International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2022; 3(2): 07-12
Received: 05-05-2022
Accepted: 09-06-2022

Viswanatha Reddy Allugunti
Chief Solutions Architect
Arohak Inc, USA

Corresponding Author:
Viswanatha Reddy Allugunti
Chief Solutions Architect
Arohak Inc, USA

Heart disease diagnosis and prediction based on hybrid machine learning model

Viswanatha Reddy Allugunti

DOI: https://doi.org/10.33545/27076571.2022.v3.i2a.49

Abstract
Heart disease is responsible for a considerable percentage of deaths that occur all over the globe, and it has developed into a major health risk for many individuals. The detection of cardiovascular diseases such as heart attacks, myocardial infarction diseases, and others like these is a crucial problem that must be met by the regular clinical data analysis. Early prevalence of heart failure may save the lives of many people. The application of machine learning (ML) can provide an efficient answer to the problem of making decisions and accurate forecasts. The application of machine learning strategies is making significant headway in the medical sector. In the work that has been presented, a unique technique to machine learning is proposed for the purpose of predicting cardiac disease. The PhysioNet Dataset was utilised for the study that was proposed, and data mining algorithms like regression and classification were utilised. Support Vector Machine, Decision Tree and Random Forest are both the machine learning approaches that are utilised here. The cutting-edge strategy for the machine learning model has been devised. Support Vector Machine, Random Forest, Decision Tree, and the Hybrid model (Hybrid of SVM, RF and DT) are the four types of machine learning algorithms that are utilised in the implementation process. The accuracy level of the heart disease prediction model using the hybrid model was found to be 88.7 percent based on the results of the experiments. The user's input parameter will be utilised to predict heart illness, which will be done with a model that is a hybrid of Decision Tree and Random Forest. The interface is built to acquire the user's input parameter.

Keywords: Heart disease, chronic disease, feature extraction, classification, machine learning

Introduction
The rise of cardiovascular illnesses as well as a growth in unhealthy lifestyle choices are two of the most significant threats to the health and well-being of humans (CVD). Latest studies out from World Health Organization (WHO) suggest that cardiovascular disease (CVD) passed cancer to become the main cause of death around the world in 2016. In that year, cancer was responsible for the deaths of 17.9 million people. As of now, around 290 million individuals in China are afflicted with some form of cardiovascular disease, making the prevention of and treatment for cardiovascular disease an essential concern for anyone concerned with their health.

The process of data mining is helpful for the analysis and comprehension of enormous amounts of data. Its primary functions are data extraction and the formulation of decisions for further application uses. Clustering, association rule, and classification are three of the main approaches that fall under the umbrella of "data mining". These data mining strategies can be put into action using any number of algorithms, which are readily available. ven though there are simulation tools accessible, such as Weka, Python programming is becoming increasingly popular due to the rise of these algorithms constructed using scikit learn packages. Therefore, the execution of data mining techniques in real time is more trustworthy than it has ever been. In the field of medical diagnosis, where human error can be decreased through the use of computer analysis and where accuracy can be enhanced, the application of machine learning is experiencing explosive growth. Using machine learning techniques, a disease can be reliably diagnosed in a very high percentage of cases. Concepts related to machine learning are used to provide predictions about diseases such as cardiovascular disease, liver illness, diabetes, and malignancies. In the medical field, regression methods such as Random forest, lasso, and logistic regressions were utilised. Classification techniques such as decision tree and naive bayes as well as SVM (Support Vector Machine) are also accessible. Machine learning algorithms are extensively employed in the realm of medical diagnostics, and are responsible for the majority of malignancy forecasts.

According to the findings of several surveys, cardiovascular diseases are responsible for roughly 17 million fatalities worldwide each year (CVD). The early discovery of disease has the potential to save other lives, and timely medication administration by patients is an important factor in lowering mortality rates. Cardiovascular disorders encompass a wide range of dangers, including but not limited to heart disease, all of them, and so on. These illnesses are now becoming increasingly prevalent, even in younger age groups, as a direct result of the shift away from active lifestyles that has been brought about by lifestyle changes. Habits such as smoking, not getting enough exercise, eating foods rich in cholesterol and junk food, and not leading a healthy lifestyle are the primary contributors to heart disease.

