Journal of Informatics and Web Engineering

Vol. 5 No. 2 (June 2026) eISSN: 2821-370X

# AI-Powered Continuous Health Monitoring for Chronic Disease Management in Malaysia: Real-Time Risk Prediction Using Wearable Sensors and Digital Biomarkers

**Umi Najiah Ahmad Razimi[1*], Yong Yoke Leng[2**], Nor Hapiza Mohd Ariffin[3] , Mohammed Hazim Alkawaz[4]**

[1,2]Faculty of Computing and Digital Technology, HELP University, Persiaran Cakerawala. Subang Bestari, Seksyen U4 40150 Shah Alam, Selangor, Malaysia

[3]MIS Department, Faculty of Business, Sohar University, 3111 Al Jamiah Street, Sohar 311, Oman

[4]Department of Computer Science, University of Mosul, Al Majmooah, 41002, Mosul, Nineveh, Iraq

**corresponding author: (yokeleng.y@help.edu.my; ORCiD: 0000-0001-8976-5787)*

***corresponding author: (uminajiah.ar@help.edu.my; ORCiD: 0000-0002-1567-4011)*

*Abstract* – Chronic Non-Communicable Diseases (NCDs), such as diabetes mellitus, hypertension, and cardiovascular disease, make up a significant portion of the mortality cases in Malaysia, as well as impose consistent demands for costs related to health care expenditure. According to the cost reports at the national level, major NCDs (cardiovascular disease, diabetes, and cancer) are responsible for RM 9.65 billion annually as direct healthcare cost [3]. However, the current approach to managing chronic diseases focuses heavily on periodic assessments in clinic setting at weekly or monthly intervals. Here, a predictive analysis framework based on AI technology, leveraging continuous monitoring with patient data collected from consumer wearables to predict potential health events up to five days ahead is proposed. A synthetic dataset was generated to simulate the behaviour of 500 patients affected by diabetes and/or hypertension, who were continuously monitored throughout a year for heart rate, heart rate variability, physical activity, sleep indices, and blood pressure features. The models are compared to determine the optimal algorithm in the case study. The LSTM model showed 93.7% accuracy in predicting five days in advance, outperforming the random forest model (91.5%), XGBoost model (90.6%), and the simple threshold rule-based approach (78.6%). The sensitivity and specificity scores of the LSTM model were 87.6% and 95.9%, respectively, while the ROC-AUC score was 0.983. According to the SHAP analysis, previous heart rate values (history for three days and seven days) have the largest contributions in predicting patient outcomes, followed by changes in heart rate variability and low activity levels. Although evaluation is based on synthetic data for methodological validation, the framework is intended for extension to real-world wearable streams to support earlier risk detection and more proactive chronic disease management.

*Keywords— Wearable Sensor, Chronic Disease Management, Digital Biomarker, Long Short-Term Memory, Predictive Analytics, Precision Health*

*Received: 5 February 2026; Accepted: 5 April 2026; Published: 16 June 2026*



Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.2.22

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

# 1. INTRODUCTION

## *1.1 The Chronic Disease Crisis in Malaysia*

Non-Communicable Diseases (NCD) currently stand as the primary health challenge which affects people in Malaysia for an extended period. The national reports demonstrate that NCDs which accounted for 74% of total fatalities in 2016 NCD fatalities showed cardiovascular disease as the most common cause of death [1]. Diabetics constitute a significant portion of the population with an estimated 18.3% of Malaysian adults having diabetes while half of them remain undetected [2]. The condition of hypertension affects approximately 6.4 million adults which represents approximately 30% of the population because it serves as a major cause of stroke heart failure and chronic kidney disease [1].

The economic effect caused by this trend directly influences the economic situation. According to the estimates made by the Ministry of Health, the yearly financial cost of major NCDs in the form of healthcare expenses equals RM 9.65 billion [3]. However, the total cost related to the issue equals RM 64.2 billion, with RM 51.8 billion attributed to productivity losses [4]. The mentioned problem creates negative impacts on the health budget, and at the same time, the consequences affect the economy in general.

The problem is likely to escalate in its severity. The information provided by the Department of Statistics Malaysia shows that the population aged 60 and older will comprise 11.6% of the Malaysian population in 2024, and later in 2040, their percentage will be 17.3% [5]. National statistics proves that 25.1% of adults in the country do not exercise, which means that behavioral problems can be noticed [2]. However, it should be stated that some gaps are observed when analyzing different groups of people. For example, higher prevalence levels have been noticed among Indians (about 25.1%), whereas among the Malaysians, the figures are around 15.25%, and the figure is 12.87% among Chinese [6]. This information helps design effective prevention and monitoring systems for the disease.

## *1.2 Limitations of Traditional Chronic Disease Management*

Patients with chronic illnesses require outpatient follow-up appointments which occur every few weeks to several months. Malaysia's clinical practice guideline for type 2 diabetes recommends reassessing glycaemic control (HbA1c) after 3–6 months following treatment initiation or adjustment, with earlier review when targets are not achieved [7]. The actual process of making decisions relies on the limited data from clinic readings which doctors collect throughout the year because they cannot see any significant deterioration that happens before the next appointment.

A second limitation exists because short patient appointments require doctors to depend on what patients remember. The self-reported medication use adherence of patients shows different results from monitored medication adherence [8] and research guidelines recommend against using self-reported data as the only method to measure patient adherence to treatment [9]. Patients who experience health deterioration recognize their condition too late to receive emergency treatment, which results in them needing expensive medical care for serious complications. National NCD reporting shows that direct cost burden increases when patients receive major complications and hospital admissions [3], [4].

The healthcare system faces double challenges because of restricted workforce availability and limited access to services. The evidence from Malaysia demonstrates that medical specialists in MOH hospitals distribute themselves unevenly across different geographical areas, which results in multiple areas having unequal access to specialist medical services [10]. The Malaysian data shows that primary care systems experience two problems because of short consultation times and high patient volume, which limits their ability to conduct assessments and counseling during each appointment [11]. The access barriers create an additional problem because Malaysian research shows that specific populations face limitations from distance and travel time and service capacity limits which according to spatial accessibility studies create access disparities that match the existing urban-rural access gaps [12], [13].

## *1.3 The Wearable Health Technology Revolution*

Wearables have become a common technology used by people today. According to IDC market projections the worldwide shipment of wearable devices will reach 537.9 million units in 2024 [14]. In Malaysia, people adopt technology according to their age and the specific situation; one study about young adults found that one-fifth of participants owned or used smartwatches [15] which shows that smartwatches reached significant market presence in certain demographic groups.

Most consumer devices depend on a limited number of sensors, yet these sensors produce multiple signals which have clinical significance. Photoplethysmography (PPG) serves as a standard method for heart rate estimation while inter-beat intervals enable HRV-type features which depend on the specific device design and signal quality [16]. Accelerometers enable users to track their activities through step counting and intensity measurement, which consumer trackers use for their sleep-wake detection system [17]. Some products also advertise cuffless blood pressure estimation, but validation and calibration remain active challenges and performance can vary across users and contexts [18].

Continuous sensing enables the detection of new phenomena which were previously unobservable. The wearables provide continuous time tracking through their minute-level data collection which produces daily activity and sleep summaries. Home-based health monitoring can decrease clinic-based measurement errors which include white coat hypertension [19]. Wearables have been tested across entire populations through their use in the Apple Heart Study which showed that irregular pulse notifications predicted atrial fibrillation with a positive predictive value of 0.84 based on ECG results [20]. Researchers have assessed wearable device capabilities to diagnose infectious diseases through actual COVID-19 detection which they achieved by merging symptom tracking with their technology [21].

### *1.4 Artificial Intelligence for Predictive Health Monitoring*

Wearables produce multivariate time series data. Simple thresholds can work for a few signals, but they break when patterns involve multiple variables and delayed effects. The standard approach for this task uses sequence models. The gated structure of LSTM networks enables users to control which information they want to keep and which information they want to lose over time [22]. This matches physiological data where risk builds gradually rather than flipping instantly.

In healthcare, ML systems have been used in healthcare to perform various related tasks. The field of healthcare research uses CGM-type data for short-horizon glucose prediction [23], ICU settings to develop early warning models for patient deterioration which includes sepsis prediction based on routinely collected variables [24], and deep neural networks to detect and classify arrhythmias from ECG data [25]. The input characteristics of these examples match each other because they involve noisy physiological signals which have time dependencies and produce clinically relevant lead times.

Interpretability serves as an essential requirement for practical implementations. The prediction system must provide its underlying reasons to clinicians who want to understand results for specific patient instances. SHAP delivers local feature attributions through Shapley-value foundation [26] implementation. The system demonstrates its ability to display recent trend contributions which resulted in high-risk prediction through specific deployment functionality.

### *1.5 Research Objectives*

The worldwide advancements in wearable health monitoring systems show progress yet Malaysia research in this area remains restricted because studies focus on local service viability and user satisfaction research instead of complete monitoring system implementation [27], [28]. The research study establishes five distinct goals to fill existing research deficiencies: (i) Integrate multi-modal wearable data (heart rate, HRV, activity, sleep, blood pressure) into comprehensive risk profiles, (ii) Train and compare LSTM, Random Forest, and XGBoost models for 5-day advance event prediction, (iii) Employ SHAP analysis to identify most predictive digital biomarkers, (iv) Design operational framework for Ministry of Health deployment, and (v) Quantify cost-effectiveness through prevented hospitalization analysis.

