Article citation info:
Shaker AH, Gharghan SK, Ibrahim IA, Ayyal AH, Chisab RF. Machine learning algorithms for predicting heart failure based on heart sounds captured by electronic stethoscope. Diagnostyka. 2026;27(3):2026312. http://doi.org/10.29354/diag/231602.

DIAGNOSTYKA, 2026, Vol. 27, No. 3

e-ISSN 2449-5220
DOI: 10.29354/diag/231602

MACHINE LEARNING ALGORITHMS FOR PREDICTING HEART FAILURE BASED ON HEART SOUNDS CAPTURED BY ELECTRONIC STETHOSCOPE

Ali Hussein SHAKER [1] *, Sadik Kamel GHARGHAN [2] *, Ibrahim Amer IBRAHIM [3], Ahmed Hammed AYYAL [4], Raad Farhood CHISAB [5]

[1] Ministry of Health, Karbala Health Directorate, Al-Hindiya Sector for Primary Healthcare, Karbala, Iraq
[2] Middle Technical University, Technical Engineering College of Artificial Intelligence, Baghdad, Iraq
[3] Biomedical Engineering Department, Al Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq
[4] Middle Technical University, Kut Technical College, Wasit, Iraq
[5] Middle Technical University, Technical Institute / Kut, Wasit, Iraq
* Corresponding author, e-mails: ali.hussein.shaker94@gmail.com, sadik.gharghan@mtu.edu.iq

Abstract

Heart failure (HF) is a critical medical illness that can be deadly if not diagnosed or predicted in time, as it results from the heart's inability to pump enough blood throughout the body. Traditional diagnostic tools such as electrocardiograms (ECGs), phonocardiograms (PCGs), and stethoscopes have been inadequate for early detection of heart failure. This study leveraged recorded heart-sound data from an electronic stethoscope, converting it into NumPy arrays using Mel-Frequency Cepstral Coefficients (MFCC) to predict heart failure using machine learning (ML) algorithms. The study utilized four ML algorithms: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost). The dataset, comprising 13 MFCC-derived audio features for 300 subjects, was split into 80% for training and 20% for testing using 10-fold cross-validation. To ensure the validity of the reported performance and prevent data leakage, strict patient-wise data separation was enforced, and no feature engineering was conducted after the split. The ML models' performance was assessed using accuracy, sensitivity, precision, specificity, F1-score, precision-recall curve, and Area Under the Curve (AUC) to identify the most accurate model. SVM and XGBoost emerged as the most accurate models in the testing set, achieving 100% accuracy across all metrics and the highest AUCs of 100% for RF, SVM, and XGBoost. Based on heart sound data and ML models, the proposed study has demonstrated superior predictive performance for HF compared with previous studies.

Keywords: electronic stethoscope, heart failure prediction, heart sounds data, machine learning, mel frequency cepstral coefficients

List of Symbols/Acronyms

AI – Artificial Intelligence;
AUC– Area Under the Curve;
AV – Atrioventricular;
BI-LSTM – Bidirectional Long Short-Term Memory;
BNP – B-Type Natriuretic Peptides;
CAD– Coronary Artery Disease;
CHF – Chronic Heart Failure;
CNN– Convolutional Neural Network;
CONV-RF – Convolutional-Random Forest;
CVD – Cardiovascular Disease;
DL – Deep Learning;
EF – Ejection Fraction;
FN – False Negative;
FP – False Positive;
GB – Gradient Boosting;
HD – Heart Disease;
HF – Heart Failure;
KNN – K-Nearest Neighbors;
MFCC – Mel Frequency Cepstral Coefficient;
ML – Machine Learning;
RF – Random Forest;
SNR – Signal to Noise Ratio;
SVM – Support Vector Machine;
TN – True Negative;
TP – True Positive;
VHD – Valvular Heart Disease;
XGBOOST – Extreme Gradient Boosting;

1. INTRODUCTION

Heart Failure (HF) is a serious health condition in which the heart fails to pump enough blood to meet the body's metabolic needs [1]. This failure can be the result of several risk factors, such as obesity,

Received 2026-03-18; Accepted 2026-08-23; Available online 2029-09-01

high blood pressure, coronary artery disease (CAD), heart infections, heart valve stenosis, and hereditary heart disease (HD) [2]. Symptoms of HF are evident, such as shortness of breath, chronic dry cough, extreme fatigue, swollen feet and legs, and heart palpitations. There is also a so-called Ejection Fraction (EF), which expresses the ability of the left ventricle to pump blood with each contraction and is measured as a percentage; if the percentage is less than 40%, it indicates that the person has HF. However, if the percentage exceeds 40%, the person does not have HF [3, 4]. Laboratory tests analyzing blood protein levels, such as natriuretic peptides (BNP and N-terminal ProBNP) [5], echocardiography [6, 7], and electrocardiography [8], are commonly used for diagnosis. HF can be treated with lifestyle changes, appropriate medications, and, in some cases, corrective surgery. HF can also be prevented by exercising routinely, eating healthy foods, and avoiding smoking and alcohol [9].

Statistics show that deaths from HF are worrying, with the disease being a leading cause of death globally [10]. A recent examination of UK primary care data indicated that the number of persons with HF increased by 23% between 2002 and 2014, from 750,125 to 920,616, constituting 1.4% of the population [11]. HF is a rapidly progressing disease that affects 26 million people worldwide and is growing [12].

Machine learning (ML) techniques can enhance HF prediction and treatment by pinpointing high-risk factors and generating precise profiles. These models assist with risk screening and classification, suggest suitable tests, alleviate doctors' workload, and conserve resources. Consequently, they can reduce mortality from HF and improve patients' quality of life. Moreover, individuals should prevent HF by maintaining a healthy lifestyle and taking care of their health.

Chowdhury et al. [13] introduced a prototype of an innovative digital stethoscope designed to monitor and diagnose heart sounds continuously in real time. It addresses the need for a portable, intelligent system for early detection of cardiovascular disease (CVD). The stethoscope captures heart sounds, filters them, and wirelessly transmits them to a computer for analysis. The system comprises two subsystems: one for sensing and one for intelligent detection. Acoustic signals are acquired, amplified, filtered, and converted by an Arduino microcontroller before being sent to a computer via Bluetooth for MATLAB processing. The system achieved 88% accuracy in classifying normal heart sounds and 97% in detecting abnormal heart sounds. It offers high accuracy in early CVD detection with low-power wireless connectivity for real-time analysis. However, potential disturbances in radio communication and the absence of comparisons with other methods in the field hinder the model's evaluation.

Omarov et al. [14] discussed the development of a digital stethoscope prototype to diagnose heart abnormalities in real time and employ ML techniques. The research proposed an early warning device that detects cardiac irregularities via acoustic changes, addressing the challenges of dynamic auscultation and reliance on echocardiography. The system comprises three subsystems: a portable electronic stethoscope, an ML decision-making unit to differentiate between normal and abnormal heart sounds, and a results presentation and visualization unit. The 15-second portable device diagnoses normal and abnormal heart sounds with over 90% accuracy, including 93.5% for normal and 93.25% for abnormal. It supports remote monitoring. It cannot determine the need for heart surgery and is expensive, which may limit its use.

Zeinali and Niaki [15] proposed analyzing stethoscope heart sounds using signal processing and ML algorithms to improve CVD diagnosis. This novel method classified heart sounds into multiple categories, unlike previous studies that classified them as normal or abnormal. The researchers used classifiers such as Gradient Boosting (GB), support vector machine (SVM), RF, and genetic algorithms for feature selection. The dataset was aggregated using K-means and DbScan. Their method achieved 87.5% accuracy on multi-class data and 98% accuracy on binary classification. GB led with 87.5% accuracy, followed by RF (81.25%) and SVM (75%). The algorithms' advanced feature selection methods performed better on multi-featured datasets than traditional methods. However, accuracy declined as the number of input variables increased.

Ali et al. [16] introduced Lu-Net, a Deep Learning (DL) based noise reduction model, to improve HD detection using stethoscope-recorded heart sound signals. Lu-Net addresses environmental and physiological noise in these signals. It employs a convolutional deep-decoding architecture with bidirectional long short-term memory (Bi-LSTM) to capture rhythmic patterns in heart sounds, outperforming current U-Net and Full Convolutional Network models, achieving significant signal-to-noise ratio (SNR) improvements. Lu-Net achieves an average SNR improvement of 5.613 dB and 1.750 dB, respectively, using only 1.32 M parameters. It also enhances classification accuracy for detecting HD in noisy environments, particularly in the PASCAL cardiac sounds dataset. However, its high computational complexity may limit its real-time applicability in resource-constrained settings.

Sinha et al. [17] proposed the convolutional-random forest (Conv-RF) model for the classification and analysis of valvular heart disease (VHD). To solve VHD classification problems using DL, the model uses a CNN-based SqueezeNet and an RF classifier. The goal was to classify seven heart valves using model strengths. The researchers used an ESP32-powered electronic stethoscope and a Raspberry Pi. The Conv-RF model improved feature extraction and classification. SqueezeNet achieved

98.65% accuracy, and Conv-RF achieved 99.37%, demonstrating high sensitivity and privacy across datasets. The system outperformed individual models and other modern approaches regarding accuracy, cost-effectiveness, portability, and suitability for real-time applications. However, it exhibited limitations in eliminating ambient noise and noise from microphone movement during auscultation.

Nishitha and Chittesh [18] introduced a novel ML-DL method to detect chronic heart failure (CHF) from heart sounds. The lack of automated methods for detecting heart sounds in CHF motivated the work. The researchers wanted to improve heart sound classification by providing a reliable detection method. ML methods with diverse features and DL methods based on signal-time spectroscopy are used. A high accuracy of 96.34% and a low error rate of 3.76% were achieved on 6 PhysioNet datasets with 947 participants and one personal dataset. The method outperforms others, highlighting its efficacy in classifying various heart sounds. However, limitations include sparse data, which can limit its applicability, and a focus on accuracy without addressing sensitivity, privacy, or other criteria.

