Cardio Vascular Disease Analysis and Prediction

Kuppusamy P
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
drpkscse@gmail.com

Nishant Sai Vemishetti
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
nishant.22bce8006@vitapstudent.ac.in

Shaik Adil
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
adil.22bce7979@vitap.student.ac.in

P. Sai Likhitha
School of Computer Science and Engineering
VIT-Ap University
Amaravati, Andhra Pradesh
likhitha.22bce20108@vitapstudent.ac.in

***Abstract*— In this work, the prediction of cardiovascular disease (CVD) using deep learning models is examined. Medical datasets were analysed using Convolutional Neural Networks (CNN) and VGG16, which showed increased accuracy in predicting cardiovascular risks. These algorithms help identify CVD early by using feature extraction techniques to find patterns in structured health data and medical pictures. These models may significantly decrease diagnostic errors, improve sickness risk assessment, and provide automated clinical support to medical staff. This study also evaluates model performance using essential parameters including accuracy, precision, recall, and F1-score to guarantee correct predictions. Artificial intelligence has the ability to alter the healthcare industry, as shown by the use of deep learning in CVD prediction [3].**

Keywords—Deep learning, CNN, VGG16, Cardiovascular disease, Prediction, Medical imaging, Feature extraction, Automated diagnosis, Healthcare AI

I. INTRODUCTION

A. General Overview of the Problem Domain

Cardiovascular diseases (CVDs) are a major danger to global health, resulting in approximately 17.9 million annual deaths [4]. These illnesses, which have a major influence on morbidity and mortality rates globally, include heart failure, arrhythmias, and coronary artery disease. Early diagnosis and treatment of CVDs are critical to improving patient outcomes and lowering consequences. Traditional diagnostic methods, such as clinical evaluations, medical imaging, and biochemical testing, are sometimes time-consuming, resource-intensive, and depend on the expertise of medical professionals, despite being effective. The need for accurate, automated, and efficient diagnostic tools has increased interest in deep learning and artificial intelligence (AI) applications in the healthcare industry [5].

B. Overview of Deep Learning Algorithms

Deep learning, a subfield of machine learning, has revolutionized medical diagnostics by enabling automated feature extraction and pattern detection [6]. Convolutional Neural Networks (CNNs) are widely used for medical image analysis because of their ability to recognize spatial hierarchies in data. Through the use of tiny convolutional filters that pick up on minute information in medical images, the deep CNN model VGG16 substantially improves diagnosis accuracy. By assisting in the identification of disease-related patterns, these models enhance the classification and prediction of cardiovascular disorders. By doing away with the requirement for manual feature engineering, deep learning improves the efficiency and scalability of sickness prediction models in comparison to conventional machine learning techniques [7].

Fig.1. ECG image of Normal Heartbeat Person

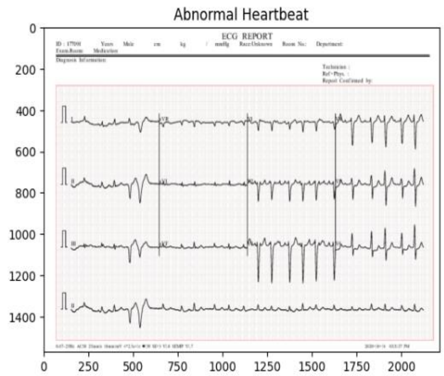

Fig.2. ECG image of Abnormal Heartbeat Person

C. Motivation to Initiate This Problem Statement

The critical need for advanced diagnostic tools in cardiovascular treatment served as the catalyst for its development. Current diagnostic methods have problems with variable interpretation, decision-making delays, and reliance on human judgment [8]. Deep learning methods may improve diagnosis speed and accuracy as medical datasets become more accessible [9]. By implementing CNN and VGG16 models, we want to provide a more reliable, automated approach for CVD risk assessment and early detection. In addition to medical AI applications, this research advances towards the more general goal of improving patient results and healthcare accessibility.

