International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 9 Issue 6, June-2019,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

UTILIZING MACHINE LEARNING AND NON-INVASIVE TECHNIQUES FOR HEART DISEASE DIAGNOSIS

Pulloori Pratibha, Research Scholar, Dept of Computer Science, Himalayan University
Dr. Kailash Karande, Research Guide, Dept of Computer Science, Himalayan University

ABSTRACT

Heart disease is considered to be the most dangerous non-communicable disease globally, responsible for a staggering number of deaths every year. According to recent statistics, approximately 17.9 million people die annually due to heart-related illnesses. Cardiovascular disease (CVD) is the umbrella term used to describe various heart conditions, and timely diagnosis and treatment are crucial for effective management and prevention of serious complications. Currently, there are two primary methods of diagnosing CVD: invasive and non-invasive. Invasive methods, such as coronary angiography, are complex, expensive, and often associated with discomfort and complications. On the other hand, non-invasive methods generate significant amounts of data, which can be categorized into three types: clinical parameters, heart signals, and heart images. Machine learning (ML) has emerged as a promising tool for diagnosing heart disease using non-invasive methods. ML frameworks can be developed based on these data categories, and involve a range of techniques such as feature selection, classification, pre-processing, segmentation, and feature extraction. These frameworks aim to automate the diagnosis process and improve the accuracy and efficiency of CVD diagnosis.

Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are the most prevalent ML methods in all frameworks. However, recent advancements in deep neural networks have shown promising results in classifying heart sound signals and cardiovascular images. The present study provides a comprehensive review of recent and prevalent ML methodologies in all frameworks, offering new researchers in the domain of machine learning a set of guidelines and avenues to automate the diagnosis process of heart disease. By leveraging these cutting-edge technologies, we can improve the accuracy, speed, and efficiency of CVD diagnosis, leading to better outcomes for patients.

Keywords: Cardiovascular disease, machine Learning, CVD diagnosis

1. INTRODUCTION

Communicable diseases such as COVID-19 and swine flu are known for their fast spread, but they typically have low mortality rates compared to non-communicable diseases, including diabetes, heart disease, liver cancer, and breast cancer. In recent years, machine learning researchers have focused on medical data mining, covering a wide range of diseases, including breast cancer, heart disease, diabetes, Parkinson's disease, hepatitis, liver disorders, lung cancer, pancreatic cancer, leukemia, and brain tumors.

Among these diseases, heart disease is considered the most dangerous and unpredictable, with the highest mortality rate among all non-communicable ailments. In 2015, more than 17.7 million people died due to different heart diseases. According to recent statistics, approximately 17.9 million people die every year worldwide due to heart problems, which equates to around 49,000 people per day. Death rates due to heart disease are even higher in underdeveloped and developing nations due to the costly diagnosis process.

Heart ailments come in different forms, including congenital heart disease, left-sided heart failure, right-sided heart failure, ischemic heart disease, myocardial infarction, arrhythmias, systemic and pulmonary hypertensive heart disease, valvular heart disease, infective endocarditis and non-infective vegetation, cardiomyopathies and myocarditis, pericardial disease, and pericardial tumors. Congenital heart disease, usually found in newborn babies, has a low mortality rate, and its diagnosis is typically less addressed by machine learning researchers.

On the other hand, machine learning-based diagnosis of cardiovascular diseases (CVDs) such as ischemic heart disease, arrhythmia, and valvular heart disease has been widely studied due to the life-threatening nature of these heart ailments. The vast literature of machine learning-based diagnosis of these diseases lies in the fact that they pose significant risks to human health.

Heart disease is a severe and widespread problem, and coronary artery disease (CAD) is one of its most common and deadly forms. CAD is caused by the build-up of atherosclerotic plaque in the arteries that supply blood to the heart. This plaque build-up can lead to a heart attack, or myocardial infarction (MI), which can be fatal. Therefore, it is crucial to detect the early

formation of atherosclerotic plaque by measuring relevant clinical parameters and taking steps to prevent its formation through appropriate medical interventions.

