Applied Mathematics and Computational Intelligence
Volume 13, No. 2, 2024 [100-112]

A Review on Predictive Model for Heart Disease using Wearable Devices Datasets

Mohd Syafiq Asyraf Suhaimi, Nor Azuana Ramli*, Noryanti Muhammad

Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan Pahang, Malaysia

* Corresponding author: azuana@umpsa.edu.my

Received: 23 November 2023
Revised: 2 April 2024
Accepted: 30 April 2024

ABSTRACT

Heart diseases were the number one killer in Malaysia based on the data from the Department of Statistics Malaysia in the previous year. Heart diseases were the principal causes of death for the population aged 41 and above. Many studies have discovered the factors that cause heart disease and ways to prevent it. Among the ways to prevent heart disease include analysis of the patient's historical data, developing predictive modeling involving statistical and machine learning techniques, and monitoring health conditions through wearable devices. This paper reviewed the predictive model applied in heart disease prediction using wearable device datasets. Artificial neural networks (ANNs) have grown in popularity in data mining and machine learning for their ability to classify input data into several categories by detecting hidden connections in the data, which is beneficial in predicting correct classifications. Other approaches, such as Naive Bayes, Support Vector Machine, and Decision Tree algorithms, are used to analyze medical data sets to forecast cardiac disease. According to the survey, ANNs are likely to be good for heart disease prediction in terms of classification accuracy on training and test datasets.

Keywords: Heart Disease, Machine Learning, Predictive Modeling, Wearable Devices,

1 INTRODUCTION

Heart disease is a condition that affects the heart or blood vessels. Smoking, high blood pressure, high cholesterol, a poor diet, a lack of exercise, and obesity can all raise the risk of various cardiac problems. Coronary artery disease (narrow or blocked coronary arteries) is the most prevalent type of heart disease, and it can cause chest discomfort, heart attacks, or stroke. Other cardiac illnesses include congestive heart failure, irregular heartbeats, congenital heart disease (heart disease that develops at birth), and endocarditis (inflamed inner layer of the heart), which is also known as cardiovascular disease [1]. Heart disease refers to a group of disorders that affect the heart. Heart illnesses include blood vessel disease, such as coronary artery disease, heart rhythm disorders (arrhythmias), congenital disabilities (congenital heart defects), heart valve disease, heart muscle disease, heart infection, and heart failure [2]. Many types of heart disease can be avoided or treated with a healthy lifestyle and early detection. This early prediction is possible with modern technologies such as predictive modeling in machine learning.

In recent years, the use of applications using mathematical and statistical models in healthcare has been increasing as data size has increased and the constraints on computational power have gradually disappeared. According to a study by [3], the number of studies using data mining in healthcare is increasing. Among them, 51.5% of studies using data mining and 39.3% of studies using machine learning methods were reported. As such, many studies on heart disease prediction use data mining and machine learning. Machine learning is a growing field of computing algorithms that aim to mimic human intelligence by learning from their surroundings. They are regarded as the workhorse in the new era of big data. Machine learning techniques have been effectively employed in various industries, including pattern recognition, computer vision, aerospace engineering, finance, entertainment, computational biology, and biological and medical applications [4, 5]. Moreover, the size of the data typically increases from day to day. The demand to comprehend big, complex, information-rich data sets has grown in all technology, industry, and research disciplines. In today's competitive world, the capacity to extract relevant knowledge concealed in massive amounts of data and act on that knowledge is becoming increasingly vital. Data mining is obtaining knowledge from data using computer-based information systems (CBIS), including innovative methodologies [5].

The healthcare business is a data-rich environment because it generates enormous volumes of data, such as electronic medical records, administrative reports, and other benchmarking findings [6]. Data mining in healthcare is mainly used to anticipate various diseases and to aid clinicians in making therapeutic decisions. Most of the studies predict cardiovascular disease using data attributes from the patients' data that is now available online. This study investigated a cardiovascular disease prediction model by considering only data that can be measured using wearable devices.

2 LITERATURE REVIEW

Global cardiovascular disease risk assessment is traditionally based on clinical risk scores that estimate the 10-year risk. However, only some of these scores capture the dynamic changes in personalized risk that closely follow lifestyle habits. Incorporating subjective lifestyle behaviors in risk assessment has been challenging; therefore, objective data derived from wearables provide a renewed opportunity to make cardiovascular disease risk assessment more accurate, comprehensive, and dynamic over a lifetime.