Recent advancements in deep learning (DL) have shown that end-to-end learning is superior to more conventional, feature-based machine learning. This was proved through a series of experiments. Figure 1 depicts the tissues or organs and arteries in addition to the four different right atrium of the heart that are related to auscultation. Also included in this illustration is the auscultation procedure. There is a distinct variation in the physiologic state of the heart associated with each of the four components. As can be seen in Figure 2, variations in blood flow inside the heart are brought on by various partial cardiac sound elements. An improvement in the integrity of the ascending aorta, overall functioning of the ventricular contraction, as well as the systemic circulation within the heart is correlated to an increase in the volume, frequency, and synchronisation of cardiovascular sound.

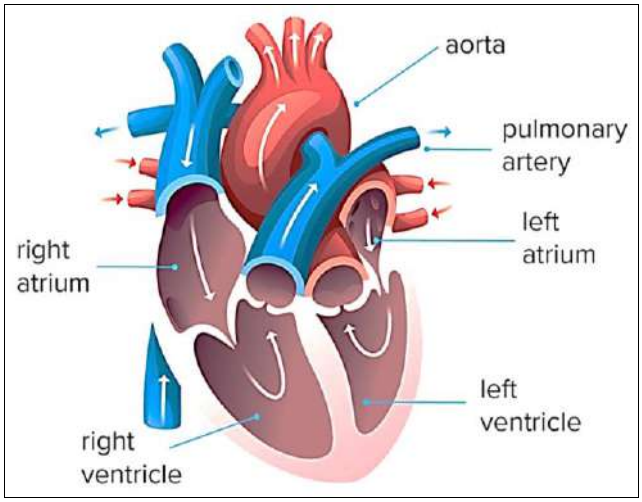

Fig 1: Sectional Schematic of the Human Heart

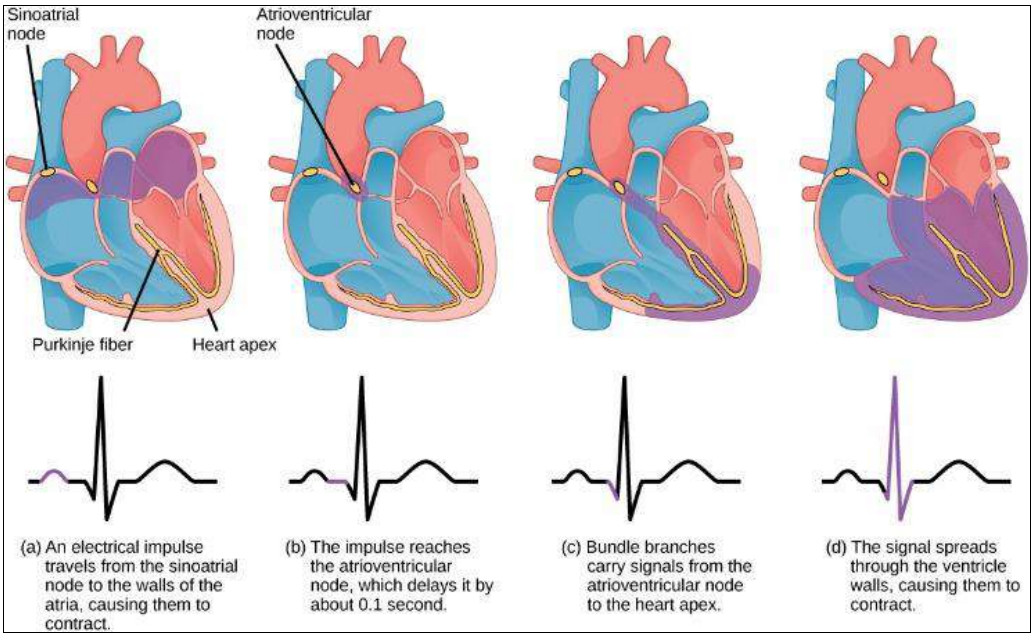

Fig 2: Partial Heart Sound Components of Coronary

Blood Flow

The purpose of this research is to determine whether or not cardiac disease may be predicted using machine learning and an automated way of medical diagnosis. Because it provides the most accurate classification results, the hybrid model is the one that we use to make our heart disease predictions. The probabilities that are obtained through one machine learning algorithms are given as input to another machine learning model, which is a revolutionary technique. This technique is known as a hybrid model. This hybrid model provides us with the findings that have been better optimised based on both the machine learning technique and the implementations that are taken into consideration.