In this paper, the heart sound data were collected from an electronic stethoscope to predict HF. Section 1 provides an introduction and an extensive review of previous studies on HF prediction, comparing various ML and DL techniques, including the number of data points, ML type, and limitations. Section 2 discusses the concept of heart sounds, data characteristics, features used, electronic stethoscopes, and data preprocessing in Python. Section 3 details each algorithm used. Section 4 elaborates on the evaluation of performance metrics. Sections 5, 6, and 7 present and compare the results based on known performance metrics, confusion matrices, the area under the curve (AUC), and precision-recall curves. The best ML algorithm prediction accuracy is also compared with previous studies. Finally, Section 8 presents a conclusion for this paper.

The contributions of this paper are as follows:

1. An electronic stethoscope system was implemented to record heart sounds, using ML algorithms and Python to predict HF in real time.
2. Heart sound data were collected from healthy and unhealthy individuals using an electronic stethoscope, then trained and evaluated using ML algorithms.
3. Four ML algorithms were assessed using various performance metrics to identify the one with the highest prediction accuracy.
4. It was found that GB, RF, and Extreme Gradient Boosting (XGBoost) achieved 100% accuracy, surpassing the other algorithms used in earlier studies and outperforming the models used in this paper.

Table 1 summarizes previous studies by their methods, accuracy, and limitations.

Table 1. Summary of related works

Study	Methodology	Accuracy	Limitations
Chowdhury et al. [13]	Arduino-based digital stethoscope, Bluetooth transmission, and MATLAB processing for intelligent detection.	88% (Normal) 97% (Abnormal)	• Potential radio communication disturbances. • Lack of comparison with other existing methods.
Omarov et al. [14]	A portable electronic stethoscope with a Machine Learning (ML) decision-making unit for a 15-second diagnosis.	93.5% (Normal) 93.25% (Abnormal)	• High cost. • Cannot determine the need for heart surgery.
Zeinali & Niaki [15]	Signal processing, K-means/DbScan for clustering, and classifiers such as Gradient Boosting (GB), RF, and SVM.	98% (Binary classification), 87.5% (Multi-category via GB)	• Accuracy declines as the number of input variables increases.
Ali et al. [16]	Lu-Net: A Deep Learning noise reduction model using convolutional deep-decoding and Bi-LSTM.	Average SNR improvement of 5.613 dB & 1.750 dB (Improves classification in noisy environments).	• High computational complexity, which may limit real-time use in resource-constrained settings.
Sinha et al. [17]	Conv-RF model: Combining CNN-based SqueezeNet with a Random Forest classifier using ESP32 and Raspberry Pi.	98.65% (SqueezeNet), 99.37% (Conv-RF)	• Limitations in eliminating ambient noise and microphone movement artifacts during auscultation.
Nishitha & Chittesh [18]	Hybrid ML (with diverse features) and DL methods based on signal-time spectroscopy.	96.34% accuracy (on PhysioNet and personal datasets)	• Sparse data may limit wider applicability. • Focuses only on accuracy without addressing sensitivity or privacy.

2. METHODOLOGY

This section presents the dataset used in this paper, outlining its main characteristics, source, and relevant details. It also provides an overview of the variables used to predict HF, which are crucial for understanding patients' conditions and improving prediction accuracy. In addition, the ML algorithms used to train and evaluate the dataset are described, highlighting their strengths and relevance to the research. These algorithms were chosen to ensure comprehensive data analysis. Furthermore, the performance of various models is evaluated and compared using metrics such as accuracy, precision, recall, and F1-score to determine the most effective algorithm for predicting HF. This evaluation offers valuable insights into the algorithms' predictive capabilities and potential real-world applications.

2.1. Concept of Heart Sound

Heart sounds can be heard with a stethoscope. These sounds result from blood flow through the heart chambers and heart valve closure during its cycle [19]. There are four main heart sounds, commonly referred to as first (S1), second (S2), third (S3), and fourth (S4). The S1 and S2 are normal heart sounds, while S3 and S4 are abnormal.

There are four main areas on the patient's chest where the electronic stethoscope is placed to record heart sounds, as illustrated in Figure 1. These areas are as follows:

1. Aortic Area: located on the right side of the sternum, in the space between the second and third ribs. The aortic valve sound is heard here.
2. Pulmonary Artery Area: It is located on the left side of the sternum, in the space between the second and third ribs. The pulmonary valve sound is heard here.
3. Tricuspid Area: It is located on the lower side of the sternum, in the space between the fourth and fifth ribs. The mitral valve sound is heard here.
4. Mitral Area: Located on the left side of the chest, between the fifth and sixth ribs, slightly to the left. The mitral valve sound is heard here.

These areas help identify and evaluate heart sounds that indicate heart valve function and diagnose potential problems such as valve stenosis and regurgitation.

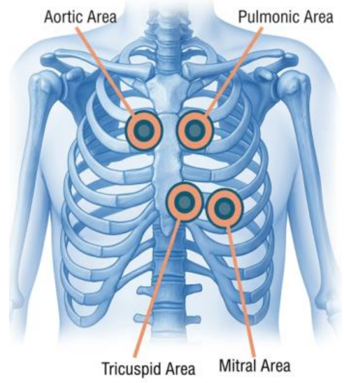


Fig. 1. Heart auscultation areas [19]

2.2. Normal Heart Sounds

The S1, commonly known as (lub), is produced by the closure of the mitral (M1) and tricuspid (T1) valves. The tricuspid valve closes immediately after the mitral valve, so the two sounds appear as a single sound. Due to the larger diameter of the atrioventricular (AV) valves, S1 has a low-pitched tone best heard with the stethoscope. The anatomical complexity of the AV valve mechanism also contributes to the longer S1 duration compared to S2 [19]. S1 represents the beginning of pumping blood from the ventricles to the arteries.

The S2, often called the "dub", occurs at the start of diastole when the semilunar valves close. S2 is characterized by a higher pitch and shorter duration, making it best heard with the diaphragm of a stethoscope because of the significant pressure differences that cause the valves to close. This sound is produced by the closure of the aortic (A2) and pulmonic (P2) valves. During inspiration, A2 and P2 can sometimes be heard as separate sounds. The splitting of S2 happens because the pulmonic valve closes later, and the aortic valve closes earlier. In the breath-in, the reduced pressure in the thoracic cavity allows more blood to flow into the pulmonary circulation from the right ventricle, prolonging the right ventricular systole and delaying the P2 sound. Additionally, the increased pulmonary vascular capacity during inspiration reduces venous return to the left atrium, shortening left ventricular systole and causing the A2 sound to occur earlier. In cases where the aortic valve closure is delayed, paradoxical splitting of S2 occurs, where P2 is heard before A2 [19].

2.3. Abnormal Heart Sounds

In certain cases, S3 (lub-dub-ta) can be detected in normal individuals. It is a low-pitched sound that occurs early in diastole and is best heard near the heart's apex, as presented in Figure 2. S3 is generated by vibrations resulting from rapid ventricular filling. In children and adults, S3 is often considered normal after exercise [19]. However, it can indicate the presence of HF in adults.

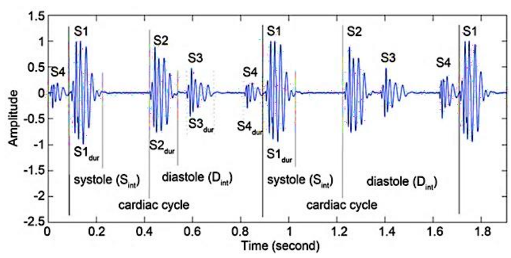


Fig. 2. Diagram of heart sounds [13]

S4 (ta-lub-dub) can occur due to atrial systole. This low-pitched sound arises during the late diastolic or presystolic phases, as depicted in Figure 2, while S4 is commonly heard in young children. In adults, it is often linked to CVD. S4 may be absent in individuals with atrial fibrillation, leading to a loss of the atrial kick. In cases of tachycardia, S4 and S3 can merge, creating a summation gallop [19]. This sound is usually heard in patients with ventricular wall stiffening, such as in cases of high blood pressure or ventricular hypertrophy.

2.4. Heart Sounds Data for HF Prediction

Figure 3 illustrates an integrated process that begins with collecting heart sound data from electronic stethoscopes via Bluetooth, which is then sent to an application on the smartphone. From there, the data is transmitted to a computer for analysis and extraction of important features using ML algorithms. These features are then used to train predictive models, which are evaluated using various performance metrics to ensure accuracy. Ultimately, the trained models are used to predict HF, contributing to early and accurate diagnosis of the condition.

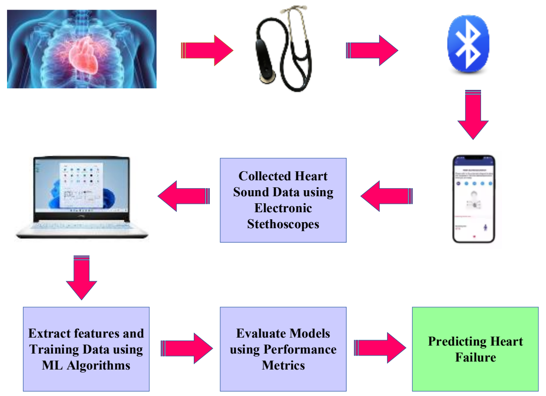


Fig. 3. Block diagram of heart-sound data collected by an electronic stethoscope for predicting heart failure

2.4.1. Feature Extraction from Heart Sounds

Variable extraction methods are crucial for effectively handling data by extracting pertinent variables. Mel Frequency Cepstral Coefficients (MFCC) [20] were employed to extract features from heart sounds in this context.