## 2. BACKGROUND AND RELATED WORK

### *2.1 Digital Biomarkers and Physiological Monitoring*

Heart Rate Variability (HRV) as Autonomic Function Biomarker: HRV measures the time intervals between consecutive heartbeats. The method serves as a non-invasive practical tool which enables evaluation of autonomic system functions through its ability to show how sympathetic and parasympathetic system activities change throughout time [29]. The time-domain measurement Standard Deviation of Normal-to-Normal (SDNN) intervals functions as a conventional time-domain measurement because it provides a single value which shows total variability throughout the entire measurement period.

The medical community uses low HRV as an indicator which shows that people experience decreased autonomic system control. The relationship between lower HRV and increased cardiovascular risks has been established through research studies which examined various population groups. The Framingham Heart Study showed that participants in the lowest HRV quartile faced a significantly greater risk of cardiovascular death compared to individuals who exhibited higher variability levels [30].

In the case of a diabetic patient, the long-term effects of high blood glucose levels on the body affect the nerves responsible for regulating heart rate and vessel function. The condition known as diabetic cardiac autonomic neuropathy becomes very clinically relevant as it increases the chances of severe cardiovascular events that may happen suddenly without any symptoms of prior complications. One crucial fact is that there is a reduction in HRV, many years before the first symptoms arise [31]. HRV analysis has become a tool that allows for the early detection of cardiovascular issues; in turn, monitoring devices based on PPG technology make it possible for a person to monitor his/her heart activity continuously.

Resting Heart Rate (RHR) Dynamics: RHR measurement during periods of complete rest without physical activity provides additional information about a person's cardiovascular system health and their body response to stress. The actual value brings useful insight when it shows consistent movement away from a person's normal resting state. Extended periods of elevation which last from multiple days to several weeks have been linked to two medical conditions and three health outcomes which include acute illness and heart failure progression and medication side effects [32]. The COVID-19 pandemic showed that wearable devices could detect baseline changes in users which happened before they showed any indications of symptoms [21]. Individual differences in baseline measurement create a situation where people experience distinct variations of what constitutes "normal" behaviour.

Physical Activity and Sedentary Behaviour: Wearables can give easy activity proxies in the form of steps and activity level. In larger cohorts based on accelerometer data, those with more daily steps have a reduced risk of mortality with increasing benefits but at a declining rate as the activity increases [33]. Another matter is the matter of sedentary activity. Prolonged periods of being sedentary increase the risk of heart disease even in people meeting the physical activity guidelines, meaning regular movement might confer extra protection [34].

Sleep Quality and Architecture: The amount and quality of sleep have significant impacts on cardiometabolic health. Short-sleeping patterns have previously been shown to be related to impaired glucose metabolism, hypertension, and neuro-hormonal alterations that result in obesity and insulin resistance [35]. Current consumer sleep trackers use accelerometery and heart rate information to estimate sleep measures. According to validation studies against polysomnography, consumer-grade sleep trackers may achieve reasonable accuracy for basic sleep-wake determination but struggle in more accurate staging tasks [36].

Blood Pressure Variability: The preferred method for taking blood pressure in home settings compared to clinic-based measures is that of reducing situational influences such as white-coat hypertension and measuring variability in a more realistic fashion [37]. It should be noted that variability per se is meaningful, as short-term blood pressure variability and variability from one visit to another have been linked to strokes and cardiovascular disease regardless of mean blood pressure values [38].

### *2.2 Machine Learning Architectures for Health Prediction*

Long Short-Term Memory Networks (LSTM): LSTMs aim to deal with temporal structures of high length, tackling the vanishing gradient problem which affects conventional RNNs whenever dependencies are present over a large range of temporal distance [22]. This is achieved using an internal cell state that is updated through the use of three gates, the forget gate which regulates what previous information is needed to be kept, the input gate which determines how new information should be stored in memory, and the output gate which specifies how much memory is shared with the next layer.

In medical time-series forecasting, this model architecture can prove useful since biological variations often tend to develop gradually rather than instantaneously, and LSTM models have found applications in CGM short-horizon forecasting, showing that a considerable portion of predictions up to 30-60 minutes ahead are within clinically acceptable error intervals [39]. On similar grounds, LSTM models for ICU monitoring have shown that sequences of vital signs can be used to predict sepsis, exhibiting AUROC values over 0.85 during real-time scenarios [40].

Random Forest and Gradient Boosting: Random Forest algorithms create decision tree ensembles using bootstrap aggregation, whereby decision trees train on randomly chosen data sets and subsets of the features for splitting nodes [41]. It intrinsically supports heterogeneous data types, manages missing data, and estimates feature importance based on the increase in out-of-bag error rates. Gradient Boosting builds decision trees successively, where each successive tree corrects mistakes made by the previous ones [42]. XGBoost is a fast implementation of gradient boosting, which uses regularization to avoid overfitting and class imbalances through sampling weights.

Personalized Baseline Modelling: Population-based baselines fail to take into account large variations between individuals because of genetic factors, physical condition, weight distribution, and drug intake. Personal baselines use initial observation periods of about 14-30 days to estimate individual baselines before computing features as deviations from those baselines [43]. Research has shown that personalized modelling enhances specificity without sacrificing sensitivity [21].

### *2.3 Explainable AI and Clinical Adoption*

Use of machine learning in health care traditionally has encountered opposition due to the “black box” nature of the technology. Most predictive models lack an explanation component, violating the principles of evidence-based medicine, which demand clear explanations [44]. The SHAP method makes use of notions from cooperative game theory to apportion credit for predictions among all relevant variables [26]. The SHAP values for any given prediction reveal the extent to which individual features contribute to the final prediction above and beyond the baseline value. The Tree Explainer algorithm computes SHAP values in tree-based machine learning models by calculating Shapley values in polynomial time [45]

### *2.4 Malaysian Context and Research Gaps*

There is relatively little published research involving Malaysian applications of wearable technology or machine learning for monitoring diseases. Bayesian networks for dengue prediction in Klang Valley were created by Raja et al. (2019), with an accuracy rate of 79-84% [46]. Moreover, Salim et al. (2021) used Random Forest for dengue prediction in Selangor based on climate-related features, reaching an accuracy rate of 68% [47]. It should be noted that none of the papers mentioned consider the continuous monitoring of diabetes and hypertension.

Some of the important limitations are: (1) the restricted geographical scope to individual cities instead of a nationwide study, (2) the limited set of features without considering biomarkers from wearables, (3) the lack of any analysis of model interpretability to promote its clinical use, (4) the absence of any operational plan of implementation at the Ministry of Health, (5) no consideration of health economics, (6) independent approaches for diabetes and hypertension management despite the 28% prevalence rate of both diseases' comorbidity, and (7) no validation for specific races [6].

### *2.5 Contemporary AI in Medicine (2025) and Alignment of This Work*

The latest research in 2025 indicates the continuing progress in using medical AI applications in various data sources and healthcare settings. According to Baker et al., there is a ResNet50 architecture model used for detecting melanomas in New Zealand, proving that deep learning pipelines still play a crucial role in applications aimed at the screening purpose [48]. Another recent development, according to Baker et al., concerns a predictive model for pandemics' forecasting in New Zealand and partner countries, showing that the field of risk forecasting still interests many scholars [49]. Moreover, Baker et al.'s study conducted in 2025 and presented at the IEEE conference in Kota Kinabalu, Malaysia, involves a framework for predicting diseases and its application issues [50].

This study aligns with these directions by focusing on continuous, longitudinal prediction rather than single-visit measurements. The proposed framework uses multimodal wearable-derived signals and engineered temporal features to estimate five-day-ahead risk, and it includes SHAP-based interpretation to support clinical review. Although the present evaluation is based on synthetic data to validate the modelling and analytics pipeline, the workflow is structured so that real wearable streams can be substituted at the data ingestion stage without changing downstream feature engineering, training, and evaluation.

## 3. METHODS

### *3.1 Study Design and Patient Cohort*

A synthetic wearable dataset was generated for 500 patients with diabetes and/or hypertension, tracked over 365 days (totalling 182,500 patient-days). The cohort includes 198 diabetes-only patients (39.6%), 160 hypertension-only (32.0%), and 142 with both conditions (28.4%). Ages were set between 40 and 80 years (mean 59.8, SD 11.9) to reflect the age range in which both conditions are most prevalent. Sex distribution was set at 55% female and 45% male. Residence was set at 60% urban, 25% suburban, and 15% rural, broadly informed by Malaysia's urban–rural population structure but intentionally oversampling non-urban groups to support equity-related analyses [51].

Synthetic data were used for four reasons. First, it avoids patient privacy and governance constraints during early-stage development. Second, the deterioration patterns and noise levels can be controlled, which makes model behaviour easier to test. Third, the dataset can be regenerated for reproducibility. Finally, the same pipeline can be applied to real wearable streams later by replacing the synthetic generator with real data ingestion, without changing the modelling workflow.

### *3.2 Wearable Sensor Data Streams and Digital Biomarkers*

Table 1 summarises the key signals captured from the wearable monitoring setup (steps, sleep, and optional home blood pressure), together with reference benchmarks used to interpret the cohort's baseline status. It also states the operational definition of an adverse health event and the resulting event frequency, providing context for the prediction task that follows.

