Research on Humanities and Social Sciences www.iiste.org
ISSN 2224-5766 (Paper) ISSN 2225-0484 (Online)
Vol.13, No.14, 2023

Predicting Cardiovascular Disorders Through Stethoscope Audio Using Convolutional Neural Network

Erika Carlos Medeiros[1*] Patricia Cristina Moser[1] Jorge Cavalcanti Barbosa Fonsêca[1]
Rômulo César Dias de Andrade[1] Fernando Ferreira de Carvalho[1,2]
1. University of Pernambuco, Caruaru, PE, Brazil
2. Cesar School, Recife, PE, Brazil
* E-mail of the corresponding author: erika.medeiros@upe.br

Abstract
Cardiovascular disorders pose a significant global health challenge, resulting in a substantial number of annual deaths. Early and accurate prediction of heart disorders is crucial to mitigate their impact on individuals and healthcare systems. In this study, we explore the potential of Convolutional Neural Network in automating heart disease prediction using spectrogram data. The dataset comprises audio recordings collected from the general public via an iPhone app and a clinical trial using a digital stethoscope. We preprocess the data to obtain spectrograms and design a Convolutional Neural Network architecture to classify heart sounds into distinct categories. The Convolutional Neural Network exhibits promising performance, achieving an accuracy of approximately 77%. Our research highlights the opportunity to leverage Convolutional Neural Network in this context, paving the way for advanced automated cardiac diagnostics.
Keywords: Cardiovascular disorders, Convolutional neural network
DOI: 10.7176/RHSS/13-14-02
Publication date: August 31st 2023

1. Introduction

Cardiovascular disorders (CVDs) stand as a formidable and widespread global health challenge, responsible for a staggering number of deaths each year. In 2019 alone, an estimated 17.9 million lives were claimed by CVDs, accounting for approximately 32% of all global death. Tragically, the majority of these fatalities occur in low- and middle-income countries, underscoring the urgent need to address this silent epidemic in vulnerable communities. Alarmingly, a significant proportion of premature deaths, occurring before the age of 70, are attributed to CVDs, amounting to 38% of all noncommunicable disease-related deaths. These statistics reinforce the critical importance of early detection and effective management strategies to mitigate the devastating impact of CVDs on individuals and healthcare systems worldwide (WHO, 2021).

Traditionally, the clinical assessment of cardiac health relies on the use of conventional stethoscopes, enabling medical professionals to listen to and interpret heart auscultations. However, the accurate diagnosis of heart conditions using conventional stethoscopes requires extensive training and experience, rendering it an operator-dependent technique with inherent limitations. Furthermore, distinguishing between subtle variations in heart sounds corresponding to different heart pathologies remains a complex and challenging task (Leng et al., 2015). As a result, there is a pressing demand for innovative and technologically-empowered solutions to enhance the diagnostic accuracy and accessibility of cardiac evaluations.

Historically, cardiac auscultation has been the primary method for identifying heart abnormalities, relying on the trained ear of medical professionals. While widely accepted, this operator-dependent technique has inherent limitations, often resulting in variations in diagnostic accuracy (Patrizio et al., 2023). However, the convergence of healthcare and cutting-edge technology has brought forth a transformative solution - the integration of Artificial Intelligence (AI) (McCarthy, 2007) into cardiac diagnostics (Houssein et al., 2022).

In recent times, the remarkable progress in technology and the advent of AI have opened new avenues for healthcare. Integrating AI into medical practices, specifically in the form of Machine Learning (ML) (Mohanty et al., 2021) and Deep Learning (DL) (Learning, 2020) algorithms, has shown significant promise in transforming collected data into valuable clinical insights (Brites et al., 2022). Harnessing the power of AI to augment cardiac auscultation could revolutionize the way we detect and diagnose heart conditions.

In this context, this study at hand leveraged data gathered from the general public through the iStethoscope Pro iPhone app and a clinical trial in hospitals, employing the digital stethoscope DigiScope (Bentley et al., 2012). The objective of this paper is to develop an automated cardiac diagnostic system using Convolutional Neural Network (CNN) (Albawi et al., 2017) architecture, leveraging data collected from the general public via the iStethoscope Pro iPhone app and a clinical trial using the digital stethoscope DigiScope. The focus is on accurate classification of normal and abnormal heart sounds by extracting essential features from spectrograms. This innovative approach holds great promise in improving healthcare by providing reliable and efficient tools for early diagnosis of abnormal cardiac conditions, ultimately benefiting patients' care and treatment.

Among the related works listed in Section 2, there is a notable absence of studies that utilize CNNs for predicting heart disorders. This study aims to explore and address this gap by investigating the application of CNNs in heart disease prediction. By harnessing the potential of CNNs to analyze complex spectrogram data, we intend to enhance the accuracy and efficiency of heart disease classification, thereby contributing to the advancement of automated diagnostic systems in the medical field.

