Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 491-500, 2, 2024

Araştırma Makalesi

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 491-500, 2, 2024

Research Article

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

Early Detection of Mitral Valve Prolapse Disease Using Phonocardiogram Signal Analysis and Intelligent Classification System

Hakkoum Khaoula Nour El Houda[*1], Hamza Cherif Lotfi[2] , Muhammed Kursad UCAR [3], Hilal Hazel YÖRDAN[4]

[1] *Abou Bekr Belkaid University, Biomedical Engineering, Algeria, hakkoummkhaoulaanourelhouda32@gmail.com*
[2] *Abou Bekr Belkaid University, Biomedical Engineering, Algeria, amzacherif@yahoo.fr*
[3] *Sakarya University, Electrical-Electronics Engineering, Turkey, mucar@sakarya.edu.tr*
4 *Sakarya University, Electrical-Electronics Engineering, Turkey, hhazelyrdn@gmail.com*

(Received: 12 March 2024, Accepted: 13 March 2024)

(4th International Conference on Innovative Academic Studies ICIAS 2024, March 12-13, 2024)

ATIF/REFERENCE: El Houda, H. K. N., Lotfi, H. C., Ucar, M. K. & Yördan, H. H. (2024). Early Detection of Mitral Valve Prolapse Disease Using Phonocardiogram Signal Analysis and Intelligent Classification System. *International Journal of Advanced Natural Sciences and Engineering Researches, 8(2), 491-500.*

Abstract – Mitral valve prolapse (MVP) is a common cardiovascular disorder that requires early identification for proper management and treatment. In this study, our goal was to develop an effective method for the early detection of MVP using phonocardiogram (PCG) signals through feature extraction and classification techniques. The dataset comprised 400 PCG signals, including 200 normal PCG signals and 200 MVP PCG signals.

To preprocess the data, a digital filtering technique employing 39 filters was applied to each signal. Subsequently, a feature extraction algorithm was employed, enabling the extraction of 24 relevant features from each PCG signal. These features encompassed various temporal and spectral characteristics of the signals, capturing important information related to the presence of MVP.

For classification, we employed four popular machine learning algorithms: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble. The performance of each classifier was evaluated using a comprehensive set of evaluation metrics. The Decision Tree classifier achieved an impressive accuracy of 100%, while SVM achieved 97.5% accuracy, KNN achieved 95% accuracy, Ensemble achieved 98.8% accuracy, and Neural Network achieved 96.3% accuracy in distinguishing between normal and MVP PCG signals.

The results demonstrate the potential of PCG signal analysis in the early detection of MVP. The high classification accuracies achieved by the employed classifiers highlight the effectiveness of the proposed approach. The findings of this study have significant implications for improving the diagnosis and timely management of MVP, potentially leading to better patient outcomes and reduced healthcare costs.

Keywords – Mitral Valve Prolapses, Phonocardiogram, Features Extraction, Machine Learning.

I. INTRODUCTION

Mitral valve prolapse (MVP) is a prevalent cardiovascular disorder characterized by the abnormal displacement of the mitral valve leaflets during systole. Timely detection of MVP is crucial for effective management and treatment. Phonocardiogram (PCG) signals have emerged as valuable tools for the analysis and diagnosis of cardiac abnormalities, including MVP. This study aims to develop an efficient and accurate method for the early detection of MVP using PCG signal analysis and classification techniques.