The MFCC [21] is an important feature extraction method in audio signal processing. Widely used in ML applications, speech recognition, voice biometrics, and speaker recognition. Below is an explanation of the key aspects of this technique:

1. Mel-Frequency: the heard sound can consist of a wide range of frequencies with unequal energy at different frequencies. MFCC [9] technology aims to mimic how humans respond to sound, focusing on low frequencies rather than high, thereby representing frequencies more accurately. For this study, we chose to use MFCC because we are primarily interested in the low-frequency components of heart sounds, where clinically relevant pathological sounds (S3 and S4) occur and where the sounds are not always easy for the human ear to hear. Hence, the low-frequency part of the spectrum is suitable for targeting this specific heart failure classification problem. However, the presence of S3 and S4 sounds is important in the clinical setting and could indicate a greater risk of heart failure.
2. Cepstral: a concept in instantaneous transformation that represents the shifts in energy transactions in audio signals. Cepstral works to reduce the effects of unwanted sounds, such as noise and frequency distortions in the sound spectrum.
3. Mel-Frequency Cepstral Coefficients: The main goal of MFCC technology is to extract coefficients that effectively represent the audio spectrum of the audio signal. These coefficients typically range from 13 to 40 and are used as input representations for ML models. In this paper, 13 coefficients are used.

The selection of 13 MFCC coefficients is based on established standards in audio and biomedical signal processing. This value was chosen because it effectively captures the spectral envelope of cardiac sounds, which is critical for accurate pathology classification. At the same time, higher-order coefficients, which are often sensitive to recording noise, are excluded. Furthermore, the feature vector size is kept at 13 to provide a compact input space relative to the sample size (300), which helps prevent overfitting. Consequently, no additional dimensionality-reduction techniques, such as Principal Component Analysis, were required, as the feature space was already appropriately constrained for the classification models.

2.4.2. Data Collection

This section collected data from databases that record normal and abnormal heart sounds. This data includes recordings of about 300 heart sounds, 150 normal and 150 abnormal. Using 10-fold cross-validation, the data were divided into two groups: training (80%) and testing (20%), as shown in Table 2. This data contains 13 MFCC features extracted from recordings, as illustrated in Table 3.

Table 2. Splitting the heart sound data

Category	Normal	Abnormal	Total
Train	120	120	240
Testing	30	30	60

One of the very important processes for testing the validity and accuracy of the system proposed in this paper is to obtain real-world data and feed it into the system to assess its ability to accurately predict HF. This data will be collected at the hospital from healthy people, consisting of heart-sound recordings from the electronic stethoscope.

All recordings were gathered from different patients, confirming that no person contributed more than one recording to the dataset. It should be noted that detailed demographic information (including age range, sex distribution, and comorbidity profiles) was not uniformly available for all subjects in the utilized dataset. This represents a limitation of the current study, as such variables may influence the spectral characteristics of heart sounds and thereby affect model generalizability. Future work should incorporate datasets with comprehensive clinical metadata to assess the impact of demographic factors on prediction accuracy.

2.5. Electronic Stethoscope

This electronic stethoscope (Model HM-9260, illustrated in Figure 4) is used to auscultate changes in heart, lung, and other organ sounds. Details of the stethoscope are given in Appendix A. The stethoscope has two operating modes: APP auscultation mode and ordinary auscultation mode. If the Bluetooth indicator's blue light remains on after the stethoscope is turned on, it indicates the stethoscope is in APP auscultation mode. Mobile app operation manual:

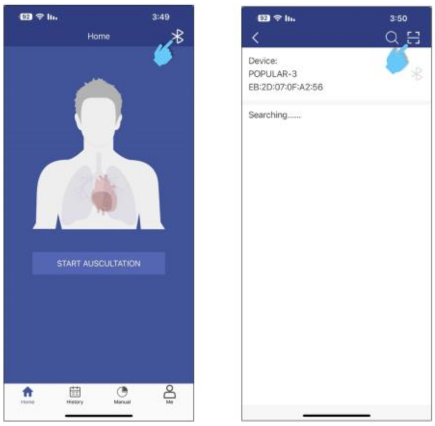
Fig. 4. Snapshot of an electronic stethoscope

1. Connect the stethoscope: press and hold the volume (+) button for over 3 seconds, turn on the Bluetooth function of the stethoscope, open the APP, click the Bluetooth icon in the upper right corner, as shown in Figure 5 (a), enter the device management page, and click the scan button in the upper right corner. Scanning the stethoscope's QR code can quickly connect, as shown in Figure 5(b). If the stethoscope has been paired before, the paired mobile phone will automatically connect to it. A smartphone application changes the auscultation mode between the heart and lungs. When the heart sounds mode is on, the stethoscope will be orange; in the lung sounds mode, it will be blue, switching between them with the power button. If the Bluetooth indicator blinks slowly blue after turning on the stethoscope, it indicates the stethoscope is in normal auscultation mode. This mode allows the user to control stethoscope positions in the App by switching between interfaces, while the power button starts recording sounds, as shown in Figure 6 (a) and (b)

   Table 3 presents the complete feature matrix of 13 MFCC coefficients for all 300 subjects. The mean and standard deviation of each coefficient are provided in the summary row. Notably, MFCC #1 exhibits the largest absolute mean value and highest variability across subjects, reflecting the dominant low-frequency energy content of cardiac sounds. MFCC features #2 to #13 exhibit progressively smaller magnitudes, consistent with the decreasing contribution of higher-order cepstral coefficients to the overall spectral envelope. The data were collected from the Kaggle dataset, which indicates that recording times for each patient ranged from 5 to 15 seconds.
2. Recording and storage of cardiopulmonary sounds: By going to the home page of the App, clicking on (Start Auscultation), as shown in Figure 7 (a), to enter the (Patient Info) interface, as shown in Figure 7 (b), fill in the patient info, and click on (Next) to enter the auscultation interface.

Table 3. Variables extracted from heart sound recordings

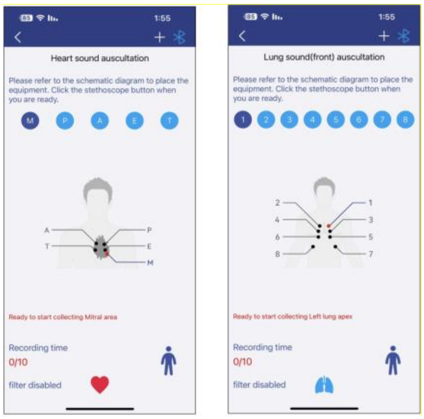
No.	MFCC													HF
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	
1	-137.82	11.639	-0.128	0.939	-0.078	0.521	0.227	1.342	0.128	0.132	-0.283	0.019	0.403	0
2	-139.81	10.891	-0.413	1.457	1.827	1.775	0.424	0.647	0.098	0.992	0.341	0.200	0.316	0
3	-170.89	12.284	-0.539	2.883	1.139	1.317	0.711	1.167	0.076	1.294	0.758	0.054	0.422	0
4	-144.37	12.281	-0.437	2.184	1.355	1.204	0.516	0.932	-0.148	1.030	0.591	0.264	0.325	0
5	-139.43	9.788	2.389	3.137	1.040	1.272	0.244	0.800	0.001	0.125	0.038	-0.406	0.187	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
295	-245.53	35.176	17.573	-2.817	12.175	6.284	-2.578	1.880	5.319	-1.076	-1.246	0.925	0.278	1
296	-243.36	35.803	18.166	-1.553	12.240	6.408	-2.082	1.729	5.380	-1.198	-1.142	0.872	-0.038	1
297	-241.68	36.384	18.400	-0.867	12.213	6.511	-1.729	1.626	5.550	-1.149	-1.001	0.999	-0.180	1
298	-245.74	35.117	17.584	-2.751	12.307	6.381	-2.587	1.968	5.363	-1.048	-1.095	1.047	0.412	1
299	-250.29	33.939	16.537	-5.526	12.251	6.100	-3.606	2.327	5.204	-0.827	-1.393	0.987	1.018	1
300	-246.55	35.108	17.556	-3.223	12.413	6.406	-2.731	2.131	5.469	-0.975	-1.304	0.929	0.378	1



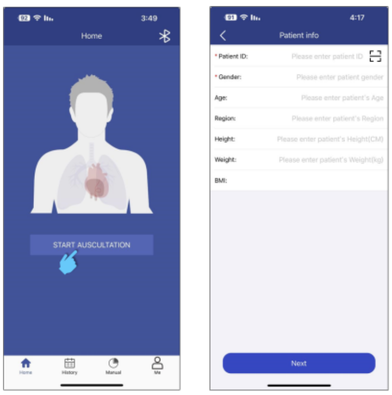
(a) (b)
Fig. 5. Connect the stethoscope (a) home page and (b) scanning for stethoscope

3. Audio analysis, annotation, and sharing

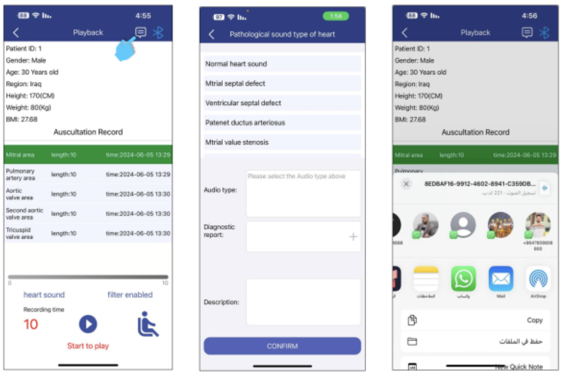
- Click the label icon in the recording or playback interface, as shown in Figure 8 (a), to pop up the pathological sound type page. Users can select the corresponding pathological sound type. Users can also fill in or query relevant record information in the diagnostic report and description columns for follow-up tracking, as shown in Figure 8(b).
- Press and hold the record on the playback page to share the recording, as shown in Figure 8(c).