In addition to the introduction, the paper comprises five main sections. The second section, "Related Work," discusses existing research on heart disease prediction using various data mining and ML techniques. It provides a comprehensive review of previous approaches and identifies the gap in the literature regarding the utilization of CNNs for heart disorder prediction. The third section, "Methodology," presents the detailed framework and methodology employed in this study, including data preprocessing, model architecture, and training procedures. The fourth section, "Discussion of Results," thoroughly analyzes and interprets the outcomes obtained from the CNN model, learning curves, and confusion matrix, shedding light on its strengths and areas of improvement. The fifth section, "Conclusions," offers a comprehensive summary of the research findings, presents key insights, and proposes future directions for enhancing the model's accuracy and efficiency. Lastly, the paper includes a "References" section that acknowledges the sources of information and prior studies that have contributed to the development of this research. Together, these sections provide a comprehensive exploration of the proposed CNN-based heart disorder prediction approach and its potential impact on automated medical diagnostics.

2. Related Work

In the section of related works, we explore research and studies that have addressed the prediction of heart disease using various ML and data mining techniques. These works have contributed valuable insights into the development of accurate and efficient prediction models, aiming to improve early diagnosis and management of heart-related conditions. By examining these related works, we can gain a comprehensive understanding of the advancements in this domain and identify the strengths and limitations of different approaches to heart disease prediction.

According to Singh and Kumar (2020), heart disease, also known as CDV, has become a leading cause of death worldwide in recent decades. Early and accurate diagnosis is crucial for effective management of the disease and requires reliable approaches. Data mining has emerged as a valuable technique for processing vast healthcare data. Researchers in this domain employ various data mining and ML techniques to analyze complex medical data and predict heart disease. This research paper focuses on heart disease attributes and presents a model based on supervised learning algorithms, including Naïve Bayes (Webb et al., 2010), Decision Tree (DT) (Suthaharan, 2016), K-nearest neighbor (KNN) (Peterson, 2009), and Random Forest (RF) (Breiman, 2001). The dataset used for this study is sourced from the Cleveland database of UCI repository and contains 303 instances with 76 attributes. Among these attributes, only 14 are considered for testing to evaluate the performance of different algorithms. The main goal of this research is to estimate the probability of developing heart disease in patients. The findings indicate that the KNN algorithm achieved the highest accuracy score in this prediction task.

According to Jindal *et al.* (2021), the research paper focuses on predicting heart disease based on various medical attributes. They developed a heart disease prediction system using ML algorithms such as Logistic Regression (LR) (Kleinbaum et al., 2002) and KNN. The proposed model showed good accuracy compared to previously used classifiers like Naive Bayes. The system enhances medical care and reduces costs, providing significant knowledge to predict patients with heart disease effectively.

Kavitha *et al.* (2021) proposed a novel ML approach for predicting heart disease using the Cleveland heart disease dataset. The study employed data mining techniques such as regression and classification, utilizing RF, DT, and a hybrid model combining both algorithms. The hybrid model achieved an impressive accuracy level of 88.7% in heart disease prediction. An interface was designed for user input, making personalized predictions accessible and user-friendly.

Brites *et al.* (2022) addresses the challenge of heartbeat sound classification for heart disease diagnosis. The dataset used contains three categories: Normal, Murmur, and Extra-systole heartbeat sounds. The proposed framework includes noise removal using band filters and fixing the sampling rate size of each sound signal. Down-sampling techniques are then applied to obtain more discriminant features and reduce the frame rate dimension without compromising results. The proposed model, based on Recurrent Neural Network (RNN) (Nelson, 2017) with Long Short-Term Memory (LSTM) (Nelson, 2017), Dropout (Nelson, 2017), Dense, and Softmax layers (Nelson, 2017), outperforms other methods and shows promising competitiveness for heartbeat sound classification.

According to Ramesh *et al.* (2022), the World Health Organization (WHO) reports that approximately 12 million deaths occur due to heart diseases worldwide. These diseases have a significant impact on the medical industry and are particularly dangerous in countries with limited resources and inadequate medical attention. Timely and accurate prediction of heart diseases is a challenge, but advances in data mining and ML techniques offer hope. In this study, various ML and DL algorithms were applied to a dataset with 303 rows and 14 attributes obtained from Kaggle. The models used include LR, Naive Bayes, KNN, Support Vector Machine (SVM)

(Steinwart & Christmann, 2008), Multi-Layer Perceptrons (MLP) (Taud & Mas, 2018), Artificial Neural Network (ANN) (Hopfield, 1988), DT, RF, XG Boost (Chang et al., 2018), and Cat Boost (Gamini, 2021), aiming to improve heart disease prediction and reduce the number of deaths associated with cardiovascular conditions.