The proposed method for predicting heart disease is based on an efficient machine learning diagnostic model that incorporates a significant novelty-based hybrid model. The risk of developing heart disease can be estimated using this hybrid model. Processing is done with the assistance of the Physio Net Dataset. Researchers working in the field of machine learning frequently consider this dataset. This dataset contains a total of 570 cases, and it has around 14 different features.

The purpose of the investigation is to categories it using a binary scale that ranges from 0 (no evidence of heart disease) to 1. (Present of heart disease). Patients are able to receive treatment depending on the results that are generated by our proposed approach. The application that is being offered assists in the process of adopting preventative measures for patients.

Related Work

Tanachat Nilanon *et al*. [1] focus on an automatic analysis of normal/abnormal phonocardiogram (PCG) recordings as component of Factors played a role in Cardiology Challenge 2016, with the goal of swiftly recognising subjects requiring an expert diagnosis. To increase the classifiers' robustness, 5 second segments were recorded and trained to detect them. The aggregate recording classification was derived by voting on the segment classifications. Features include spectrograms and Prevalence in the general population Cepstrum coefficients.

And over 26 million people worldwide have chronic heart disease. This increases cardiovascular patient mortality, leading to over 1 million hospitalizations each year in North America And Europe. Avoiding, tracking, encouraging early identification, and minimizing hospitalizations or living dangers improves the adherence to treatment. Martin Gjoreski *et al*. [2] used machine-learning to diagnose chronic cardiac problems from heart sounds. The system filters, segments, extracts functionality, and learns. The method was evaluated using 122 individuals' leave-one-subject data.

One of the most widespread dangers to people's health all across the world is coronary artery disease. Controlling cardiac sounds through various methods is an important part of a non-screening aid that plays a significant role in the diagnosis of cardiac illnesses. Suyi Li and colleagues [3] conducted research to examine the most recent developments in computerised heart sensing that have occurred during the past five years. Cardiac sound theory, its correlation to cardiology or low cardiac output, and core means of process and interpretation of cardiac frequency components such as de-noising, edge detection, retrieval and identification of characteristics, including the use of of in-depth machine learning in the analysis of cardiac sound are all important topics. Cardiac sound theory was developed to explain the relationship between cardiac sounds and coronary or cardiovascular disorders.

Jain *et al*. [4] developed the DWT PCG signal denotation technique with the intention of simplifying the antiquated discrete wavelet transformation function. This was accomplished by using a mother-size "Coif-5' wavelet in conjunction with a regressive, non-linear, intermediary functional process and evolutionary algorithms (SWT). A modified strategy can cut down on lower information coefficients while simultaneously improving the efficiency of noise removal.

The most recent core toning denoising solution was developed by Mondal *et al*. [5] and it makes use of a hybrid time - frequency paradigm. The coefficients that appeared the most frequently in the wavelet tree were selected, and the SVD algorithm was used to remove the noise component of the cardiovascular signal.

In the context of Internet of Things (IoT) research, Faheem Akhtar *et al*. [6] presented a method for a sensor-based smartphone device with both the potential of identifying heartbeats; the sensed heartbeat is attributed into to the framework, which further needs to perform some computations to determine whether or not an individual is truly in a consistent level or require specialized emergency care. In addition to developing an application that can associated with a group throughout the improvement from such a preliminary cardiorespiratory risk monitoring and forecasting framework by developing a comprehensive introduction to the research to aid one another in their efforts, one of the motivating factors for this investigation is to compare and contrast various existing technologies. Another primary intention of this research is to make a comparison various existing technologies.