(a) (b)
Fig. 6. Modes of auscultation in a stethoscope (a) heart and (b) lung sounds



(a) (b)
Fig. 7. Recording and storage of cardiopulmonary sounds (a) home page and (b) patient information



(a) (b) (c)
Fig. 8. Audio analysis: (a) playback sounds, (b) auscultation record, and (c) recording sharing

3. ARTIFICIAL INTELLIGENCE ALGORITHMS

Rapid advances in AI algorithms and systems signify a crucial technological leap in medicine [22]. This progress is not confined solely to medical applications. These algorithms are actively utilized to gain insights into diseases and disorders by analyzing available data [23]. Beyond medicine, their applicability extends to enhancing the understanding of human behavior and various systems, spanning disciplines such as AI and light-based research [24]. This technology is poised to profoundly impact multiple aspects of life, including engineering and healthcare. Anticipated outcomes encompass significant advancements in scientific and medical domains driven by the evolution of AI algorithms [25]. Figure 9 illustrates the flowchart of the proposed research. The process begins with importing essential libraries and heart sound data from the computer. Next, data preprocessing steps are undertaken to find missing values and detect outliers, ensuring data quality. The dataset is split into training and testing sets, with 80% used for training and 20% for testing using 10-fold cross-validation. Various ML models, including RF, SVM, XGBoost, and GB, are trained and validated on the data. The trained models are used to predict HF. The performance of these models is evaluated using metrics such as accuracy, precision, recall, and F1 score. Additionally, results are visualized using a confusion matrix, an AUC curve, and a precision-recall curve to understand model performance better. The process concludes with a comprehensive evaluation of the predictive models.

The cause behind the use of ML in this work is:

- The limited dataset size (300 samples) makes deep learning susceptible to overfitting without extensive augmentation.
- Conventional ML models offer superior interpretability compared to deep neural networks.
- The study serves as a foundational benchmark that quantifies the maximum achievable

performance with MFCC features before exploring more complex architectures.

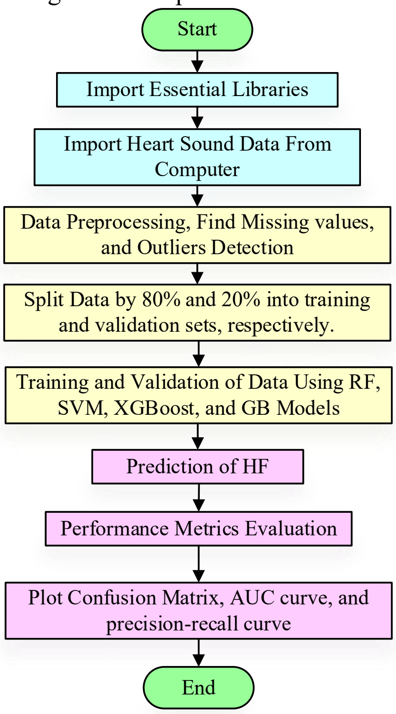


Fig. 9. General flowchart of the proposed study

3.1. Random Forest

The RF is a model consisting of several decision trees developed by Breiman [26], where each tree is built from a random subset of the training data. The samples comprising this training data subset are randomly selected with replacement. Thus, some samples can be used repeatedly in the same tree. When a tree is divided, a random subset of features is considered. In the testing phase, each decision tree in the forest predicts the result of the selected test sample, and the dominant vote determines the final result. The RF achieves high prediction accuracy by avoiding overfitting issues that plague single-tree models. Each decision tree is built from a random subset of the data features. For classification, a specific number of features is usually chosen to divide each node in the tree, and this number is often fixed. For example, if the dataset has 36 features, six features may be selected to split each node, and this cannot be changed. It is also customary to use all available features to split the node in regression problems [27], as shown in Figure 10.

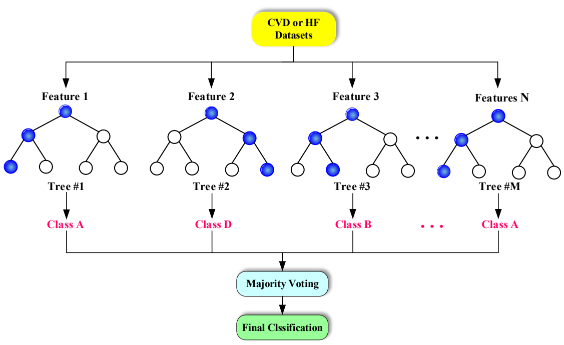


Fig. 10. Random forest model

3.2. Support Vector Machine

SVM is a statistically supervised ML technology for classification and regression [28]. It was originally proposed by Vapnik [29, 30] and Curtis in 1995. Although their principles and derivations differ from those of ANNs, some authors consider SVM a special type of ANN. However, many authors reject this claim due to fundamental differences between the two technologies.

Native SVM technology aims to solve the problem of binary classification, where there is only a classification between two categories, and the problem of multi-class classification. Binary linear SVM classification is based on calculating optimal super-level decision boundaries that separate one category from another using the training dataset.

There are two types of optimizations:

- It is possible to achieve ideal separation of training data categories, and the optimal fixed margin is used, with the excess-level decision limits chosen to maximize the distance from the excess level to the nearest training data point.
- Alternatively, suppose the ideal classification is not desirable or impossible. In that case, the optimal soft margin is used, where the reduction in the misclassification rate is traded off against increasing the distance to the nearest properly classified training point.

Decision limits in SVM classification are computed from the training dataset and are defined by the so-called support vectors, which alone yield the same decision limits. After the support plane is determined, the SVM classifier is ready for use with a new dataset, where the input class is determined by the decision region the input vector lies in [30]. Figure 11 shows the classification of the SVM corresponding to two different patterns: (A) a linearly separable pattern, where green circles are completely separated from red squares by the decision surface. (B) a nonlinear pattern, where no dividing surface can separate all green circles from red squares [30].

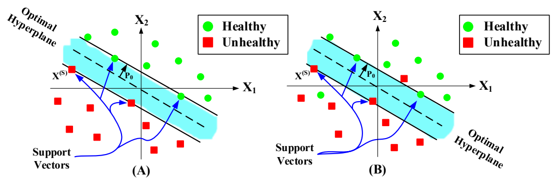


Fig. 11. Support vector machine model

3.3. Gradient Boosting and Extreme Gradient Boosting

GB and XGBoost are used to create ensemble models via gradient descent, enhancing the performance of weak learners [31]. However, XGBoost goes beyond the basic GB approach by optimizing the system and refining the algorithms. It was initially developed by Chen and Guestrin [32] and has since been further developed by others.

XGBoost is part of the Distributed ML Community [33].

In GB, the model is built in stages, with each stage adding a weak classifier to improve performance without altering the previous classifiers. This iterative process continues, with each new classifier focusing on areas where the previous ones performed poorly. The basic flow of the GB algorithm is illustrated in Figure 12. Initially, a decision tree is used to estimate y1, and subsequent trees are fitted based on the residuals from the previous steps (y - y1, y - y1 - y2), and the process continues. This iterative approach efficiently reduces algorithmic error [33].

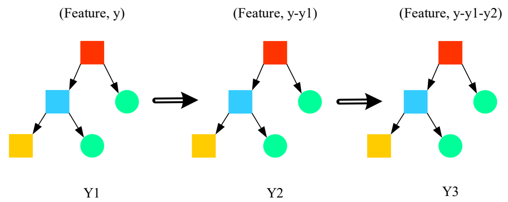


Fig. 12. Gradient boosting model

XGBoost [34], an open-source GB algorithm, has recently gained popularity, particularly in data science competitions. It operates on the principle of GB, aggregating predictions from weak learners to create a robust one. This method, as outlined by Chen and Guestrin in 2016, involves additive training. However, to achieve optimal performance, selecting the right parameters is crucial. Among these parameters are colsample_bytree (which determines the ratio of columns used in constructing each tree), subsample (dictating the ratio of training instances used), and grounds (setting the maximum number of boosting iterations), as shown in Figure 13. Users must carefully choose these parameters when applying the XGBoost model [35]. Details of the algorithm's hyperparameters are given in Appendix B.

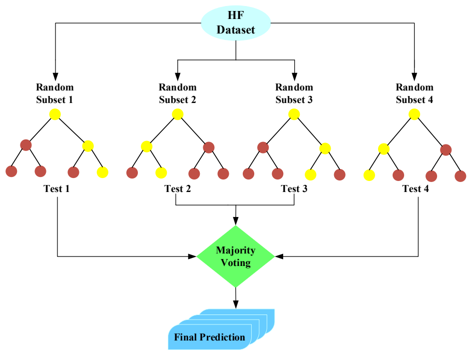


Fig 13. Extreme gradient boosting model

4. PERFORMANCE METRICS EVALUATION

Several criteria are derived from the confusion matrix to evaluate the performance of the RF, SVM, GB, and XGBoost algorithms, including accuracy, sensitivity, specificity, precision, and F1-score [26]. In addition, the Area under the curve, training time, and test time are used. Figure 14 illustrates the various performance metrics used in ML to assess predictive models, highlighting the confusion matrix and its related metrics.

The Confusion matrix contains four components, as follows:

- True Positives (TP): the number of positive samples that are expected to be positive [36].
- True Negatives (TN): the number of negative samples that are expected to be negative [37].
- False Positives (FP): the number of negative samples is incorrectly predicted as positive [38].
- False Negatives (FN): the number of positive samples is expected to be negative [39].

Performance Metrics: These metrics are very important to test the effectiveness and performance of ML algorithms, and they are as follows:

- Accuracy: The ratio of correct predictions to total predictions is represented by Equation (1) [27-29].