Diagnosis of heart disease can be carried out using either invasive or non-invasive methods. While coronary angiography (CA) is considered the gold standard for diagnosing heart disease, it is a highly complex, expensive, and risky procedure that requires the expertise of trained medical professionals. Invasive procedures like CA carry the risk of complications such as dissection of the artery, arrhythmia, kidney problems, paralysis, and even death. Moreover, continuous imaging and screening are required in CA, which results in high operational costs. This high cost of diagnosis is often not feasible, especially in underdeveloped and developing countries, and hence, the acceptability of invasive methods is not universal, and many patients in such countries often avoid it.

Machine learning has emerged as a promising tool for the early detection of heart disease. Automated detection of CVDs is an important application area of machine learning, as early detection of heart disease can save many lives. Machine learning-based models for diagnosis are usually termed as clinical decision support systems (CDSS), which can benefit not only heart patients but also physicians and governments. By using machine learning-based CDSS, physicians can make more accurate and timely diagnoses, while governments can better allocate resources to combat heart disease. This work aims to promote the use of CDSS to support clinical decision-making in heart disease diagnosis and management.

2. Objective of the Study

The major objective of the study is to analyse the diverse frameworks that are being used in the machine language for the treatment of Heart disease.

3. Methodologies of Machine Learning used in Heart Disease Diagnosis

This section delves into various models and methods associated with different machine learning frameworks. The latest state-of-the-art techniques and methodologies are discussed and compared across all ML frameworks. Prior to delving into specific technique details, a general definition of machine learning is presented. Machine learning can be broadly categorized into

four categories: supervised, unsupervised, reinforcement learning, and active learning. In the medical domain, supervised and unsupervised machine learning have significant literature.

The present study focuses on classification, which is a supervised machine learning approach. Classification is the process of constructing or learning a mapping or relation using experience from training data and then classifying the testing data by the same model or mapping or relation. Equation 1 provides a concise yet complete definition of a machine learning model that utilizes classification.

Let D be a data set described by a feature set A = {A1, A2, …Am} and a sample set S = {S1, S2, …Sn} with m number of features and s number of samples. Let us further assume that a machine learning framework can be defined by a model L that predicts the class C with accurate prediction (Q). The prediction is evaluated by a performance measure (P) in such a way that the error (Err) is minimized over all training samples during the training phase and all testing samples during the testing phase.

Overall, this section provides a comprehensive understanding of the different machine learning frameworks and how they can be utilized in the context of classification to develop accurate models for diagnosing heart disease. The study highlights the importance of minimizing errors during both the training and testing phases, and how performance measures can be used to evaluate the accuracy of predictions. By gaining a deeper understanding of these concepts, researchers can develop more efficient and effective models to improve the diagnosis and treatment of heart disease.

$$P \overset{C}{\rightarrow} min\left[Err\left\{c - Q\left[L\left(\sum_{j=1}^{n}\sum_{i=1}^{m} D(Ai, Sj)\right)\right]\right\}\right] \quad (1)$$

3.1 Use of Recent Methodologies in ML framework

Pre-processing is an essential step in all kinds of frameworks for diagnosing heart disease, although the specific functions of pre-processing may vary between frameworks. A common strategy to fill missing values is to use instance mean, while normalization is often performed using min-max normalization as reported in a recent survey on data pre-processing for heart disease classification.

The next crucial step in the framework is feature selection, which accelerates the process of building a machine learning (ML) model, resulting in increased efficiency and accuracy. Principal component analysis (PCA) has been used in many recent studies of heart disease prediction for feature selection. A recent survey also provided a detailed overview of data pre-processing for heart disease classification.

The subsequent stage is classification, which is achieved using various classifiers, such as decision tree, Bayesian classifier, artificial neural network, and support vector machines. Zhou et al. (2021) used feature weight-based feature selection and classification through decision tree. Several methodologies, including k-means and ReliefF, are used in pre-processing.

Karaolis et al. (2010) evaluated the risk of coronary heart disease using decision tree, and classification rules were extracted by identifying the most important factors in risk prediction. Alizadehsani et al. (2013) introduced an informative dataset for predicting heart disease, and sequential minimum optimization (SMO) bagging, a training algorithm of SVM, ANN, and Naïve Bayes algorithms, were trained and tested using 10-fold cross-validation. It was found that SMO bagging had the highest accuracy.