Zheng *et al*. [7] provided a cutting-edge method that is based on a combination of improved SVD and Compression Sensing (CS). This system is able to preserve the heart's natural morphological properties. This strategy, in contrast to more conventional methodologies such as DWT and subjective decay, is capable of achieving a higher SNRR (EMD). The heart-sound signals that were designated were most closely connected with the heart-sound signals that were recorded originally.

An adaptive denoising technique that was introduced by Deng and Han [8]. [Citation needed] In comparison to the more conventional wavelet method, the de-noising impact achieved by their newly developed algorithm was significantly superior.

The term congestive heart failure (CHF) refers to an inadequate blood filling function of the ventricular pump, which may be the cause of an inadequate quantity of heart output that does not fulfil the requirements of the body's metabolism. Heart rate variability (HRV), which is based on RR interval, has been shown to be an established and effective predictor of CHF. Particularly when linked with cellular telephones and smart watches, short-term HRV has seen widespread adoption as a tool for monitoring the wellbeing of patients in a variety of healthcare applications. Long short-term memory (LSTM) was coupled with the CHF inception detection method by Ludi Wang *et al*. [9], who were motivated to do so by the inception module developed by GoogleNet.

In today's world, heart failure is a common disease that frequently results in potentially life-threatening circumstances (HF). Nearly 26 million people are diagnosed with this sort of sickness each year in the United States alone. It can be challenging, from both the perspective of the cardiac consultant and the surgeon, to accurately predict when a patient will experience heart failure. At the moment, there exist categorization and forecasting model that really can help the medical profession and can readily explain how when and how to use a medical data. These models can be found online. The purpose of the research carried out by Fahd Saleh Alotaibi and colleagues [10], which makes use of the UCI cardiovascular disease dataset, is to improve the accuracy of the HF prediction. Several different machine learning strategies were utilised in order to accomplish the goals of comprehending the data and estimating the likelihood of HF in a medical database. In addition, the findings and the comparison analyses showed that the original study improved the prior average score in the heart disease prediction. The coupling of the machine learning algorithm discussed in this paper with clinical decision support systems will be vital for evaluating heart failure as well as any other illness in the future using live data acquired from patients.

Proposed Methodology

In this section we will discuss about the proposed Hybrid Model for Heart disease Prediction model. We have built a Python based Hybrid ML model for the detection of Heart disease Prediction. For better analysis of the results, we also compare the results of our hybrid algorithms with the standard state of art algorithm. For evaluation of the

obtained results we use Accuracy, F1-Score, Precision and Recall as the performance parameters.

Hybrid SVM+RF+DT Model

For the detection of abusive text and slang language we developed a Hybrid SVM+RF+DT Model. The use of a classification algorithm does not result in increased productivity in terms of effectiveness and detection rate, which is why the use of a hybrid classifier has been advocated. At the beginning of the process, an SVM classifier is utilised for the purpose of anomaly identification. In this stage, the dataset is divided into normal and attack classes. The RF/DT classifier is put to use as a misuse detection mechanism in the second stage, at which point it sorts the attacks into a variety of categories. Figure 1 illustrates the proposed framework, which is made up of three separate modules: a module for preprocessing datasets, a module for classifying data, and a module for evaluating how well the classification worked. Their common characteristics are that they were created for the implementation of visual function approximation issues as well as for the generation of textual annotations from picture sequences, both of which are discussed below. The fundamental architecture of the proposed system is depicted in Figure 3, which includes the input data, attribute collection, sequencing learning, and output layer, all of which are represented by rectangles.

