

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Prediction of Myocardial Infarction Based on Non-ECG Sleep Data Combined with Domain Knowledge

CHANGYUN LI^{1,2}, YONGHAN ZHAO^{1,2}, QIHUI MO^{1,2}, ZHIBING WANG^{1,2}, XI XU^{1,2}

¹School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China

²Hunan Key Laboratory of Intelligent Information Perception and Processing Technology, Hunan University of Technology, Zhuzhou 412007, China

Corresponding author: Yonghan Zhao (e-mail: M22077500018@stu.hut.edu.cn).

This work was supported by the School of Computer at Hunan University of Technology.

ABSTRACT Prediction of myocardial infarction (MI) is crucial for early intervention and treatment. Machine learning has increasingly been applied in the realm of disease prediction. This study explores the feasibility of utilizing easily obtainable heart rate (HR) and respiratory rate (RR) data collected during nocturnal sleep, in conjunction with clinical characteristics and medical domain knowledge, to predict MI. Data for this investigation were sourced from the Sleep Heart Health Study (SHHS) program in the United States, which was categorized into MI and non-MI groups based on the occurrence or absence of MI during follow-up, involving a total of 488 participants. Multiple features related to HR and RR were extracted and integrated with clinical features; three algorithms—MLP, SVM, and XGBoost—were employed for model construction. The findings indicated that the MLP model exhibited superior performance, achieving an accuracy rate 71.1%. Furthermore, three medical rules age, HR, and RR were incorporated into the MLP model to mitigate the limitations of small sample sizes. The experiments demonstrate that the model's accuracy reaches its optimal level by combining the age rule, improving to 73.1%. The findings indicate that leveraging non-cardiac electrophysiological data obtained during sleep alongside medical domain knowledge can significantly enhance the accuracy of early predictions regarding cardiac MI while offering novel insights for its prevention and diagnosis.

INDEX TERMS myocardial infarction, machine learning, sleep, non-ECG, domain knowledge guidance

I. INTRODUCTION

MYOCARDIAL Infarction (MI) is a serious and enduring condition characterized by the heart's inability to circulate sufficient blood to meet the body's demands. It represents a significant challenge in global public health, being the leading cause of death worldwide. The World Heart Federation reports that cardiovascular diseases account for approximately 20.5 million deaths annually, constituting nearly one-third of all global fatalities [1]. As a critical condition within the spectrum of cardiovascular diseases, both morbidity and mortality rates associated with MI are on the rise globally, particularly in China, where its incidence continues to escalate year by year, exacerbating an already severe situation [2]. In light of this pressing issue, there is an urgent need for more effective early intervention and treatment strategies.

In recent years, advancements in information technology have led to a remarkable increase in the application of machine-learning techniques within the medical diagnostic

field. Computers can efficiently process vast amounts of data while minimizing manual errors and enhancing accuracy; thus, machine learning has proven reliable for disease diagnosis. Numerous studies have been conducted utilizing machine learning methodologies to predict cardiovascular diseases [3] [4], [5] yielding promising results.

Traditional methods for cardiovascular disease (CVD) prediction that rely on electrocardiogram (ECG) data [6]–[11] or various physiological marker analyses [12], [13] are widely utilized in clinical settings; however, they exhibit significant limitations. These approaches typically depend on specialized medical equipment, are complex and costly to implement, and have a limited collection timeframe that does not facilitate monitoring of long-term physiological trends in individuals. Patients are often assessed only when symptoms become apparent or after a disease has developed, which hampers the timely capture of early risk signals. Furthermore, myocardial infarction (MI) can occur suddenly, making it challenging to

effectively prevent this condition solely through short-term monitoring data.

To address the limitations of traditional methods, in this study, we introduce a novel approach for predictive analysis of myocardial infarction (MI) utilizing heart rate and respiratory rate data collected during nighttime sleep. Sleep is one of the most critical physiological activities in humans and is closely linked to fundamental bodily systems, including the cardiovascular system [14]. Research indicates that resting heart rate and respiratory rate correlate with cardiovascular diseases such as MI [15], [16]. Unlike conventional data collection methods like ECGs, portable devices can continuously monitor physiological parameters during sleep. This method not only enhances accessibility but also allows for the capture of long-term trends within individuals. Such an approach facilitates continuous, non-invasive data collection in daily life, thereby aiding in detecting early risk signals associated with MI.

In addition, the challenges and high costs associated with data collection from MI patients have resulted in a scarcity of publicly available databases, particularly those featuring small study samples in prior research. Based on the challenges above, this study integrates medical domain knowledge, extracts data features, and incorporates constraint rules into the machine learning model. This approach utilizes straightforward if-then logic to guide model training. The objective is to investigate whether heart rate and respiratory rate data during sleep can enhance the prediction of myocardial infarction (MI), thereby addressing the limitations associated with small sample sizes. Simultaneously, this methodology aims to improve the model's interpretability and reliability. In summary, this study aimed to analyze the prediction of myocardial infarction (MI) using non-electrocardiogram (non-ECG) data, specifically heart rate and respiratory rate during nocturnal sleep. By integrating knowledge from the medical field, the findings may contribute to significant advancements in the early prediction of MI.

The primary contributions of this study can be summarized as follows:

1) By concentrating on non-cardiac data sources, this study represents the first systematic application of heart rate and respiratory rate data during nocturnal sleep for myocardial infarction (MI) prediction. This approach circumvents the limitations associated with traditional MI prediction methods and significantly reduces data collection costs.

2) The exploration of long-term, continuously monitored sleep data for MI prediction offers a more comprehensive understanding of individual physiological trends and early risk indicators.

3) In conjunction with medical domain knowledge, three rule constraints—age, heart rate, and respiratory rate—were proposed and integrated into the machine learning model. This integration enhanced the model's interpretability and predictive performance.