### *3.3 Feature Engineering for Temporal Pattern Recognition*

Table 2 summarises the 33 temporal and derived features engineered from the raw vitals to capture multi-day deterioration patterns across five categories: lag features, rolling-window statistics, trend-based features, derived biomarker ratios, and personalised deviations from baseline.

Blood pressure data were available for only 30% of patients, creating systematic missingness. Missing values were imputed using the cohort median (137.1 mmHg for SBP), while a binary indicator variable allowed models to learn differential reliance on BP when available. After feature engineering, the dataset comprised 50,336 patient-days with an event rate 26.5%.

Table 1. Summary of Wearable Sensor Domain, Measurement Protocols, and Cohort Characteristics

| Domain | Sensor / Modality | Measurement Definition | Baseline / Thresholds | Cohort Statistics / Rationale |
|---|---|---|---|---|
| **Heart Rate Monitoring** | PPG | RHR is measured daily in the morning upon waking to reflect baseline cardiovascular demand. | Individual baseline computed from the first 30 days (mean ± SD per patient) | Mean baseline RHR: **72.5 ± 8.5 bpm**, approximating healthy adult norms with realistic inter-individual variation |
| **HRV** | PPG-derived SDNN | SDNN computed from 5-minute recordings during quiet rest following standard protocols [29] | Deterioration thresholds: **<40 ms** (reduced HRV), **<30 ms** (severely reduced; associated with elevated cardiovascular risk)[30] | Mean baseline HRV: **45.5 ± 13.9 ms**; enables tracking gradual autonomic decline over weeks to months |
| **Physical Activity** | 3-axis accelerometer | Daily step count and sedentary time estimation | Activity categories: sedentary (<5,000), low (5,000–7,500), moderate (7,500–10,000), active (>10,000), based on commonly used pedometer indices [52] | Mean daily steps: 6,692 ± 2,101, indicating generally low activity in the synthetic cohort. |
| **Sleep Monitoring** | Accelerometer-based sleep tracking | Total sleep time, sleep efficiency, and nocturnal disruption frequency | Healthy benchmarks: 7–9 h sleep, ≥85% efficiency, 2–3 awakenings/night | Mean sleep: **6.4 ± 1.2 h**; efficiency **78.7 ± 9.1%**; disruptions **4.5 ± 1.7/night**, indicating poor sleep quality |
| **Blood Pressure Measurement** | Smart BP monitors (subset) | Morning and evening systolic/diastolic readings capturing circadian variation | Optimal SBP <120 mmHg | 30% adoption reflects real-world usage; mean SBP **137.1 ± 19.4 mmHg**, typical of treated hypertensive patients |
| **Adverse Health Event Definition** | Clinical criteria | High-risk deterioration requiring medical intervention | Diabetes: severe hypo/hyperglycaemia, DKA; Hypertension: SBP ≥180 or DBP ≥120; CVD: MI, HF exacerbation, severe arrhythmia | Events occurred in **27.2% of patient-days**; a 5-day horizon provides actionable lead time for outpatient intervention |

Table 2. Feature Engineering Strategy: Temporal and Derived Features

| Feature Category | Number of Features | Variables Involved | Temporal Scope | Description / Rationale |
|---|---|---|---|---|
| **Lagged Features** | 9 | RHR, HRV, and daily steps | 1-day, 3-day, and 7-day lags | Historical values enable the model to distinguish sustained deterioration (e.g. elevated heart rate persisting for a week) from transient stress caused by single-day spikes. |
| **Rolling Window Statistics** | 12 | RHR, HRV, steps, sleep duration | 7-day and 14-day moving windows | Moving averages reduce day-to-day noise while preserving underlying trends. Comparison between 7-day and 14-day windows identifies accelerating deterioration versus stable chronic dysfunction. |
| **Temporal Trends** | 2 | RHR, HRV | 7-day rolling window | Linear regression slopes capture directional change over time. Increasing RHR slopes and decreasing HRV slopes indicate worsening physiological status, with slope magnitude quantifying deterioration speed. |
| **Derived Biomarkers** | 5 | RHR, HRV, steps, sleep | Multi-window derived metrics | Physiologically motivated composite indicators: HRV decline rate, activity decline percentage, sleep disruption increase, HR-to-activity ratio, and a composite stress score (0–4) reflecting cumulative physiological stress. |
| **Personalised Deviation Features** | 5 | RHR, HRV, steps, sleep, blood pressure | Deviation from 30-day baseline | Differences between current values and individual baseline capture personalised risk, detecting deviations relative to a patient's own norm rather than population-level thresholds. |

### *3.4 Machine Learning Model Architectures*

Table 3 provides a consolidated view of the predictive approaches evaluated, including the input representation used for each model, the main architectural choices or structure, and the training decision strategy. Presenting these configurations side-by-side clarifies what differs between models (sequence learning vs engineered features vs rules) and supports fair comparison under a consistent experimental protocol.

Table 3. Summary of Predictive Models, Input Representations, and Training Configurations

| Model | Input / Data Representation | Architecture / Structure | Key Hyperparameters | Training / Decision Strategy | Notes |
|---|---|---|---|---|---|
| **LSTM Network** | 14-day sequences; input shape: *(batch_size, 14 timesteps, 33 features)* | LSTM (96 units, return_sequences=True) → Dropout (0.4) → LSTM (48 units) → Dropout (0.3) → Dense (24 units, ReLU) → Output (1 unit, sigmoid) | Learning rate: 0.001 | Adam optimiser, binary cross-entropy loss, batch size 64, early stopping (patience = 12 epochs) | Designed to capture temporal dependencies while balancing model capacity and overfitting risk |
| **Random Forest Classifier** | 33 features with rolling statistics and lagged values | Ensemble of decision trees | 150 trees, max depth = 12, min samples split = 30, min samples leaf = 15 | Balanced class weights to address 27% event rate imbalance | Temporal patterns encoded implicitly via engineered features |
| **Gradient Boosting (XGBoost)** | 33 features with rolling statistics and lagged values | Sequentially boosted decision trees | 150 estimators, max depth = 6, learning rate = 0.08, min samples split = 30 | Gradient boosting optimisation | Handles missing values and class imbalance effectively |
| **Threshold-Based Baseline** | Current-day vital signs only | Rule-based decision system | Thresholds: RHR > 90 bpm, HRV < 35 ms, steps < 4,000, sleep < 5.5 hours | High-risk alert triggered if ≥ 2 of 4 criteria are met | No temporal modelling; serves as a clinical baseline comparator |

### *3.5 Model Training Protocol and Evaluation Metrics*

Using patient-wise splitting means that the time-based information will be kept, and there is no leakage of data. In this case, we will have the following split for 500 patients according to the following ratio: 350 (70%) for training, 75 (15%) for validation, and 75 (15%) for test evaluation. All data for 365 days of one patient goes into the corresponding set. After removing the first 14 days to extract the feature data, we end up with 35,200, 7,392, and 7,744 data points for training, validation, and test correspondingly. Event rates are relatively consistent: training 26.6%, validation 25.8%, test 27.1%.

Feature scaling (z-score normalization) was performed using statistics from the training set only and applied to both the validation and testing datasets. The performance measures considered for evaluation were accuracy, sensitivity, specificity, precision, F1 score, and ROC-AUC. These were targeted as follows: accuracy >85%, sensitivity >80% (detect at least 80% of events), specificity >85% (limit false alarms), F1-score >0.75, and ROC-AUC >0.90.

*3.6 Explainability Analysis via SHAP*

SHAP value decomposes the prediction output as the sum of each feature's contributions [26]. In our experiments with the Random Forest model, we applied the TreeExplainer method to calculate SHAP values efficiently and accurately for tree-based ensembles [45]. To rank features according to their importance, we analysed the average absolute SHAP values of 1,000 test samples. A high value of |SHAP| for a certain feature indicates that it is likely to be a crucial biomarker. Individual explanations quantify each feature's contribution to specific predictions, enabling personalised clinical guidance.

*3.7 Synthetic Data Generation Methodology*

Instead of generating discrete event days randomly, continuous patient-specific health risk scores evolving smoothly over time were created. A baseline risk was determined by disease severity for each patient as follows:

- Comorbid diabetes and hypertension patients - 0.35,
- Hypertension-only patients - 0.25,
- Diabetes-only patients - 0.30.

This base risk was then modulated by sinusoidal functions with different frequencies representing physiological cycles. Total daily risk = base_risk + sum_of_sinusoids, clipped to [0,1] range.

Each physiological variable degraded proportionally to risk score, creating biological coherence:

- RHR = baseline_rhr + risk × 20,
- HRV = baseline_hrv × (1 - risk × 0.5),
- steps = baseline_steps × (1 - risk × 0.4),
- sleep = baseline_sleep - risk × 2.0,
- blood pressure = baseline_bp + risk × 35.

Risks above 0.48 were defined as the risk threshold, tuned to achieve an event rate of approximately 27%. This approach guarantees realistic temporal patterns whereby vitals gradually worsen as risk increases. With the pattern, a detectable multi-day deterioration trajectory can be created for effective LSTM learning.

*3.8 Implementation Platform*

The system was implemented using Python 3.9+ with standard machine learning libraries, namely:

- TensorFlow 2.13+ (LSTM)
- scikit-learn 1.3+ (Random Forest, XGBoost)
- SHAP 0.43+
- Pandas
- NumPy
- matplotlib
- seaborn

A standard laptop CPU with 8GB+ RAM is required to run the system, with a runtime of approximately 15-20 minutes. The use of a fixed random seed ensures reproducibility across runs. The complete code is provided in Appendix A.