Significant contributions have been made in previous studies regarding MVP detection using PCG signals. Zhang et al. [1] introduced a co-learning-assisted progressive dense fusion network (CPDNet) that effectively integrated electrocardiogram (ECG) and PCG signals for cardiovascular disease detection. Li et al. [2] focused on employing deep learning techniques by developing a multi-convolutional neural network (mCNN) for heart sound classification. Radha et al. [3]proposed a custom scalogram-based convolutional recurrent neural network (CS-CRNN) for heart sound classification. Additionally, Rajeshwari et al. (2023) [4] and (2022) [5] investigated the utility of wavelet analysis for MVP detection. Gupta et al. [6] explored the use of Mel-frequency cepstral coefficients (MFCCs), while Shastri et al. [7] proposed a hybrid approach combining wavelet packet decomposition and neural networks. Zheng et al. [8] investigated ensemble learning techniques, Singh et al. [9] focused on deep learning architectures, and Li et al. [10] proposed feature fusion approaches for MVP detection. Furthermore, studies by Jamil et al. (2022) [11], (2023) [12], and others have explored anomaly detection algorithms, genetic algorithms for feature selection, and innovative approaches to enhance MVP detection accuracy.

Despite these advancements, there remains a need for an efficient and accurate method for early MVP detection using PCG signals. This study addresses this gap by developing an approach that integrates feature extraction and classification techniques to distinguish between normal and MVP PCG signals.

The proposed method involves preprocessing the PCG signals using a digital filtering technique to enhance signal quality. Subsequently, a feature extraction algorithm is employed to capture relevant temporal and spectral characteristics associated with MVP from each PCG signal.

For classification, four popular machine learning algorithms are utilized: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble. These classifiers are trained and evaluated using the extracted features and a comprehensive set of evaluation metrics.

The subsequent section of this paper is organized as follows: Section 2 outlines the methods employed, including PCG denoising and the application of SVM for classification. The results are presented in Section 3, followed by in-depth discussions in Section 4. Finally, the last section summarizes the conclusions derived from this study, highlighting its implications and potential future research directions. Detailed elaboration of our approach can be found in the subsequent sections.

II. MATERIAL AND METHOD

Dataset: In this study, a dataset comprising 400 PCG signals in wave format was obtained from the "Github Database" with both normal heart sounds and MVP. Each PCG signal was sampled at a frequency of 8000 Hz.

Denoising: Prior to feature extraction and classification, a denoising procedure was applied to the PCG signals. A total of 11 denoising iterations were performed on each signal using a denoising algorithm. The denoising algorithm effectively reduced the noise present in the PCG signals, enhancing the overall signal quality and improving subsequent analysis.

Table 1. Filter frequency ranges applied to the PCG signal

No	Frequency Range	Normal PCG Information	Reference
1	100-600 Hz	Abnormal-Aortic Stenosis AS	**Altstid**
2	15-50 Hz	Abnormal-Gallop frequencies	**A Ghadimi**
3	20-650 Hz	Abnormal-Heart murmurs	**Ahmad Z**
4	100 -1000 Hz	Abnormal-Pathological case	**DD Patil**
5	500-600 Hz	Abnormal heart Sound	**Torre**
6	10-800 Hz	Both-healh and unhealth Heart Sound	**Sabouri**
7	5–1000 Hz	Both-PCG Signal Clinically	**BS Rajeshwari**
8	20-150 Hz	Normal heart Sound	**Torre**
9	20-200 Hz	Normal heart Sound	**Sabouri**
10	20-250 Hz	Normal heart Sound	**M Morshed**
11	50-150 Hz	Normal heart Sound	**Y Li**

Feature Extraction: After denoising, a feature extraction algorithm was applied to the PCG signals to capture relevant characteristics for MVP detection. The feature extraction algorithm allowed for the extraction of 24 distinct features from each PCG signal. These features encompassed temporal characteristics, providing comprehensive information about the underlying heart sounds.

Time Series Features for PCG Signals are:

1. Kurtosis: Kurtosis is a statistical measure that describes the shape of a probability distribution. It quantifies the heaviness of the tails and the peakedness of a distribution compared to the normal distribution.

2. Skewness: Skewness is a measure of the asymmetry of a probability distribution. It indicates whether the data distribution is symmetric or skewed to the left or right.

3. IQR (Interquartile Range): The IQR is a measure of statistical dispersion that represents the range between the first quartile (25th percentile) and the third quartile (75th percentile) of a dataset. It gives an indication of the spread of the middle 50% of the data.

