaining relatively stable regardless of whether the input features were coarse (e.g., L0) or highly precise (e.g., L4). This challenges the prevailing belief that fine-grained, numerical input is essential for reliable prediction, and instead highlights the robustness of AI models in extracting meaningful patterns even from abstracted or qualitative representations. Notably, some hybrid combinations of granularity levels achieved strong performance, suggesting complementary information across levels. However, a few dual-level combinations performed poorly, indicating the importance of careful feature design and semantic alignment when integrating multi-granular data. These findings support the development of human-centered, adaptive decision-support systems that allow for flexible data input without compromising accuracy, ultimately enhancing inclusivity, interpretability, and user well-being.

8 Limitations and Future Work

This study has some limitations that should be acknowledged. First, the model was developed and validated using the PIMA Indian Diabetes dataset, which, although widely used, has inherent constraints. The dataset is relatively small, contains only female participants, and includes a limited number of clinical attributes (e.g., glucose, BMI, age, blood pressure). As a result, the generalizability of the model across broader and more diverse populations has not been fully validated. Additionally, the use of a Decision Tree classifier, which, although interpretable and ease of implementation to align with the study's focus on validating rough versus precise data, is prone to overfitting and may not capture complex non-linear relationships among features.

Future work will focus on addressing these limitations by incorporating larger and more diverse datasets to improve model robustness and applicability. Another key direction will involve collecting lifestyle and self-reported behavioural data and integrating that with clinical indicators to create a more holistic and personalised predictive model. In addition to the classifiers, future research should explore deep learning models and large language models, which may offer superior predictive accuracy and better handle complex feature interactions.

9 Conclusion

Diagnosis and management of chronic diseases such as diabetes are multifaceted problems. One of the key challenges herein is the lack of data to help with early diagnosis, ongoing management, and potential reversal and transformation of the individual. We do have some precise data on individuals' health stored in health platforms, but not all the attributes are relevant, and the non-relevant attributes are

captured by these platforms. So, we have both a problem of missing data as well as irrelevant data. However, the proliferation of smart devices is helping us to capture some of the relevant data regarding aspects such as exercise and sleep in a reasonably precise fashion. However, other aspects related to lifestyle choices and behaviours, such as diet and emotions, are not captured at all or can only be captured in a rough and approximate fashion.

Hence, the key question we asked at the beginning of our research journey was whether we could use rough and precise data of individuals in a complementary fashion to diagnose and manage chronic diseases and use these to transform the individual.

Our Measure, Model, Manage framework (section 4) at the abstract level aligns with existing research. However at the detail level it contradicts and thereby enhances current approaches which focus on precise data to measure, model, and manage chronic diseases such as diabetes. Our approach takes a holistic view of the individual and allows for the inclusion of rough and precise data to model and manage their chronic diseases. Furthemore we designed and implemented an GenAI-driven granularity adaptation framework (section 5) that helps us to work with different granularities of rough and precise data using inverse reasoning.

Our study demonstrates that diabetes risk can indeed be accurately predicted using not only high-precision biometric inputs (L4) but also rough, qualitative data (L0) guided by the Measure, Model, Manage framework, and their various combinations. Through systematic experimentation across single and multi-level granularity configurations using a typical machine learning model - Random Forest, we found that combinations involving both coarse and fine-grained features often yield comparable performance to purely precise inputs. This suggests that user-friendly, low-burden input formats can be viable for health risk prediction without sacrificing accuracy. Our contributions include a rule-based granularity mapping framework, an extensible evaluation pipeline across levels, and empirical evidence supporting hybrid input strategies. Future research may explore dynamic personalisation of granularity based on user preference or context, integration with longitudinal tracking data, and real-world deployment to assess psychological and behavioural outcomes over time.

These findings suggest potential future directions for our research (a) build a holistic profile of an individual using rough and precise data (b) use AI systems to approximate rough data and transform data into more useable forms both rough and precise through reasoning (c) build descriptive, predictive, and prescriptive apps and systems using rough and precise data to diagnose, manage, and transform individuals and (d) finally evaluate these concepts, processes, and systems in the real world.

10 References

Abdel-Rahman, N., Manor, O., Valinsky, L., Mosenzon, O., Calderon-Margalit, R., & Roberman, S. (2022). What is important for people with type 2 diabetes? A focus group study to identify relevant aspects for Patient-Reported Outcome Measures in diabetes care. *PLOS ONE*, 17(11), e0277424. https://doi.org/10.1371/journal.pone.0277424

Bloom, D. E., Chen, S., Kuhn, M., McGovern, M. E., Oxley, L., & Prettner, K. (2020). The economic burden of chronic diseases: Estimates and projections for China, Japan, and South Korea. *The Journal of the Economics of Ageing*, 17, 100163.