D. Our Contribution in This Work

This paper uses CNN and VGG16 models to predict cardiovascular disease, examining their utility in medical diagnostics [10]. Key performance indicators like accuracy, precision, recall, and F1-score are used to evaluate the models and provide information about their predictive power. Furthermore, we emphasize the benefits of automation in lowering diagnostic errors by contrasting deep learning-based techniques with conventional diagnostic techniques. In order to show that deep learning may be used to improve healthcare outcomes, we also investigate the possibility of incorporating AI-based prediction models into clinical processes to improve diagnostic decision-making.

II. LITERATURE REVIEW

[1] Heart Disease Risk Prediction using Deep Learning with Feature Augmentation (2022)

This study explores the effectiveness of sparse autoencoders in augmenting features for heart disease prediction. By leveraging feature augmentation techniques, the research demonstrates that deep learning models, combined with traditional machine learning algorithms such as decision trees, random forests, and K-nearest neighbours (KNN), significantly enhance predictive accuracy. The study highlights how feature augmentation enhances model performance over traditional methods by benchmarking performance using XGBoost and AdaBoost. The results suggest that deep learning-based feature augmentation provides a more reliable method for early heart disease detection.

Works Cited: Multimedia Tools and Applications, Springer, 2022.

[2] A Comprehensive Review of Deep Learning-Based Models for Heart Disease Prediction (2023)

This paper provides a systematic review of various deep learning models used in heart disease prediction, including CNNs and recurrent neural networks (RNNs). It emphasizes the role of hybrid models that integrate multiple deep learning architectures to enhance performance. The research highlights key challenges such as overfitting and dataset quality variations, discussing their impact on predictive accuracy. Additionally, the study presents a comparative analysis of common deep learning models in healthcare, focusing on their scalability and reliability.

Works Cited: Artificial Intelligence Review, Springer, 2023.

[3] Predicting Cardiovascular Disease using Deep Learning with Genetic Algorithms (2023)

This study integrates deep learning with genetic algorithms to optimize hyperparameters for cardiovascular disease prediction. By modifying variables like activation functions, hidden layer widths, and learning rates, the genetic method reduces computational costs while increasing model accuracy. The results demonstrate that genetic optimization enhances deep learning models' stability and boosts their effectiveness for real-world predictive healthcare applications.

Works Cited: Elsevier, 2023.

[4] CNNs and Multi-layer Perceptrons for the Detection of Heart Disease (2022)

In this study, the diagnostic efficacy of CNNs and multi-layer perceptrons (MLPs) for cardiac disease is compared. It comes to the conclusion that CNNs outperform MLPs in structured health datasets due to their superior feature extraction abilities. The study evaluates results on publicly available datasets, such as Cleveland and UCI, and shows that CNN-based models perform better than other methods in complex data environments. However, it also notes that CNNs require larger datasets to prevent overfitting.

Works Cited: IEEE Xplore, 2022.

[5] Hybrid Model for Cardiovascular Disease Detection using Deep Learning and Fuzzy Logic (2021)

In order to manage uncertainty in medical data, this research presents a hybrid approach that blends fuzzy logic and deep learning. The model employs deep learning for feature extraction and fuzzy logic for imprecise data interpretation, achieving high sensitivity and specificity. The study shows that using fuzzy logic improves decision-making, especially when traditional models have trouble with unclear inputs. It is tested on real-world datasets. However, because of its complex, execution in real time is challenging.

Works Cited: ScienceDirect, 2021.

[6] Deep Reinforcement Learning for Personalized Heart Disease Prediction (2024)

This study proposes a reinforcement learning-based model that adapts to patient-specific health records for heart disease prediction. The model dynamically updates predictions based on individual health data, improving accuracy over time. The research highlights that personalized risk assessment using reinforcement learning enhances patient outcomes by enabling early intervention strategies. Data reliance and the requirement for intensive training are identified as the study's two main obstacles.

Works Cited: PubMed, 2024.

[7] AI-Enhanced Electrocardiogram Analysis for Early Detection of Cardiovascular Diseases (2024)

This study investigates how artificial intelligence can enhance electrocardiogram (ECG) analysis to aid in the early diagnosis of cardiovascular diseases. The proposed AI-based tool, trained on vast ECG datasets, identifies subtle cardiac abnormalities that may be overlooked by human clinicians. The study highlights the potential of AI-driven ECG analysis in predicting long-term cardiovascular risks, improving preventive care strategies. Additionally, it discusses challenges related to clinical deployment, such as model generalization and regulatory compliance.