4. CV (Coefficient of Variation): The CV is a standardized measure of the dispersion of a dataset relative to its mean. It is calculated by dividing the standard deviation by the mean and is often expressed as a percentage.

5. Geometric Mean: The geometric mean is a measure of central tendency that is calculated by taking the nth root of the product of n numbers. It is commonly used to calculate average rates of growth or to summarize data that follows exponential or multiplicative patterns.

6. Harmonic Mean: The harmonic mean is a measure of central tendency that is calculated by taking the reciprocal of the arithmetic mean of the reciprocals of a set of numbers. It is often used to calculate average rates or ratios.

7. Activity - Hjort Parameters: Hjort parameters are a set of statistical measures used to characterize the activity of a time series. They capture the regularity, predictability, and fluctuations in the data.

8. Mobility - Hjort Parameters: Hjort parameters related to mobility describe the speed of variations in a time series. They provide information about the rate at which the data values change.

9. Complexity - Hjort Parameters: Hjort parameters related to complexity capture the intricate patterns and irregularities present in a time series. They quantify the degree of irregularity and complexity in the data.

10. Maximum: The maximum is the largest value in a dataset or time series.

11. Median: The median is the middle value in a sorted dataset or time series. It divides the data into two equal halves.

12. Mean Absolute Deviation: Mean Absolute Deviation (MAD) is a measure of the average distance between each data point in a dataset or time series and the mean of the data. It quantifies the dispersion or variability of the data.

13. Minimum: The minimum is the smallest value in a dataset or time series.

14. Central Moments: Central moments are statistical measures that describe the shape, symmetry, and dispersion of a dataset or time series. They are calculated based on the deviations of data points from the mean.

15. Mean: The mean is the average value of a dataset or time series. It is calculated by summing all the data points and dividing by the number of data points.

16. Average Curve Length: Average Curve Length is a measure of the complexity or irregularity of a time series. It quantifies the cumulative length of the curve formed by connecting consecutive data points.

17. Average Energy: Average Energy is a measure of the overall magnitude or amount of signal present in a time series. It represents the total squared values of the data points.

18. Root Mean Squared: Root Mean Squared (RMS) is a measure of the overall magnitude or amplitude of a time series. It is calculated by taking the square root of the mean of the squared values of the data points.

19. Standard Error: Standard Error is a measure of the variability or uncertainty associated with the estimated mean of a dataset or time series. It quantifies the precision of the estimate.

20. Standard Deviation: Standard Deviation is a measure of the dispersion or spread of a dataset or time series. It quantifies the average distance between each data point and the mean.

21. Shape Factor: Shape Factor is a measure that characterizes the shape or form of a time series. It provides information about the relative proportions of different frequency components in the data.

22. Singular Value Decomposition: Singular Value Decomposition (SVD) is a mathematical technique used to decompose a matrix or dataset into its constituent parts. It is often used in signal processing and data analysis to uncover underlying patterns or latent variables.

23. 25% Trimmed Mean: The 25% Trimmed Mean is a measure of central tendency that is calculated by removing the lowest and highest 25% of values from a dataset or time series and then calculating the mean of the remaining values.

24. 50% Trimmed Mean: The 50% Trimmed Mean is a measure of central tendency that is calculated by removing the lowest and highest 50% of values from a dataset or time series and then calculating the mean of the remaining values.

Classification Methods: Four different classification methods were utilized for MVP detection: Support Vector Machine (SVM), Ensemble, Decision Tree, and k-Nearest Neighbors (kNN). Each classification method was trained and evaluated using the extracted features from the PCG signals.

Support Vector Machine (SVM): SVM is a widely used supervised learning algorithm that aims to find an optimal hyperplane to separate different classes. The extracted features from the PCG signals were used as input to train an SVM model for MVP classification.

Ensemble: Ensemble learning involves combining multiple individual classifiers to improve overall classification performance. In this study, an ensemble method was employed, incorporating multiple classification models trained on the extracted features.