A study by [7] created a heart rate monitor system that leverages Apple Watch to collect user data. Apple has provided the Apple Watch Application Programming Interface (API) so developers can readily obtain bodily information. The API is known as the "Health Kit." The researchers can only obtain the heart rate from the data set since the Apple Watch keeps bodily information in the data set. K-means were used to cluster the heart rate after filtering the data. When the heart rate is abnormal, the system will notify the user. The researcher began by categorizing heart rate patterns into three categories: high, normal, and low, and then included K- means as one characteristic. Then, 12 characteristics were added to forecast the outcome with the highest effect. Finally, the researchers created a 24-hour pie chart to depict heart rate fluctuations. Users may track their 24-hour heart rate fluctuations. Figure 1 depicts the flowchart.

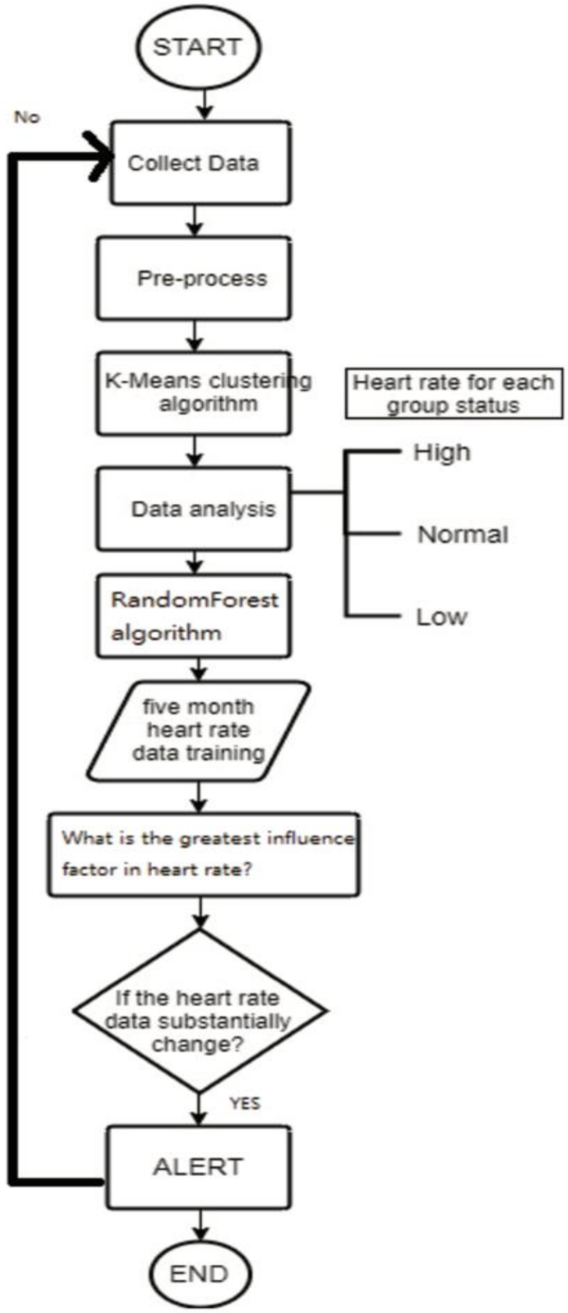

Figure 1 : System flowchart [7]

Based on the 12 parameters used, the researchers developed a Random Forest algorithm. The settings determine the 12 characteristics and 500 trees. The "High-temperature" parameter has the most significant impact, as seen in Table 1. The error value is 0.1116993, according to mean squared error (MSE) calculations. The senior may stay at home while the temperature is high, but their heart will be weaker than previously. It should be cautious of the temperature, which varies significantly from day to night.

Table 1 : The importance parameters part of heart rate [7]

Parameter	Value
Kmeans_group	139938.56
steps	183558.33
High_temp	230426.16
Low_temp	125756.13
Temp_status	19433.60
humidity	87405.96

Another study by [8] improved heart rate variability (HRV) measures using consumer wearables and machine learning. The researchers' goal was to minimize smartwatch HRV readings by employing more sensor information to provide accurate HRV measurements and to increase the reliability and quality of the root means square of successive differences (RMSSD). Data from a chest-based heart rate monitor (Firstbeat Bodyguard 2) was compared to data from one of the most popular consumer smartwatches, which has an accelerometer and an optical heart rate sensor capable of monitoring inter-beat intervals.