II. RELATED WORK

In recent years, machine learning has increasingly been applied in the domain of disease risk prediction, particularly for cardiovascular diseases. Traditionally, methods for predicting cardiovascular disease have relied on the analysis of electrocardiograms (ECG) or physiological markers. Than et al. [17] utilized a machine-learning approach known as the Myocardial Ischemic Injury Index to provide an individualized and objective assessment of MI likelihood. This algorithm incorporates factors such as age, gender, and high-sensitivity cardiac troponin I concentrations to evaluate MI risk in individual patients. Wang et al. [18] utilized readily accessible routine blood and biochemical test data to develop machine learning models for the diagnosis of cardiovascular disease, while also investigating the distinctive hematological characteristics associated with this condition. Yilmaz et al. [13] explored the role of hematological predictors in forecasting acute myocardial infarction through interpretable artificial intelligence (AI). Ibrahim et al. [6] employed ECG data alongside relevant ancillary information to predict MI episodes and introduced a framework that emphasizes interpretability within machine learning models for predicting MI, underscoring its significance in cardiovascular disease prediction. Tadesse et al. [10] proposed an end-to-end deep learning methodology aimed at predicting both the presence and timing of MI through multi-lead ECG fusion. Additionally, Chakraborty et al. [19] conducted a comparative study examining various methods for detecting and predicting MI while proposing a deep learning technique that demonstrated superior performance. Chicco et al. [20] implemented an enhanced random forest classifier to predict heart failure based on gene expression data from patients with ST-segment elevation myocardial infarction (STEMI). Hu et al. [21] developed a dynamic risk score utilizing machine learning techniques to forecast cardiogenic shock episodes among patients experiencing acute decompensated heart failure and MI. Li et al. [22] constructed an MI prediction model employing three algorithms—XGBoost, LightGBM, and Random Forest—targeting patient hospitalization records and clinical data; this model achieved commendable performance outcomes. Tripathy et al. [23] investigated the application of ensemble learning methods for the prediction of early heart disease by integrating electrocardiogram (ECG) data with clinical features. They employed stacking and voting techniques to combine multiple machine learning classifiers, ultimately achieving an accuracy rate 82.6% using the voting method. Bhatt et al. [24] introduced a k-modes clustering approach, utilizing models such as Random Forest, Decision Tree Classifier, Multi-Layer Perceptron (MLP), and XGBoost. The results indicated that the MLP model achieved the highest accuracy at 87.28%. These studies collectively demonstrate the effectiveness of machine learning techniques in predicting cardiovascular diseases. However, the application of ECG data for predicting cardiovascular disease (CVD) still presents certain limitations. It primarily serves as a tool to assist in clinical diagnosis rather than effectively monitoring

long-term physiological trends in individuals. Additionally, it struggles to capture early risk signals and fails to facilitate true early detection and treatment.

Currently, researchers have begun to investigate the use of sleep data for CVD prediction. Zhang et al. [9] examined the relationship between sleep heart rate variability (HRV) and long-term CVD outcomes, demonstrating that a machine learning model integrating sleep HRV with other clinical features shows promise for early prediction of CVD outcomes. Yang Ye et al. [25] proposed a two-layer machine learning modeling approach aimed at exploring how nonlinear kinetic analysis of HRV signals during nocturnal sleep can more effectively identify potential risks to the cardiovascular system. Jahromi et al. [26] investigated the application of deep convolutional neural networks (CNNs) for analyzing electrocardiogram (ECG) signals collected during nocturnal sleep, aiming to predict coronary heart disease (CHD). They examined how ECG signals from various sleep stages could serve as predictors for CHD, achieving an accuracy rate of 75%. Each of these studies underscored the significant potential of sleep data in predicting cardiovascular diseases (CVD). However, acquiring HRV data necessitates higher precision equipment and more complex processing algorithms compared to heart rate and respiratory rate data. This complexity consequently increases measurement costs.

We are increasingly concerned that the integration of prior domain knowledge into machine learning models is emerging as a significant research focus. The incorporation of such knowledge can enhance both the interpretability and predictive performance of these models, demonstrating considerable potential in the realm of medical diagnosis [27]. Singh et al. [28] proposed a fuzzy rule-based classification system for diagnosing coronary artery disease, which notably improves the overall classification accuracy of the system. Nie et al. [29] introduced a deep reinforcement learning approach guided by prior knowledge to address thoracic disease diagnosis, effectively tackling small-sample learning challenges and enhancing model generalization capabilities. Furthermore, Yue et al. [30] presented an innovative method for bladder tumor segmentation that integrates clinical logic rules about bladder tumors and bladder walls into Deep Convolutional Neural Networks (DCNNs). This approach achieves accurate segmentation results even with a limited number of annotated images. Zhang et al. [31] introduced a generic rule loss constraint aimed at embedding domain knowledge, demonstrating significant improvements in model performance on cardiovascular disease classification datasets when training data is scarce.

During the course of our study, we found that resting heart rate is an independent risk factor for myocardial infarction (MI), as demonstrated by Sharashova et al. [15]. Furthermore, the mean nocturnal respiratory rate was identified as an independent predictor of long-term cardiovascular and all-cause mortality in a community-based population study [16]. In the present investigation, we aimed to predict the risk of MI using easily obtainable data on heart rate and respiratory

rate during sleep. To achieve this objective, we first processed the dataset to extract heart rate and respiratory frequency data and performed necessary data cleaning. Subsequently, leveraging medical domain knowledge, we engineered features to identify relevant variables and incorporated clinical factors such as age and gender. Next, multiple machine learning base models were trained using the processed data, followed by an evaluation of the best-performing model. Finally, this optimal base model was enhanced with rules derived from medical domain expertise to assess whether this improved methodology would yield more accurate results and superior model performance.

III. MATERIALS AND METHODS

The research data utilized in this study were sourced from the open-access database of the Sleep Heart Health Study (SHHS) (URL: <https://sleepdata.org/datasets/shhs>) [32], [33]. This cohort study was designed to investigate whether sleep breathing disorders serve as risk factors for cardiovascular disease. Between November 1, 1995, and January 31, 1998, a number of individuals aged 40 years or older participated in baseline polysomnography (PSG) conducted at home using a Compumedics SleepWatch polysomnograph. Follow-up visits continued until 2011 to document the occurrence and latency of myocardial infarction (MI) events.

The specific experimental procedure is illustrated in Fig. 1 and includes the following steps: 1) sample screening; 2) data preprocessing; 3) feature extraction; 4) base model construction; 5) domain knowledge embedding; 6) model performance evaluation.

A. PARTICIPANTS

In the present study, which included 488 participants, we focused on individuals who had not experienced a myocardial infarction (MI) at baseline. Participants were subsequently divided into MI and non-MI groups based on whether an MI occurred during later follow-up. We excluded those with lower sleep efficiency (less than the mean minus twice the standard deviation of all subjects) in the study. Additionally, it has been reported that benzodiazepines, tricyclic antidepressants, and non-tricyclic antidepressants may influence heart rate variability [34]. Therefore, participants who used such medications within two weeks prior to polysomnography (PSG) monitoring were also excluded. To ensure comprehensive data availability for our analysis, we eliminated samples with significant missing data and removed those containing extreme outliers. Following this screening process, we retained all eligible cases; 244 samples remained in the MI group. Conversely, the non-MI group comprised 3,513 samples—resulting in a substantial imbalance between positive and negative cases—which could adversely affect classifier performance [35]. This situation is due to learning algorithms' tendency to predict all minority classes as majority classes to achieve higher accuracy rates. To mitigate this issue, we employed K-Means undersampling [36] for the non-