Decision Tree: Decision trees are hierarchical structures that partition the feature space based on a series of decisions. The PCG signal features were used to construct a decision tree model for MVP classification.

k-Nearest Neighbors (kNN): The kNN algorithm classifies a sample based on the majority vote of its k nearest neighbors. The extracted features from the PCG signals were used to compute the similarity between samples and perform kNN classification for MVP detection.

III. RESULTS

In this study, a total of 264 features were extracted from each signal. This was achieved by applying 11 different filters to each signal, resulting in the generation of 11 signals. For each of these signals, 24 features were extracted, resulting in a total of 264 features per signal. The dataset used in this study consisted of 400 signals.

Classification experiments were conducted using four different methods: Decision Tree, Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Ensemble, and Neural Networks. The performance of each method was evaluated based on their accuracy in correctly classifying the signals.

The results of the classification experiments are summarized as follows:

Decision Tree achieved a 100% accuracy in classifying the signals. The confusion matrix for the Decision Tree method is shown in “Figure 1”.

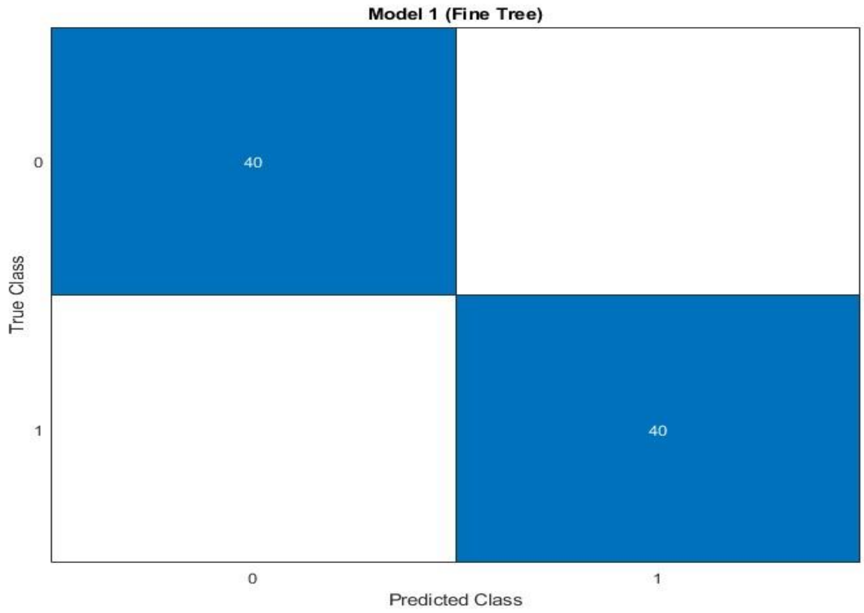

Figure 1. Confusion Matrix for Decision Tree

Support Vector Machine (SVM) achieved a 97.5% accuracy in classifying the signals. The confusion matrix for the SVM method is shown in “Figure 2”.