The raw data from the heart rate monitor was collected using the First Beat SPORTS Individual program. Data was collected from the consumer smartwatch using a special smartphone software and the wearable's Bluetooth streaming functionality. A healthy person wore both devices simultaneously for 72 hours and collected data. Following pre-processing, around 200,000 observations were left, which included heart rate monitor inter-beat interval, consumer smartwatch inter-beat interval, and consumer smartwatch three-axis accelerometer data. The researchers divided the dataset into 80% for training and 20% for testing.

The first trial resulted in a reduction in the root-mean-squared error (RMSE) between heartbeats. The first experiment reduced RMSE between heart rate monitor and consumer smartwatch RMSSDs on the test set from 48.89 to 28.50. After correcting, the statistically significant correlation between heart rate monitor and consumer wristwatch RMSSDs in the test set increased from $r = 0.37$, $p=0.001$ before to $r = 0.58$, $p=0.001$.

In [9], a machine learning and Internet of Things (IoT)-based cardiac disease identification system using Higher Order Boltzmann Deep Belief Neural Network (HOBDBNN) was built. The IoT was utilized to collect health information from patients via embedded sensor devices. The gadgets recorded patients' electrocardiogram (ECG), blood pressure, chest pain typologies, cholesterol levels, vascular information, lowest and maximum heart rates, and angina and depression symptoms. These data were processed by the automated heart disease prediction system, which employs an improved machine learning algorithm. All these steps are summarized as shown in Figure 2.

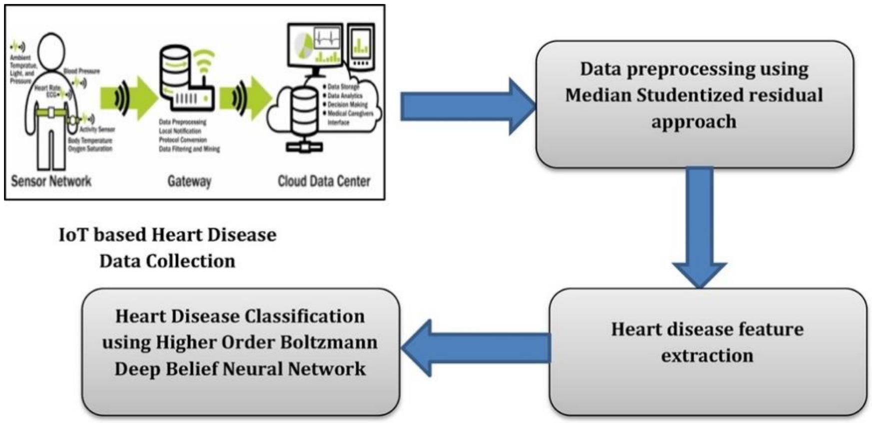

Figure 2: Heart disease prediction steps and structure [9]

The MATLAB tool was used to create the HOBDBNN approach-based illness prediction system. During the implementation phase, the system splits the total quantity of data into 70% for training and 30% for testing to confirm the heart disease prediction system's efficiency. Approximately 1500 patient details were gathered, and analysed according to the stated protocol to assess the system's efficiency. Following the deployment of the suggested device, the system's efficiency was assessed using a variety of performance measures, including F-measure, sensitivity, specificity, loss function, and receiver operating characteristic (ROC) curve.

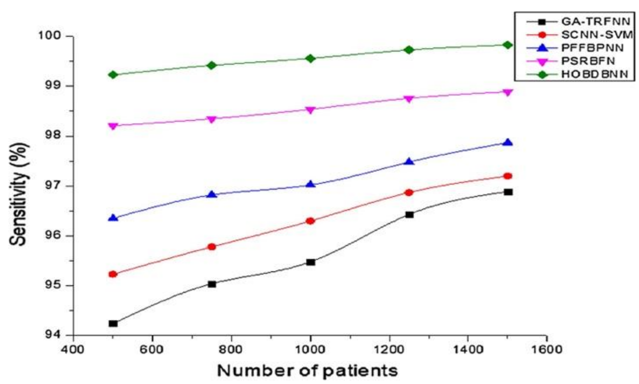

Figure 3: Sensitivity [9]

Based on Figure 3, the HOBDBNN technique outperforms other classifiers such as GA-TRFNN (95.6%), SCNN-SVM (96.27%), PFFBPNN (97.01%), and PSRBFN (98.55%).