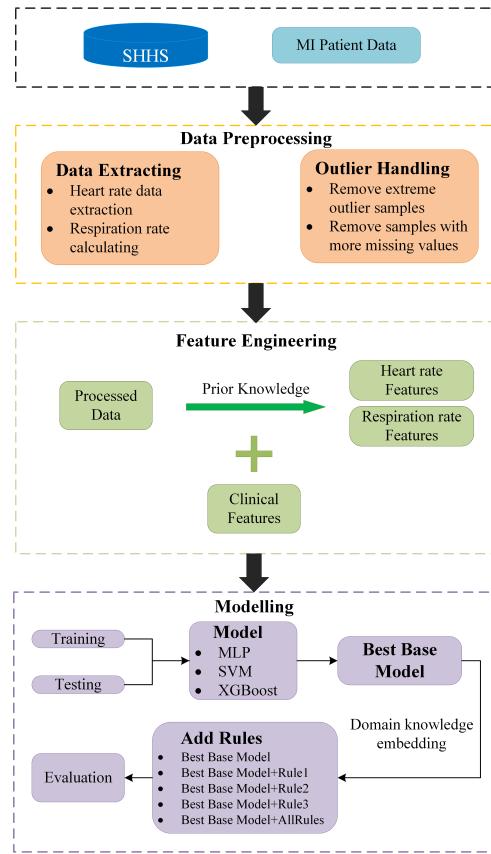


FIGURE 1. Proposed framework for predicting myocardial infarction.

MI group to match its sample size with that of the MI group. The MI and non-MI groups comprised a total of 488 samples.

The SHHS study protocol received approval from the institutional review board of each participating center. Each participant provided signed informed consent prior to their involvement in the study. All methods adhered strictly to relevant guidelines and regulations. The current investigation analyzed de-identified data from the SHHS database and did not involve any protocols requiring approval from an institutional review board or ethics committee.

B. DATA PROCESSING

The collected PSGs contained several signals, among which the 'H.R.' channel represented the heart rate signal. This signal was resampled at a frequency of 1 Hz and presented as a time series of heart rate data.

For the respiratory frequency data, we obtained the raw waveforms captured by the chest strap respirometer from the 'THOR RES' channel in the PSGs of retained participants. We employed a peak-finding method to determine respiration counts [37], where two peaks correspond to one complete breath. However, as illustrated in Fig. 2, significant noise within the signal complicated this process, resulting in numerous minor extreme points near each peak. In this study, we utilized an extreme point detection method to identify peaks;

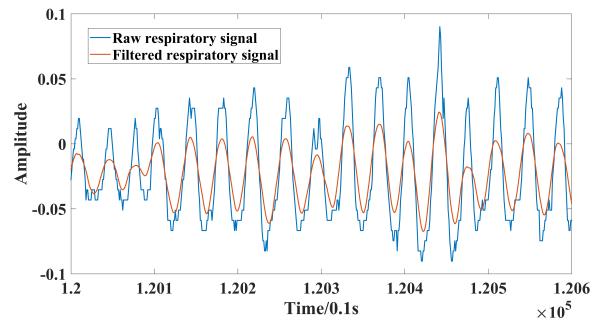


FIGURE 2. The first filtered respiratory signal within a one-minute segment was intercepted, with time on the horizontal axis and the data sampled at 10 Hz for a 1-minute-long segment of 600 data points; the vertical axis is the amplitude.

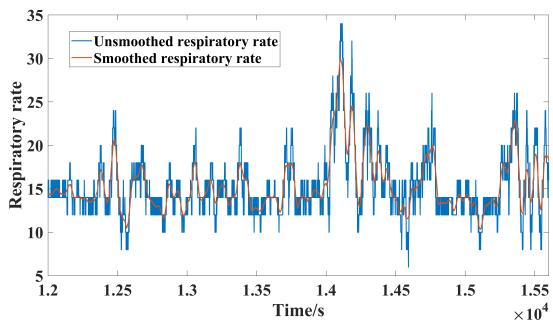


FIGURE 3. Respiratory rate within a one-hour segment was intercepted, with time in seconds on the horizontal axis and respiratory rate in beats/minutes on the vertical axis.

nonetheless, noise interference led to an overestimation of erroneous extreme points being counted as valid peaks. We applied a sliding mean filter for signal smoothing to mitigate this issue. The filtered signal was then divided into overlapping cycles of 30 seconds each with 1-second intervals between them. Consequently, twice the number of respiratory events occurring within each 30-second cycle yielded the respiratory frequency; we considered the average respiratory frequency during these cycles as an estimate for instantaneous respiratory rates for the first second of each cycle.

After calculating the initial respiration rate, the results underwent further smoothing. This step aims to eliminate minor fluctuations introduced by window sliding and potential signal distortions resulting from the initial smoothing process. This additional smoothing enhances the stability and continuity of the time series data for respiratory frequencies, thereby reducing errors associated with the discrete nature of the computational method itself (e.g., step-size effects due to window sliding). The outcomes of this processing are illustrated in Fig. 3, where time-series data were collected for respiratory frequency estimates expressed in units per second.

Due to the variability in sleep duration among individuals, data from participants who exhibited excessively long sleep durations were excluded based on the typical sleep length observed in the majority of subjects. This exclusion was

TABLE 1. Features from heart rate and respiratory rate data

Feature	Definition
Mean_HR	Mean of heart rate
Std_HR	Standard deviation of heart rate
Min_HR	Minimum heart rate
Mid_HR	Median heart rate
Max_HR	Maximum value of heart rate
VC_HR	Coefficient of variation, standard deviation of heart rate divided by the mean value
HR70	Percentage of time heart rate is greater than 70
HR55	Percentage of time when heart rate is less than 55
AROC_HR	Mean rate of change of heart rate
Peaks_HR	Number of extremes in heart rate
Trans_HR	Number of heart rate transitions from higher (lower) than average to lower (higher) than average
Mean_RR	Mean value of respiratory rate
Std_RR	Standard deviation of respiratory rate
Min_RR	Minimum value of respiratory rate
Mid_RR	Median respiratory rate
Max_RR	Maximum value of respiratory frequency
VC_RR	Coefficient of variation, standard deviation of respiratory frequency divided by mean value
RR20	Percentage of time when respiratory rate is greater than 20
RR12	Percentage of time with respiratory frequency less than 12
AROC_RR	Mean rate of change of respiratory frequency
Peaks_RR	Number of extremes in respiratory rate
Trans_RR	Number of transitions from higher (lower) than average to lower (higher) than average respiratory rate

implemented without significantly compromising the overall integrity of the sleep data. Ultimately, we retained each participant's heart rate and respiratory rate measurements for 25,200 seconds (7 hours).

C. FEATURE EXTRACTION

Based on prior experience and expertise in the medical domain, a total of 22 features were extracted from the processed time series data of heart rate and respiratory rate, comprising 11 distinct features. This information is presented in Table 1.

The aforementioned features were supplemented with multiple clinical characteristics, including age, gender, waist-to-hip ratio, body mass index (BMI), height, current smoking status, history of smoking (ever smoked or not), as well as histories of diabetes and hypertension. Consequently, the feature vector for each sample was constructed using a total of 31 distinct features.