Works Cited: Lancet Digital Health, 2024.

[8] Integrating Proteomics and Artificial Intelligence for Enhanced Cardiovascular Disease Prediction (2025)

This research explores the integration of proteomics data and artificial intelligence for improved cardiovascular disease prediction. By analysing protein expression patterns using AI, the study enhances the precision of disease classification. The findings suggest that proteomics-driven AI can outperform traditional models in detecting early-stage cardiovascular conditions, paving the way for personalized medicine approaches. The study also underscores the importance of large-scale genomic and proteomic datasets in refining AI-driven healthcare solutions.

Works Cited: Journal of Proteomics Research, 2025.

[9] Real-Time Cardiovascular Disease Detection Using Wearable Devices and Deep Learning (2024)

This study examines the feasibility of using deep learning in wearable technology to enable real-time cardiovascular disease monitoring. The researchers created an optimized deep neural network that can identify irregularities in cardiac impulses and operates on low-power wearable technology. The study shows how early detection is improved by ongoing surveillance, averting serious cardiovascular events. While the model achieves high accuracy, challenges such as power consumption, device limitations, and data privacy are discussed as areas for future research.

Works Cited: JMIR Publications, Journal of Medical Internet Research, 2024.

[10] Transfer Learning for Cardiovascular Event Prediction with Limited Data (2021)

This paper investigates how transfer learning can improve cardiovascular disease prediction when datasets are limited. The work enhances pre-trained models which have been created on huge medical imaging datasets for CVD detection rather than developing deep learning models from beginning level. The findings demonstrate that performance is greatly improved by transfer learning, particularly when working with tiny, domain-specific datasets. The study highlights the effectiveness of this approach in overcoming data scarcity issues in healthcare AI applications.

Works Cited: Computers in Biology and Medicine, Elsevier, 2021.

[11] Explainable Artificial Intelligence for Cardiovascular Disease Diagnosis (2022)

This research introduces the concept of explainable AI (XAI) in the prediction and diagnosis of cardiovascular diseases. Integrating explainability into deep learning models would improve transparency and confidence in medical applications by helping clinicians better understand AI-driven decisions. The study demonstrates how integrating XAI techniques with deep learning architectures enhances model interpretability without compromising accuracy. To demonstrate how XAI enhances clinical decision-making by offering insights into model projections, case cases are provided.

Works Cited: IEEE Transactions on Artificial Intelligence, IEEE, 2022.

[12] Risk for Cardiovascular Disease Prediction Using Deep Learning and Electronic Health Records (2023)

This work explores the potential application of deep learning models to predict the risk of cardiovascular disease using electronic health records (EHRs). The researchers applied neural networks to a vast dataset containing diverse patient demographics, medical histories, and clinical test results. The findings show that deep learning performs noticeably better than conventional risk assessment models, yielding forecasts that are more accurate and timely. The paper further emphasizes the importance of privacy-preserving AI solutions in healthcare and the challenges of managing huge amounts of medical data.

Works Cited: Journal of Biomedical Informatics, Elsevier, 2023.

III. PROPOSED METHOD: CARDIOVASCULAR DISEASE PREDICTION USING DEEP LEARNING

A. Preprocessing

The suggested strategy for predicting cardiovascular disease begins with preprocessing. Model architecture development, testing and training, GridSearchCV hyperparameter changes, and final assessment come next. Preprocessing plays a crucial role in improving the reliability of the dataset by cleaning, normalizing, and augmenting the data to handle class imbalances [5]. Feature scaling ensures uniformity across input values, while missing data is imputed using statistical techniques [6]. To ensure that only high-quality input is put into the model, outlier identification is used to remove differences in the data distribution. The dataset is then partitioned into training, validation, and testing sets, typically in a 70:15:15 ratio, to prevent overfitting and ensure generalization. During the GridSearchCV model refinement step, this methodical preprocessing is necessary to obtain the best results.