In summary, pre-processing is a critical step in all frameworks for diagnosing heart disease, with specific functions that may vary between frameworks. Feature selection and classification are the subsequent stages, where PCA and various classifiers such as decision tree, Bayesian classifier, artificial neural network, and support vector machines are employed. Several methodologies are used in pre-processing, including k-means and Relief, while SMO bagging has been found to have the highest accuracy in predicting heart disease.

Table 1 some studies that shows on the disparity of machine learning framework A for heart disease

Data set	No. of features	No of samples	classifiers	Acc.	Sens.	Spec.	Other measures
SPECT(heart), SPECTF, Statlog	23 44 13	267 267 270	DT with feature weight	67.44, 71.14, 72.74,	--	--	F1-Score 42.89, 46.56, 45.11
Cleveland heart	303	14	FAMD-LR FAMD-kNN FAMD-SVM FAMD-DT FAMD-RF	91.80 90.16 91.80 81.96 93.44	92.85 96.42 100 96.96 90.90	93.93 87.87 90.90 100 96.96	F1-score 91.22 89.65 91.80 78.43 92.59
Cleveland heart	14	303	Hybrid RF with Linear Model	88.4	92.8	82.6	Precision 90.1 F-Measure 90
Statlog(Heart)	13	270	SLFN(ELM) with CSO	86.12	--	--	--
SPECTF SAheart	44 9	267 462		78.61 75.42			
CHS	355	5888	GA-SVM	93.3	99.5	87.1	--
Cleveland heart,Hungarian, Switzerland,VA Long Beach (UCI)	920	14	Fuzzy Boosting with PSO	85.76	90.02	82.31	86.48
Statlog(heart)	13	270	ANN-CAPSO	81.85	74.63	90.21	AUC 0.876
Z-Alizadeh Sani	54	303	Bagging SMO,	94.08	96.30	96.55	--
Dept of Card.PGH, Cyprus	14	1500	DT	75	73	71	--

3.2 Methodologies on evolving ECG

The machine learning (ML) Framework-B involves several more steps as compared to the ML Framework-A. It consists of several stages such as pre-processing, segmentation, feature extraction, feature selection, and classification. In the current study, pre-processing and

segmentation have been considered together. In the case of electrocardiogram (ECG) data, the primary task of pre-processing is to detect and attenuate frequencies that are associated with artifacts. Indiscriminate and adaptive filters may sometimes distort the actual morphology of signals; therefore, wavelet transforms are being used as recent trends in pre-processing. High pass filter, low pass filter, band rejection filter, base line wander, and notch filter are some of the methods used in the pre-processing of ECG data in recent studies. Prior to segmentation, normalization and QRS complex enhancement are done. Segmentation involves transforming a signal into smaller segment signals so that it can be analyzed in a better way. Feature extraction is a crucial phase of ECG classification, and it involves extracting typical ECG features such as duration of P wave, PQ/PR/QT interval, QRS width, amplitudes of P/T/QRS and ST level. However, the most common feature used in machine learning is RR interval. The Pan-Tompkins algorithm is usually used to analyze R peaks, which is ultimately required for RR interval and QRS detection. Some recent studies have presented temporal features and morphological features as two major kinds of extracted features. Temporal features are based on RR intervals, while S-transform and wavelet transform-based features are morphological features. Various methods such as power spectral density, Welch's method, periodogram, Fourier discrete transform, Hamming window, and series of logarithms of signals are used to perform feature extraction. Feature selection is the next stage in Framework-B, which aims to select the most relevant features. The selected features accelerate the classification process and improve the accuracy as well. In a study by Song et al. (2006), the performance of the support vector machine (SVM) classifier was enhanced with linear discriminant analysis (LDA) by classifying ECG signals, achieving a maximum accuracy of 99.88%. Genetic algorithm was used as a feature selection algorithm by encoding genes as 0 to reject a given feature while 1 to accept the given feature. SVM, k-NN, probabilistic neural network, and radial basis function neural network classification are used for ECG signals after feature selection. In another study by Dalal and Vishwakarma (2020), the kernel extreme machine is optimized with genetic algorithm. Deep learning methodologies are also proving their competence for ECG classification nowadays. Huang et al. (2019) proposed a 2-dimensional convolution neural network (2-D CNN) based ECG classification system and achieved a significant classification accuracy of 99.00% in comparison with 1-D CNN (90.93%). Deep neural network has reduced the pre-processing task to a great extent.