## 4. RESULTS

*4.1 Dataset Characteristics*

Table 4 shows descriptive statistics for the synthetic wearable data set. This cohort shows a typical physiological profile of high-risk chronic disease patients. The mean RHR of 78.6±9.5 bpm is higher than the adult norm (~72 bpm), reflecting cardiovascular stress due to diabetes, hypertension, and medications. The mean cardiac variability of

38.5±12.5 ms is lower than the healthy threshold (45 ms), indicating autonomic dysfunction such as diabetic neuropathy and hypertensive arterial stiffness.

Table 4. Dataset Characteristics (N=500 Patients, 365 Days)

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| Age (years) | 59.8 | 11.9 | 40.0 | 80.0 |
| RHR (bpm) | 78.6 | 9.5 | 55.0 | 110.0 |
| HRV - SDNN (ms) | 38.5 | 12.5 | 15.0 | 85.0 |
| Daily Steps | 6,692.0 | 2,101.3 | 1,000.0 | 15,000.0 |
| Sleep Duration (hrs) | 6.4 | 1.2 | 4.0 | 10.0 |
| Sleep Disruptions (per night) | 4.5 | 1.7 | 2.0 | 13.0 |
| Sleep Efficiency (%) | 78.7 | 9.1 | 60.0 | 100.0 |
| Sedentary Time (hrs) | 9.6 | 2.3 | 4.0 | 16.0 |
| Systolic BP (mmHg)* | 137.1 | 19.4 | 90.0 | 200.0 |
| Event Rate (%) | 27.2 | — | — | — |

*30% of patients with BP monitoring

Low physical activity levels, with an average of 6,692±2,101 steps per day, marked the cohort as "low activity." Sleep quality was also poor: average 6.4±1.2 hours of sleep (less than the recommended 7–9 hours), sleep efficiency 78.7±9.1% (<85%), and night awakenings 4.5±1.7 times (higher than 2–3 times normal). Average sedentary time of 9.6±2.3 hours exceeded the 8-hour threshold. The event rate of 27.2% indicates a high-risk period during which composite health deterioration reached a critical level, due to the selection of a high-risk cohort, a broad event definition, and continuous monitoring for 365 days.

### *4.2 Model Performance Comparison*

Table 5 shows the performance metrics of all models on the test set. LSTM achieved the highest accuracy of 93.7%, outperforming Random Forest (91.5%), XGBoost (90.6%), and threshold baseline (78.6%). The 15.1 percentage point advantage of LSTM over simple threshold rules indicates that machine learning outperforms traditional clinical decision algorithms.

Table 5. Model Performance (5-Day Lead Time)

| Model | Accuracy | Sensitivity | Specificity | Precision | F1-Score | ROC-AUC |
|---|---|---|---|---|---|---|
| LSTM | 0.937 | 0.876 | 0.959 | 0.887 | 0.882 | 0.983 |
| XGBoost | 0.906 | 0.783 | 0.951 | 0.857 | 0.818 | 0.966 |
| Random Forest | 0.915 | 0.883 | 0.927 | 0.818 | 0.849 | 0.971 |
| Threshold | 0.786 | 0.432 | 0.917 | 0.660 | 0.522 | 0.675 |

Sensitivity varied between models. LSTM achieved 87.6%, identifying 1,829 out of 2,087 events and missing 258 (false negatives). Random Forest was slightly higher at 88.3% but with lower specificity (92.7% vs. LSTM's 95.9%). XGBoost was lower at 78.3%, missing more events but still having high specificity (95.1%). The baseline threshold was only 43.2% sensitive, missing 57% of the deteriorating patients—not suitable for clinical use that requires high sensitivity.

The specificity pattern supports the sensitivity findings. The LSTM achieved 95.9% specificity, correctly identifying 5,411 out of 5,643 non-events, with only 232 false positives (4.1%), important for avoiding alert fatigue. The model accuracy was 88.7%, meaning that nearly 9 out of 10 high-risk alerts represented real, actionable events.

The ROC-AUC metric shows excellent discrimination for the machine learning models: LSTM 0.983 (nearly perfect), Random Forest 0.971, and XGBoost 0.966. The baseline threshold of only 0.675, almost on par with random guesses (0.50), shows inadequate performance. The advantage of LSTM may be due to its temporal modelling—combining multi-day patterns such as high heart rate 7 days ago, low HRV 3 days ago, and decreased activity yesterday is better predictor of events than today's readings alone.

*4.3 ROC Curve Analysis*

Figure 1 shows the ROC curves for all four models. The LSTM achieved near-perfect discrimination (AUC=0.983), with 90% sensitivity achieved before any false positives occurred. Random Forest (AUC=0.971) and XGBoost (AUC=0.966) also performed very well, almost on par with the LSTM. The threshold baseline (AUC=0.675) was much weaker, with only 45% sensitivity at a 5% FPR, missing more than half of the events despite the equivalent false alarm rate.

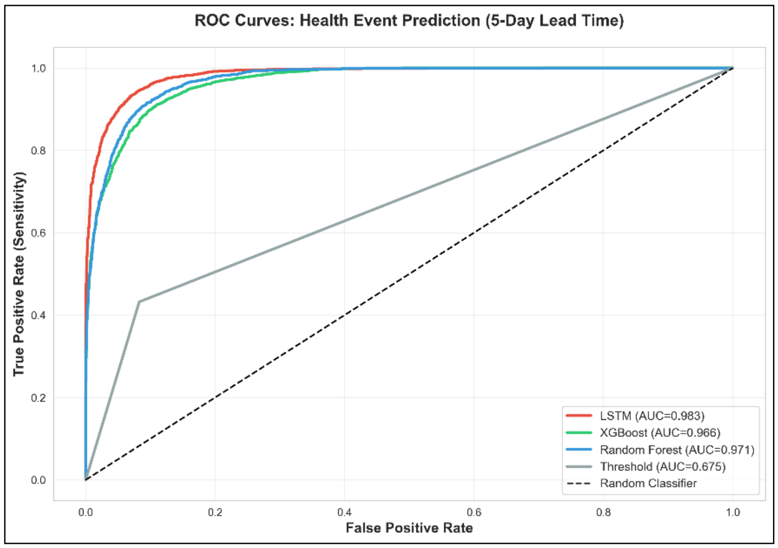


Figure 1. ROC Analysis

Threshold selection involves a balance between sensitivity (patient safety) and specificity (resource utilization). A conservative threshold (0.35) gives high sensitivity (95%) but more false alarms (8%). A moderate threshold (0.50) is balanced (sensitivity 87.6%, FPR 4.1%), while an aggressive threshold (0.65) reduces false alarms (2%) but also decreases sensitivity (78%).

*4.4 Confusion Matrix Analysis*

Figure 2 shows the LSTM confusion matrix with 7,730 predictions divided into four categories. True negatives (5,411) are low-risk true events that are correctly predicted, representing 95.9% of true non-events, allowing patients to avoid unnecessary testing or interventions. False positives (232) are high-risk false non-events (4.1%), which result in unnecessary alerts, but this rate is low and indicates acceptable specificity for clinical use. False negatives (258) occurred when actual high-risk events were predicted to be low, representing 12.4% of all events. These errors are the most clinically risky because patients do not receive timely intervention. In contrast, true positives (1,829) are events that were correctly predicted to be high-risk, with 87.6% of events detected. These patients received intervention 5 days before the crisis, potentially avoiding emergency admission to hospital. The overall accuracy of 93.7% provides a brief overview, but the confusion matrix shows the performance in more detail. The model identifies stable patients very well (specificity 95.9%) and detects impending events well (sensitivity 87.6%). This profile is suitable for clinical use, as minimizing false negatives is important, while a moderate false positive rate is acceptable because the proposed intervention is non-invasive.

*4.5 SHAP Feature Importance Analysis*

Figure 3 shows the top 15 predictive features by absolute SHAP value. The two most important features, rhr_lag_7 and rhr_lag_3 (SHAP=0.126), represent resting heart rates 7 and 3 days ago, respectively. This suggests that multi-day heart rate patterns are more predictive of cardiac deterioration than single-day readings. The model identifies that sustained tachycardia over several days indicates deterioration in cardiac function.