D. BASE MODEL CONSTRUCTION

We employed three algorithms to identify the most effective model algorithm: MLP, SVM, XGBoost, and CNN. These were utilized for model construction, with the classification

TABLE 2. Hyperparameters Utilized by the Model

Method	Hyperparameters
MLP	hidden_dim1=128,hidden_dim2=64,Activation='Relu', learning_rate=0.0001, optimizer='Adam'
XGBoost	n_estimators=100,max_depth=3,learning_rate=0.1, random_state=42
SVM	C=1.0, kernel='rbf', gamma='scale'
CNN	filter=16,32,kernel_size=3,activation='Relu', optimizer='Adam',learning_rate=0.0001

feature being whether or not non-MI patients experienced myocardial infarction (MI) during the follow-up period after baseline. To ensure the generalizability of the model, we implemented five-fold cross-validation to obtain average values for each metric used in evaluating model performance. The hyperparameters for each model are presented in Table. 2. The experiments were conducted on a Windows 10 system utilizing PyCharm as the programming environment, based on Python version 3.8, and equipped with an Intel(R) Xeon(R) CPU E3-1225 v5 @ 3.30 GHz processor alongside 32.0 GB of RAM.

E. DOMAIN KNOWLEDGE EMBEDDING

Considering the limited amount of data available in this study, we aimed to enhance the model's performance and generalization on small sample datasets by incorporating domain knowledge. The overall loss function of the model comprises two components: task classification loss and rule loss. For task classification loss, we employed binary cross-entropy as the chosen metric. The total loss function is presented below.

$$L_{\text{total}} = L_{\text{task}} + \alpha \cdot L_{\text{rule}} \quad (1)$$

Given that the objective of this study was to predictively identify the risk of developing early myocardial infarction (MI) using easily obtainable non-cardiac electrophysiological data, rather than relying on clinically specialized medical information, we emphasized the significance of certain characteristics and integrated this with domain knowledge from medical research. Factors known to be associated with the onset of myocardial infarction include age (AGE), heart rate (HR), and respiratory rate (RR). The incidence of cardiovascular disease increases significantly after the age of 60 years [38]. At resting heart rates exceeding 70 beats per minute, the risk of myocardial infarction (MI) progressively escalates with higher resting heart rates [15]. A respiratory rate between 12 and 20 breaths per minute at rest is considered normal in adults; a respiratory rate above 20 breaths per minute is classified as shortness of breath. Additionally, an elevated nocturnal respiratory rate may be an independent predictor of long-term cardiovascular disease [16].

Consequently, we propose three rule constraints:

- 1) If AGE > 60, then the risk of developing MI is elevated;
- 2) If HR > 70, then the risk of MI is increased;

3) If $RR > 20$, then the risk of MI is heightened.

We have introduced a rule loss function to embedding the three rules above into the model, as described below: [31]:

$$L_{rule}(\Theta; x, \tilde{x}) = 1(x_j > \tau_j) \cdot \max(f(x; \Theta) - f(\tilde{x}; \Theta) + \delta, 0) \quad (2)$$

For each rule, the features must first be normalized, after which the features x_j are subjected to random perturbations to obtain \tilde{x} . The perturbation is constrained within the range (0,1), allowing for small variations in feature values to be effectively modeled. The parameter δ serves as a hyperparameter that regulates the strictness of the rule. This formula indicates that when the condition $x_j > \tau_j$ is met, the probability of predicting a sample as having MI should increase if x_j rises. Conversely, a loss defined by $\max(f(x; \Theta) - f(\tilde{x}; \Theta) + \delta, 0)$ is imposed to penalize any violations of this rule.

IV. RESULTS

We employed several metrics, including Accuracy, Precision, F1 Score, Sensitivity, and Specificity, to evaluate the performance of the models. Additionally, we utilized the Bootstrap method to calculate 95% confidence intervals (CI) for all performance evaluation metrics to assess the robustness of the models [12], [18]. Initially, four baseline models—MLP, SVM, XGBoost, and CNN—were evaluated without incorporating any additional rules. The results are presented in Table. 3 . As illustrated in Figure 4, which visualizes the comparison of model effects, it is evident that the MLP model achieved the highest accuracy at 0.711 with a corresponding 95% confidence interval of (0.691, 0.734). Both XGBoost and SVM exhibited similar levels of accuracy at 0.703 (0.676–0.733) and 0.701 (0.672–0.730), respectively; however, CNN demonstrated less satisfactory performance with an accuracy of 0.695 (0.654–0.721). Meanwhile, upon examining the various model evaluation metrics, it is evident that the MLP model outperforms the XGBoost, SVM, and CNN models across three key indices: Accuracy, F1-score, and Sensitivity. Additionally, it demonstrates superior robustness. Ultimately, through a comprehensive assessment, the MLP model has been identified as the optimal base model for this study.

The features used in this study, such as clinical features and HR and RR features, were analyzed, and the importance of the features was assessed by using the Random Forest [39], as illustrated in Fig. 5. Age and waist-hip ratio emerged as the two most critical features, corroborating findings in [9]. Additionally, HR and RR features—such as the percentage of time with an RR less than 12 and the maximum values of both HR and RR—also substantially enhanced model performance. Other HR and RR features exhibited a more balanced level of importance while significantly influencing the risk of potential myocardial infarction (MI). Although a history of hypertension [40]and diabetes [41] has been associated with an increased risk of cardiovascular disease (CVD) prevalence, and gender [42], [43] is recognized as a potential risk factor

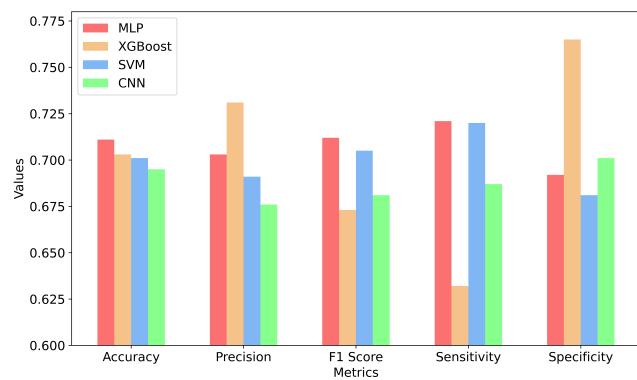


FIGURE 4. Comparative visualization of base model effects

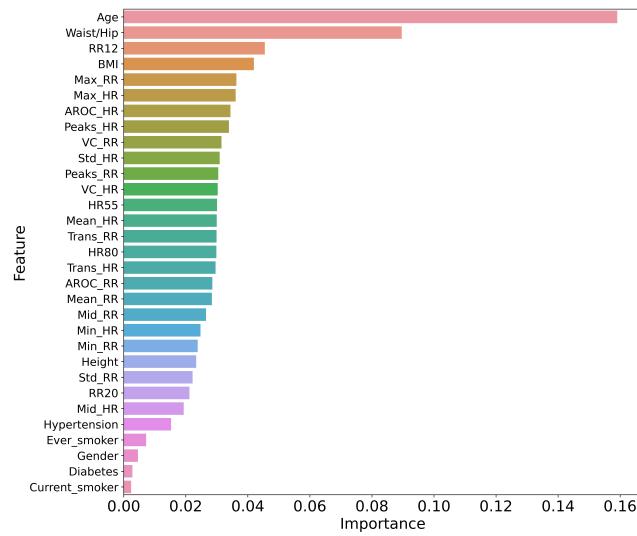


FIGURE 5. The importance of features in the base model. Features suffixed with HR and RR are denoted as heart rate and respiratory rate, respectively