According to Disease *et al.* (2023), ML is widely utilized in various fields, including healthcare. It has the potential to play a crucial role in predicting locomotor abnormalities, heart ailments, and other medical conditions. Early detection of such conditions can provide valuable insights to doctors, enabling personalized diagnosis and treatment. The focus of the project is to employ ML algorithms for predicting potential heart conditions in humans. They compare the performance of different classifiers, such as DT, Naive Bayes, LR, SVM, and RF. Additionally, They propose an ensemble classifier (Dietterich, 2000) that combines the strengths of strong and weak classifiers, utilizing a large number of training and validation samples. The goal is to enhance the accuracy and efficiency of our predictive model for heart disease detection.

In conclusion, the section on related works has provided a comprehensive overview of the existing research on heart disease prediction and the various data mining and ML techniques employed in this domain. From the literature review, it is evident that heart diseases pose a significant global health challenge, and the accurate and timely prediction of these conditions is crucial for effective medical interventions. While previous studies have explored different approaches, there remains an opportunity to leverage the power of CNNs in this context, as seen less frequently in the literature. In the next section, we will delve into the methodology used in this study, detailing the data preprocessing, model architecture, and evaluation procedures, aiming to develop an innovative and accurate automated cardiac diagnostic system.

3. Methodology

In this study, we followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework (Eckerson et al., 2000) to automate the classification of heart sound recordings using a combination of CNN. The CRISP-DM process consists of six main phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

a) Business Understanding:

The primary objective of this study was to develop a robust and accurate system capable of automatically classifying different cardiac pathologies from heart sound recordings. Early detection of cardiac disorders through automated systems could significantly impact healthcare, leading to timely interventions and improved patient outcomes.

b) Data Understanding:

The heart sound dataset consisted of 585 audio files, each labeled with one of four classes: Normal, Murmur, Extra Heart Sound, and Artifact. Each class represents specific heart sound characteristics and may have additional background noises.

The dataset was used in Gomes and Pereira (2012). The audio files had varying lengths, ranging from 1 second to 30 seconds. We conducted exploratory data analysis to gain insights into the data distribution, class imbalances, and patterns.

The Normal category contains recordings of normal, healthy heart sounds, with variations in background noises and occasional random noise. The heartbeats in this category typically exhibit a clear "lub dub, lub dub" pattern, with the time from "lub" to "dub" shorter than the time from "dub" to the next "lub" when the heart rate is less than 140 beats per minute (Gomes & Pereira, 2012).

The Murmur category represents heart murmurs, characterized by a whooshing, roaring, rumbling, or turbulent fluid noise at specific temporal locations: between "lub" and "dub" or between "dub" and "lub." Heart murmurs can be a symptom of various heart disorders, some of which can be serious. It is essential to differentiate murmurs from normal heart sounds, as they occur between "lub" and "dub" or between "dub" and "lub," not coinciding with these events (Gomes & Pereira, 2012).

The Extra Heart Sound category includes additional heart sounds, such as "lub-lub dub" or "lub dub-dub." While extra heart sounds may not always indicate disease, they can be essential signs of underlying health conditions, and early detection could be beneficial. These sounds are crucial to detect as they may not be adequately identified using ultrasound (Gomes & Pereira, 2012).

The Artifact category contains various sounds like feedback squeals, echoes, speech, music, and noise. In this category, heart sounds are typically not discernible, and the temporal periodicity at frequencies below 195 Hz is minimal or absent. Distinguishing this category from the others is essential to ensure data quality and provide instructions to reattempt data collection if necessary (Gomes & Pereira, 2012).

Finally, the Extrasystole category involves sounds that indicate irregular heartbeats with extra or skipped heartbeats, e.g., "lub-lub dub" or "lub dub-dub." Extrasystoles can be normal occurrences in adults and common in children but may also be linked to heart diseases in specific cases (Gomes & Pereira, 2012).