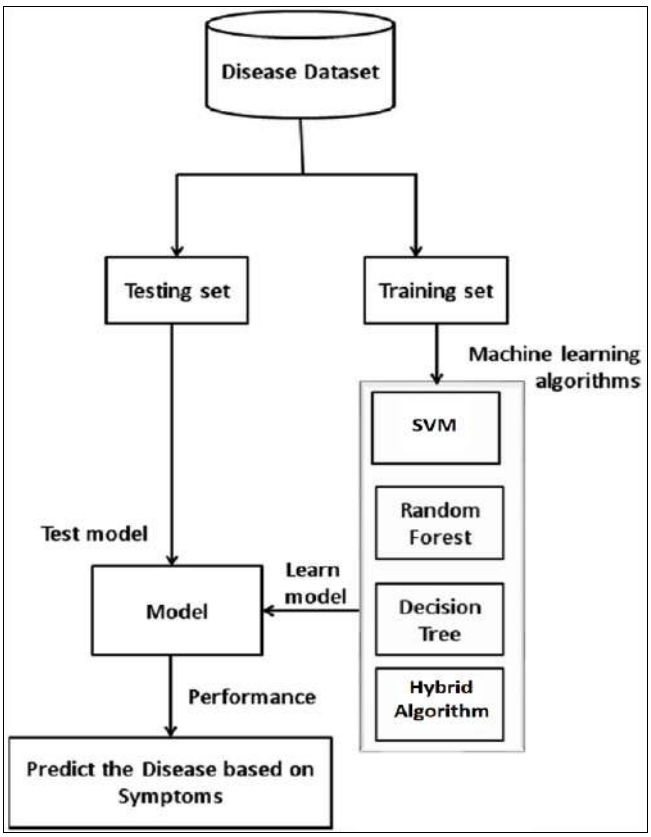

Fig 3: System Architecture for Heart disease Prediction model

Result and Analysis

Dataset Description

PhysioNet Dataset

The archive for the 2016 Physio Net Cardiology Challenge includes six datasets that range in length from five seconds to little over two minutes. (The six research teams that are participating in the contest) provide a combined total of 3,153 movies of heart sounds. In either a clinical or a non-clinical environment, recordings were made with both healthy persons and pathological patients, and these recordings were made from a variety of places across the body. The areas of the aorta, lungs, tricuspid valve, and mitral valve; however, each of these could be one of nine other sites, including all four frequent locations. Both the stable subject and the patient pathology samples contain both adults and children in varying stages of development. There is a possibility that each patient supplied anything from one to six recordings. The dataset, on the other hand, does not provide a lot of information on the person whose record it is. Both recordings were frequency checked at 2000 Hz, and the "Wav" file format was used to preserve them. Each record has only one PCG lead guitar track. Due to the fact that the recordings were made in largely uncontrolled environments, some recordings have been altered by a variety of different types of noise, including voice, stethose, breathing, and gut tone. In addition, determining whether such recordings should be classified as malignant or benign was extremely challenging, if not impossible.

Result Comparision

Following is a table that compares the results of several different classifiers applied to the PhysioNet dataset. The authors make use of a variety of classifier and features extraction strategies, some of which include deep learning techniques, machine learning algorithms, and traditional classification algorithms including SVM, Adaboost, Logistic Regression, and kNN, among others. The comparative analysis is shown in figure 3, which has been provided for your convenience.

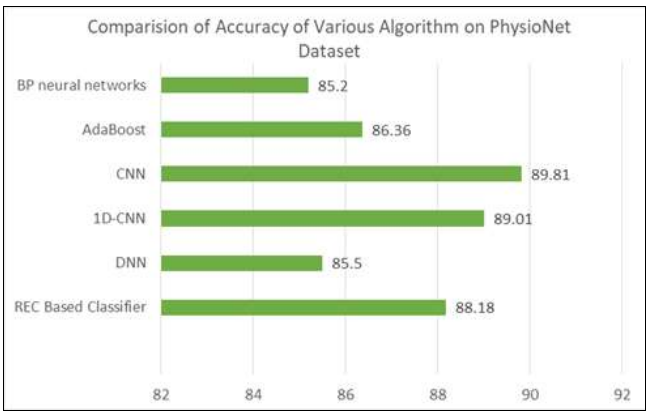

Fig 3: Comparison of Various Existing Algorithm from

Literature Review

The hybrid model that was proposed is evaluated using the PhysioNet dataset, which would be composed of three separate sets. The suggested classifier is evaluated on the training samples and then tested on the testing samples, with each set being divided into 80 percent training instances and 20 percent testing samples respectively. The performance evaluation, including overall accuracy, precision, and recall, is carried out, and analysis is carried out on SVM, DT, and RF. This analysis is then compared with the conventional methodologies and hybrid SVM-RF-DT approach. The results of classification using SVM, DT, RF, and hybrid SVM-RF-DT classifiers are tabulated in Table 1. The comparative analysis of the results is shown in figure 4.