3.3 Machine Learning framework-C

Cardiovascular imaging involves multiple imaging techniques, such as X-ray (Computed Tomography (CT)), Echocardiography, Cardiac Computed Tomography (CCT), Cardiac Magnetic Resonance (CMR), Nuclear Imaging, Single Photon Emission Computed Tomography (SPECT), scintigraphy, and Positron Emission Tomography (PET) [76]. Automated cardiovascular image analysis has revolutionized the field by accelerating the diagnosis of heart problems. Echocardiography is an imaging technique that is commonly used due to its low cost. During pre-processing, the image is smoothened by filtering the noise, and unnecessary artifacts are removed while selecting the region containing the heart image. Next, features such as mean, standard deviation, entropy, and texture features are extracted, followed by image classification. Scintigraphy and SPECT detect perfusion defects through correlation calculation in rest and stress images, using color thresholding to identify perfusion. Segmentation and feature extraction are performed, followed by feature ranking, where features are ranked according to their discriminative property, and the best compact extracted feature subset is selected. Abnormal images indicate the presence of heart disease. Echocardiography images mainly involve chamber quantification, ejection fraction and strain measurement, valve images, and overall function. Ultrasound imaging in echocardiography can also capture the image of arteries suffering from atherosclerosis, which can ultimately cause coronary heart disease. Ultrasound imaging is a common, low-cost, and highly reliable non-invasive method in AI and ML. CMR is used to assess the function and cardiac volumes with better accuracy. A study by Tan et al. (2018) used an ANN-based fully automated short axis and long axis information to segment the left ventricle image, leading to the detection of CVD with improved efficacy. The performance measure used in the study is the Jaccard index, which measures the similarity of image objects. AI-based CCT methodologies are used to detect quantification of artery plaques, blood flow, and coronary artery calcium scoring. In a study, AdaBoost outperformed Naïve Bayes and Random Forest in terms of accuracy, sensitivity, specificity, and ROC curve to build an ML model to detect obstructive coronary artery stenoses. PET and SPECT use common methodologies to detect cardiovascular abnormalities using cardiovascular imaging. A recent study used stress and rest images of 192 patients to create a heart image dataset for public use like UCI. A knowledge-

based classification model and a deep learning-based model were used to detect perfusion abnormalities with SPECT imaging. The sensitivity observed was 100% in both models, while accuracy was higher in the deep learning model. However, shallow features outperformed deep features, which were created by SVM. In summary, the study achieved better accuracy with the combination of classical ML models and deep learning models. While it is beyond the scope of this paper to provide a detailed discussion of all cardiovascular image classification methods, recent studies are summarized in Table 2 for cardiovascular image and heart disease prediction. The table includes the reference number of research papers, techniques used for classification, type of imaging, accuracy, sensitivity, specificity, and other performance measures such as p-value, AUC, model building time, and Jaccard index. The results are considered statistically significant if the p-value is less than or equal to 0.05, and classification is considered better if the AUC value is near 1 and poor if it is below 0.5.

Table 2 Recent studies on the methodologies

Techniques	Type of Imaging	Accuracy	Sensitivity /Specificity	Other performance measure
Deep Learning SVM	SPECT	93	100/86	4.03 sec processing time
Random Forest	Echocardigraphy	--	--	AUC .99 p<0.001
Neural network regression	CMR	--	--	Jaccard index 77±0.11, p<0.001
BPNN	Echocardiography	87.5	--	--
AdaBoost,PCA	CCT	70	79/64	--
Naïve Bayes, PCA	Scintigraphy	81.3	83.7/79.2	P<0.05

4. DISCUSSIONS

Recent studies highlight the significant contribution of machine learning in automating the diagnosis of heart disease. Three types of data - clinical and physiological parameters, signals, and images - can be used to build ML models. Framework-A has simpler pre-processing and works efficiently with physiological and clinical parameters and ECG signals using classifiers

like ANN, SVM, KNN, DT, and Bayesian. Framework-B and C focus on feature extraction for raw ECG and PCG signals. Modern studies use envelope and duration features for heart sound segmentation to improve heart disease prediction. SVM and ANN excel in ECG classification while deep neural networks are preferred for PCG and cardiovascular imaging. Although there are many ML techniques available, there is still room for improvement in accuracy and other performance measures.