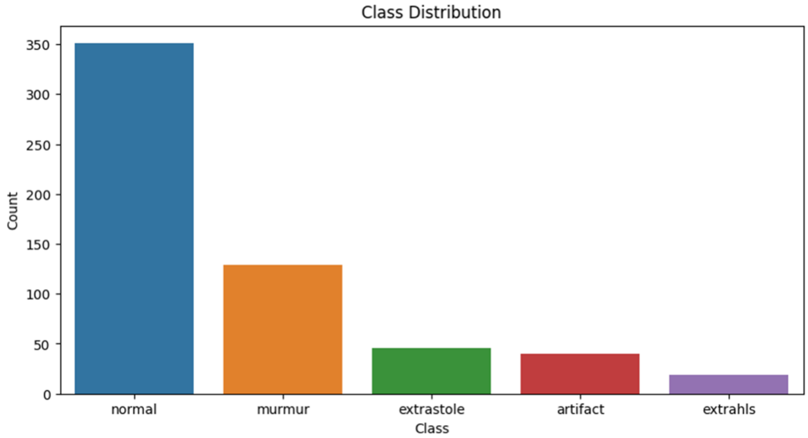
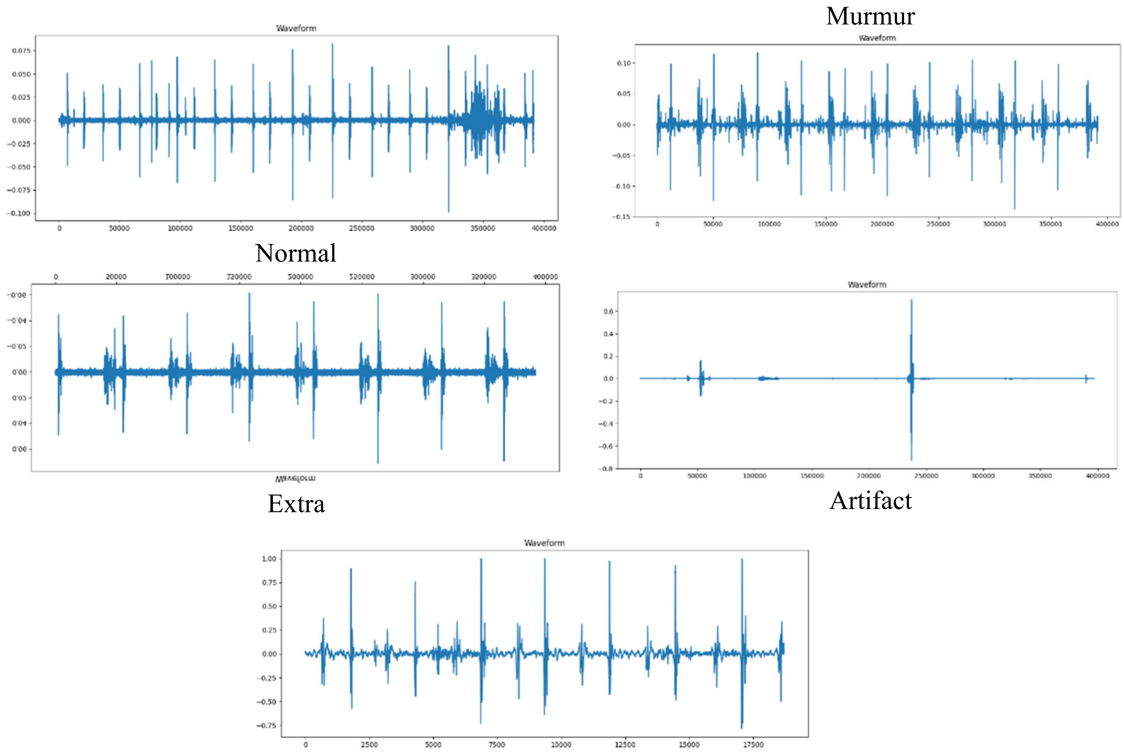


Figure 1. Class Distribution

As depicted in Figure 1, the distribution of samples among these classes is as follows: Normal class contains 351 samples, Murmur class has 129 samples, Extrasystole class has 46 samples, Artifact class includes 40 samples, and the Extra Heart Sound class (Extrahls) comprises 19 samples. It is evident from this distribution that the Normal class has the highest number of samples, indicating a relatively balanced representation of this category. On the other hand, the Extra Heart Sound and Artifact classes have notably fewer samples, suggesting a potential class imbalance that may impact model performance.



Extrasystole

Figure 2. Waveform shapes of the different heart sound classes

Figure 2 illustrates the waveform shapes of the different heart sound classes. The waveforms provide visual representations of the acoustic signals captured from heart sound recordings, highlighting the distinctive patterns and variations that characterize each cardiac condition. Analyzing the waveform shapes enables a better understanding of the frequency components and temporal characteristics associated with normal heart sounds, murmurs, extrastole, and other cardiac abnormalities.

c) Data Preparation:

To ensure uniformity and extract relevant features for classification, we conducted essential preprocessing steps on the raw audio data. Firstly, the audio files were loaded and converted into Mel spectrograms. Mel spectrograms provide a frequency representation of sound over time, allowing us to capture the distinctive patterns and characteristics of heart sounds (Wyse, 2017). Additionally, we applied a power-to-decibel transformation to enhance the representation of the spectrograms. Since the generated spectrograms had varying lengths, we employed padding or truncation to fix their length at 1000. This standardization enabled us to achieve consistent input sizes for the CNN model. The final preprocessed spectrogram data, along with their corresponding labels, served as the input and target data for training and evaluating the CNN. In the context of addressing the class imbalance observed in the dataset, we employed the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) to balance the distribution of samples among the different classes. As detailed in Figure 1, the dataset initially presented an uneven distribution. Recognizing the potential impact of this class imbalance on model performance, we applied SMOTE exclusively to the training dataset. By doing so, we generated synthetic samples for the minority classes, thereby augmenting their representation and achieving a more equitable class distribution. This preprocessing step aims to enhance the robustness and effectiveness of our machine learning model in effectively capturing patterns across all classes.