Table 1: Classification Report of Machine Learning Algorithms

	Accuracy	Precision	Recall	F1-Measure
DT	86.97	86.66	74.98	71.46
RF	85.88	81.98	79.08	78.65
SVM	88.72	85.67	81.46	74.44
Hybrid SVM-DT-RF	94.96	93.05	93.31	90.61

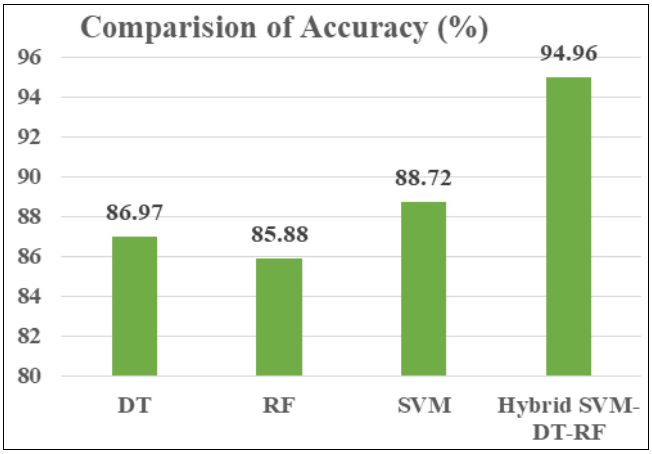

Fig 4: Comparative Analysis of the Accuracy

Conclusion

Because of the rising number of people who pass away from heart diseases, it is imperative that a method be devised that is able to reliably and effectively predict the occurrence of heart diseases. The objective of the research was to identify the machine learning algorithm that was the most effective at diagnosing cardiac problems. The development of new technologies in the areas of machine learning and internet of things (IoT) could play a significant part in the creation of such a smart system. In this study, our primary objective is to construct a hybrid classification process for the development of predicting heart disease. Additionally, we will cover the various research that has been conducted in the past addressing the monitoring and diagnosis of cardiac failure. In this study, we present the empirical evaluation of various previous researchers on the applying machine learning approaches for the diagnosis of cardiac abnormalities. These methods have been used to detect cardiac problems. In addition to that, we offered our analytic findings regarding the accuracy of these works. The Support Vector Machine, Decision Tree, and Random Forest Classifier are the algorithms that were employed in the construction of the model that was presented. Compared to SVM, which has an accuracy of 88.72 percent, Random Forest, which has an accuracy of 85.88 percent, and Decision Tree, which has an accuracy of 86.97 percent, our model has a 96.94 percent accuracy. The work can be improved in the future by creating an application that is based just on proposed algorithm and by using a large - scale dataset in comparison to the one that was used in this analysis. This will assist to just provide improved outcome and assist medical professionals in the accurate and efficient prediction of heart disease.