5. CONCLUSION

This paper provides a concise yet informative overview of the different types of data used for heart disease prediction. The review focuses on contemporary supervised machine learning techniques and identifies three major workflows, namely ML Framework-A, ML Framework-B, and ML Framework-C. Feature selection is found to be crucial in ML Framework-A, while feature extraction plays a key role in ML Framework-B and C. Among the classifiers used for heart disease diagnosis with clinical and physiological parameters, SVM and ANN are found to perform better than others in ML Framework-A. For machine learning methodologies that use raw ECG in ML Framework-B, SVM outperforms other classifiers. In PCG classification (ML Framework-B), convolution neural network (CNN) based deep neural networks are found to work well and outperform other classifiers. ML-based heart disease prediction using cardiovascular imaging (ML Framework-C) is a diverse image classification problem that depends on the imaging techniques used. SVM, ANN, and deep learning methods work well for cardiovascular imaging, including echocardiography, CMR, and SPECT. Performance measures depend on the type, size, and features of the dataset used for building the ML model. The review also highlights the key observation that the hybridization of metaheuristic approaches improves the classification process in terms of time and accuracies when used for heart disease diagnosis.

References

- Z. Jiang and S. Choi, “A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope,” Expert Syst. Appl., vol. 31, no. 2, pp. 286–298, Aug. 2006, doi: 10.1016/j.eswa.2005.09.025.

- J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," IEEE Trans. Biomed. Eng., vol. 62, no. 4, pp. 1169–1178, Apr. 2015, doi: 10.1109/TBME.2014.2377695. A. Gavrovska, G. Zajić, V. Bogdanović, I. Reljin, and B. Reljin, "Identification of S1 and S2 Heart Sound Patterns Based on Fractal Theory and Shape Context," Complexity, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/1580414.
- M. Feldman and S. Braun, "Description of Free Responses of SDOF Systems via the Phase Plane and Hilbert Transform: The Concepts of Envelope and Instantaneous Frequency," vol. 3089, p. 973, 1997.
- H. Liang, S. Lukkarinen, and I. Hartimo, Heart Sound Segmentation Algorithm Based on Heart Sound Envelogram, vol. 24. 1997, p. 108.
- Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support Vectors Machine-based identification of heart valve diseases using heart sounds," Comput. Methods Programs Biomed., vol. 95, no. 1, pp. 47–61, Jul. 2009, doi: 10.1016/j.cmpb.2009.01.003.
- S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a durationdependent hidden Markov model," Physiol. Meas., vol. 31, no. 4, pp. 513–529, Apr. 2010, doi: 10.1088/0967- 3334/31/4/004.
- S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," Expert Syst. Appl., vol. 37, no. 12, pp. 8019–8026, Dec. 2010, doi: 10.1016/j.eswa.2010.05.088.
- F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. Abdullah Ramaiah, "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification," Comput. Biol. Med., vol. 43, no. 10, pp. 1407–1414, Oct. 2013, doi: 10.1016/j.compbiomed.2013.06.016.
- H. Her and H. Chiu, "Using time-frequency features to recognize abnormal heart sounds," in 2016 Computing in Cardiology Conference (CinC), Sep. 2016, pp. 1145–1147.

- W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," Biomed. Signal Process. Control, vol. 32, pp. 20–28, Feb. 2017, doi: 10.1016/j.bspc.2016.10.004.
- W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," Expert Syst. Appl., vol. 84, pp. 220–231, Oct. 2017, doi: 10.1016/j.eswa.2017.05.014.
- Z. Abduh, E. A. Nehary, M. Abdel Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers," Biomed. Signal Process. Control, vol. 57, p. 101788, Mar. 2020, doi: 10.1016/j.bspc.2019.101788.