d) Modeling:

In the Modeling phase, we designed a CNN tailored to analyze the heart sound spectrogram data. The CNN architecture, as shown in Section 4, comprised several key layers, starting with convolutional layers, which were responsible for detecting specific patterns and features in the spectrogram data. MaxPooling layers were then employed to downsample and reduce the spatial dimensions of the extracted features, enhancing computational efficiency and preventing overfitting. The flatten layer transformed the output of the convolutional layers into a one-dimensional vector, ready to be fed into dense layers for the final classification. Within the dense layers, the CNN learned complex representations and relationships among the extracted features, enabling accurate categorization of heart sounds into different cardiac categories for diagnostic purposes. During training, the CNN optimized its parameters using the Adam optimizer (Tato & Nkambou, 2018) while minimizing the sparse categorical cross-entropy loss function, thereby honing its ability to make precise predictions. The evaluation of the CNN's performance was based on its accuracy (Monico et al., 2009) in classifying heart sound data, reflecting its proficiency in distinguishing various cardiac conditions.

e) Evaluation:

During the evaluation process, we divided the dataset into training (80%) and test sets (20%). The primary metric used to assess the models' performance was accuracy, which represents the percentage of correctly classified instances in the test set. By focusing on accuracy, we gained an overall understanding of how well the models were able to make correct predictions for the cardiac diagnostic task.

Additionally, we constructed confusion matrices (Beauxis-Aussalet & Hardman, 2014) to visualize the models' performance across all classes. The confusion matrices provide insights into the number of true positive, true negative, false positive, and false negative predictions for each class, allowing us to identify specific areas where the models may be struggling to make accurate classifications for certain cardiac pathologies.

In this methodology section, the study follows the CRISP-DM framework to develop an automated cardiac diagnostic system using CNN architecture. Initial phases define the objective and explore the heart sound dataset. Data Preparation involves transforming audio into Mel spectrograms, while Modeling introduces a CNN tailored for this task. Evaluation assesses the model's accuracy and visualizes performance through confusion matrices. This approach bridges expert medical assessment with accessible diagnostic tools, forming the foundation for the subsequent detailed discussion of results.

In the next section, we will delve into the outcomes yielded by the modeled CNN, discussing the achieved accuracy and analyzing the confusion matrix. Furthermore, we will provide an in-depth exposition of the neural network architecture employed, spotlighting its layers and functionalities within the realm of cardiac sound classification. These insights will furnish a comprehensive understanding of the model's performance and structure.

4. Discussion of Results

In this section, we will explore the results obtained from the CNN model architecture.

The CNN architecture consists of several layers. Figure 3 depicts the architecture of the CNN.

Below is an explanation of the function of each layer:

a) Conv2D (Convolutional) Layer: This layer performs convolutional operations on the input spectrogram data. It uses 32 filters of size (3, 3) to scan the input spectrogram and extract local patterns, such as edges and textures. Each filter learns to detect specific features from the input data. The output of this layer is a feature map with 32 channels, representing different learned patterns;

b) MaxPooling2D Layer: The MaxPooling layer is a downsampling operation that reduces the spatial dimensions of the feature map obtained from the Conv2D layer. It applies a pooling filter of size (2, 2) to the feature map and keeps the maximum value within each pooling window. MaxPooling helps in reducing the computational complexity and prevents overfitting by capturing the most important information from the feature map;

c) Conv2D Layer (Second Convolutional Layer): Similar to the first Conv2D layer, this layer performs convolutional operations using 64 filters of size (3, 3). It further learns higher-level features from the previously extracted patterns in the feature map;

d) MaxPooling2D Layer (Second MaxPooling Layer): This layer performs another downsampling operation on the feature map using a pooling filter of size (2, 2). It further reduces the spatial dimensions and captures more abstract and robust features;
e) Flatten Layer: The Flatten layer converts the 2D feature map obtained from the last MaxPooling2D layer into a 1D vector. This process "flattens" the data to be fed into the subsequent Dense layers;
f) Dense Layer (First Dense Layer): The first Dense layer is a fully connected layer with 128 neurons. It takes the flattened 1D vector as input and performs matrix multiplication with learned weights to extract higher-level representations of the features.
g) Dropout Layer: Dropout is a regularization technique that randomly drops a certain fraction (in this case, 50%) of the neurons' outputs during training. This helps in preventing overfitting by forcing the model to learn redundant features and enhances the model's generalization capability;
h) Dense Layer (Second Dense Layer): The last Dense layer consists of 5 neurons, representing the number of output classes (heart disorder categories). It applies the Softmax activation function to convert the output scores into probabilities, indicating the likelihood of the input spectrogram belonging to each class.