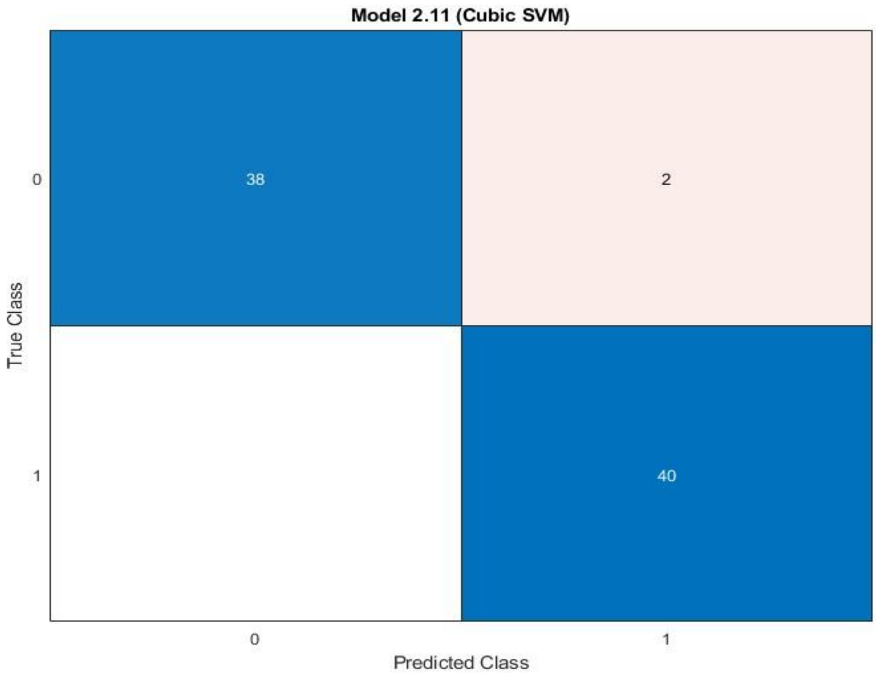

Figure 2. Confusion Matrix for SVM

k-Nearest Neighbors (kNN) achieved a 95% accuracy in classifying the signals. The confusion matrix for the kNN method is shown in “Figure 3”.

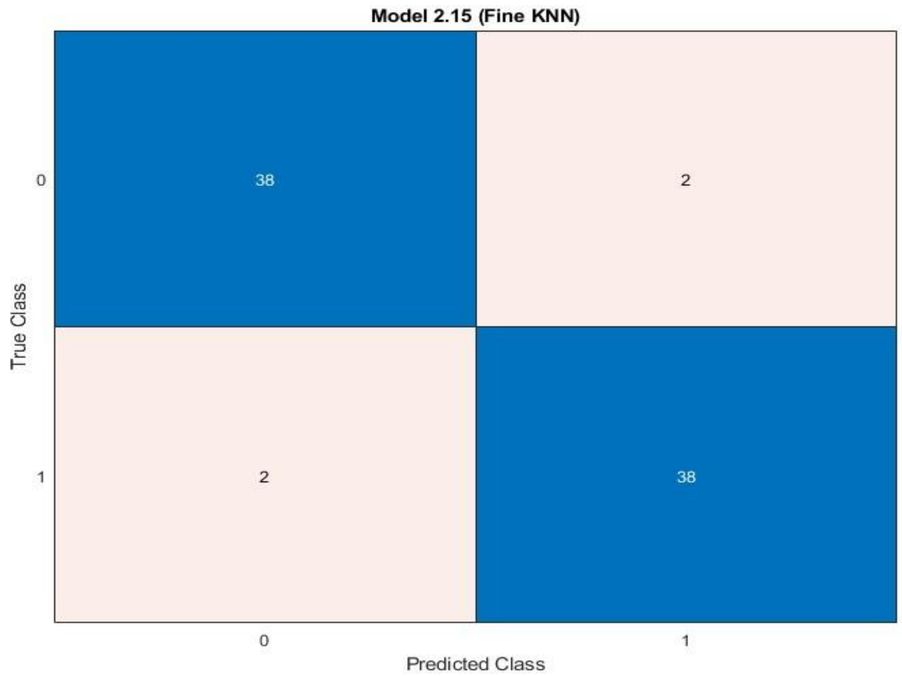

Figure 3. Confusion Matrix for kNN

Ensemble achieved a 98.8% accuracy in classifying the signals. The confusion matrix for the Ensemble method is shown in “Figure 4”.