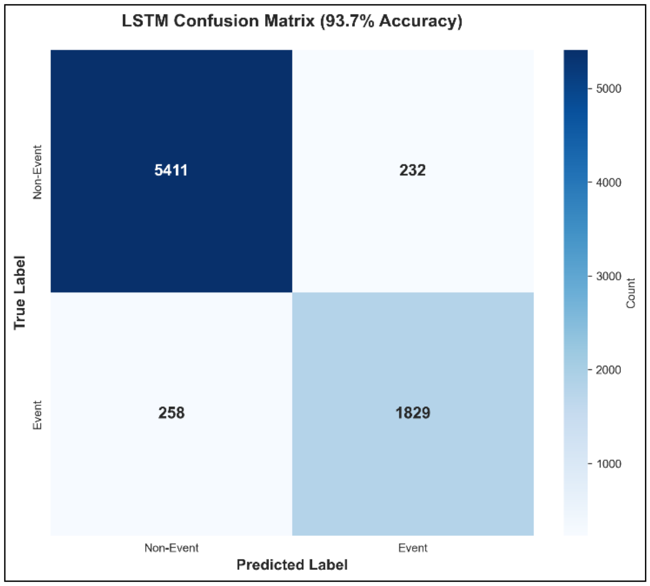


Figure 2. Confusion Matrix for LSTM

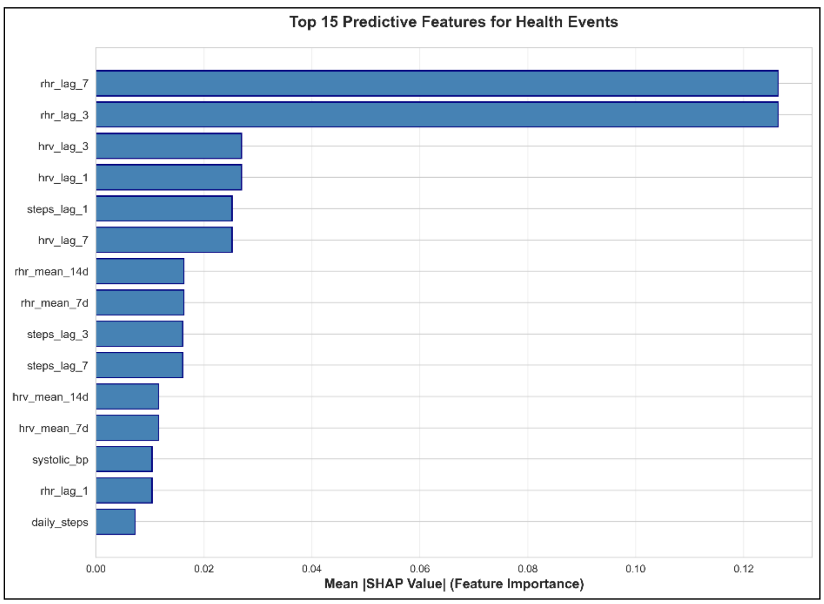


Figure 3. SHAP Importance

HRV-related features rank 3–6, including hrv_lag_3 (0.027), hrv_lag_1 (0.027), steps_lag_1 (0.025), and hrv_lag_7 (0.025). Recent HRV (1-3 days old) slightly more predictive than the weekly measures, which points to a strong short to medium term HRV decrease. HRV SHAP values (0.025-0.027) are approximately 5 times less than the rhr_lag, which implies that HRV is significant but not as important as heart rate. Lag physical activity (steps_lag_1, -3, -7) has an association with 0.016 -0.025 SHAP, with the latest day being more predictive. Rolling mean heart rate (rhr_mean_7d, rhr_mean_14d) also has a moderate contribution.

SHAP score is reliable to cardiovascular pathophysiology. High heart rate signifies high cardiac stress and oxygen requirements, and low heart rate indicates dysfunction in the autonomic system. Reduced physical activities are an indicator of disability caused by cardiac failure or metabolic decompensation. The analysis of the temporal patterns reveals that the degradation process is not sudden but takes time per day to weeks and adheres to the course of chronic disease.

SHAP analysis makes it possible to do patient-specific interventions. High-risk patients because of continuously increased heart rates (rhr_lag_7 and rhr_lag_3) will be advised to take cardiac appointments in the form of ECG, echocardiogram, and medication review. Patients who are considered at risk because of acute reduction of activity (steps_lag1) may require musculoskeletal assessments or depression screening.

### *4.6 Patient Health Trajectory Visualization*

Figure 4 demonstrates the physiological trend during the 20 days before the period of high-risk. At baseline, the vitals of patient were comparatively stable: HRV 32-56 ms (average 45 ms), heart rate 67-78 bpm, number of steps daily is between 5,300–8,300, and sleep duration was 5.5–7.8 hours. At this time, vital signs fluctuated day to day by a small margin without any distinct trend.

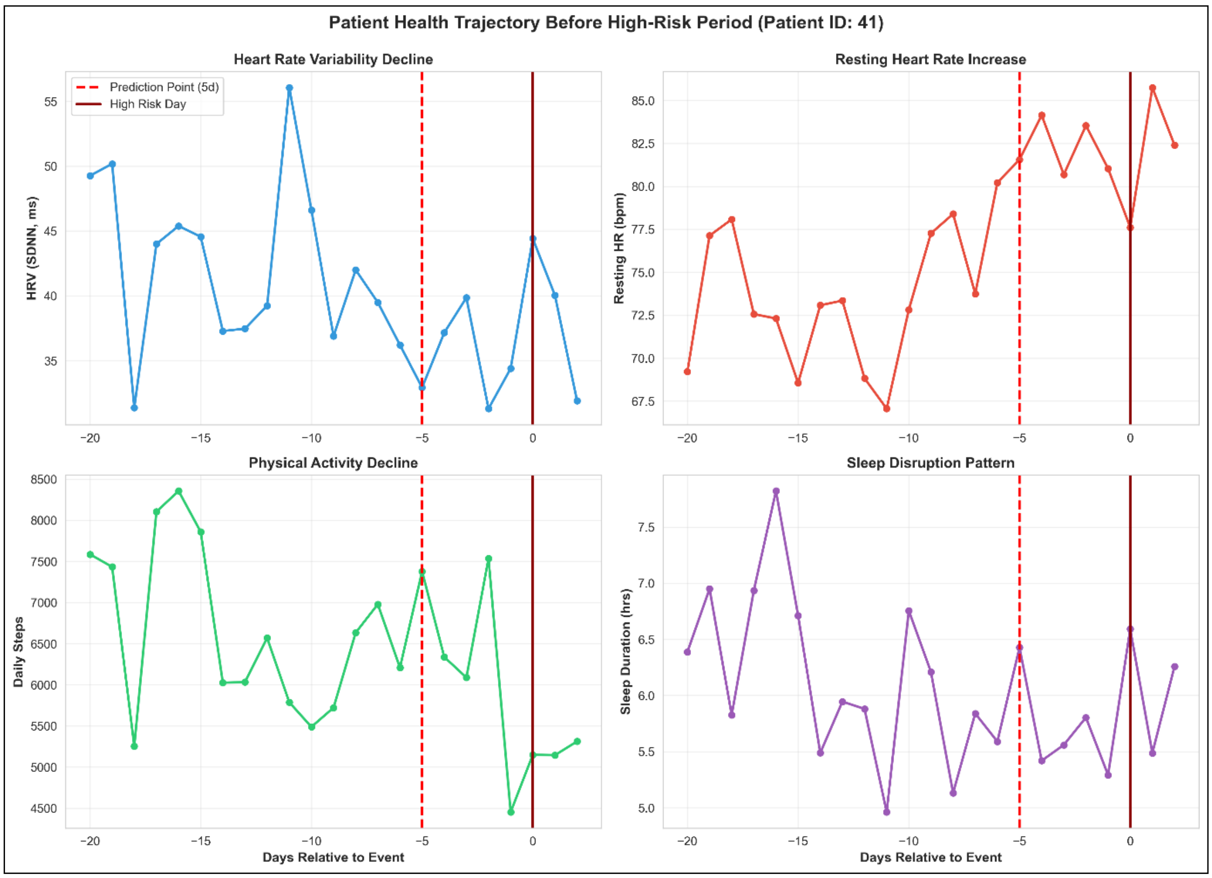


Figure 4. Patient Trajectory

Between days -10 and -5, the deterioration was significant. HRV was reduced by 50 ms to 35 ms (15 ms reduction), the rate of the RHR went up to 80 bpm, and the amount of physical activity dropped to about 6,000 steps, compared to the previous amount of about 7,500. At the prediction point (-5), HRV reached 33 ms, RHR was 83 bpm, daily steps were 7,300, and sleep was 5.5 hours, which gave a result of a high-risk warning at 87 percent.

Between -5 to 0, the progression was downwards towards the high-risk days. HRV remained low (30–35 ms), heart rate was elevated (80-88bpm), activity reduced, and the sleep quality was low. By day 0, HRV remained low, heart rate was 82 bpm, and the steps decreased 5,300. The overall risk score was below the threshold, which suggest severe deterioration and needs intervention.

It has been visualized that the erosion is a process that takes days towards an impending crisis. At day -5, early signals can be intervened with: medications can be changed by doctors, patients can become more active and monitor their diets, or telehealth. This helps assess the necessity of hospital visits in-person or preventing the hospitalization. No single biomarker is sufficient to provide warning; predictive power comes from a combination of signals such as reduced HRV, elevated heart rate, decreased activity, and insufficient sleep.

## 5. DISCUSSION

### *5.1 Principal Findings and Clinical Implications*

This research established that the decline of chronic illnesses can be identified within five days by continuous monitoring using artificial intelligence in wearable devices. The accuracy level and sensitivity of this approach were 93.7% and 87.6% respectively. The comparisons of the performances revealed that the LSTM model gave improved results compared to the Random Forest (91.5%), the XGBoost (90.6%) and the traditional methods which rely on clinical rules (78.6%). This observation indicates that information-driven learning techniques are very beneficial in addressing the challenges of physiological patterns as opposed to predetermined rule-based systems.

Clinically, this sensitivity of value of 87.6% indicates the capacity of the system to detect most of the patients at risk of deteriorating i.e. virtually nine in every ten patients are likely to be detected at a very young stage. Simultaneously, high specificity (95.9) aids to decrease the number of false alarms, hence, the risk of alarm fatigue in a real healthcare setting is minimized. The analysis of feature attribution based on the SHAP method indicated that alterations in RHR during the previous three to seven days had the greatest role to play in the determination of the prediction. Comparatively, the variability of heart rate and level of physical activity also played a role in the model performance although the contribution was relatively low. Physiologically, this is quite understandable since cardiovascular degradation can manifest itself in terms of chronically increased heart rate, loss of autonomic control, and impaired capacity to sustain physical exercise.