Each layer in the CNN architecture plays a crucial role in extracting relevant features from the input spectrograms and transforming them into meaningful predictions for heart disorder classification.

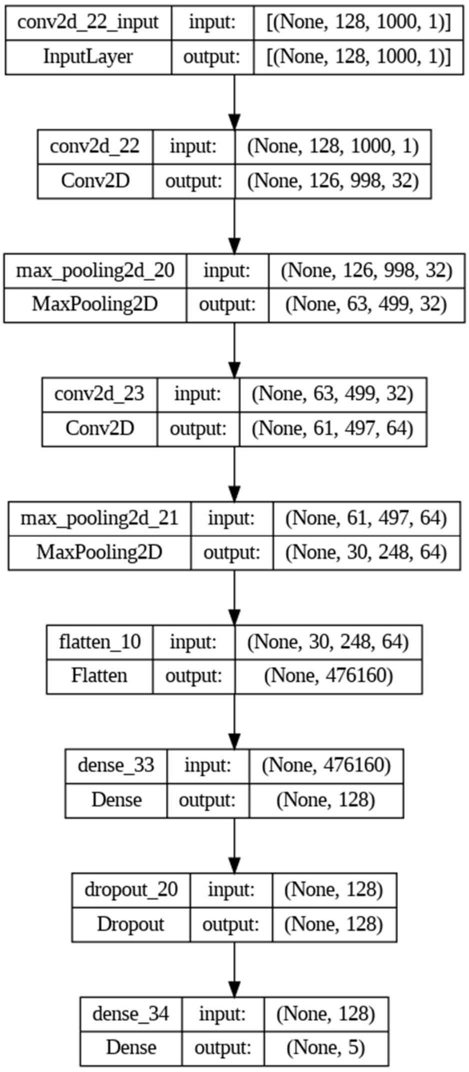


Figure 3. The architecture of the CNN

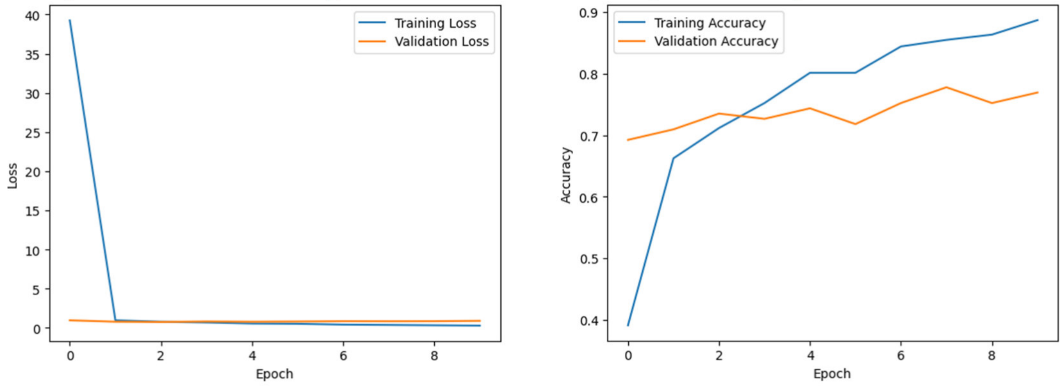
The learning curve plot in Figure 4 shows the training and validation performance of the CNN model over 10 epochs. The x-axis represents the number of epochs, and the y-axis represents the loss and accuracy values. From the results, we can observe the following trends:

a) Loss and Accuracy: As the number of epochs increases, both the training and validation loss decrease, and the accuracy increases. This indicates that the model is learning and improving its performance on both the training and validation datasets;

b) Overfitting: Initially, in the first few epochs, there is a significant gap between the training and validation accuracy, with the training accuracy being higher. This gap starts to narrow down as the epochs progress, suggesting that the model is becoming more generalized and less prone to overfitting;

c) Convergence: The loss and accuracy values seem to stabilize after a few epochs, indicating that the model has reached a point of convergence, and further training may not lead to substantial improvements;

d) Generalization: The final validation accuracy of approximately 0.77 (77%) suggests that the model can predict heart disorders with reasonable accuracy on unseen data. However, there is still room for improvement, and further optimization or architectural adjustments might lead to better results.



Loss Learning Curve Accuracy Learning Curve

Figure 4. Learning Curves

Overall, the learning curve indicates that the CNN model is learning from the data and making progress in predicting heart disorders. However, it is essential to fine-tune the model and experiment with hyperparameters to achieve even better performance and potentially address any limitations in its current state.