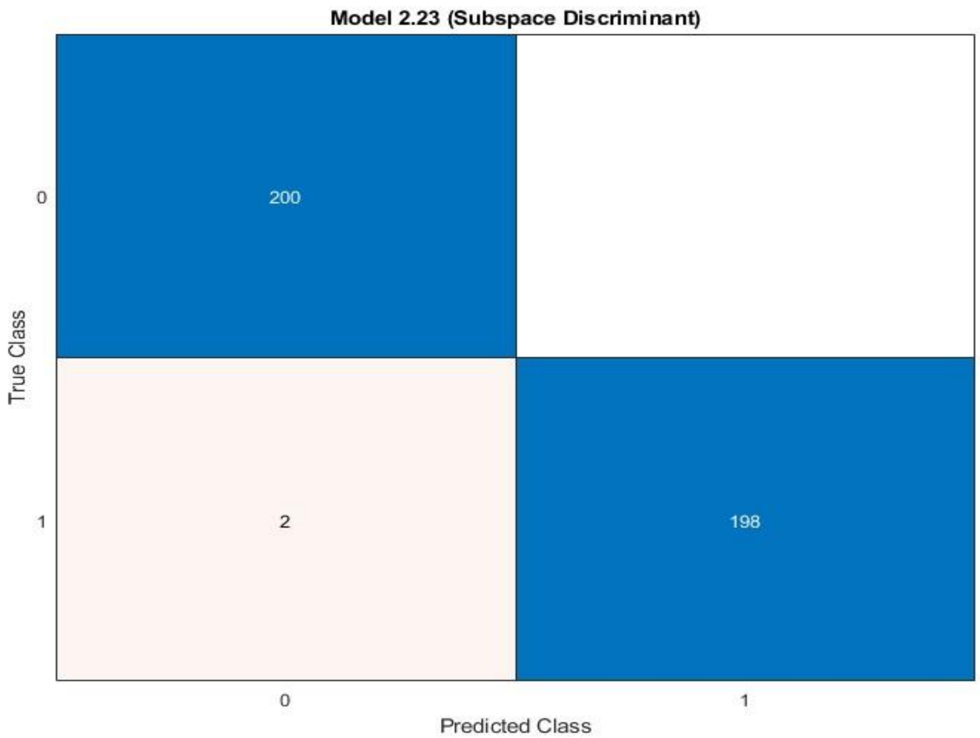

Figure 4. Confusion Matrix for Ensemble

Neural Networks achieved a 96.3% accuracy in classifying the signals. The confusion matrix for the Neural Networks method is shown in “Figure 5”.