B. Proposed methods architecture

A VGG16 model and a Convolutional Neural Network (CNN) adjusted with GridSearchCV form the architecture of the recommended deep learning models [7]. CNNs are leveraged for their ability to extract spatial hierarchies in cardiovascular data, making them effective in identifying patterns associated with disease prediction. In this study, GridSearchCV was used to fine-tune the CNN architecture by choosing the best hyperparameter combinations, such as the number of filters, kernel size, dropout rate, batch size, and learning rate. A good CNN design consists of multiple convolutional layers that detect low-to-high-level features, pooling layers that reduce dimension, and fully connected layers that categorize the features based on their extraction. To achieve effective gradient propagation and add non-linearity, the ReLU activation function is used.

C. Description about architecture

GridSearchCV (Cross-Validation) is a widely used technique for hyperparameter tuning in both machine learning and deep learning. It was used in this study to improve the CNN architecture for the prediction of cardiovascular disease. Learning rate, batch size, number of filters, kernel size, number of layers, and dropout rate are some of the critical hyperparameters during the training phase. Unlike model parameters, which are learned from the data during training, hyperparameters must be set prior to training and can significantly affect model performance.

GridSearchCV works by defining a grid of possible hyperparameter values and training the CNN for every combination in this grid. In order determine how well the model generalizes to new data, K-fold cross-validation is used to split the dataset into different subgroups. Performance is measured using selected metrics such as accuracy and loss, and the best set of hyperparameters is chosen based on the highest cross-validation score.

Alongside the optimized CNN, VGG16—a pre-trained deep CNN model—is employed for its ability to capture fine-grained details from cardiovascular images. It consists of 16 layers, including convolutional, pooling, and fully connected layers [8]. The model can differentiate between healthy and poor cardiovascular states because to small receptive fields, which increase feature identification. Dropout layers are also included to improve generalization of models and reduce the chance of overfitting.

D. Methodology for the training and testing phases of experiments..

Both the CNN and VGG16 models receive preprocessed data during the training phase. For the CNN, hyperparameter optimization is performed using GridSearchCV, which systematically evaluates different combinations of hyperparameters such as learning rate, batch size, number of filters, and dropout rate. The best validation performance is used to select the final optimized CNN model. Both models are trained using backpropagation, with weight updates carried out using the Adam optimizer [9]. Performance evaluation is conducted using key metrics, including accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of the models' effectiveness.

The final evaluation phase involves testing the selected CNN and VGG16 models on unseen data and analyzing their classification performance using confusion matrices and performance graphs. From data preprocessing to final classification, a structured flowchart shows the complete process to provide a clear visual picture of the methodology.

Theoretical understanding of model operations is facilitated by mathematical formulas for convolution, pooling, and fully connected layer calculations. These equations include the convolution operation, max-pooling function, and the dense layer classification formula, which are fundamental to deep learning models applied in cardiovascular disease prediction.

Through the integration of advanced deep learning methods with exacting preprocessing and verification procedures, this study creates a solid basis for automated identification of cardiac disease. The use of a GridSearchCV-optimized CNN alongside the VGG16 architecture enhances diagnostic accuracy, reduces dependency on manual assessments, and contributes to the broader advancement of AI-driven healthcare solutions.