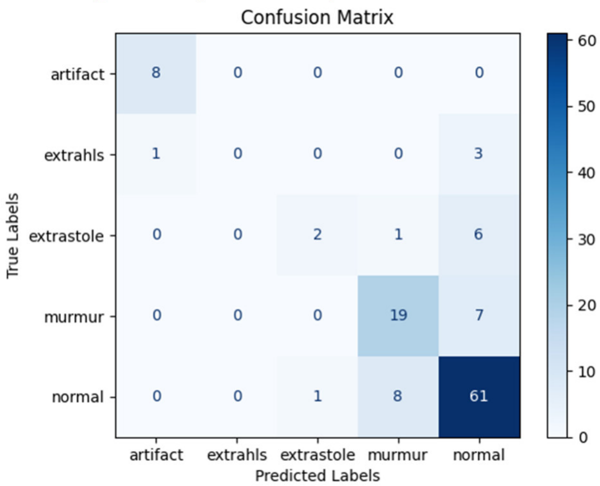


Figure 5. Confusion Matrix

The confusion matrix presented in Figure 5 provides a comprehensive evaluation of the CNN model's performance for each heart sound class. In the "artifact" class, all 8 instances were correctly classified, resulting in a true positive count of 8 and no misclassifications. For the "extrahls" class, 3 instances were correctly classified, but 1 instance was incorrectly classified as "extrastole." Moving on to the "extrastole" class, 6 instances were correctly classified, but 2 instances were misclassified as "murmur" and 1 as "normal." The "murmur" class had 19 instances correctly classified, but it also had 7 instances misclassified as "extrastole" and 0 instances misclassified as "normal." Lastly, the "normal" class, with the highest number of instances (70), had 61 correctly classified instances. However, it also had 8 instances misclassified as "extrastole" and 1 instance misclassified as "murmur." Overall, the confusion matrix reveals that the CNN model performed well for certain classes like "artifact" and "extrahls," but faced challenges in distinguishing between "murmur" and "normal" classes. The misclassifications in these classes might be attributed to the similarities in the acoustic patterns of the spectrograms, warranting further investigation and refinement of the model for enhanced performance.

Overall, the confusion matrix shows that the CNN model performed well for certain classes such as "artifact" and "extrahls," but faced challenges in distinguishing between "murmur" and "normal" classes. The misclassifications in these classes might be due to similarities in the spectrogram patterns, leading to confusion during classification. To further enhance the model's performance, additional fine-tuning and data augmentation techniques can be explored.

In summary, the CNN model exhibited promising performance in predicting heart disorders, with a final validation accuracy of approximately 77%. The learning curve demonstrated the model's learning progress and potential for generalization, while the confusion matrix highlighted its strengths in certain classes and challenges in distinguishing between others. To further enhance the model's accuracy and address misclassifications, additional fine-tuning, hyperparameter optimization, and data augmentation techniques can be explored. Overall, this study lays the foundation for leveraging CNNs in heart disorder classification and contributes valuable insights to the field of automated medical diagnostics.

5. Conclusions

In conclusion, this study successfully achieves its proposed objective, which is to explore the potential of CNNs in automating heart disorder prediction through the analysis of spectrogram data derived from audio recordings obtained using stethoscopes. The presented CNN architecture showcases significant promise, attaining an accuracy of approximately 77% in effectively classifying diverse cardiac pathologies. By harnessing the power of spectrograms, the CNN adeptly captures essential acoustic patterns and temporal characteristics inherent to each heart sound class, thus laying the foundation for the development of advanced automated diagnostic systems.

This research stands out by utilizing authentic audio data sourced from heart sound recordings obtained during real-world clinical examinations using stethoscopes. This direct connection to practical healthcare scenarios adds a layer of relevance and applicability to the study, given the enduring significance of stethoscopes in cardiac evaluations. The successful integration of CNNs in this specific medical context introduces fresh possibilities for enhancing the accuracy and efficiency of heart disorder detection, potentially revolutionizing the diagnosis and management of cardiac conditions in healthcare settings.

The diversity of the dataset, encompassing five distinct classes including normal, murmur, extra heart sound, artifact, and extrasystole, plays a pivotal role in the success of the CNN model. This model's ability to accurately categorize heart sounds underscores the significant potential of CNNs in cardiac diagnostics, offering accessible and precise tools for early detection of heart disorders. Notably, the CNN approach's computational efficiency is a crucial advantage, making it suitable for real-time applications and environments with limited computational resources.

As this study advances the application of CNNs in heart disorder classification, it is important to acknowledge certain limitations and chart a course for future enhancements. The research methodology, guided by the CRISP-DM framework, ensures methodical and thorough execution of the classification task. Despite this rigorous approach, limitations related to dataset size and class imbalances were identified. These constraints may influence the model's ability to generalize and perform optimally on unseen data. To address these limitations and further enhance the CNN model's accuracy, future research directions could encompass data augmentation techniques, transfer learning, and the incorporation of more comprehensive and diverse datasets.