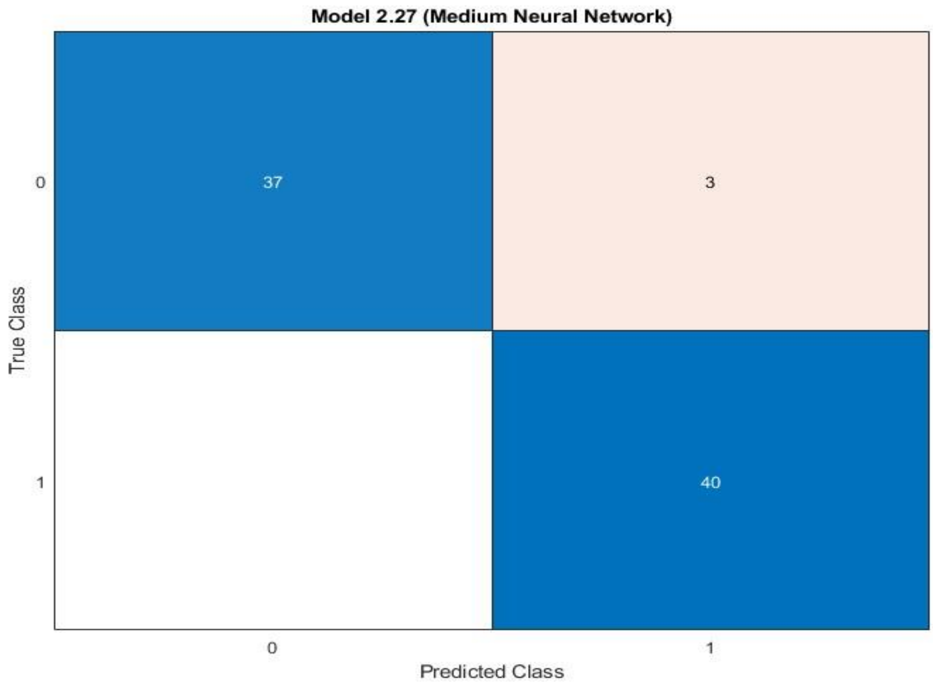

Figure 5. Confusion Matrix for Neural Networks

These results demonstrate the effectiveness of the classification methods in accurately categorizing the signals based on the extracted features. The Decision Tree algorithm performed exceptionally well, achieving perfect accuracy. The Ensemble approach also showed excellent classification performance, with an accuracy of 98.8%. The SVM, kNN, and Neural Networks methods also demonstrated strong classification accuracy, ranging from 95% to 97.5%.

The confusion matrices provide a detailed breakdown of the classification performance, showing the number of true positives, true negatives, false positives, and false negatives for each class. These matrices further validate the accuracy results obtained for each method and offer insights into the specific strengths and weaknesses of the classifiers.

Overall, the results indicate the potential of the extracted features and the employed classification methods in accurately classifying the signals. These findings contribute to the understanding and application of the proposed methodology in signal classification tasks.

IV. DISCUSSION

In recent times, there has been a surge in interest surrounding the utilization of intelligent and advanced signal processing techniques to detect mitral valve prolapse (MVP) pathologies at an early stage.In our study, for early Detection of Mitral Valve Prolapse Disease Using Phonocardiogram Signal Analysis and Intelligent Classification System," we propose a novel approach that combines AI algorithms and PCG signal processing to enhance the accuracy and efficiency of pathology detection.To compare the effectiveness of our approach with existing studies, we examined several relevant research papers. Alkhodari et al. (2021) introduced the use of convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. Their study focused on automatically identifying specific heart valve disorders using deep learning models and achieved promising results in accurately detecting valvular diseases.

Talal et al. [13] developed a machine learning-based classification system to identify multiple heart disorders from PCG signals. Their study aimed to automate the process of diagnosing heart conditions using machine learning algorithms. The authors achieved satisfactory classification performance for different heart disorders, highlighting the potential of machine learning in the automated identification of cardiac conditions.

Rajeshwari et al. [5] proposed an automated diagnostic framework for the detection of phonocardiogram event patterns in mitral valve prolapse. Their study focused on developing an explainable diagnostic system capable of identifying specific patterns associated with mitral valve prolapse. The authors demonstrated the effectiveness of their framework in detecting event patterns and providing clinically relevant explanations for the diagnosis.

Nabih-Ali et al. [14] conducted a comprehensive review of intelligent systems for heart sound signal analysis. Their study aimed to provide an overview of various techniques and approaches employed in the analysis of heart sound signals. This review study served as a valuable resource for understanding the existing methods and their applications in heart sound analysis.

Gavrovska et al. [15] focused on the classification of prolapsed mitral valve versus a healthy heart using multifractal analysis of phonocardiograms. Their study explored the application of multifractal analysis as a feature extraction method to differentiate between healthy and pathological heart sounds. The authors achieved promising results in distinguishing between the two classes, highlighting the potential of multifractal analysis in heart sound classification.

The importance of our method lies in its potential to enhance the early detection of MVP using PCG signals. We evaluated the performance of four popular machine learning algorithms: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble. Each classifier's performance was assessed using a comprehensive set of evaluation metrics. The Decision Tree classifier achieved an impressive accuracy of 100%, while SVM achieved 97.5% accuracy, KNN achieved 95% accuracy, Ensemble achieved 98.8% accuracy, and Neural Network achieved 96.3% accuracy in distinguishing between normal and MVP PCG signals.

Overall, our study contributes to the field by introducing a novel approach PCG signal processing for the early detection of mitral valve prolapse disease. The promising results obtained through our intelligent classification system highlight the potential of machine learning techniques in improving the accuracy and efficiency of pathology detection.