The key practical aspect of this research is that all major physiological indicators in the study can be recorded through the commercially available smartwatches that cost RM500 to RM2,000. This realisation draws attention to the possibility of deploying device-based monitoring at a larger scale and not relying on costly and highly specialised clinical tools that generally have high maintenance costs. Besides, the five-day prediction window provides sufficient lead time to implement early intervention, including medication changes, virtual visits, lifestyle changes, or even to schedule clinical follow-ups. On the whole, these results prove the importance of a more proactive approach to disease management by shifting the focus from reactive treatment towards proactive prevention to minimize the risk of deterioration of the patient.

### *5.2 Comparison to Existing Literature*

The achievement of our LSTM model reach to 93.7% accuracy, it shows significantly higher than previous studies of wearable health AI in Malaysia. In comparison, Salim et al. (2021) achieved 68% with Random Forest for dengue prediction using climate data alone [47], Majeed et al. (2023) reported 78% accuracy with LSTM without integrating biomarkers or specific conditions [53], and Raja et al. (2019) obtained 79–84% with Bayesian methods for dengue prediction in Klang Valley [46]. While these studies demonstrate the potential of health AI in Malaysia, none focused on continuous monitoring of individuals for diabetes or hypertension.

Internationally, our results are consistent with high-performance wearable health systems. The Apple Heart Study with 419,297 participants showed an 84% positive predictive value for atrial fibrillation detection [20], when compare it to our accuracy which is 88.7% this indicates a slight increment. DeepGluco achieved 91% accuracy for predicting glucose levels 30 minutes ahead [54], comparable to the accuracy of our LSTM even though our prediction spanned 5 days instead of 30 minutes. COVID-19 detection via wearable devices achieved 85–90% accuracy 1–2 days before symptoms appeared [55], indicating effective early warning even though the lead time was shorter than our 5-day window.

Since the use of synthetic data are well known to be interpreted with caution. Therefore, it is expected that in real-world application, the accuracy may drop to 85–90% versus 93.7%. This is due to factors such as inconsistent device use, sensor noise, medication non-adherence, unpredictable acute events, and other confounds. However, 85% accuracy with a 5-day warning is still significantly better than quarterly episodic monitoring that provides no early warning.

### *5.3 Operational Deployment Framework*

Data Infrastructure: Real-time data flow from wearable devices to the cloud requires API connections with devices such as Apple Health, Samsung Health, Fitbit and Google Fit for authorized data access. This data is continuously processed on cloud servers and integrated with Electronic Health Records according to the FHIR standard. Data quality also needs to be monitored to detect missing data, unusual readings and implausible patterns.

Malaysia's MySejahtera digital health system, developed during the COVID-19 pandemic, has shown that large-scale adoption of health technology is feasible and socially acceptable [56]. The app has been downloaded over 32 million times, is able to handle high concurrent usage, and has successfully built public trust. Therefore, wearable monitoring can be built on this existing system as it has a proven architecture and a large user base.

Alert Protocol Tiering: A tiered response system is used to match risk levels with resource use. For low risk (40–60%), patients receive educational SMS messages. Moderate risk (60–80%) will be encouraged to undergo a telehealth session within 48 hours. While high risk (over 80%) requires a face-to-face clinic visit within 24 hours. This approach helps ensure that health resources are used more effectively according to the patient's risk level.

Clinical Workflow Integration: It is an important element in the success of its integration with the existing clinical practices. It is essential to use the same platform to make sure that predictions from wearable devices are displayed directly in the electronic health record at the time of appointment. A systematic record system also capable of ensuring that every clinical response to each piece of information is documented. This is one of the steps that aid the system in evaluating the system. whether the interventions made are really effective in averting hospital admissions.

Stakeholder Training: Patients should be trained on how to use the wearable gadgets along with understanding the health information. Community health workers are very particularly important to provide strong support, especially to the elderly patient as well as those from rural populations. In the meantime, clinicians should be trained intensively in order to assess the outcomes critically so not to be over-dependent on the outcomes of AI systems. At the same time, the Ministry of Health must come up with clear guidelines on payment, quality standards and data management to ensure the safety, ethical, and sustainable application of this technology.

### *5.4 Health Economic Analysis*

Cost-Benefit Model (500-Patient Cohort, Annual):

*Monitoring Costs:* Wearable devices RM 83,000/year (amortized over 3-year lifespan), cloud infrastructure RM 50,000, clinical staff (2 FTE monitoring nurses) RM 120,000. Total Annual Cost: RM 253,000

*Benefits from Prevented Hospitalizations:* Event rate 27.2% × 500 patients = 136 events annually. LSTM sensitivity 87.6% × 136 = 119 events detected. Intervention success rate (conservative 30%) × 119 = 36 prevented hospitalizations. At RM 20,000 average cost per hospitalization: RM 720,000 direct savings

Return on Investment: (RM 720,000 benefits - RM 253,000 costs) / RM 253,000 = 185% annual ROI

National Scaling Scenario: Malaysia has 3.9 million diabetics + 6.4 million hypertensives with substantial overlap. Assume 10% highest-risk patients (1 million) would benefit most from intensive monitoring:

- Annual monitoring cost: 1,000,000 × RM 253 = RM 253 million
- Prevented hospitalizations: 1,000,000 × 0.272 × 0.876 × 0.30 = 71,510 events
- Direct savings: 71,510 × RM 20,000 = RM 1.43 billion
- Net benefit: RM 1.18 billion annually

This conservative analysis (30% intervention success) demonstrates compelling economic case. Even if only 20% of detected deterioration is actually prevented, net savings exceed RM 800 million annually.

*5.5 Ethical Considerations*

Privacy and Data: End-to-end encryption mechanisms are used to protect data security during transmission. This is because the transmission of sensitive personal health data continues through continuous monitoring. In addition, strict access controls and an informed consent process that comply with the Personal Data Protection Act (PDPA) are also implemented to ensure privacy and legal compliance.

Algorithmic Bias and Fairness: Among problems that exist in training the model is the use of urban population data where it may not work well against rural population data. Therefore, the level of model performance needs to be tested using more varied data including groups of different areas of residence, age and gender. In addition, the diversity of training data also needs to be increased including more data from groups with less data. This step can ensure that the model will be fairer and work well for all populations.

Autonomy and Consent: Each patient must give their consent to participate in this monitoring voluntarily. They must also be assured that all information will be provided in a clear and easy to understand manner. Among the rights granted to patients is that they can determine the frequency of notifications and the level of sensitivity of the information they will receive. In addition, third parties such as insurance companies or employers do not have the authority to mandate this monitoring.

Alert Burden and Mental Health: Continuous monitoring has the potential to be misinterpreted as a health problem. A graded alert system is a very effective solution, where immediate notification triggers will only take effect if a high-risk situation is detected. In addition, providing clear explanations to users as well as mental health screening to identify concerns is important to maintain a balance between patient safety and psychological well-being.

*5.6 Limitations and Future Directions*

Synthetic Data Limitations: The accuracy is expected to decrease to 85-90% due to the inherent complexity when real patient data is introduced. In particular, a domain shift may occur because real Malaysian patient data are likely noisier and more heterogeneous than the current synthetic dataset. Therefore, prospective clinical trials are important where 1,000 random patients are assigned to an intervention group (wearable monitoring + AI alerts) versus a control group (standard care).  Full deployment will proceed after 12-month of follow-up by measuring hospitalization rates, emergency visits, and quality of life.

Limited Outcome Specificity: This model is only able to detect serious health conditions in general. If separate models or multiple classifications were used, the exact cause of the problem, for example diabetes, high blood pressure or heart problems, could be identified more clearly and more appropriate treatment could be given.

Missing Contextual Factors: There are several factors that were not included in this study, namely medication adherence, dietary intake, psychosocial stress and genetic factors. This is because the process of integrating information from electronic health records including diagnostic data, medications and laboratory reports has the potential to increase the level of complexity and most importantly concerns about the privacy of the data.

Causal Inference Gap: This model only focuses on the ability to make predictions but is unable to detect the actual cause of a study. Therefore, a randomized controlled trial (RCT) that serves as a benchmark for whether the group that received the warning and follow-up action is comparable to the group that did not receive warning can prove whether early warning actually prevents health events or simply detects inevitable deterioration.

Future Research Directions:

- Real-world validation trial with 1,000 patients, randomized controlled, 12-month follow-up
- Multi-sensor expansion: continuous glucose monitors, ECG patches, smart medication dispensers
- Outcome-specific models: separate classifiers for diabetic emergencies, heart failure, hypertensive crises
- Reinforcement learning: optimize intervention selection for maximum event prevention
- Federated learning: privacy-preserving training where devices learn locally, share only model updates
- Patient-facing decision support: mobile app translating SHAP explanations into actionable recommendations

## 6. CONCLUSION

The clinical potential of AI-based continuous health monitoring for chronic disease management using consumer wearable sensors in this study demonstrates both technical and clinical potential. With 93.7% accuracy and 87.6% sensitivity using LSTM networks, adverse health events were predicted 5 days in advance. This clearly outperformed traditional episode monitoring (without early warning) and threshold-based clinical rules (78.6% accuracy). Based on SHAP interpretation, it revealed lagged heart rate patterns and HRV trends as the strongest predictors, thus providing biological validation, clinical interpretation and actionable guidance that are important for adoption.