for myocardial infarction (MI) and coronary heart disease. These characteristics contributed minimally to the models in the present study. Additionally, smoking status and lifetime smoking did not demonstrate statistical significance. This situation may arise due to the existence of redundant information among data characteristics or features.

In order to investigate the independent predictive capability of HR and RR during sleep for myocardial infarction (MI), we excluded the relevant features of HR and RR from the optimal base model, which is a multilayer perceptron (MLP). Subsequently, we constructed a comparative model utilizing only clinical features and assessed its performance against that of the full MLP model. The results are presented in Table. 4. When HR and RR features were omitted, the model's accuracy was found to be 68.7%, reflecting a relative decrease of 2.4%. Additionally, all other performance indicators exhibited varying degrees of decline. This suggests that HR and RR during sleep possess significant potential for MI prediction.

We incorporated prior domain knowledge into the model

TABLE 3. Comparison of base model effects

	MLP	XGBoost	SVM	CNN
Accuracy (95% CI)	0.711 (0.691,0.734)	0.703 (0.676,0.733)	0.701 (0.672,0.730)	0.695 (0.654,0.721)
Precision (95% CI)	0.703 (0.658,0.748)	0.731 (0.675,0.785)	0.691 (0.639,0.748)	0.676 (0.595,0.734)
F1-Score (95% CI)	0.712 (0.662,0.761)	0.673 (0.620,0.738)	0.705 (0.660,0.740)	0.681 (0.612,0.719)
Sensitivity (95% CI)	0.721 (0.665,0.773)	0.632 (0.550,0.720)	0.720 (0.687,0.743)	0.687 (0.629,0.736)
Specificity (95% CI)	0.692 (0.627,0.747)	0.765 (0.719,0.821)	0.681 (0.640,0.724)	0.701 (0.664,0.743)

Note: 95% CI: 95% confidence intervals (Lower-Upper bound).

TABLE 4. The performance of two prediction models consisting of different feature vectors

	Clinical characteristics and HR, RR metrics	Only clinical characteristics
Accuracy	0.711	0.687
Precision	0.703	0.673
F1 Score	0.712	0.689
Sensitivity	0.721	0.715
Specificity	0.692	0.659

Note: All models were based on 5-fold cross validation, but only reported the average value of each metric in this table.

to enhance model performance, particularly given that the experiments were conducted on a limited sample dataset. We analyzed three results in the optimal baseline model MLP for each of the three priori rules discussed in section III, part E: using one rule alone (Rule_Age, Rule_HR, Rule_RR) and all three rules combined (Rule_Multi).

The average values of the multiple metrics were computed using five-fold cross-validation. Additionally, 95% confidence intervals were calculated for the evaluation metrics of model performance. The results are presented in Table. 5 Visualization of these evaluation outcomes, as illustrated in Fig. 6, indicates that the MLP model exhibited varying degrees of improvement with the addition of each individual rule. In particular, the model's performance exhibits a more significant improvement when age is utilized as a constraint rule, in comparison to the separate addition of heart rate and respiratory rate rules. This results in an accuracy of 0.731 (0.715, 0.766), reflecting an enhancement of 2.0%.

However, when all three rules were integrated into the model simultaneously, while there was some improvement over the base model's results, this enhancement did not surpass that achieved by any single rule alone; rather, it was slightly inferior to those obtained from individual rule incorporation. This observation suggests that employing multiple priori rules to guide model learning may have introduced conflicts within the learning process. Certain samples may align with one rule but not others, complicating efforts for simultaneous optimization across all established guidelines.

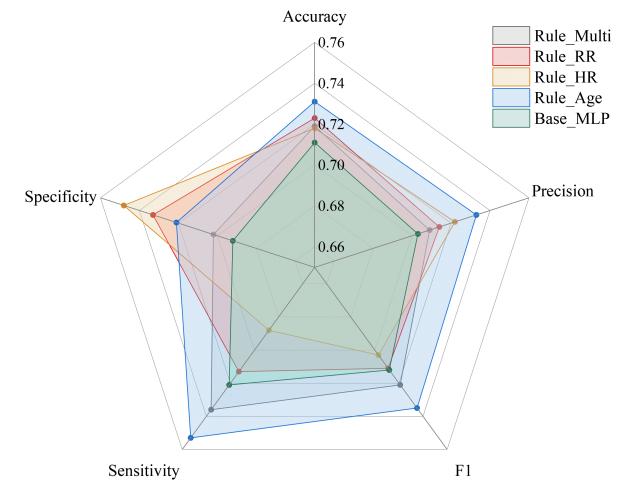


FIGURE 6. Performance of MLP-based models without embedded rules vs. models embedded with different rules

V. DISCUSSION AND CONCLUSIONS

Myocardial infarction (MI) represents a significant global public health challenge, making risk assessment essential for early intervention and therapeutic strategies. Previous studies have predominantly relied on clinical medical data obtained from specialized medical devices for prediction, often utilizing short-term data collected during wakefulness. This approach is not only costly but also it presents challenges in data collection due to limited availability and inherent constraints. In this study, we leveraged the open-source dataset from the SHHS project to acquire easily obtainable non-cardiovascular metrics such as heart rate and respiratory rate during sleep through effective data processing. Our primary research objective was to predict MI by integrating patients' clinical information with relevant medical domain knowledge for feature extraction and model development. We conducted experiments using four foundational models: MLP, SVM, XGBoost, and CNN. Our findings indicated that the MLP model exhibited relatively strong predictive performance, achieving an accuracy of up to 71.1%. To enhance model performance further, we incorporated medical domain knowledge by introducing three constraint rules about age, heart rate, and respiratory rate into the baseline MLP model individually. The results demonstrated improved performance