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

- Sensitivity: calculated as the percentage of true positive cases out of all predicted positive cases, is expressed in Equation (2) [30, 31].

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

- Specificity denotes the ratio of true negative cases to all actual negative cases, demonstrated in Equation (3) [32, 33].

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

- Precision, indicating the proportion of true positive cases to all predicted positive cases, is revealed in Equation (4) [34].

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- F1-score: a harmonic mean of sensitivity and precision [26], is calculated as shown in Equation (5) [35, 40].

$$F1-Score = 2 * \frac{(Precision * Sensitivity)}{(Precision + Sensitivity)} \quad (5)$$

- Furthermore, Equation (6) provides the AUC, which assesses the model's ability to differentiate between classes [41].

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \quad (6)$$

TPR and FPR are the true-positive and false-positive rates, respectively. Moreover, train time is the time spent training models, measured in seconds. Test time is the total time required to finish testing models, in seconds.

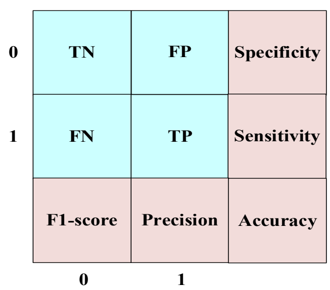


Fig. 14. Confusion matrix

5. RESULTS AND DISCUSSION

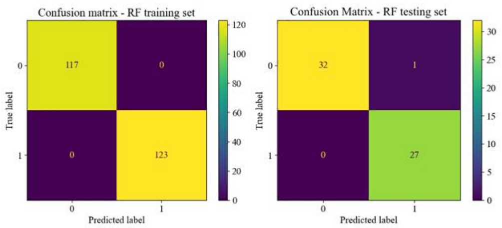
The results for the heart sound data will be displayed in this section. Four ML algorithms were adopted because they achieved the highest accuracy in predicting HF using clinical and heart-sound data from the electronic stethoscope.

The high accuracy values obtained by GB (100% on the training set), XGBoost (98.3-100% on the test set), and RF (100% on the training set) are justified by several factors. First, strict patient-wise separation was enforced throughout the 10-fold cross-validation: no audio recordings from the same subject were present in both the training and testing folds simultaneously, thereby eliminating the risk of data leakage. Second, MFCC features were extracted independently for each fold's training partition; no global normalization or feature selection was applied across train and test splits. Third, the balanced dataset (150 normal, 150 abnormal) eliminates class imbalance as a confounding factor. Fourth, we acknowledge that 300 subjects constitute a moderate-sized dataset, and external validation on an independent, multi-center dataset is necessary to confirm generalizability — this is a primary limitation of the current work and is recommended as a priority for future research.

Environmental noise and inter-patient physiological variability are acknowledged as critical challenges in real-world deployment, and adaptive ensemble strategies (e.g., noise-robust boosting and Bayesian hyperparameter optimization) are proposed as promising extensions.

5.1. Performance of ML Based on Confusion Matrix

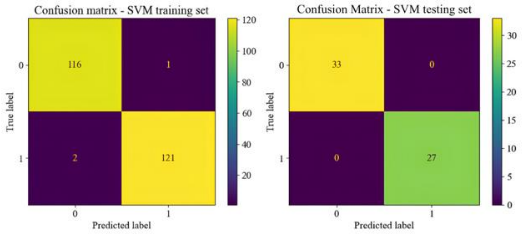
Figure 15 presents the confusion matrix for the RF algorithm on the training and testing sets. The RF correctly predicted all samples in the training set, classifying 117 as class 0 and 123 as class 1, with no errors. On the testing set, the RF maintained strong performance, correctly predicting 32 samples as class 0 and 27 as class 1, while making only one error by classifying one sample as class 1 when it was class 0. In short, this model performs well on both sets.



(a) (b)

Fig. 15. Confusion matrix for RF in (a) training and (b) testing sets for heart sound data

Figure 16 illustrates the confusion matrix for the SVM model for both sets. In the training set, the model correctly classified 116 samples as class 0 and 121 samples as class 1, with two errors: one class 1 sample misclassified as class 0 and one class 0 sample misclassified as class 1, showing good performance with some minor errors. On the testing set, the model performed excellently, correctly classifying all samples with 0 errors, while 33 samples were correctly predicted as class 0 and 27 as class 1. This reflects the strong performance of the SVM in testing with some minor errors in training, indicating a balanced and efficient model.

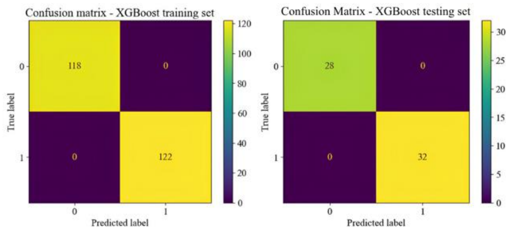


(a) (b)

Fig. 16. Confusion matrix for SVM in (a) training and (b) testing sets for heart sound data

Figure 17 presents the confusion matrix for the XGBoost algorithm on both sets. In the training set, the XGBoost algorithm correctly classified all samples: 118 as class 0 and 122 as class 1, with no errors, indicating perfect performance. In the test set, this algorithm performed excellently, correctly classifying 28 samples as class 0 and 32 as class 1 without errors. This reflects the model's strong performance across both sets, indicating that it is very effective and maintains a good balance without testing errors.

Figure 18 shows the confusion matrix for the GB algorithm. The matrix on the left presents the training set with two classes: "0" and "1". This indicates that 117 cases were correctly classified as class 0, 123 as class 1, and no cases were misclassified. The matrix on the right presents the testing set for classes "0" and "1". This indicates that 32 cases were correctly predicted to be class "0". At the same time, there was only one classification error, where one case from class "0" was incorrectly predicted as class "1". However, many correct predictions remained, with 27 cases correctly identified as class "1" and no cases of class "1" misclassified as class "0".

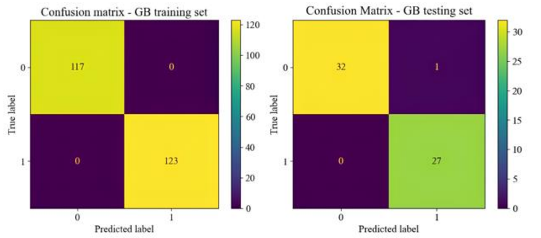


(a) (b)

Fig. 17. Confusion matrix for XGBoost in (a) training and (b) testing sets for heart sound data

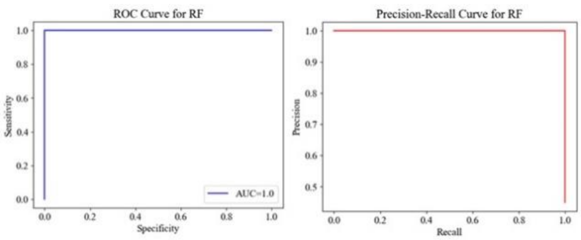
5.2. Performance of ML Based on Curves

Figures 19, 20, 21, and 22 illustrate the ROC and Precision-Recall curves for RF, SVM, XGBoost, and GB algorithms. ROC curves plot sensitivity versus specificity; the GB algorithm achieves an AUC of 98.5%, indicating very good but not perfect performance, while the RF, SVM, and XGBoost algorithms achieve an AUC of 100%, indicating perfect performance. Precision-recall curves plot precision versus recall; all algorithms show precision around 1 across most of the recall range, with a slight dip at the end for the GB algorithm, indicating excellent performance in classifying positive samples. Generally, the RF, XGBoost, and SVM algorithms perform perfectly in both curves, while the GB algorithm shows near-ideal but imperfect performance.



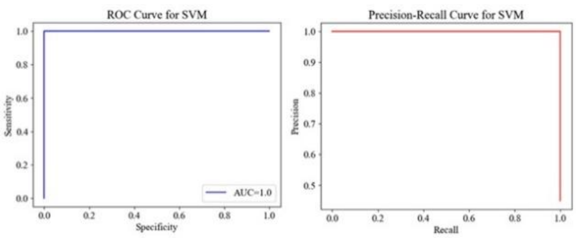
(a) (b)

Fig. 18. Confusion matrix for GB in (a) training and (b) testing sets for heart sound data



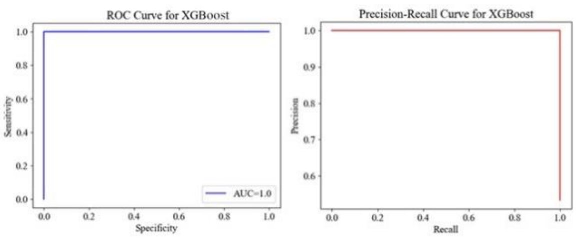
(a) (b)

Fig. 19. Performance of (a) AUC curve and (b) precision-recall curve for RF in heart sound data



(a) (b)

Fig. 20. Performance of (a) AUC curve and (b) precision-recall curve for SVM in heart sound data

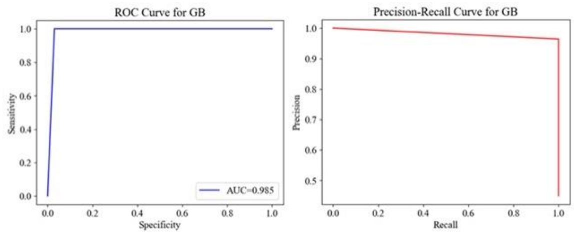


(a) (b)

Fig. 21. Performance of (a) AUC curve and (b) precision-recall curve for XGBoost in heart sound data

5.3. Performance of ML Based on Real-World Testing

The system must be tested in the real world, specifically using hospital data, to assess its accuracy and robustness in predicting HF from heart sounds and clinical data. Therefore, a graphical interface was created to notify the system of this data. Figure 23 illustrates the graphical interface for downloading heart-sound data, where a single WAV file is uploaded for prediction. It is worth noting that the algorithm used for prediction is GB, as it achieved the highest prediction accuracy across both clinical and cardiac sound data. Figures 24 and 25 display the results of the person's prediction. It is worth noting that these graphical interfaces are displayed on the computer.