With the increasing rate of chronic diseases in Malaysia where 73% of deaths due to NCD and 3.9 million people have diabetes and 6.4 million people have hypertension, this study has the potential to overcome the crisis and indirectly reduce the burden of RM 9.65 billion, which is currently incurred due to the use of scalable technology. The low-cost solution proposed in this study can benefit the health economy where net savings can reach RM 1.18 billion if 10% of high-risk patients (1 million Malaysians) are monitored. Furthermore, it promises a return on investment of up to 185% even under conservative assumptions.

However, the success of this study still requires coordination across the five domains: (1) data infrastructure—real-time wearable-to-cloud pipelines, API integration, FHIR-compliant exchange, (2) clinical integration—EHR embedding, graduated alert tiers, nurse-led population monitoring, (3) digital literacy—patient and clinician training, community health worker support, (4) equity measures—subsidised devices, performance audits across demographics, rural connectivity solutions, (5) continuous validation—prospective randomised trial, ongoing performance monitoring, model retraining.

The existence of the MySejahtera digital health system in Malaysia proves that large-scale adoption of health technology is feasible. Developing a monitoring system through wearable devices based on this platform can leverage existing infrastructure, many users and public trust. The technology and requirements are available, and the economic benefits are clear. The main challenge now is to translate the concept into actual implementation through collaboration between the Ministry of Health, healthcare providers, technology companies, universities and communities. Chronic disease management is now at a critical juncture between whether to continue to rely on reactive treatment during a crisis, or to shift to proactive AI-based monitoring that can detect problems earlier, reduce hospital admissions and improve quality of life. Therefore, action needs to be taken immediately.

## ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for the suggestions to improve the paper.

## FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

## AUTHOR CONTRIBUTIONS

Umi Najiah Ahmad Razimi: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Yong Yoke Leng: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation, Project Administration, Writing – Review & Editing;
Nor Hapiza Mohd Ariffin: Project Administration, Supervision, Writing – Review & Editing;
Mohammed Hazim Alkawaz: Project Administration, Supervision, Writing – Review & Editing.

## CONFLICT OF INTERESTS

No conflict of interests were disclosed.

## ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Derived data supporting the findings of this study are available from the corresponding author on request.

## REFERENCES

[1] Ministry of Health Malaysia, Salt Reduction Strategy to Prevent and Control Non-Communicable Diseases for Malaysia 2021–2025. Putrajaya, Malaysia: Ministry of Health Malaysia, 2021. [Online]. Available: https://www.moh.gov.my/moh/resources/Penerbitan/Rujukan/NCD/Garam/Salt_Reduction_Strategy_To_Prevent_And_Control_NCD_For_Malaysia_2021-2025_26082021.pdf. Accessed: 23 Jan. 2026

[2] World Health Organization and Ministry of Health Malaysia, Prevention and Control of Noncommunicable Diseases in Malaysia: The Case for Investment. 2024. [Online]. Available: https://www.who.int/malaysia/internal-publications-detail/prevention-and-control-of-noncommunicable-diseases-in-malaysia. Accessed: 23 Jan. 2026.

[3] Ministry of Health Malaysia, The Direct Health-care Cost of Noncommunicable Diseases in Malaysia. Putrajaya, Malaysia: Ministry of Health Malaysia, 2022. [Online]. Available: https://www.moh.gov.my/moh/resources/Penerbitan/Rujukan/NCD/NCD_Laporan/HEALTH-COST_of_NCDs-7a-WEB.pdf. Accessed: 23 Jan. 2026.

[4] World Health Organization, Regional Office for the Western Pacific, "Transforming health: Malaysia's RM 30 billion NCD strategy," 16 Dec. 2024. [Online]. Available: https://www.who.int/westernpacific/newsroom/feature-stories/item/transforming-health–malaysia-s-rm-30-billion-ncd-strategy. Accessed: 23 Jan. 2026.

[5] Department of Statistics Malaysia, "Peluncuran MyAgeing Dashboard (Press Statement)," 06 Sep. 2024. [Online]. Available: https://www.dosm.gov.my/uploads/content-downloads/file_20240906121413.pdf. Accessed: 23 Jan. 2026.

[6] S. Akhtar, A. Nasir, M. Abbas, and M. Asghar, "Prevalence of type-2 diabetes and prediabetes in Malaysia: A systematic review and meta-analysis," *PLOS ONE*, vol. 17, no. 2, pp. e0263139, 2022, doi: 10.1371/journal.pone.0263139.

[7] Ministry of Health Malaysia, Academy of Medicine Malaysia, Malaysia Endocrine & Metabolic Society, Family Medicine Specialists Association of Malaysia, and Diabetes Malaysia, Clinical Practice Guidelines: Management of Type 2 Diabetes Mellitus, 6th ed. Putrajaya, Malaysia, Dec. 2020. [Online]. Available: https://mems.my/wp-content/uploads/2021/03/CPG-T2DM_6th-Edition-2020_210226.pdf. Accessed: 23 Jan. 2026.

[8] H. Thirumurthy *et al.*, "Differences between self-reported and electronically monitored adherence among patients receiving antiretroviral therapy in a resource-limited setting," *AIDS*, vol. 26, no. 18, pp. 2399–2403, Nov. 2012, doi: 10.1097/QAD.0b013e328359aa68.

[9] M. J. Stirratt *et al.*, "Self-report measures of medication adherence behavior: Recommendations on optimal use," *Translational Behavioral Medicine*, vol. 5, no. 4, pp. 470–482, Dec. 2015, doi: 10.1007/s13142-015-0315-2.

[10] H. Ismail *et al.*, "Mapping the Locations of Medical Specialists in the Ministry of Health's Hospitals in Malaysia by Specialty, Subspecialty and Area of Interest," *Malaysian Journal of Medical Sciences*, vol. 31, no. 3, pp. 1–17, Jun. 2024, doi: 10.21315/mjms2024.31.3.1.

[11] B. A. Ahmad, K. Khairatul, and A. Farnaza, "An assessment of patient waiting and consultation time in a primary healthcare clinic," *Malaysian Family Physician*, vol. 12, no. 1, pp. 14–21, 2017.

[12] S. N. Abd Wahab, N. H. Mohd Satar, and M. Tumin, “Urban poor: Evidence of barriers in accessing public clinics in Kuala Lumpur, Malaysia,” *Malaysian Journal of Public Health Medicine*, vol. 22, no. 1, pp. 193–204, 2022, doi: 10.37268/mjphm/vol.22/no.1/art.1148.

[13] J. Ab Hamid, M. H. Juni, R. Abdul Manaf, S. N. Syed Ismail, and P. Y. Lim, “Spatial accessibility of primary care in the dual public–private health system in rural areas, Malaysia,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, Art. no. 3147, 2023, doi: 10.3390/ijerph20043147.

[14] International Data Corporation (IDC), “IDC forecasts continued growth for wearables but growth will be uneven across product categories,” 26 Sep. 2024. [Online]. Available: https://my.idc.com/getdoc.jsp?containerId=prUS52615024. Accessed: 23 Jan. 2026.

[15] S. A. A. Seman, “A structural investigation on behavioural intention among young adult users toward smartwatch adoption in Malaysia,” *Malaysian Journal of Consumer and Family Economics*, vol. 28, pp. 467–482, 2022.

[16] K. B. Kim, and H. J. Baek, “Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions,” *Electronics*, vol. 12, no. 13, Art. no. 2923, 2023, doi: 10.3390/electronics12132923.

[17] D. R. Bassett Jr, L. P. Toth, S. R. LaMunion, and S. E. Crouter, “Step counting: A review of measurement considerations and health-related applications,” *Sports Medicine*, vol. 47, no. 7, pp. 1303–1315, Jul. 2017, doi: 10.1007/s40279-016-0663-1.

[18] A. E. Schutte, “Wearable cuffless blood pressure tracking: when will they be good enough?” *Journal of Human Hypertension*, vol. 38, no. 9, pp. 669–672, 2024, doi: 10.1038/s41371-024-00932-3.

[19] D. L. Cohen, and J. R. Townsend, “How significant is white coat hypertension?” *Current Hypertension Reports*, vol. 12, pp. 165–167, 2010, doi: 10.1007/s11906-010-0115-2.

[20] M. V. Perez *et al.*, “Large-scale assessment of a smartwatch to identify atrial fibrillation,” *The New England Journal of Medicine*, vol. 381, no. 20, pp. 1909–1917, 2019, doi: 10.1056/NEJMoa1901183.

[21] G. Quer *et al.*, “Wearable sensor data and self-reported symptoms for COVID-19 detection,” *Nature Medicine*, vol. 27, no. 1, pp. 73–77, 2021, doi: 10.1038/s41591-020-1123-x.

[22] S. Hochreiter, and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[23] W. P. T. M. van Doorn *et al.*, “Machine learning-based glucose prediction with use of continuous glucose and physical activity monitoring data: The Maastricht Study,” *PLOS ONE*, vol. 16, no. 6, pp. e0253125, 2021, doi: 10.1371/journal.pone.0253125.

[24] S. Nemati, M. M. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. J. Buchman, “An interpretable machine learning model for accurate prediction of sepsis in the ICU,” *Critical Care Medicine*, vol. 46, no. 4, pp. 547–553, 2018, doi: 10.1097/CCM.0000000000002936.