Another study was conducted by [10], where an IoT-based cardiac disease identification system based on machine learning was developed. The system used the Waikato Environment for Knowledge Analysis (WEKA) to analyze heart disease data obtained from the cloud. The information was analyzed using a Support Vector Machine (SVM), classifying normal and cardiovascular disease traits with up to 97.53 percent accuracy. With the assistance of an IoT device, the system was created to collect heart disease details such as body temperature, blood pressure, heartbeat, and humidity level. The acquired data were compiled and analyzed in the health care center using the SVM learning technique. This method has been used repeatedly to identify heart problems correctly. Although this system detects heart problems quickly, it cannot manage real-time cardiac information when there is a large volume of data in the system.

[11] conducted the LINK-HF research (Multisensor Non-invasive Remote Monitoring for Prediction of Heart Failure Exacerbation) to determine the accuracy of machine learning analytics in a remote patient monitoring system in predicting HF readmission to the hospital. Secondary goals included assessing subject adherence to research protocols.

Figure 4 displays the analytical platform sensitivity at an 85 percent specificity at a discriminating threshold of 0.03. The clinical alert's specificity of 85 percent was chosen to illustrate where it may be anchored for clinical usage. This specificity is helpful for clinical decision-making and is similar to the specificity reported by the MultiSENSE algorithm, which is now sold for clinical usage.

Positive Window Methodology	**Sensitivity**	**Specificity**
Ten-day positive window		
HF hospitalization	76.0%	84.8%
Unplanned nontrauma hospitalization	68.6%	84.7%
Event-specific window		
HF hospitalization	87.5%	86.0%
Unplanned nontrauma hospitalization	77.1%	86.0%
HF indicates heart failure.		

Figure 4: Analytical platform sensitivity and specificity by positive window type and hospitalization type [11]

[12] created a smartphone-based platform solution for wearable cardiovascular disease (CVD) diagnosis capable of real-time ECG capture and display, feature extraction, and beat classification. Researchers created two proof-of-concept prototypes to prove that the suggested methodologies are appropriate for high-end and low-end smartphones and to undertake real-time performance assessments on real devices.

The study used a plug-in-based graphical user interface (GUI) platform and a machine-learning-based platform. In constructing and testing our plug-in-based real-time CVD monitoring system, researchers employed an Alive Bluetooth ECG heart monitor, an Amoi E72 Microsoft Windows Mobile 5 Smartphone, MATLAB R2008b (for initial algorithm design and validation), and Microsoft Visual Studio 2008. The Machine-Learning-Based Platform was built using an Alive Bluetooth ECG heart monitor and an HTC Microsoft Windows Mobile 6 Smartphone. Before moving to LabVIEW, researchers used MATLAB R2008b to create and test our ANN-based ECG processing and CVD classification algorithms.

Table 2 summarizes the findings. Except for the 81% prediction accuracy for fusion of paced and normal beat (PFUS), which could be attributed to the varying morphologies of fusion complexes, which change depending on the portion of the ventricles depolarized by each of the activation fronts, a prediction accuracy of greater than 90% was achieved[13].

Table 2: Predication accuracy of normal and four abnormal beats (Unit in percentage) [13]

Train	A		B		C		Average
Test	B	C	A	C	A	B	
Normal	99.7	100	99.7	95.5	99.5	100	99
RBBB	98.2	98.7	97.8	98.6	97.2	97.6	98
PVC	93.3	93.2	88.0	95.5	92.0	93.3	92.6
PACE	96.4	97.1	95.6	96.4	95.3	95.5	96
PFUS	81.8	83.3	71.4	91.7	85.7	72.7	81

*RBBB: right bundle branch block beat, PVC: premature ventricular contraction, PACE: paced beat, and PFUS: fusion of paced and normal beat.