References

1. Nilanon Tanachat, Purushotham Sanjay, Liu Yan. Normal / Abnormal Heart Sound Recordings Classification Using Convolutional Neural Network, 2016. 10.22489/CinC.2016.169-535.
2. Gjoreski Martin, Simjanoska Monika, Gradišek Anton, Peterlin Ana, Gams Matjaz. Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers, 2017, 14-19. 10.1109/IE.2017.19.
3. Li Suyi, Li Feng, Tang Shijie, Xiong Wenji. A Review of Computer-Aided Heart Sound Detection Techniques. BioMed Research International, 2020. 1-10. 10.1155/2020/5846191.
4. Jain PK, Tiwari AK. An adaptive method for shrinking of wavelet coefficients for phonocardiogram denoising, in 2016 IEEE International Conference on Digital Signal Processing (DSP), 1–5, IEEE, Beijing, 2016.
5. Mondal A, Saxena I, Tang H, Banerjee P. A noise reduction technique based on nonlinear kernel function for heart sound analysis, IEEE Journal of Biomedical and Health Informatics. 2018;22(3):775–784,
6. Akhtar F, Shakeel A, Nazir A, Wajahat A, Ullah F, Qureshi S. A Study and Application Development on Monitoring Cardio-Vascular Attack using Internet of Thing (IoT), 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020, 1-5. doi: 10.1109/iCoMET48670.2020.9073897.
7. Zheng Y, Guo X, Jiang H, Zhou B. An innovative multilevel singular value decomposition and compressed sensing based framework for noise removal from heart sounds, Biomedical Signal Processing & Control, 2017;38:34-43.
8. Deng SW, Han JQ. Adaptive overlapping-group sparse denoising for heart sound signals, Biomedical Signal Processing and Control. 2018;40:49-57.
9. Wang Ludi, Zhou Xiaoguang. Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals. Sensors. 2019;19:1502. 10.3390/s19071502.
10. Alotaibi Fahd. Implementation of Machine Learning Model to Predict Heart Failure Disease. International Journal of Advanced Computer Science and Applications. 2019, 10. 10.14569/IJACSA.2019.0100637.
11. Qiao P, Yu Z, Jingyi Z, Zhuo C. A method for diagnosing heart sounds in adolescents based on wavelet analysis and random forest, 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 2020, 69-74.
12. Shuo Zhang, Ruiqing Zhang, Shijie Chang, Chengyu Liu, Xianzheng Sha. A Low-Noise-Level Heart Sound System Based on Novel Thorax-Integration Head Design and Wavelet Denoising Algorithm. Micromachines. 2019;10(12);885.
13. Kora P, Rajani A, Chinnaiah MC, Swaraja K, Meenakshi K. IoT Based Wearable Monitoring structure for detecting Abnormal Heart. International Conference on Sustainable Energy and Future Electric Transportation (SEFET), 2021;pp:1-4, doi: 10.1109/SeFet48154.2021.9375787.
14. Gupta K, Kaul P, Kaur A. An Efficient Algorithm for Heart Attack Detection using Fuzzy C-means and Alert using IoT. 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT), 2018, 1-6. doi: 10.1109/CIACT.2018.8480157.
15. Li F, Liu M, Zhao Y, *et al.* Feature extraction and

classification of heart sound using 1D convolutional neural networks. EURASIP J Adv. Signal Process, 2019, 59. https://doi.org/10.1186/s13634-019-0651-3
16. Singh Mandeep, Cheema, Amandeep. Heart Sounds Classification using Feature Extraction of Phonocardiography Signal. International Journal of Computer Applications. 2013;77:13-17. 10.5120/13381-1001.
17. Gjoreski M, Gradišek A, Budna B, Gams M, Poglajen G. Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds, in IEEE Access. 2020;8:20313-20324. doi: 10.1109/ACCESS.2020.2968900.
18. Abduh Z, Nehary EA, Wahed MA, Kadah YM. Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and stacked auto-encoder deep neural network, Journal of Medical Imaging and Health Informatics. 2019;9(1):1-8.
19. Wu JMT, Tsai MH, Huang YZ, *et al*., Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model, Applied Sol. Computing. 2019;78:29-40.
20. Shi K, Schellenberger S, Michler F, *et al*., Automatic signal quality index determination of radar-recorded heart sound signals using ensemble classification, IEEE Transactions on Biomedical Engineering, 2019, 1.
21. Nogueira DM, Ferreira CA, Gomes EF, Jorge AM. Classifying heart sounds using images of motifs, MFCC and temporal features, Journal of Medical Systems, 2019;43(6):168.
22. Meintjes A, Lowe A, Legget M. Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks, in 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Honolulu, HI, 2018, 409-412.
23. Kay E, Agarwal A. Drop Connected neural networks trained on time-frequency and inter-beat features for classifying heart sounds, Physiological Measurement, 2017;38(8):1645–1647.
24. Li L, Wang X, Du X, Liu Y, Liu C, Qin C. Classification of heart sound signals with BP neural network and logistic regression, in Chinese Automation Congress (CAC), 2017, IEEE, Jinan, China, 2017, 7380-7383
25. Potes C, Parvaneh S, Rahman A, Conroy B, America PN, Solutions AC. Ensemble of feature-based and deep learning based classifiers for detection of abnormal heart sounds," in Computing in Cardiology Conference, IEEE, Vancouver, BC, Canada. 2016;43:621-624.