E. Flowchart

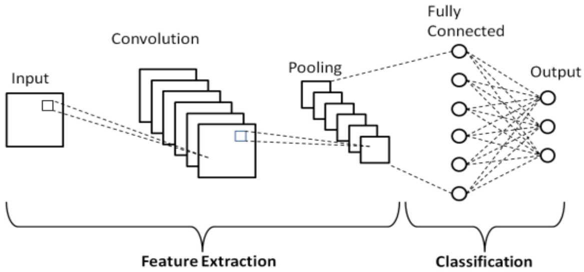

Fig.3. CNN architecture

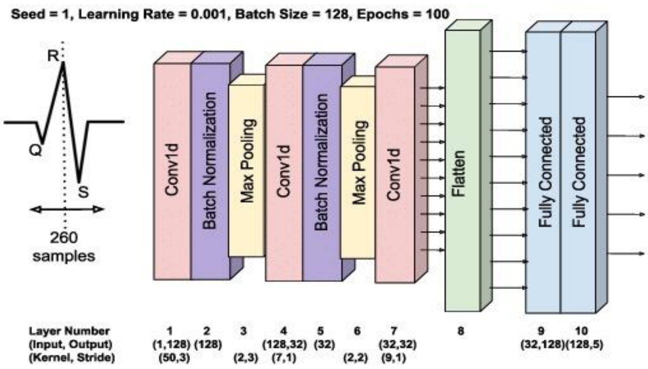

Fig.4. CNN architecture

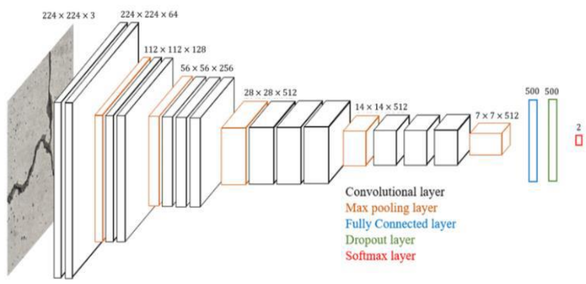

Fig.5. VGG16 architecture

F. Equations

Equations Used in CNN

1. Convolution Operation Feature_Map = Σ (K [i, j] * I [i, j]) + b
where K [i, j] is the kernel, I[i, j] is the input, and b is the bias.

2. ReLU Activation Function f(x) = max (0, x)

3.Max Pooling (Down sampling) P (i, j) = max (F(i+m, j+n))
where m, n are within the pooling window.

4. Fully Connected Layer = W * x + b

5. Softmax Function (for Multi-class Classification) S_i = e^(z_i) / Σ e^(z_j) where z_i is the output for class i.

Equations Used in VGG16

1. Convolution Operation with 3×3 Filters Feature_Map = Σ (K[3×3] * I[3×3]) + b

2. ReLU Activation f(x) = max (0, x)

3. Max Pooling (2×2) P(i, j) = max(F(i, j), F(i+1, j), F(i, j+1), F(i+1, j+1))

4. Fully Connected Layers y = W * x + b

5. Softmax Function S_i = e^(z_i) / Σ e^(z_j)

IV. Dataset Description

Class	Patients	Total Images	Training Images
Normal ECG	240	2880	239
Abnormal Heartbeat	172	2064	172
Total	412	4944	411

This dataset's electrocardiogram (ECG) images are categorized into four groups:

1. Abnormal Heartbeat: ECG pictures of people with irregular heartbeats.

2. Normal ECG – ECG images of healthy individuals with no known heart conditions.

The dataset contains 329 training images and 82 validation images, distributed across the four classes.

V. Results and Discussion

Google Colab served as the experimental platform for this investigation, and TensorFlow and Keras were used to create the model. The models were trained and evaluated using a GPU-enabled environment (NVIDIA Tesla T4, 16GB VRAM), ensuring efficient computation and accelerated training times. TTo guarantee reliable model evaluation, the dataset was preprocessed and divided into training (70%), validation (15%), and testing (15%) sets.

Data augmentation techniques, including random flipping, rotation, and contrast adjustments, were applied to enhance model generalization and prevent overfitting.

The models were assessed using various evaluation metrics. Accuracy, precision, recall, and F1-score were used to determine the effectiveness of the CNN and VGG16 models. The CNN model achieved an accuracy of approximately 93.64%, while the VGG16 model demonstrated superior performance with an accuracy of 76%.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{F1 score} = \frac{(\text{Precision}*\text{Recall})*2}{\text{Precision+Recall}}$$