A study by [14] proposed an intelligent healthcare monitoring system (SHMS). The structure is divided into different layers to thoroughly describe the information groundwork of each stage in the proposed system. Finally, the structure of the ensemble deep learning model and ontology is presented, which the SHMS employs to predict heart disease in patients and recommend dietary plans and activities. The system used a wireless body sensor network (WBSN) based on medical sensors to collect internal and external physiological data, such as an ECG, an electroencephalogram (EEG), an electromyogram (EMG), heart rate, blood pressure (BP), position, activities, respiration rate, and blood sugar, oxygen saturation, and cholesterol levels of the patient for daily health monitoring.

Table 3: Comparison results for the proposed model against other classifiers using general and specific feature weights [14]

Classifier model	General feature weighting method						Specific feature weighting method					
	Pre (%)	Rec (%)	Acc (%)	FM (%)	RMSE	MAE	Pre (%)	Rec (%)	Acc (%)	FM (%)	RMSE	MAE
SVM	81.0	69.8	69.8	66.6	0.54	0.30	87.5	81.5	84.4	84.5	0.39	0.15
Logistic regression	70.5	70.3	70.3	70.2	0.53	0.30	89.2	95.2	92.2	92.2	0.22	0.11
MLP	81.5	81.1	81.1	81.1	0.38	0.24	93.3	85.3	89.3	89.3	0.27	0.10
Random forest	74.4	70.8	70.7	69.4	0.45	0.43	87.4	87.4	87.3	87.4	0.34	0.13
Decision tree	75.5	75.7	75.4	75.5	0.40	0.30	84.6	77.7	77.6	77.6	0.39	0.31
Naive Bayes	83.1	82.4	82.4	82.4	0.35	0.19	88.8	78.5	83.4	83.4	0.38	0.36
The proposed model	84.9	84.9	84.9	84.9	0.32	0.25	98.2	96.4	98.5	97.2	0.21	0.12

The results of the proposed system and existing classifiers utilizing both general and feature weighting strategies are shown in Table 3. The outcomes of all classifiers in the general feature weighting technique must be revised compared to earlier experimental results from feature selection. Only the multilayer perceptron (MLP) decision tree and suggested model have gained accuracy, while the SVM, logistic regression, Random Forest, and Naïve Bayes have deteriorated. Furthermore, the RMSE and mean absolute error (MAE) of MLP and the decision tree were reduced, but the error rates of other classifiers increased. According to the results, the general feature weighting approach may not identify the proper feature importance for all classes. However, using an uncertain combination operation may set the feature importance for differentiation, lowering the accuracy of prediction models, and increasing the error rate.

[15] investigated the proposed intelligent healthcare monitoring system (SHMS) structure in depth. The overall structure of the SHMS was outlined first. The structure was structured into layers to completely define each stage's information foundation in the proposed system. Finally, the SHMS used the structure of the ensemble deep learning model and ontology to forecast heart disease in patients and prescribe dietary programs and activities.