[25] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019, doi: 10.1038/s41591-018-0268-3.

[26] S. M. Lundberg, and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30. Red Hook, NY, USA: Curran Associates, 2017.

[27] N. I. Mohamed Nazar, N. S. Ab Rahman, N. E. Alias, S. Zaini, T. K. Tg Mohd Kamil, N. A. Jamani, and M. H. Elnaem, “Impak Sihat: A telehealth system development and feasibility evaluation to empower rural population in Malaysia on the quality use of medicines,” *PLOS Digit. Health*, vol. 4, no. 8, pp. e0000937, Aug. 2025, doi: 10.1371/journal.pdig.0000937.

[28] D. A. Abdullah, M. A. Wazdi, H. M. Azri, and A. M. Ishak, "Satisfaction with virtual consultation services in Selangor, Malaysia: A cross-sectional study," *Cureus*, vol. 16, no. 6, pp. e63347, Jun. 2024, doi: 10.7759/cureus.63347.

[29] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996.

[30] H. Tsuji, M. G. Larson, F. J. Venditti, Jr., E. S. Manders, J. C. Evans, C. L. Feldman, and D. Levy, "Impact of reduced heart rate variability on risk for cardiac events: The Framingham Heart Study," *Circulation*, vol. 94, no. 11, pp. 2850–2855, 1996, doi: 10.1161/01.CIR.94.11.2850.

[31] A. I. Vinik, M. Erbas, and T. S. Casellini, "Diabetic cardiac autonomic neuropathy, inflammation and cardiovascular disease," *Journal of Diabetes Investigation*, vol. 4, no. 1, pp. 4–18, 2013, doi: 10.1111/jdi.12042.

[32] K. Fox *et al.* , "Resting heart rate in cardiovascular disease," *Journal of the American College of Cardiology*, vol. 50, no. 9, pp. 823–830, 2007, doi: 10.1016/j.jacc.2007.04.079.

[33] B. del Pozo Cruz, J. del Pozo Cruz, M. A. Ahmadi, F. B. Ortega, I.-M. Lee, and E. Stamatakis, "Association of daily step count and intensity with all-cause mortality in US adults," *JAMA Internal Medicine*, vol. 182, no. 7, pp. 776–784, 2022, doi: 10.1001/jamainternmed.2022.2321.

[34] A. Pandey *et al.*, "Continuous dose-response association between sedentary time and risk for cardiovascular disease: A meta-analysis," *JAMA Cardiology*, vol. 1, no. 5, pp. 575–583, 2016, doi: 10.1001/jamacardio.2016.1567.

[35] A. J. Cooper, F. N. Cappuccio, P. D'Elia, P. Strazzullo, and M. A. Miller, "Sleep duration predicts cardiovascular outcomes: A systematic review and meta-analysis of prospective studies," *European Heart Journal*, vol. 32, no. 12, pp. 1484–1492, 2011, doi: 10.1093/eurheartj/ehr007.

[36] B. Lechat, N. A. Naik, K. A. Reynolds, A. A. A. H. H. Sethi, and D. J. Eckert, "New and emerging approaches to better define sleep and circadian physiology and diagnostics," *Frontiers in Neuroscience*, vol. 15, Art. no. 751730, 2021, doi: 10.3389/fnins.2021.751730.

[37] K. Kario *et al.* , "Prognostic significance of home blood pressure monitoring in the general population," *Journal of the American College of Cardiology*., vol. 73, no. 4, pp. 402–414, Feb. 2019, doi: 10.1016/j.jacc.2018.10.076.

[38] P. M. Rothwell *et al.*, "Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension," *The Lancet*, vol. 375, no. 9718, pp. 895–905, Mar. 2010, doi: 10.1016/S0140-6736(10)60308-X.

[39] J. Daniels, P. Herrero, and P. Georgiou, "A multitask learning approach to personalized blood glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436–445, Jan. 2022, doi: 10.1109/JBHI.2021.3100558.

[40] A. van Wyk, A. Khojandi, and A. Kamaleswaran, "A hierarchical analysis of real-time data: A sepsis case study," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 978–986, 2019, doi: 10.1109/JBHI.2018.2874706.

[41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[42] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[43] P. Kasl *et al.* , "Utilizing wearable device data for syndromic surveillance: A fever detection approach," *Sensors (Basel)*, vol. 24, no. 6, Art. no. 1818, 2024, doi: 10.3390/s24061818.

[44] A. Duran, and K. Jongsma, "Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI," *Journal of Medical Ethics*, 2021, doi: 10.1136/medethics-2020-106820.

[45] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[46] D. B. Raja *et al.*, "Artificial intelligence model as predictor for dengue outbreaks," *Malaysian Journal of Public Health Medicine*, vol. 19, no. 2, pp. 103–108, 2019, doi: 10.37268/mjphm/vol.19/no.2/art.176.

[47] N. A. Mohd Salim *et al*, "Prediction of dengue outbreak in Selangor, Malaysia using machine learning techniques," *Scientific Reports*, vol. 11, Art. no. 939, 2021, doi: 10.1038/s41598-020-79193-2.

[48] O. Baker, Z. Ziran, M. Mecella, and K. Subaramaniam, "AI-driven melanoma detection in New Zealand: A ResNet50-based approach," *Informatics in Medicine Unlocked*, vol. 58, Art. no. 101697, 2025, doi: 10.1016/j.imu.2025.101697.

[49] O. Baker, Z. Ziran, M. Mecella, K. Subaramaniam, and S. Palaniappan, "Predictive modelling for pandemic forecasting: A COVID-19 study in New Zealand and partner countries," *International Journal of Environmental Research and Public Health*, vol. 22, no. 4, Art. no. 562, 2025, doi: 10.3390/ijerph22040562.

[50] O. Baker, K. Subaramaniam, A. S. Shibghatullah, Z. A. Shaffiei, and A. S. S. Amir Hamzah, "A collaborative framework for disease prediction using machine learning," in *Proceeding 2025 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, 2025, pp. 650–654, doi: 10.1109/IICAIET67254.2025.11265309.

[51] Department of Statistics Malaysia, Key Findings of Population and Housing Census of Malaysia 2020: Urban and Rural (STATS ALERT #259), Putrajaya, Malaysia, Dec. 23, 2022. [Online]. Available: https://www.dosm.gov.my/portal-main/release-document-log?cat=stats_alert&release_document_id=18519&type=pdf. Accessed: 23 Jan. 2026.

[52] C. Tudor-Locke, and D. R. Bassett, "How many steps/day are enough? Preliminary pedometer indices for public health," *Sports Medicine*, vol. 34, no. 1, pp. 1–8, 2004, doi: 10.2165/00007256-200434010-00001.

[53] M. A. Majeed, H. Z. M. Shafri, Z. Zulkafli, and A. Wayayok, "A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, Art. no. 4130, 2023.

[54] A. Ahmed, S. Aziz, A. Abd alrazaq, F. Farooq, M. Househ, and J. Sheikh, "The effectiveness of wearable devices using artificial intelligence for blood glucose level forecasting or prediction: Systematic review," *Journal of Medical Internet Research*, vol. 25, Art. no. e40259, 2023.

[55] B. Singh *et. al.* , "Real-world accuracy of wearable activity trackers for detecting medical conditions: Systematic review and meta-analysis," *JMIR mHealth and uHealth*, vol. 12, Art. no. e56972, 2024.

[56] M. Alshami, R. Abdulghafor, and A. Aborujilah, "Extending the unified theory of acceptance and use of technology for COVID-19 contact tracing application by Malaysian users," *Sustainability*, vol. 14, no. 11, Art. no. 6811, 2022.

## BIOGRAPHIES OF AUTHORS

**Yong Yoke Leng** received her PhD in Computing from Sunway University in 2019. She is an active researcher in computer science, with publications in reputable journals and presentations at various conferences. Her research interests include computer vision, pattern recognition, data mining, and charting-based forecasting, where she combines pattern recognition with computational finance techniques. She can be contacted at yokeleng.y@help.edu.my.

| | |
|---|---|
|  | **Umi Najiah** is a Lecturer specializing in analytics, machine learning, image processing, and artificial intelligence at HELP University. She is actively involved in research focusing on data-driven modeling and pattern analysis. Her work emphasizes the application of machine learning and image processing techniques for real-world problem solving. She has experience in teaching and supervising student projects in related areas. She can be contacted at email:uminajiah.ar@help.edu.my. |
|  | **Nor Hapiza Mohd Ariffin** is a Senior Lecturer with expertise in Customer Relationship Management (CRM), Strategic Information System Planning (SISP), and performance evaluation support systems). Her research interests also include human capital development, spiritual information systems, and online distance learning. She is actively involved in academic teaching, research, and the development of information systems to support organizational and educational effectiveness. She can be contacted at NAriffin@su.edu.om. |
|  | **Mohammed Hazim Alkawaz** is an academic specializing in image processing, machine learning, and multimedia. He earned his B.Sc. from the University of Mosul (2010) and his M.Sc. & Ph.D. from Universiti Teknologi Malaysia (2013,2015) receiving Best Postgraduate Student Awards. He served at Management and Science University in several academic roles. Since 2023, he has been with the University of Mosul and heads the Computer Science Department. His research spans cybersecurity Artificial intelligence and computer graphics. He can be contacted via email at mohammed.ameen@uomosul.edu.iq. |