(a) (b)

Fig. 22. Performance of (a) AUC curve and (b) precision-recall curve for GB in heart sound data

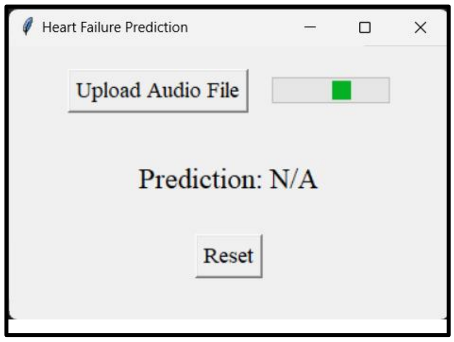
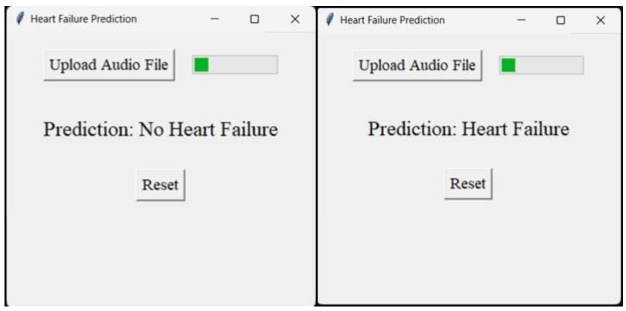


Fig. 23. Graphical interface for HF prediction



(a) (b)

Fig. 24. Graphical interface for HF prediction results (a) healthy person and (b) unhealthy person

6. RESULTS COMPARISON OF ML ALGORITHMS

Table 4 presents the performance metrics for the ML algorithms applied to the training dataset: RF, SVM, GB, and XGBoost. The SVM algorithm's accuracy was 98.7%, sensitivity 98.3%, specificity 99.1%, precision 99.2%, and F1-score 98.7%. As for the RF, GB, and XGBoost algorithms, all metrics were 100%.

Table 5 illustrates the performance metrics of the same algorithms, with values that differ slightly. The

RF and GB algorithms achieve accuracies of 98.3%, sensitivities of 100%, specificities of 96.9%, precisions of 96.4%, and F1-scores of 98.1%. In contrast, the SVM and XGBoost algorithms achieved 100% across all metrics.

Table 4. Performance metrics of the ML models adopted for heart sounds on the training set

ML models	ACC (%)	SEN (%)	SPE (%)	PRE (%)	F1-score (%)
SVM	98.7	98.3	99.1	99.2	98.7
XGBoost	100	100	100	100	100
RF	100	100	100	100	100
GB	100	100	100	100	100

ACC: Accuracy; SEN: Sensitivity; SPE: Specificity; PRE: Precision.

Table 5. Performance metrics of the ML model for heart sounds on the testing set

ML models	ACC (%)	SEN (%)	SPE (%)	PRE (%)	F1-score (%)
RF	98.3	100	96.9	96.4	98.1
GB	98.3	100	96.9	96.4	98.1
SVM	100	100	100	100	100
XGBoost	100	100	100	100	100

ACC: Accuracy; SEN: Sensitivity; SPE: Specificity; PRE: Precision.

Tables 4 and 5 show that the XGBoost, GB and RF algorithms achieved excellent performance on the training set, whereas XGBoost and SVM achieved the best results on the testing set.

While both models achieved 100% training accuracy, GB's testing accuracy dropped to 98.3%, while XGBoost maintained 100% accuracy on both the training and testing sets. The gap indicates that XGBoost is better at generalizing to unseen data, likely because of its built-in regularization mechanisms and more effective optimization strategy, which decrease the risk of overfitting. However, for GB, the modest performance loss suggests a weak bias-variance trade-off, as the model has slightly higher variance for out-of-sample predictions. However, the drop is not significant, indicating strong predictive ability in the case of GB. The manuscript has been modified to state that both methods performed well, but that XGBoost had better generalization performance than GB.

7. COMPARISON WITH PREVIOUS WORKS

Table 6 summarizes the performance of various studies that have used ML models to analyze heart sound data to predict heart conditions. It compares different studies based on the number of data samples, the type of stethoscope used, the ML model applied, and key performance metrics such as accuracy, sensitivity, specificity, precision, and F1 score. Most studies used digital or electronic stethoscopes, and model accuracy varies, with some, such as Sinha et al. (2023), achieving very high accuracy (99.37%) using a Conv-RF. In contrast, other studies, such as Zeniali and Niaki (2022), reported lower accuracy (87.5%) using a GB model.

Table 6. Comparison results with previous studies

Authors/ year	No. of Data	Type of stethoscope	ML Model	ACC %	SEN %	SPE %	PRE %	F1-SCO %
Chowdhury et al. [13]/2019	3126	Digital stethoscope	N/A	88 for normal and 97 for abnormal, classifying	N/A	N/A	N/A	N/A
Omarov et al. [14]/2022	745	Digital stethoscope	KNN	93.5 for normal and 93.25 for abnormal heart sounds	97.12 for normal and 97.64 for abnormal heart sounds	N/A	N/A	N/A
Ali et al. [16]/2023	764	Digital stethoscope	Bi-LSTM	N/A	N/A	N/A	N/A	N/A
Sinha et al. [17]/2023	1000	Electronic stethoscope	Conv-RF	99.37	99.5	98.9	N/A	N/A
Nishitha and Chittesh [18]/2023	947	N/A	DL	96.34	N/A	N/A	N/A	N/A
Kalimuthu and Hemanth [42]/2025	4,430	Digital stethoscope	SVM, KNN, NB	98.63	91.18	100	N/A	91.84
This paper	300	Electronic stethoscope	XGBoost, SVM	100	100	100	100	100

ML: machine learning; ACC: accuracy; SEN: sensitivity; SPE: specificity; PRE: precision; F1-SCO: F1-score.

The last row represents the results of the current study, which achieved 100% across all metrics using GB, XGBoost, and RF models on data collected with an electronic stethoscope, demonstrating superior performance compared to the previous studies listed.

In our proposed models, the data was split into 240 training samples and 60 test samples, with no overlap between the two sets, ensuring no data leakage. With high-quality heart sound recordings obtained with an electronic stethoscope, the proposed model can achieve high classification performance. Furthermore, our results agree with those reported in the literature. As can be seen in Table 6, some previous studies have achieved comparable performance. For instance, Sinha et al. [17] reported an accuracy of 99.37%, which is close to the outcome achieved in this study.

8. CONCLUSION

Cardiovascular diseases are among the leading causes of death worldwide, but early diagnosis and prediction can ensure timely and appropriate care; one of these diseases is HF. This study used recorded heart sounds from an electronic stethoscope and data from the Kaggle database to predict HF. The dataset comprised 13 features (columns) and 300 patients (rows). These features result from converting heart sound data to NumPy arrays using the MFCC technique. The data was split into training (80%) and testing (20%) sets using 10-fold cross-validation. Four ML models were used: RF, SVM, GB, and XGBoost. This study was conducted in the Visual Studio Code environment using Python. Various evaluation methods were employed, including confusion matrices, Receiver Operating Characteristic (ROC) curves, and precision-recall curves. The results demonstrated high test-set prediction accuracy: XGBoost and SVM achieved 100%, while GB and RF achieved 98.3%. The study highlights the need for further research to address limitations and enhance diagnostic and prediction methods. Future work could explore additional AI models, expand the dataset, and clinical data demographics to improve the accuracy of HF prediction.

The previous studies used for comparison employed digital/electronic stethoscopes with characteristics similar to those of the device used in the present study. As for the datasets, our study has fewer data than previous studies; nevertheless, the proposed method performed better. This demonstrates the efficiency and reliability of the proposed approach and shows that the suggested method can achieve high classification accuracy even with limited training data.

Heart sound signals convey important diagnostic information; however, additional information, including demographic data, medical history, laboratory measurements, imaging, etc., can complement the signal and may enhance predictive performance, model robustness and clinical utility. Hence, multimodal models that integrate heart sound signals with extensive clinical data are regarded as a significant research area for the future.

Future studies will validate the proposed model using larger, clinically representative, imbalanced datasets of heart sounds that better reflect the actual prevalence of heart failure worldwide. This validation will help determine how generalizable, robust, and prone to false positives the model is in realistic clinical settings. In addition, as multi-center clinical datasets are incorporated into the approach, its reliability and applicability will be further enhanced.