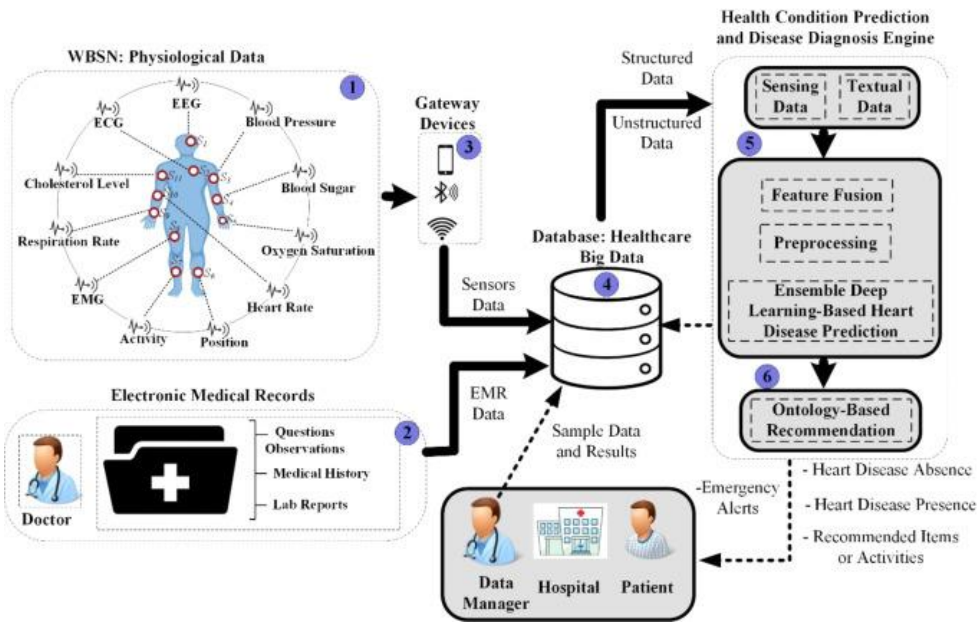

Figure 5: The smart healthcare monitoring system's framework for predicting heart illness [15]

For daily health monitoring, the system employs a wireless body sensor network (WBSN) based on medical sensors to collect internal and external physiological data such as ECG, EEG, EMG, the patient's heart rate, BP, position, activities, respiration rate, and blood sugar, oxygen saturation, and cholesterol levels as shown in Figure 5. The proposed SHMS approach, as illustrated in Figure 6, employs four consecutive layers: data collection, data fusion and feature extraction, data preprocessing, and illness prediction and recommendation.