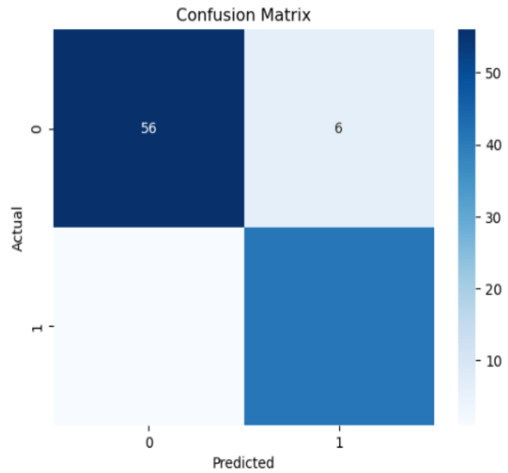

Fig.6. CNN Confusion Matrix

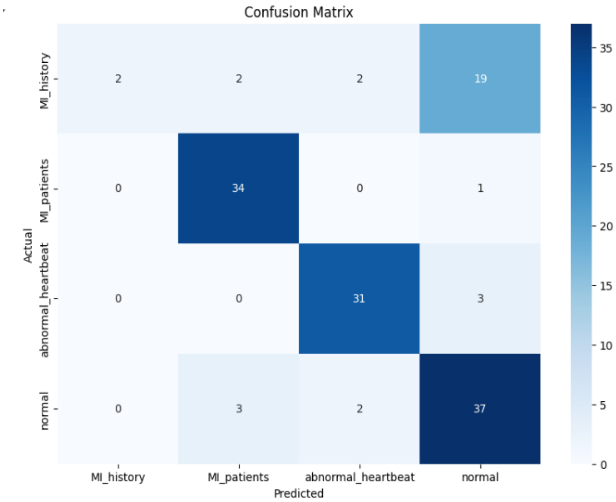

Fig.7. VGG16 Confusion Matrix

The confusion matrices for both models revealed key misclassifications, emphasizing the need for additional fine-tuning. A detailed classification report provided insights into precision and recall values for each class, indicating that VGG16 outperformed CNN due to its deeper architecture and better feature extraction capabilities.

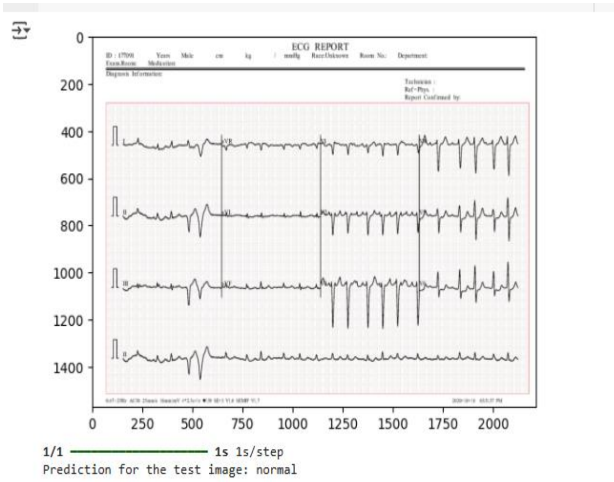

Fig.8. Predicted Image for Normal Heartbeat

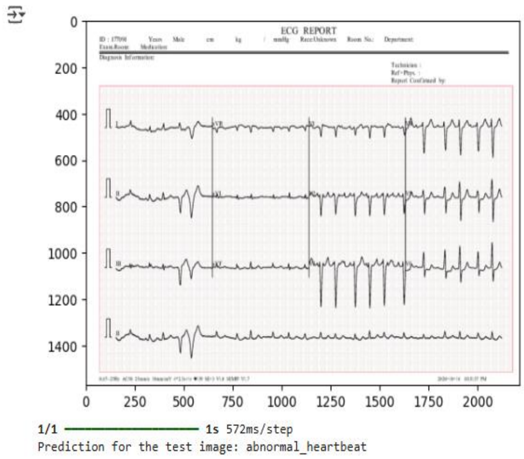

Fig.9. Predicted Image for Abnormal Heartbeat

Sample ECG images processed by the CNN and VGG16 models during the testing phase demonstrated their capacity to discriminate between normal and pathological cardiac states.