TABLE 5. Comparison of the performance of models embedded with different rules

	Base_MLP	Rule_Age	Rule_HR	Rule_RR	Rule_Multi
Accuracy (95% CI)	0.711 (0.691, 0.734)	0.731 (0.715, 0.766)	0.718 (0.690, 0.749)	0.723 (0.693, 0.759)	0.719 (0.677, 0.759)
Precision (95% CI)	0.703 (0.658, 0.748)	0.733 (0.683, 0.786)	0.722 (0.687, 0.756)	0.714 (0.663, 0.764)	0.709 (0.660, 0.775)
F1-Score (95% CI)	0.712 (0.662, 0.761)	0.735 (0.689, 0.788)	0.703 (0.680, 0.719)	0.711 (0.689, 0.731)	0.721 (0.660, 0.775)
Sensitivity (95% CI)	0.721 (0.665, 0.773)	0.753 (0.694, 0.809)	0.688 (0.629, 0.736)	0.713 (0.630, 0.810)	0.736 (0.663, 0.804)
Specificity (95% CI)	0.692 (0.627, 0.747)	0.721 (0.663, 0.783)	0.748 (0.686, 0.806)	0.733 (0.685, 0.781)	0.702 (0.657, 0.749)

Note: 95% CI: 95% confidence intervals (Lower-Upper bound).

across all instances after embedding these rules; notably, the best outcomes were observed when incorporating the age rule, which yielded an accuracy of 73.1%(0.715,0.766).

The methodology presented in this study, which utilizes non-cardiac sleep data, mitigates the challenges and expenses associated with data collection. Furthermore, it facilitates the continuous monitoring of users' physiological trends over time. At the same time, we recognize that at this stage, within the research context of various fields, there are instances where data collection poses challenges and available data is limited. Although tree-based models such as Random Forest or XGBoost employed in this study can automatically identify associations within the data, they may struggle to capture weak associations between features and targets in small-sample scenarios. By incorporating medical domain knowledge into the modeling process to mitigate reliance on large datasets for model learning, it is possible to achieve favorable prediction outcomes even with small samples while simultaneously enhancing the interpretability of the model.

Throughout our investigation, we identified that developing a multi-rule embedding model remains a considerable challenge within this field of research. Conflicts or inconsistencies among multiple rules can hinder the model's ability to satisfy all conditions simultaneously. Consequently, these rules may lead to substantial errors in predictions, particularly near the boundary values of the data. Furthermore, while this study highlights the potential of non-cardiac sleep data for predicting myocardial infarction (MI) and emphasizes the incorporation of domain knowledge to address challenges related to limited data volume, it is important to note that the overall size of the dataset remains relatively constrained. This limitation may influence the generalization capability of the model.

The myocardial infarction (MI) prediction model proposed in this study, developed using non-cardiac data collected during sleep alongside medical domain knowledge, offers a novel perspective on MI prediction and holds significant value for the prevention and diagnosis of MI. Nevertheless, this study has certain limitations, including the challenges associated with embedding multiple rules and the issue of small data samples. In future research, it is essential to design reasonable rule priorities and logical relationships while ensuring coordination among these rules. Additionally, exploring flexible rule embedding techniques—such as weight adjustment or

priority assignment—may serve as a promising avenue for improvement. Furthermore, we need to validate our conclusions using a larger dataset to achieve more robust and reliable results.

Currently, with the rise of smart home technology, there is an expectation for its integration into devices such as smart mattresses. For instance, smart mattresses have the capability to monitor users' heart rates, respiratory rates, and other physiological indicators in real time. This data can then be transmitted to the cloud for analysis via a wireless network. The myocardial infarction (MI) prediction model proposed in this study can be incorporated into smart mattresses to facilitate early warning systems for MI. Specifically, when abnormal heart rate or respiratory rate patterns are detected, the system can automatically issue alerts to prompt users to seek timely medical attention. However, despite the promising applications of smart home devices in MI prediction, concerns regarding data privacy and security warrant careful consideration. Future research must focus on developing more secure technologies for data transmission and storage in order to safeguard user privacy effectively.

ACKNOWLEDGMENT

The Sleep Heart Health Study (SHHS) was supported by National Heart, Lung, and Blood Institute cooperative agreements U01HL53916 (University of California, Davis), U01HL53931 (New York University), U01HL53934 (University of Minnesota), U01HL53937 and U01HL64360 (Johns Hopkins University), U01HL53938 (University of Arizona), U01HL53940 (University of Washington), U01HL53941 (Boston University), and U01HL63463 (Case Western Reserve University). The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute (R24 HL114473, 75N92019R002).