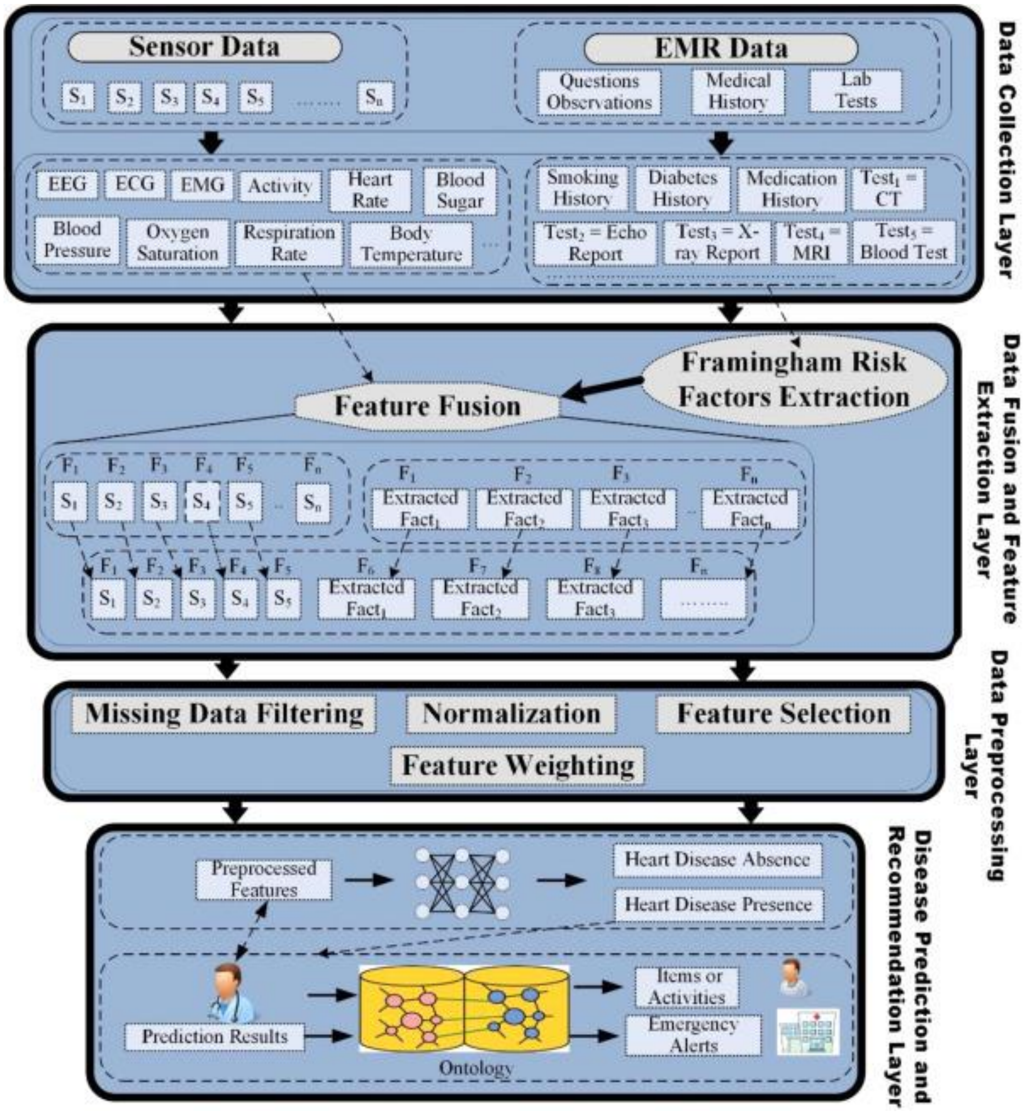

Figure 6: The heart disease prediction and diagnostic engine's information framework [15]

Figure 7 shows the experimental results for the proposed model and the other six classifiers based on the feature selection strategy, including precision, recall, accuracy, F-measure, RMSE, and MAE.

Classifier model	Before the proposed feature selection approach						After the proposed feature selection approach					
	Pre (%)	Rec (%)	Acc (%)	FM (%)	RMSE	MAE	Pre (%)	Rec (%)	Acc (%)	FM (%)	RMSE	MAE
SVM	68.1	63.3	63.2	60.0	0.60	0.36	81.9	71.8	71.8	69.3	0.53	0.28
Logistic regression	56.6	56.3	56.3	56.2	0.64	0.43	73.8	73.8	73.7	73.7	0.50	0.28
MLP	60.2	60.0	60.0	60.0	0.57	0.40	77.9	77.7	77.6	77.6	0.39	0.25
Random forest	63.2	63.3	63.2	63.2	0.48	0.46	78.5	73.8	73.7	72.6	0.45	0.42
Decision tree	56.9	56.3	56.3	54.4	0.50	0.47	74.8	74.8	74.8	74.7	0.41	0.30
Naive Bayes	67.1	66.5	66.5	66.4	0.49	0.38	80.5	80.5	80.4	80.5	0.34	0.22
The proposed model	72.8	72.2	72.2	72.1	0.43	0.31	84.5	82.5	83.5	83.5	0.32	0.25

Figure 7: Results of the proposed model's comparison with other classifiers before and after feature selection [15]

All the studies reviewed in this paper is summarized as in Table 4 below:

Table 4: A review of numerous articles on the prediction of heart disease using wearable devices.

Year	Authors	Paper Title	Wearable Devices Used	Methods Used	Accuracy Obtained
2008	B. Surawicz et. al	Chou's Electrocardiography in Clinical Practice: Adult and Pediatric	Wireless Body Sensor Network (WBSN)	MLP Decision Tree Proposed Model	93.30% 87.40% 98.20%
2017	F. Ahmed et. al	An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients	IoT-based cardiac disease identification system based using Waikato Environment for Knowledge Analysis (WEKA)	Support Vector Machine (SVM)	97.53%

2017	Jose AntonioGutiérrez-Gnecchi et. al	DSP-based arrhythmia classification using wavelet transform and probabilistic neural network	Alive Bluetooth ECG heart monitor, an Amoi E72 Microsoft Windows Mobile 5 Smartphone, MATLAB R2008b (for initial algorithm design and validation), and Microsoft Visual Studio 2008	Probabilistic Neural Network (PNN)	92.90%
2018	M.C. Chen et. al	Combining Smartwatch and Environments Data for Predicting the Heart Rate	Apple Watch (using Apple Watch Application Programming Interface)	K-Means	88.83%
2019	Z. Al-Makhadmeh et. al	Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach	Internet of Things (IoT)-based cardiac disease identification system	Higher Order Boltzmann Deep Belief Neural Network (HOBDBNN) GA-TRFNN SCNN-SVM PFFBPNN PSRBFN	99.55% 95.60% 96.27% 97.01% 98.55%
2019	Martin Maritsch et. al	Improving heart rate variability measurements from consumer smartwatches with machine learning	Chest-based heart rate monitor (Firstbeat Bodyguard 2)	Convolutional Neural Network (CNN)	71.50%
2020	J. Stehlik et. al	Continuous Wearable Monitoring Analytics Predict Heart Failure Hospitalization	Multisensor Non-invasive Remote Monitoring for Prediction of Heart Failure Exacerbation	Multivariate Change Index (MCI)	85.00%