Source of funding: *This research received no external funding.*

Author contributions: *research concept and design, A.H.S.; Collection of data, S.K.G., I.A.I; Data analysis and interpretation, I.A.I., A.H.A.; Writing the article, A.H.S, S.K.G.; Critical revision of the article, R.F.C.; Final approval of the article, A.H.S., S.K.G.*

Declaration of competing interest: *The author declares no conflict of interest.*

REFERENCES

1. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak. 2020;20(1):16. https://doi.org/10.1186/s12911-020-1023-5.
2. Ma LY, Chen WW, Gao RL, Liu LS, Zhu ML, Wang JY, Wu ZS, Li HJ, Gu DF, Yang YJ, Zheng Z, Hu SS. China cardiovascular diseases report 2018: an updated summary. J Geriatr Cardiol. 2020;17(1):1-8. http://doi.org/10.11909/j.issn.1671-5411.2020.01.001.
3. Gevaert AB, Kataria R, Zannad F, Sauer AJ, Damman K, Sharma K, Shah SJ, Van Spall HGC. Heart failure with preserved ejection fraction: recent concepts in diagnosis, mechanisms and management. Heart. 2022; 108(17):1342-1350. http://doi.org/10.1136/heartjnl-2021-319605.
4. Nauta JF, Hummel JM, van der Meer P, Lam CSP, Voors AA, van Melle JP. Correlation with invasive left ventricular filling pressures and prognostic relevance of the echocardiographic diastolic parameters used in the 2016 ESC heart failure guidelines and in the 2016 ASE/EACVI recommendations: a systematic review in patients with heart failure with preserved ejection fraction. Eur J Heart Fail. 2018;20(9):303-1311. http://doi.org/10.1002/ejhf.1220.
5. Mueller C, McDonald K, de Boer RA, Maisel A, Cleland JGF, Kozhuharov N, Coats AJS, Metra M, Mebazaa A, Ruschitzka F, Lainscak M, Filippatos G, Seferovic PM, Meijers WC, Bayes-Genis A, Mueller T, Richards M, Januzzi Jr JL. Heart Failure Association of the European Society of, "Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. Eur J Heart Fail. 2019;21(6):715-731. http://doi.org/10.1002/ejhf.1494.
6. Kusunose K, Haga A, Abe T, Sata M. Utilization of artificial intelligence in echocardiography. Circ J.

2019;83(8):1623-1629.
http://doi.org/10.1253/circj.CJ-19-0420.

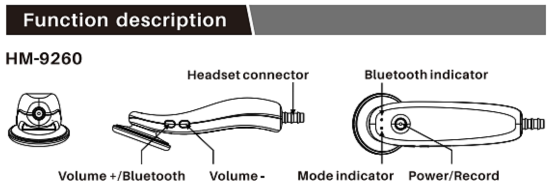
7. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract. 2018;5(4):R115-R125. http://doi.org/10.1530/ERP-18-0056.
8. Goto S, Mahara K, Beussink-Nelson L, Ikura H, Katsumata Y, Endo J, Gaggin KH, Shah SJ, Itabashi Y, MacRae CA, Deo RC. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. Nat Commun. 2021;12(1):2726. http://doi.org/10.1038/s41467-021-22877-8.
9. Li F, Zhang Z, Wang L, Liu W. Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning. Front Physiol. 2022;13:1084420. http://doi.org/10.3389/fphys.2022.1084420.
10. Li F, Tang H, Shang S, Mathiak K, Cong F. Classification of heart sounds using convolutional neural network. Applied Sciences. 2020;10(1). http://doi.org/10.3390/app10113956.
11. Taylor CJ, Ordonez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, Hobbs FDR. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 2019;364:1223. http://doi.org/10.1136/bmj.l223.
12. Tse G, Zhou J, Woo SWD, Ko CH, Lai RWC, Liu T, Liu T, Leung KSK, Li A, Lee S, Li KHC, Lakhani I, Zhang Q, Multi-modality machine learning approach for risk stratification in heart failure with left ventricular ejection fraction </= 45. ESC Heart Failure. 2020;7(6):3716-3725. http://doi.org/10.1002/ehf2.12929.
13. Chowdhury MEH, Khandakar A, Alzoubi K, Mansoor S, Tahir AM, Reaz MBI, Al-Emadi N. Real-time smart-digital stethoscope system for heart diseases monitoring. Sensors (Basel). 2019;19(12):2781. http://doi.org/10.3390/s19122781.
14. Omarov B, Saparkhojayev N, Shekerbekova S, Akhmetova O, Sakypbekova M, Kamalova G, Alimzhanova Z, Tukenova L, Akanova Z. Artificial intelligence in medicine: real time electronic stethoscope for heart diseases detection. Computers, Materials & Continua. 2022;70(2):2815-2833. http://doi.org/10.32604/cmc.2022.019246
15. Zeinali Y, Niaki STA. Heart sound classification using signal processing and machine learning algorithms. Machine Learning with Applications. 2022;7. http://doi.org/10.1016/j.mlwa.2021.100206.
16. Ali SN, Shuvo SB, Al-Manzo MIS, Hasan A, Hasan T. An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise. IEEE Access. 2023;11: 87887-87901. http://doi.org/10.1109/access. 2023.3292551.
17. Sinha Roy T, Roy JK, Mandal N. Conv-random forest-based IoT: A deep learning model based on CNN and random forest for classification and analysis of valvular heart diseases. IEEE Open Journal of Instrumentation and Measurement. 2023;2:1-17. http://doi.org/10.1109/ojim.2023.3320765.
18. Nishitha MK, Chittesh M. Perception of heart failures from heart sounds using neural networks. Journal of Engineering Sciences. 2023;14(07):769-780.
19. Alrabie S, Barnawi A. HeartWave: A multiclass dataset of heart sounds for cardiovascular diseases detection. IEEE Access. 2023;11:118722-118736. http://doi.org/10.1109/access. 2023.3325749.
20. Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Networks. 2020;130:22-32. http://doi.org/10.1016/j.neunet.2020.06.015.
21. Kui H, Pan J, Zong R, Yang H, Wang W. Heart sound classification based on log Mel-frequency spectral coefficients features and convolutional neural networks. Biomedical Signal Processing and Control. 2021;69:102893. https://doi.org/10.1016/j.bspc.2021.102893.
22. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics. 2023;13(17):2760, 2023. https://doi.org/10.3390/diagnostics13172760.
23. Myszczynska MA, Ojamies PN, Lacoste AM, Neil D, Saffari A, Mead R, Hautbergue GM, Holbrook JD, Ferraiuolo L. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nature Reviews Neurology. 2020;16(8): 440-456. https://doi.org/10.1038/s41582-020-0377-8.
24. Kutluyarov RV, Zakoyan AG, Voronkov GS, Grakhova EP, Butt MA. Neuromorphic photonics circuits: contemporary review. Nanomaterials. 2023; 13(24):3139. https://doi.org/10.3390/nano13243139.
25. Asaad Yaseen G, Gharghan SK, Mutlag AHM, Rosdiadee N. Wireless sensor network-based artificial intelligent irrigation system: Challenges and limitations. Journal of Techniques. 2023;5(3):26-41. http://doi.org/10.51173/jt.v5i3.1420.
26. Zeini HA, Al-Jeznawi D, Imran H, BernardoLFA, Al-Khafaji Z, Ostrowski KA. Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil. Sustainability. 2023;15(2):1408. http://doi.org/10.3390/su15021408.
27. Khozeimeh F, Sharifrazi D, Izadi NH, Joloudari JH, Shoeibi A, Alizadehsani R, Tartibi M, Hussain S, Sani ZA, Khodatars M, Sadeghi D, Khosravi A, Nahavandi S, Tan RS, Acharya UR, Islam SMS. RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. Scientific Reports. 2022; 12(1):11178. http://doi.org/10.1038/s41598-022-15374-5.
28. Golpour P, Ghayour-Mobarhan M, Saki A, Esmaily H, Taghipour A, Tajfard M, Ghazizadeh H, Moohebati M, Ferns GA. Comparison of support vector machine, naive bayes and logistic regression for assessing the necessity for coronary angiography. International Journal of Environmental Research and Public Health. 2020;17(18):6449. http://doi.org/10.3390/ijerph17186449.
29. Elsedimy EI, AboHashish SMM, Algarni F. New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. Multimedia Tools and Applications. 2023;83(8):23901-23928. http://doi.org/10.1007/s11042-023-16194-z.
30. Ruiz-Gonzalez R, Gomez-Gil J, Gomez-Gil FJ, Martinez-Martinez V. An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis.

Sensors (Basel). 2014;14(11):20713-35. http://doi.org/10.3390/s141120713

31. Chandrasekhar N, Peddakrishna S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes. 2023;11(4).http://doi.org/10.3390/pr11041210.
32. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. 2020;54(3):1937-1967. http://doi.org/10.1007/s10462-020-09896-5.
33. Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information Sciences. 2022;34(7):4514-4523. http://doi.org/10.1016/j.jksuci.2020.10.013.
34. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms. 2023;16(2). http://doi.org/10.3390/a16020088.
35. Sahin EK Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. Geocarto International. 2020;37(9):2441-2465. http://doi.org/10.1080/10106049.2020.1831623.
36. Aldelemy A, Abd-Alhameed RA. Binary classification of customer's online purchasing behavior using machine learning. Journal of Techniques. 2023;5(2):163-186. http://doi.org/10.51173/jt.v5i2.1226.
37. Ozcan M, Peker S. A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Analytics. 2023;3. http://doi.org/10.1016/j.health.2022.100130.
38. Shuvo SB, Ali SN, Swapnil SI, Al-Rakhami MS, Gumaei A. CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. IEEE Access. 2021;9:36955-36967. http://doi.org/10.1109/access. 2021.3063129.
39. Zhou X, Wang X, Li X, Zhang Y, Liu Y, Wang J, Chen S, Wu Y, Du B, Wang X, Sun X, Sun K. A novel 1-D densely connected feature selection convolutional neural network for heart sounds classification. Annals of Translational Medicine. 2021;9(24):1752. http://doi.org/10.21037/atm-21-4962.
40. Alshboul O, Shehadeh A, Almasabha G, Almuflih AS. Extreme gradient boosting-based machine learning approach for green building cost prediction. Sustainability. 2022;14(11). http://doi.org/10.3390/su14116651.
41. Saeedbakhsh S, Sattari M, Mohammadi M, Najafian J, Mohammadi F. Diagnosis of coronary artery disease based on machine learning algorithms support vector machine, artificial neural network, and random forest. Advanced Biomedical Research. 2023;12:51. http://doi.org/10.4103/abr.abr_383_21.
42. Kalimuthu M, Hemanth C. Preliminary study on real-time phonocardiogram signal acquisition and analysis using machine learning and IoMT for digital stethoscope. 2025;13:68682-68709. http://doi.org/10.1109/ACCESS.2025.3560763.