In conclusion, this study not only accomplishes its initial objective but also contributes to the evolving landscape of AI-driven medical diagnostics. By harnessing the capabilities of CNNs in the realm of cardiac assessments, this innovative approach holds tremendous potential to transform the detection and diagnosis of heart conditions, ultimately leading to more timely interventions and enhanced patient care.

References

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Beauxis-Aussalet, E., & Hardman, L. (2014). Visualization of confusion matrix for non-expert users. *IEEE Conference on Visual Analytics Science and Technology (VAST)-Poster Proceedings*.

Bentley, P., Nordehn, G., Coimbra, M., Mannor, S., & Getz, R. (2012). *Classifying Heart Sounds Challenge*. http://www.peterjbentley.com/heartchallenge/

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Brites, I. S. G., Silva, L. M., Barbosa, J. L. V., Rigo, S. J., Correia, S. D., & Leithardt, V. R. Q. (2022). Machine Learning and IoT Applied to Cardiovascular Diseases Identification Through Heart Sounds: A Literature Review. *Lecture Notes in Networks and Systems*, *414 LNNS*, 356–388. https://doi.org/10.1007/978-3-030-96293-7_32

Chang, Y. C., Chang, K. H., & Wu, G. J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing Journal*, *73*, 914–920. https://doi.org/10.1016/j.asoc.2018.09.029

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/https://doi.org/10.1613/jair.953

Dietterich, T. G. (2000). Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*, 1–15. https://doi.org/https://doi.org/10.1007/3-540-45014-9_1

Disease, H., Using, P., Learning, M., Kumar, U. R., Pradesh, A., Pradesh, A., Pradesh, A., Pradesh, A., Raju, N. S., & Pradesh, A. (2023). *Journal of Engineering Sciences*. *14*(04).

Eckerson, W. W., Hanlon, N., & Barquin, R. (2000). The crisp-dm model: the new blueprint for data mining. *Journal of Data Warehousing*, *5*, 15.

Gamini, P. (2021). *Detection of Credit Card Fraudulent Transactions using Boosting Algorithms*. *February*.

Gomes, E. F., & Pereira, E. (2012). Classifying heart sounds using peak location for segmentation and feature construction. *Aistats*, *1*, 1–5.

Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*, *4*(5), 3–10. https://doi.org/10.1109/101.8118

Houssein, E. H., Ibrahim, I. E., Hassaballah, M., & Wazery, Y. M. (2022). Integration of Machine Learning and Optimization Techniques for Cardiac Health Recognition. In *Integrating meta-heuristics and machine learning for real-world optimization problems* (pp. 121–148). Springer.

Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, *1022*(1). https://doi.org/10.1088/1757-899X/1022/1/012072

Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, 1329–1333. https://doi.org/10.1109/ICICT50816.2021.9358597

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression*. Springer.

Learning, D. (2020). Deep learning. *High-Dimensional Fuzzy Clustering*.

Leng, S., Tan, R. S., Chai, K. T. C., Wang, C., Ghista, D., & Zhong, L. (2015). The electronic stethoscope. *Biomedical Engineering Online*, *14*(1), 1–37.

McCarthy, J. (2007). *What is artificial intelligence*.

Mohanty, S. N., Nalinipriya, G., Jena, O. P., & Sarkar, A. (2021). *Machine Learning for Healthcare Applications*. Wiley. https://books.google.com.br/books?id=uQsmEAAAQBAJ

Monico, J. F. G., Dal Poz, A. P., Galo, M., Dos Santos, M. C., & De Oliveira, L. C. (2009). Acurácia e precisão: revendo os conceitos de forma acurada. *Boletim de Ciências Geodésicas*, *15*(3), 469–483.

Nelson, D. M. Q. (2017). *Uso de redes neurais recorrentes para previsão de séries temporais financeiras*.

Patrizio, H. A., Phyu, R., Kim, B., & Brolis, N. V. (2023). Utilization of Simulation to Teach Cardiac Auscultation: A Systematic Review. *Cureus*, *15*(7).

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, *4*(2), 1883.

Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132–148.

Singh, A., & Kumar, R. (2020). Heart disease prediction using machine learning algorithms. *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 452–457.

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Suthaharan, S. (2016). Decision tree learning. In *Machine Learning Models and Algorithms for Big Data Classification* (pp. 237–269). Springer.

Tato, A., & Nkambou, R. (2018). *Improving adam optimizer*.

Taud, H., & Mas, J. F. (2018). Multilayer perceptron (MLP). In *Geomatic Approaches for Modeling Land Change Scenarios* (pp. 451–455). Springer. https://doi.org/10.1007/978-3-319-60801-3_27

Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of Machine Learning*, *15*, 713–714. https://doi.org/10.1007/978-1-4899-7502-7_581-1

WHO. (2021). *WORLD HEALTH ORGANIZATION*. Cardiovascular Diseases. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Wyse, L. (2017). Audio spectrogram representations for processing with convolutional neural networks. *ArXiv Preprint ArXiv:1706.09559*.