2020	Farman Ali et. al	A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion	Wireless Body Sensor Network (WBSN)	SVM Logistic Regression MLP Random Forest Decision Tree Naïve Bayes	71.8% 73.7% 77.6% 73.7% 74.8% 80.4%

3 CONCLUSION

This research addressed and surveyed many techniques and approaches to wearables and prediction algorithms used to identify cardiac disease and performance measures to evaluate the accuracy of different classifiers. Furthermore, for classifications, researchers employed several methodologies, such as artificial neural networks, Naïve Bayes, and various types of wearables to collect data. According to the survey, ANNs are likely to be suitable for heart disease prediction in terms of classification accuracy on training and test datasets. It is mainly used for heart disease prediction with a large dataset. Finally, most of the studies evaluated the performance of the classifiers using sensitivity, specificity, and accuracy.

ACKNOWLEDGEMENT

This paper was created within the project "Development of a predictive algorithm for heart disease using an ensemble of logit model and nearest neighbour tree method based on smartwatch data". Project registration number RDU210328 under UMPSA Internal Grant.

REFERENCES

[1] NCI Dictionary of Cancer Terms. (2022). National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/heart-disease\

[2] Heart disease - Symptoms and causes. (2021, February 9). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118

[3] T. Anand, R. Pal , S.K. Dubey, "Data mining in healthcare informatics: Techniques and applications," in *Proceedings - 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 4023-4029.

[4] I. El Naqa & M. J. Murphy, "What Is Machine Learning?" in *Machine Learning in Radiation Oncology: Theory and Applications*, I. El Naqa, R. Li & M. J. Murphy, Eds. Switzerland: Springer Cham, 2015, pp. 3–11.

[5] G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, "The use of computer-based information systems

by German managers to support decision making," *Inf. Manag.*, vol. 41, no. 6, pp. 763–779, 2004.

[6] N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta, "Knowledge Management in Healthcare," vol. 63, pp. 5–18, 2005.

[7] M.C. Chen, R.C. Chen and Q. Zhao, "Combining Smartwatch and Environments Data for Predicting the Heart Rate," in *Proceedings - 2018 IEEE International Conference on Applied System Invention (ICASI)*, 2018, pp. 661-664.

[8] Martin Maritsch, Caterina Bérubé, Mathias Kraus, Vera Lehmann, Thomas Züger, Stefan Feuerriegel, Tobias Kowatsch, and Felix Wortmann, "Improving heart rate variability measurements from consumer smartwatches with machine learning," in *Proceedings - 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*, 2019, pp. 934–938.

[9] Z. Al-Makhadmeh & A. Tolba, "Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach," *Measurement*, vol. 147, 106815, 2019.

[10] F. Ahmed, "An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients," *International Journal of Computer Applications*, vol. 164, no.6, pp. 36–40, 2017.

[11] J. Stehlik, C. Schmalfuss, B. Bozkurt, J. Nativi-Nicolau, P. Wohlfahrt, S. Wegerich, K. Rose, R. Ray, R. Schofield, A. Deswal, J. Sekaric, S. Anand, D. Richards, H. Hanson, M. Pipke & M. Pham, "Continuous Wearable Monitoring Analytics Predict Heart Failure Hospitalization, *Circulation: Heart Failure*, vol.13, no. 3, 2020. [Online].

[12] J. J. Oresko, H. Duschl, A.C. Cheng, "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 3, pp. 734-740, 2010.

[13] Gutiérrez-Gnecchi, J. A., Morfin-Magaña, R., Lorias-Espinoza, D., Tellez-Anguiano, A. D. C., Reyes-Archundia, E., Méndez-Patiño, A., & Castañeda-Miranda, R. (2017). DSP-based arrhythmia classification using wavelet transform and probabilistic neural network. Biomedical Signal Processing and Control, 32, 44–56.

[14] B. Surawicz and T. K. Knilans, *Chou's Electrocardiography in Clinical Practice: Adult and Pediatric*. Philadelphia, PA: Saunders Elsevier, 2008, pp. 607

[15] Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, Kyung-Sup Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, 2020, pp. 208-222.