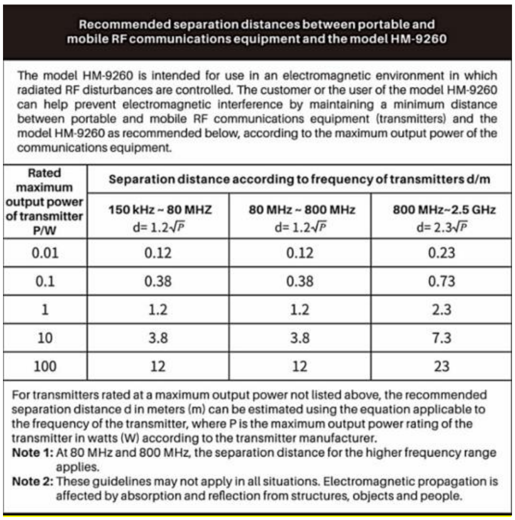
Appendix A: Electronic Stethoscope Specification

Model : HM-9260



Guidance and manufacturer's declaration - electromagnetic immunity			
The model HM-9260 is intended for use in the electromagnetic environment specified below. The customer or the user of the model HM-9260 should assure that it is used in such an environment.			
Immunity test	**IEC 6060 test level**	**Compliance level**	**Electromagnetic environment-guidance**
Electrostatic discharge (ESD) IEC 61000-4-2	±6kV contact ±8kV air	±6kV contact ±8kV air	Floors should be wood, concrete or ceramic tile. If floors are covered with synthetic material, the relative humidity should be at least 30 %.
Electrical fast transient/burst IEC 61000-4-4	± 2kV for power supply lines ±1 kV for Input/output lines	Not applicable	Not applicable
Surge IEC 61000-4-5	±1kV differential mode ±2kV common mode	Not applicable	Not applicable
Voltage dips, short interruptions and voltage variations on power supply input lines IEC 61000-4-11	<5%U_T(>95% dip in U_T) for 0.5 circle 40%U_T(60% dip in U_T) for 5 circle 70%U_T(30% dip in U_T) for 25 circle <5%U_T(95% dip in U_T) for 5 sec	Not applicable	Not applicable
Power frequency (50/60 Hz) magnetic field IEC 61000-4-8	3A/m	3A/m, 50/60Hz	Power frequency magnetic fields should be at levels characteristic of a typical location in a typical commercial or hospital environment.
Note: U_T is the a.c. mains voltage prior to application of the test level			

Guidance and manufacturer's declaration - electromagnetic immunity			
The model HM-9260 is intended for use in the electromagnetic environment specified below. The customer or the user of the model HM-9260 should assure that it is used in such an environment.			
Immunity test	**IEC 6060test level**	**Compliance level**	**Electromagnetic environment-guidance**
Conducted RF IEC 61000-4-6	3 V 150 kHz~80 MHZ	3 V	Portable and mobile RF communications equipment should be used no closer to any part of the model HM-9260, including cables,than the recommended separation distance calculated from the equation applicable to the frequency of the transmitter. **Recommended separation distance** d= $1.2\sqrt{P}$ d= $1.2\sqrt{P}$ 80 MHz~800 Mhz d= $2.3\sqrt{P}$ 800 MHz~2.5 GHz where 'P' is the maximum output power rating of the transmitter in watts (W) according to the transmitter manufacturer and d is the recommended separation distance in meters (m). Field strengths from fixed RF transmitters, as determined by an electromagnetic site survey, [a] should be less than the compliance level in each frequency range. [b] Interference may occur in the vicinity of equipment marked with the following symbol:
Radiated RF IEC 61000-4-3	3 V/m 80 MHz~2.5 GHz	3 V/m	
Note 1: At 80MHz and 800MHz, the higher frequency range applies. **Note 2:** These guidelines may not apply in all situations. Electromagnetic propagation is affected by absorption and reflection from structures, objects and people.			
1. Field strengths from fixed transmitters, such as base stations for radio (cellular/cordless) telephones and land mobile radios, amateur radio, AM and FM radio broadcast and TV broadcast cannot be predicted theoretically with accuracy. To assess the electromagnetic environment due to fixed RF transmitters, an electromagnetic site survey should be considered. If the measured field strength in the location in which the model HM-9260 is used exceeds the applicable RF compliance level above, the model HM-9260 should be observed to verify normal operation. If abnormal performance is observed, additional measures may be necessary, such as reorienting or relocating the model HM-9260 . 2. Over the frequency range 150 kHz to 80 MHz, field strengths should be less than 3 V/m.			



Recommended separation distances between portable and mobile RF communications equipment and the model HM-9260			
The model HM-9260 is intended for use in an electromagnetic environment in which radiated RF disturbances are controlled. The customer or the user of the model HM-9260 can help prevent electromagnetic interference by maintaining a minimum distance between portable and mobile RF communications equipment (transmitters) and the model HM-9260 as recommended below, according to the maximum output power of the communications equipment.			
Rated maximum output power of transmitter P/W	Separation distance according to frequency of transmitters d/m		
	150 kHz ~ 80 MHZ $d=1.2\sqrt{P}$	80 MHz ~ 800 MHz $d=1.2\sqrt{P}$	800 MHz~2.5 GHz $d=2.3\sqrt{P}$
0.01	0.12	0.12	0.23
0.1	0.38	0.38	0.73
1	1.2	1.2	2.3
10	3.8	3.8	7.3
100	12	12	23

For transmitters rated at a maximum output power not listed above, the recommended separation distance d in meters (m) can be estimated using the equation applicable to the frequency of the transmitter, where P is the maximum output power rating of the transmitter in watts (W) according to the transmitter manufacturer.
Note 1: At 80 MHz and 800 MHz, the separation distance for the higher frequency range applies.
Note 2: These guidelines may not apply in all situations. Electromagnetic propagation is affected by absorption and reflection from structures, objects and people.

Appendix B: Hyperparameters of ML algorithms.

Method	ML Algorithm	Hyperparameter
HF using Heart Sounds	RF	Number of estimators=10, max depth=2
	SVM	Kernel: RBF
	GB	Number of estimators=25, max depth=3
	XGBoost	Number of estimators=8, max depth=3

Ali Hussein SHAKER received the B.Sc. degree in Medical Electronic Instrumentation Engineering from Al-Hussain University College, Iraq, in 2018, and the M.Sc. degree in the Department of Medical Instrumentation Techniques, Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq, in 2024. He is currently a Medical Devices Engineer specialist in the Karbala Health Directorate, Ministry of Health, Iraq. His current research interests include biomedical signal processing, machine learning-based medical diagnosis, Internet of Medical Things (IoMT), embedded systems for healthcare applications, and medical data analytics.
e-mail: ali.hussein.shaker94@gmail.com

Sadik Kamel GHARGHAN (Member, IEEE) received the B.Sc. degree in electrical and electronics engineering and the M.Sc. degree in communication engineering from the University of Technology, Iraq, in 1990 and 2005, respectively, and the Ph.D. degree in communication engineering from Universiti Kebangsaan Malaysia (UKM), Malaysia, in 2016. He is currently a Professor at the Technical Engineering College of Artificial Intelligence, Middle Technical University, Baghdad, Iraq. His research interests include energy-efficient wireless sensor networks, biomedical sensors, microcontroller applications, WSN localization using artificial intelligence and optimization algorithms, indoor and outdoor path-loss modeling, harvesting techniques, wireless power transfer, machine-learning-based disease diagnosis, and drones for medical and agricultural applications.
e-mail; sadik.gharghan@mtu.edu.iq

Ibrahim Amer IBRAHIM was born in Baghdad, Iraq, in 1975. He received the B.E. degree in electrical engineering from the University of Baghdad, Baghdad, Iraq, in 1997, and the M. Sc. degrees in electrical engineering/ Electronic and communication Engineering from University of Baghdad, Baghdad, Iraq, in 2000 and a his Ph.D. from University of Malaya, Kula lumpur, Malaysia, in 2019, he joined the Department of Biomedical Engineering, University of Baghdad, as a Lecturer in 2005, and in 2012 became a Senior Lecturer, He gets it assistant professor degree at 2021. His current research interests include power electronics, electronic circuits, biomedical electronic projects and biosignal processing.
e-mail; ibrahimamer@kecbu.uobaghdad.edu.iq

Ahmed Hameed AYYAL received his bachelor's degree in Computer Engineering Techniques from Dijlah University, College of Engineering, Iraq, in 2019, and his Master's degree in Computer Engineering Techniques from the Electrical Engineering Technical College/Middle Technical University in Baghdad in 2024. He currently works as a computer and information technology engineer and instructor at the Ministry of Higher Education and Scientific Research – Middle Technical University/Kut Technical College, and as an external lecturer specializing in medical devices and biomedical engineering at the Kut Technical Institute. His current research interests include biomedical engineering and the potential application of various artificial intelligence algorithms to develop effective solutions in this vital field; medical diagnostics based on machine and deep learning; the Internet of Medical Things; embedded systems for healthcare applications; and medical data analytics.
e-mail; bbc0062@mtu.edu.iq

Raad Farhood CHISAB was born in Baghdad, Iraq, in 1975. He received a B.Sc. in Electrical Engineering and an M.Sc. in Control and Computer Engineering from the College of Engineering at the University of Baghdad. In 2015, the author obtained his Ph.D. in Electronic and Communication Engineering from SHIATS University, India. Since 2005, the author has worked as a lecturer, researcher, and training supervisor at the Ministry of Higher Education

and Scientific Research–Middle Technical University in Baghdad, Iraq. His research interests include signal processing, image processing, wireless communication, mobile technology, wireless sensor networks, IOT, 4G, and 5G. The author has published about 20 papers in national and international journals.
e-mail; raad.farhood@mtu.edu.iq