REFERENCES

- [1] C. J. L. Murray, "World heart report 2023," *Nature Medicine*, vol. 28, no. 10, pp. 2019–2026, Oct. 2022. [Online]. Available: <https://www.nature.com/articles/s41591-022-01990-1>
- [2] M. LIU, X. HE, X. YANG, and Z. WANG, "Interpretation of Report on Cardiovascular Health and Diseases in China 2023," *Chinese General Practice*, p. 1.
- [3] M. A. Naser, A. A. Majeed, M. Alsabah, T. R. Al-Shaikhli, and K. M. Kaky, "A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges," *Algorithms*, vol. 17, no. 2, p. 78, Feb. 2024. [Online]. Available: <https://www.mdpi.com/1999-4893/17/2/78>
- [4] S. Hussain, S. Ahmad, and M. Wasid, "Artificial intelligence-driven intelligent learning models for identification and prediction of cardiomeurological disorders: A comprehensive study," *Computers in Biology and Medicine*, vol. 184, p. 109342, Jan. 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0010482524014276>
- [5] A. Muhamremović and J. Kevrić, "Cardiovascular Health in AI: A Comprehensive Overview to Acute Myocardial Infarction Prediction," in *MEDICON'23 and CMBEIH'23*, A. Badnjević and L. Gurbeta Pokvić, Eds. Cham: Springer Nature Switzerland, 2024, vol. 93, pp. 628–635, series Title: IFMBE Proceedings. [Online]. Available: <https://link.springer.com/10.1007/978-3-031-49062-0-66>
- [6] L. Ibrahim, M. Mesinovic, K.-W. Yang, and M. A. Eid, "Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values," *IEEE Access*, vol. 8, pp. 210410–210417, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9268965/>
- [7] B. S. Raghukumar, B. Naveen, and D. Lachikarathman, "Predicting the Myocardial Infarction from Predictive Analytics Through Supervised Machine Learning," *SN Computer Science*, vol. 4, no. 4, p. 369, Apr. 2023. [Online]. Available: <https://link.springer.com/10.1007/s42979-023-01775-4>
- [8] X. Yan, L. Zhang, J. Li, D. Du, and F. Hou, "Entropy-Based Measures of Hypnopompic Heart Rate Variability Contribute to the Automatic Prediction of Cardiovascular Events," *Entropy*, vol. 22, no. 2, p. 241, Feb. 2020. [Online]. Available: <https://www.mdpi.com/1099-4300/22/2/241>
- [9] L. Zhang, H. Wu, X. Zhang, X. Wei, F. Hou, and Y. Ma, "Sleep heart rate variability assists the automatic prediction of long-term cardiovascular outcomes," *Sleep Medicine*, vol. 67, pp. 217–224, Mar. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1389945719316466>
- [10] G. A. Tadesse, H. Javed, K. Weldemariam, Y. Liu, J. Liu, J. Chen, and T. Zhu, "DeepMI: Deep multi-lead ECG fusion for identifying myocardial infarction and its occurrence-time," *Artificial Intelligence in Medicine*, vol. 121, p. 102192, Nov. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0933365721001858>
- [11] P. Li, Y. Hu, and Z.-P. Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," *Biomedical Signal Processing and Control*, vol. 66, p. 102474, Apr. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1746809421000719>
- [12] H. Sadr, A. Salari, M. T. Ashoobi, and M. Nazari, "Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models," *European Journal of Medical Research*, vol. 29, no. 1, p. 455, Sep. 2024. [Online]. Available: <https://eurjmedres.biomedcentral.com/articles/10.1186/s40001-024-02044-7>
- [13] R. Yilmaz, F. H. Yagin, A. Raza, C. Colak, and T. C. Akinci, "Assessment of Hematological Predictors via Explainable Artificial Intelligence in the Prediction of Acute Myocardial Infarction," *IEEE Access*, vol. 11, pp. 108 591–108 602, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/10268904/>
- [14] M. De Zambotti, J. Trinder, A. Silvani, I. M. Colrain, and F. C. Baker, "Dynamic coupling between the central and autonomic nervous systems during sleep: A review," *Neuroscience & Biobehavioral Reviews*, vol. 90, pp. 84–103, Jul. 2018. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0149763417306577>
- [15] E. Sharashova, T. Wilsgaard, E. B. Mathiesen, M.-L. Løchen, I. Njølstad, and T. Brenn, "Resting heart rate predicts incident myocardial infarction, atrial fibrillation, ischaemic stroke and death in the general population: the Tromsø Study," *Journal of Epidemiology and Community Health*, vol. 70, no. 9, pp. 902–909, Sep. 2016. [Online]. Available: <https://jech.bmjjournals.org/lookup/doi/10.1136/jech-2015-206663>
- [16] M. Baumert, D. Linz, K. Stone, R. D. McEvoy, S. Cummings, S. Redline, R. Mehra, and S. Immanuel, "Mean nocturnal respiratory rate predicts cardiovascular and all-cause mortality in community-dwelling older men and women," *European Respiratory Journal*, vol. 54, no. 1, p. 1802175, Jul. 2019. [Online]. Available: <http://erj.ersjournals.com/lookup/doi/10.1183/13993003.02175-2019>
- [17] M. P. Than, J. W. Pickering, Y. Sandoval, A. S. Shah, A. Tsanas, F. S. Apple, S. Blankenberg, L. Cullen, C. Mueller, J. T. Neumann, R. Twerenbold, D. Westermann, A. Beshiri, N. L. Mills, On behalf of the MI Collaborative, P. M. George, A. M. Richards, R. W. Troughton, S. J. Aldous, A. R. Chapman, A. Anand, J. Greenslade, W. Parsonage, J. Boeddinghaus, K. Wildi, T. Nestelberger, P. Badertscher, S. Du, J. Huang, S. W. Smith, N. A. Sörensen, and F. Ojeda, "Machine Learning to Predict the Likelihood of Acute Myocardial Infarction," *Circulation*, vol. 140, no. 11, pp. 899–909, Sep. 2019. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.119.041980>
- [18] Z. Wang, Y. Gu, L. Huang, S. Liu, Q. Chen, Y. Yang, G. Hong, and W. Ning, "Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data," *Cardiovascular Diabetology*, vol. 23, no. 1, p. 351, Sep. 2024. [Online]. Available: <https://cardiab.biomedcentral.com/articles/10.1186/s12933-024-02439-0>
- [19] A. Chakraborty, S. Chatterjee, K. Majumder, R. N. Shaw, and A. Ghosh, "A Comparative Study of Myocardial Infarction Detection from ECG Data Using Machine Learning," in *Advanced Computing and Intelligent Technologies*, M. Bianchini, V. Piuri, S. Das, and R. N. Shaw, Eds. Singapore: Springer Singapore, 2022, vol. 218, pp. 257–267, series Title: Lecture Notes in Networks and Systems. [Online]. Available: <https://link.springer.com/10.1007/978-981-16-2164-2-21>
- [20] D. Chicco and L. Oneto, "An Enhanced Random Forests Approach to Predict Heart Failure From Small Imbalanced Gene Expression Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, pp. 2759–2765, Nov. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9274513/>
- [21] Y. Hu, A. Lui, M. Goldstein, M. Sudarshan, A. Tinsay, C. Tsui, S. Maidman, J. Medamana, N. Jethani, A. Puli, V. Nguy, Y. Aphinyanaphongs, N. Kiefer, N. Smilowitz, J. Horowitz, T. Ahuja, G. I. Fishman, J. Hochman, S. Katz, S. Bernard, and R. Ranganath, "A dynamic risk score for early prediction of cardiogenic shock using machine learning," Mar. 2023, arXiv:2303.12888 [cs]. [Online]. Available: <http://arxiv.org/abs/2303.12888>
- [22] Y.-W. Li, X.-D. Yang, Y.-C. Cai, and X.-Y. Kong, "Artificial intelligence-based prediction of acute myocardial infarction mortality risk," in *2023 2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)*. Zhuhai, China: IEEE, Oct. 2023, pp. 164–169. [Online]. Available: <https://ieeexplore.ieee.org/document/10396320/>
- [23] T. Tripathy, M. Das, S. Mohapatra, S. K. Behera, and G. C. Maihi, "Heart Disease Prediction Using Data Analysis and Effective Machine Learning Techniques," in *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*. Bhubaneswar, India: IEEE, Mar. 2024, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/10530842/>
- [24] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, p. 88, Feb. 2023. [Online]. Available: <https://www.mdpi.com/1999-4893/16/2/88>
- [25] Y. Yang, X. Yan, F. Hou, and L. Pan, "[Study on the prediction of cardiovascular disease based on sleep heart rate variability analysis]," *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, vol. 38, no. 2, pp. 249–256, Apr. 2021, place: China.
- [26] A. F. Jahromi, S. Naghsh, A. A. Safavi, and A. Sharafkhaneh, "Predicting Coronary Heart Disease from Sleep-Stage-Specific Electrocardiograms Using Convolutional Neural Networks," in *2024 10th International Conference on Control, Instrumentation and Automation (ICCI)*. Kashan, Iran, Islamic Republic of: IEEE, Nov. 2024, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/document/10768153/>
- [27] C. Sirocchi, A. Bogliolo, and S. Montagna, "Medical-informed machine learning: integrating prior knowledge into medical decision systems," *BMC Medical Informatics and Decision Making*, vol. 24, no. S4, p. 186, Jun. 2024. [Online]. Available: <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02582-4>
- [28] N. Singh and P. Singh, "Medical Diagnosis of Coronary Artery Disease Using Fuzzy Rule-Based Classification Approach," in *Advances in Biomedical Engineering and Technology*, A. A. Rizvanov, B. K. Singh, and P. Ganasala, Eds. Singapore: Springer Singapore, 2021, pp. 321–