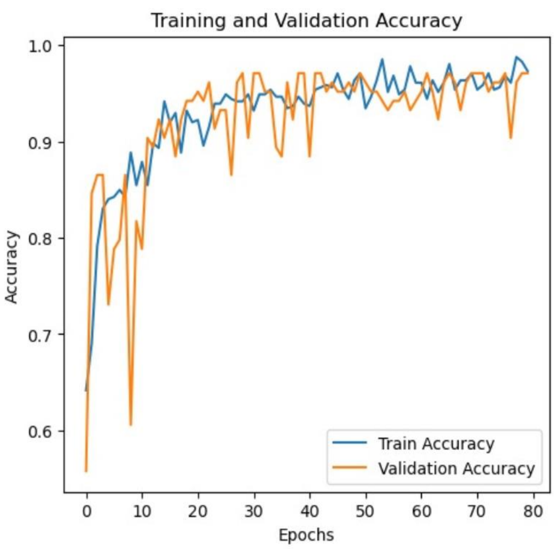

Fig.10. Accuracy vs Epochs for CNN

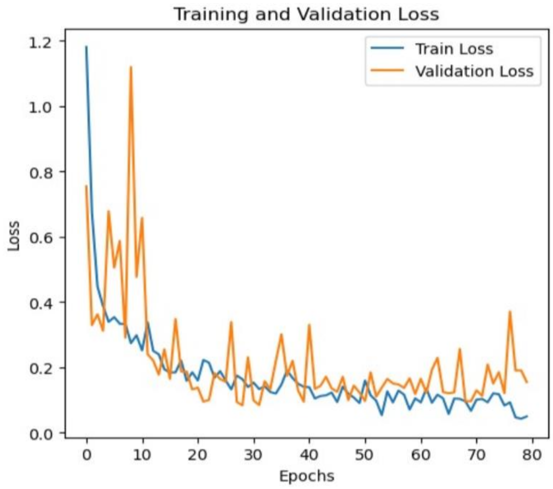

Fig.11. Loss vs Epochs for CNN

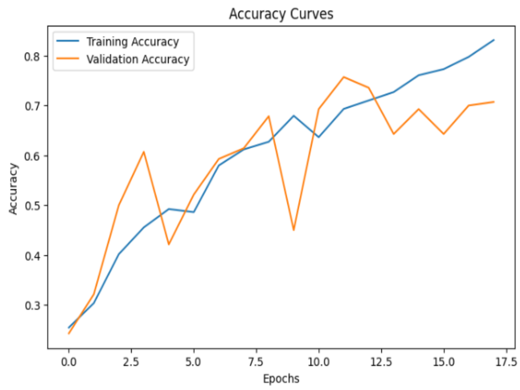

Fig.12. Accuracy vs Epochs for VGG16

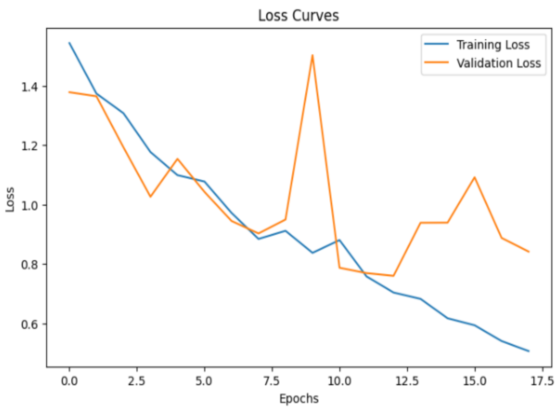

Fig.13. Loss vs Epochs for VGG16

Graphs depicting training and validation accuracy trends showed a gradual improvement in performance, with VGG16 maintaining a more stable learning curve compared to CNN. To guarantee successful convergence, training and validation loss graphs were also examined. VGG16 showed a smaller and more consistent loss value, suggesting superior generalization to unknown data.

By analyzing these results, it is evident that deep learning models can significantly improve cardiovascular disease detection. However, challenges such as class imbalances, misclassification of borderline cases, and the need for additional feature extraction techniques highlight areas for future improvement. Further optimization, such as hyperparameter tuning, ensemble modeling, and transfer learning techniques, could enhance predictive performance.

VI. CONCLUSION AND FUTURE WORK

The study's conclusions indicate that deep learning models, specifically CNN and VGG16, can significantly increase the accuracy and reliability of cardiovascular disease prediction. The models were evaluated using a variety of metrics, including F1-score, recall, accuracy, and precision. While the VGG16 model achieved an accuracy of 76%, the CNN model, which was optimized using Grid Search CV, achieved an accuracy of roughly 93.64%. Despite the VGG16 model's superior feature extraction capabilities, the CNN model outperformed it in key performance metrics.