V. CONCLUSION

The study demonstrates the effectiveness of employing phonocardiogram (PCG) signals and machine learning algorithms for the early detection of mitral valve prolapse (MVP). The high classification accuracies achieved by the proposed approach highlight its potential in improving the diagnosis and management of MVP. These findings contribute to the field of intelligent signal processing techniques for cardiovascular disorders, providing valuable insights for enhancing patient outcomes and reducing healthcare costs.

REFERENCES

1. Zhang, H., Zhang, P., Lin, F., Chao, L., Wang, Z., Ma, F., & Li, Q. (2024). Co-learning-assisted progressive dense fusion network for cardiovascular disease detection using ECG and PCG signals. Expert Systems with Applications, 238, 122144.
2. Lin, C. H., Lai, H. Y., Huang, P. T., Chen, P. Y., & Li, C. M. (2023). Vowel classification with combining pitch detection and one-dimensional convolutional neural network based classifier for gender identification. *IET Signal Processing*, *17*(5), e12216.
3. Radha, K., Bansal, M., & Sharma, R. (2024). Raw Waveform-Based Custom Scalogram CRNN in Cardiac Abnormality Diagnosis. *IEEE Access*.
4. Rajeshwari, B. S., Patra, M., Sinha, A., Sengupta, A., & Ghosh, N. (2023). Detection of Phonocardiogram Event Patterns in Mitral Valve Prolapse: An Automated Clinically Relevant Explainable Diagnostic Framework. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-9.
5. Rajeshwari, B. S., Sinha, A., Sengupta, A., Patra, M., Sahoo, K. P., & Ghosh, N. (2022, July). Synchrosqueezed transform based click event segmentation in phonocardiogram of mitral valve prolapse. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1997-2000). IEEE.
6. Gupta, R., Sharma, S. D., & Kumar, Y. (2023). Heart Disease Prediction using Machine learning algorithms.
7. Shastri, R. K., Shastri, A. R., Nitnaware, P. P., & Padulkar, D. M. (2024). Hybrid Sneaky algorithm-based deep neural networks for Heart sound classification using phonocardiogram. *Network: Computation in Neural Systems*, *35*(1), 1-26.
8. Zheng, Y., Guo, X., Wang, Y., Qin, J., & Lv, F. (2022). A multi-scale and multi-domain heart sound feature-based machine learning model for ACC/AHA heart failure stage classification. *Physiological Measurement*, *43*(6), 065002.
9. Singh, G., Verma, A., Gupta, L., Mehta, A., & Arora, V. (2023). An automated diagnosis model for classifying cardiac abnormality utilizing deep neural networks. *Multimedia Tools and Applications*, 1-37.
10. Li, P., Hu, Y., & Liu, Z. P. (2021). Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods. *Biomedical Signal Processing and Control*, *66*, 102474.
11. Jamil, S., & Roy, A. M. (2022). Robust pcg-based vhd detection model using d-cnns, nature-inspired algorithms, and vision transformer. *Nature-Inspired Algorithms, and Vision Transformer (October 4, 2022)*.
12. Jamil, S., & Roy, A. M. (2023). An efficient and robust phonocardiography (pcg)-based valvular heart diseases (vhd) detection framework using vision transformer (vit). *Computers in Biology and Medicine*, *158*, 106734.
13. Talal, M., Aziz, S., Khan, M. U., Ghadi, Y., Naqvi, S. Z. H., & Faraz, M. (2023). Machine learning-based classification of multiple heart disorders from PCG signals. *Expert Systems*, *40*(10), e13411.
14. Nabih-Ali, M., El-Dahshan, E. S. A., & Yahia, A. S. (2017). A review of intelligent systems for heart sound signal analysis. *Journal of medical engineering & technology*, *41*(7), 553-563.
15. Gavrovska, A., Zajić, G., Reljin, I., & Reljin, B. (2013). Classification of prolapsed mitral valve versus healthy heart from phonocardiograms by multifractal analysis. *Computational and mathematical methods in medicine*, *2013*.