- 330, series Title: Lecture Notes in Bioengineering. [Online]. Available: <http://link.springer.com/10.1007/978-981-15-6329-4-27>
- [29] W. Nie, C. Zhang, D. Song, L. Zhao, Y. Bai, K. Xie, and A. Liu, "Deep Reinforcement Learning Framework for Thoracic Diseases Classification via Prior Knowledge Guidance," Jun. 2023, arXiv:2306.01232 [cs, eess]. [Online]. Available: <http://arxiv.org/abs/2306.01232>
- [30] X. Yue, X. Huang, Z. Xu, Y. Chen, and C. Xu, "Involving logical clinical knowledge into deep neural networks to improve bladder tumor segmentation," *Medical Image Analysis*, vol. 95, p. 103189, Jul. 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1361841524001142>
- [31] W. Zhang, F. Liu, C. M. Nguyen, Z. L. Ou Yang, S. Ramasamy, and C.-S. Foo, "Training neural networks with classification rules for incorporating domain knowledge," *Knowledge-Based Systems*, vol. 294, p. 111716, Jun. 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0950705124003514>
- [32] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, and P. W. Wahl, "The Sleep Heart Health Study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, Dec. 1997.
- [33] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, Oct. 2018. [Online]. Available: <https://academic.oup.com/jamia/article/25/10/1351/5026200>
- [34] C. M. Licht, E. J. De Geus, R. Van Dyck, and B. W. Penninx, "Longitudinal Evidence for Unfavorable Effects of Antidepressants on Heart Rate Variability," *Biological Psychiatry*, vol. 68, no. 9, pp. 861–868, Nov. 2010. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0006322310007596>
- [35] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 1–6, Jun. 2004. [Online]. Available: <https://dl.acm.org/doi/10.1145/1007730.1007733>
- [36] N. S. Kumar, K. N. Rao, A. Govardhan, K. S. Reddy, and A. M. Mahmood, "Undersampled K -means approach for handling imbalanced distributed data," *Progress in Artificial Intelligence*, vol. 3, no. 1, pp. 29–38, Aug. 2014. [Online]. Available: <http://link.springer.com/10.1007/s13748-014-0045-6>
- [37] N. Harrington, Q. M. Bui, Z. Wei, B. Hernandez-Pacheco, P. N. DeYoung, A. Wassell, B. Duwaik, A. S. Desai, D. L. Bhatt, P. Agnihotri, R. L. Owens, T. P. Coleman, and K. R. King, "Passive longitudinal weight and cardiopulmonary monitoring in the home bed," *Scientific Reports*, vol. 11, no. 1, p. 24376, Dec. 2021. [Online]. Available: <https://www.nature.com/articles/s41598-021-03105-1>
- [38] G. L. Booth, M. K. Kapral, K. Fung, and J. V. Tu, "Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study," *The Lancet*, vol. 368, no. 9529, pp. 29–36, Jul. 2006. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0140673606689678>
- [39] M. Saarela and S. Jauhainen, "Comparison of feature importance measures as explanations for classification models," *SN Applied Sciences*, vol. 3, no. 2, p. 272, Feb. 2021. [Online]. Available: <https://doi.org/10.1007/s42452-021-04148-9>
- [40] D. Levy, "The Progression From Hypertension to Congestive Heart Failure," *JAMA: The Journal of the American Medical Association*, vol. 275, no. 20, p. 1557, May 1996. [Online]. Available: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1996.03530440037034>
- [41] W. B. Kannel and D. L. McGee, "Diabetes and cardiovascular risk factors: the Framingham study." *Circulation*, vol. 59, no. 1, pp. 8–13, Jan. 1979. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/01.CIR.59.1.8>
- [42] W. Castelli, "Epidemiology of coronary heart disease: The Framingham study," *The American Journal of Medicine*, vol. 76, no. 2, pp. 4–12, Feb. 1984. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/0002934384909525>
- [43] F. E. Kuhn, "Coronary Artery Disease in Women: Risk Factors, Evaluation, Treatment, and Prevention," *Archives of Internal Medicine*, vol. 153, no. 23, p. 2626, Dec. 1993. [Online]. Available: <http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.1993.00410230028006>

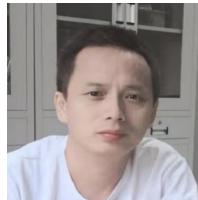


CHANGYUN LI was born in Hunan, China, in 1972. He received the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2007. He is currently the Dean of the College of Information and Intelligence, Hunan Agricultural University, and a full professor of the College of Computer Science, Hunan University of Technology. His research interests include industrial big data analysis, industrial equipment fault diagnosis, intelligent information

perception and processing technology, the Internet of Things, and software methodology. He has published many research articles in international conferences and journals, such as *Journal of Information and Communication Technology (JICT)*, *Journal of Software (JSW)*, and *The Journal of Chemical Physics (JCP)*.



YONGHAN ZHAO was born in Shandong, China, in 1999. He received the B.S. degree in software engineering from Hunan University of Technology, Zhuzhou, China, in 2021, where he is currently pursuing the M.S. degree in computer science and technology. Her research interests include healthcare data science and artificial intelligence.



ZHIBING WANG was born in Hunan, China, in 1974. He received the M.S. degree in computer science and technology from the Hunan University of Technology, Zhuzhou, China, in 2010. He is currently an Associate Professor with the School of Computer Science, Hunan University of Technology. His research interests include industrial Internet, trusted software, and knowledge graph. He has published many research articles in international conferences and journals, such as *Journal of Information and Communication Technology (JICT)*, *Journal of Software (JSW)*, and *International Conference on Computer Integrated Manufacturing and Automation (ICDMA)*.



QIUIHUI MO was born in Hunan, China, in 1998. He received the B.S. degree in software engineering from Southwest Minzu University, Zhuzhou, China, in 2021. He is currently pursuing the M.S. degree in computer science and technology at Hunan University of Technology, Zhuzhou, China. His research interests include big data and cloud computing.



XI XU was born in Hunan, China, in 1978. He received the Ph.D. degrees in electronics and communication engineering from Guizhou University, Guiyang, China. He is currently a lecturer at the School of Computer Science, Hunan University of Technology. His research interests include Internet of Things technology, signal processing, and radio frequency communication technology. He has published many research articles in international conferences and journals, such as *Cluster Computing*, *International Conference on Networking and Digital Society (ICNDS)*.