Brendel, A. B., Lembcke, T. B., & Kolbe, L. M. (2022). Towards an integrative view on design science research genres, strategies, and pivotal concepts in information systems research. ACM SIGMIS Database: *The DATABASE for Advances in Information Systems*, 53(4), 9-23.

Britton, A., Brunner, E., Kivimaki, M., & Shipley, M. J. (2012). Limitations to functioning and independent living after the onset of coronary heart disease: what is the role of lifestyle factors and obesity?. *The European Journal of Public Health*, 22(6), 831-835.

Chien, B. C., Hu, C. H., & Ju, M. Y. (2009). Learning fuzzy concept hierarchy and measurement with node labeling. *Information Systems Frontiers*, 11, 551-559.

Gil-Herrera, E., Aden-Buie, G., Yalcin, A., Tsalatsanis, A., Barnes, L. E., & Djulbegovic, B. (2015). Rough set theory based prognostic classification models for hospice referral. *BMC Medical Informatics and Decision Making*, 15, 1-18.

Grande-Ramírez, J. R., Meza-Palacios, R., Lasserre, A. A. A., Flores-Asis, R., & Vázquez-Rodríguez, C. F. Intelligent fuzzy system to assess the risk of type 2 diabetes and diagnosis in marginalized regions. *IAES International Journal of Artificial Intelligence*, 13, 1935-1944.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.

K Shahade, A., & V Deshmukh, P. (2024). Enhancing Diabetes Prediction Using Ensemble Machine Learning Model. *International Journal of Computing and Digital Systems*, 17(1), 1-13.

Maass, W., & Storey, V. C. (2021). Pairing conceptual modeling with machine learning. *Data & Knowledge Engineering*, 134, 101909.

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. (2014). A machine learning approach to improving dynamic decision making. *Information Systems Research*, 25(2), 239-263.

Magliano, D. J., & Boyko, E. J.; IDF Diabetes Atlas 10th edition scientific committee. (2021). Global picture [10th ed.] In *IDF Diabetes Atlas* (10th ed.). *International Diabetes Federation*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK581940/

Mosa, A. S. M., Thongmotai, C., Islam, H., Paul, T., Hossain, K. T., & Mandhadi, V. (2022). Evaluation of machine learning applications using real-world EHR data for predicting diabetes-related long-term complications. *Journal of Business Analytics*, 5(2), 141-151.

Nodoust, A., & Ghatari, A. R. (2024). Assessment of a Hybrid Machine Learning Algorithm in Healthcare Management for Predicting Diabetes Disease. *International Journal of Supply and Operations Management*, 11(4), 462-482.

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems Development in Information Systems Research. *Journal of Management Information Systems*, 7(3), 89–106.

Pawlak, Z. (2002). Rough set theory and its applications. *Journal of Telecommunications and Information Technology*, 7-10.

Pekel Özmen, E., & Özcan, T. (2020). Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm. *Journal of Forecasting*, 39(4), 661-670.

Pima Indians Diabetes Dataset. (2019). *Pima Indians Diabetes Dataset* (PIDD) documentation [Dataset]. Kaggle. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

Piri, S., Delen, D., Liu, T., & Zolbanin, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decision Support Systems*, 101, 12-27.

Rawat, V., & Suryakant, S. (2019). A classification system for diabetic patients with machine learning techniques. *International Journal of Mathematical, Engineering and Management Sciences*, 4(3), 729-744.

Rosman, L., Gehi, A., & Lampert, R. (2020). When smartwatches contribute to health anxiety in patients with atrial fibrillation. *Cardiovascular Digital Health Journal*, 1(1), 9-10.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, 35(1), 37-56.

Serrano, L. P., Maita, K. C., Avila, F. R., Torres-Guzman, R. A., Garcia, J. P., Eldaly, A. S., ... & Forte, A. J. (2023). Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. *The Permanente Journal*, 27(4), 100.

Svedbo Engström, M., Leksell, J., Johansson, U.-B., & Gudbjörnsdottir, S. (2016). What is important for you? A qualitative interview study of living with diabetes and experiences of diabetes care to establish a basis for a tailored Patient-Reported Outcome Measure for the Swedish National Diabetes Register. *BMJ Open*, 6(3), e010249. https://doi.org/10.1136/bmjopen-2015-010249

World Health Organization. (2023). *Noncommunicable diseases*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

Ye, M., Wu, X., Hu, X., & Hu, D. (2013). Anonymizing classification data using rough set theory. *Knowledge-Based Systems*, 43, 82-94.

Zhou, W., Liu, X., Bai, H., & He, L. (2024). Intelligent medical diagnosis and treatment for diabetes with deep convolutional fuzzy neural networks. *Information Sciences*, 677, 120802.

Copyright