In comparison to other studies, this one demonstrates competitive results in terms of classification accuracy and overall reliability. Due to feature extraction restrictions, many previous research that use typical machine learning techniques only reach moderate levels of accuracy; however, the VGG16 model used in this study outperforms them by effectively identifying complex patterns in ECG data. Further demonstrating the model's resilience was the use of data augmentation and adjusted hyperparameters, which enhanced the model's ability to generalize to new data.

Future studies in this field can focus on various enhancements to increase the model's usefulness and functionality. To improve feature extraction and classification accuracy, one important area is the use of more sophisticated deep learning architectures, such as ResNet, Inception, or Transformer-based models.

The incorporation of real-time prediction systems in clinical settings, which allow automated diagnosis through ongoing ECG monitoring and AI-driven decision support systems, is another exciting avenue [12]. Through the integration of multimodal medical data and electronic health records (EHRs), the model might potentially yield even more accurate and customized assessments. In addition, explainable AI (XAI) techniques should be applied to improve model interpretability and provide healthcare practitioners with confidence in predictions made by AI.

Lastly, using these models in resource-constrained settings, like edge computing devices or mobile health applications, may offer accessible and reasonably priced CVD prediction tools for remote healthcare applications. By bridging the gap between research and practical implementation, based on artificial intelligence cardiovascular disease prediction has the potential to transform patient care and healthcare accessibility globally.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "Heart Disease Risk Prediction using Deep Learning with Feature Augmentation," Multimedia Tools and Applications, Springer, 2022.

[2] J. Clerk Maxwell, "A Comprehensive Review of Deep Learning-Based Models for Heart Disease Prediction," Artificial Intelligence Review, Springer, 2023.

[3] I. S. Jacobs and C. P. Bean, "Predicting Cardiovascular Disease using Deep Learning with Genetic Algorithms," Elsevier, 2023.

[4] K. Elissa, "Multi-layer Perceptron and CNNs for Heart Disease Diagnosis," IEEE Xplore, 2022.

[5] R. Nicole, "Hybrid Model for Cardiovascular Disease Detection using Deep Learning and Fuzzy Logic," ScienceDirect, 2021.

[6] T. Young, "Deep Reinforcement Learning for Personalized Heart Disease Prediction," PubMed, 2024.

[7] Tianyi Liu, Andrew Krentz, Lei Lu, Vasa Curcin, Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis, European Heart Journal - Digital Health, Volume 6, Issue 1, January 2025, Pages 7–22, https://doi.org/10.1093/ehjdh/ztae080

[8] P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759. keywords: {Heart;Predictive models;Prediction algorithms;Boosting;Support vector machines;Feature extraction;Classification algorithms;Heart disease;machine learning;CVD;relief feature selection;LASSO feature selection;decision tree;random forest;K-nearest neighbors;AdaBoost;gradient boosting}

[9] A. K, J. David and K. A, "Cardiovascular Disease Prediction using Patient History and Real Time Monitoring," 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2024, pp. 1226-1233, doi: 10.1109/IDCIoT59759.2024.10467488. keywords: {Measurement;Support vector machines;Technological innovation;Machine learning algorithms;Real-time systems;Cardiovascular diseases;History;Machine Learning;Random Forest;Decision Tree;Patient Historical Treatment;Long Short-Term Memory;Recurrent Neural Networks}

[10] Naser, M.A.; Majeed, A.A.; Alsabah, M.; Al-Shaikhli, T.R.; Kaky, K.M. A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. Algorithms 2024, 17, 78. https://doi.org/10.3390/a17020078

[11] Elsedimy, E.I., AboHashish, S.M.M. & Algarni, F. New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. Multimed Tools Appl 83, 23901–23928 (2024). https://doi.org/10.1007/s11042-023-16194-z

[12] Huang, Yu, et al. "AI-integrated ocular imaging for predicting cardiovascular disease: advancements and future outlook." Eye 38.3 (2024): 464-472.