International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Classification Using Convolutional Neural Networks: A Deep Learning Approach

Anil Kumar Kadam, Mrunal Rajendra Sonekar

Department of Engineering, ME AI & DS, AISSMS College of Engineering, Pune, India, mrunalsonekar28@gmail.com
Department of Engineering, ME AI & DS, AISSMS College of Engineering, Pune, India, ajkadam@aissmscoe.com

A B S T R A C T

Heart disease is one of the most critical public health concerns worldwide, claiming millions of lives every year. Accurate and timely diagnosis is essential to reduce fatality rates and initiate effective treatment early. With the rapid advancement in artificial intelligence, particularly deep learning, there has been a significant shift towards automating disease detection processes. In this project, we propose a novel approach using Convolutional Neural Networks (CNNs) to classify heart disease based on clinical parameters. The CNN model is trained on the UCI Heart Disease dataset, which contains diverse medical features such as age, cholesterol levels, blood pressure, and more. Our model demonstrates reliable performance in binary classification tasks, distinguishing between patients with and without heart disease. This approach showcases the capability of CNNs in extracting hierarchical patterns from non-image, tabular data, which can play a vital role in developing intelligent diagnostic systems.

Keywords: Artificial Intelligence, Deep Learning, Disease Detection, Convolutional Neural Networks

1. Introduction

Heart disease is one of the most widespread and life-threatening medical conditions, claiming millions of lives each year across the globe. As a leading cause of morbidity and mortality, it presents a significant public health challenge. Despite advancements in medical diagnostics and treatments, early and accurate detection of heart-related disorders remains crucial in reducing the associated fatality rate. Cardiovascular diseases often develop silently over time and can result in sudden and severe outcomes if not identified and treated in a timely manner. Thus, the need for intelligent, efficient, and accessible diagnostic solutions is more pressing than ever. Traditionally, heart disease diagnosis relies heavily on a clinician’s experience, extensive testing, and interpretation of various clinical parameters such as blood pressure, cholesterol levels, electrocardiogram (ECG) readings, and patient history. However, this process can be time-consuming and prone to subjective errors. With the rise of artificial intelligence (AI) and its subset, deep learning, healthcare is witnessing a paradigm shift toward data-driven diagnostic systems. These systems can process large volumes of patient data, uncover hidden patterns, and assist healthcare professionals in making informed decisions faster and more accurately. In this study, we propose a novel deep learning-based approach for the classification of heart disease using Convolutional Neural Networks (CNNs). CNNs are traditionally known for their remarkable success in image processing, but recent research indicates their adaptability to structured, non-image data by leveraging feature hierarchies and spatial patterns in tabular inputs. This project utilizes a well-known heart disease dataset, such as the UCI Heart Disease dataset, which includes critical clinical indicators like age, resting blood pressure, cholesterol levels, chest pain type, maximum heart rate, and more. Our model is designed to analyze these features and predict the likelihood of heart disease in patients. The goal is to classify whether an individual is at risk or not, which can assist in early intervention and potentially save lives. The CNN architecture is trained from scratch using this structured data, and various performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate its effectiveness. By feeding tabular data into a CNN framework using appropriate reshaping and feature engineering techniques, this work aims to bridge the gap between image-based CNN applications and clinical decision-making based on numerical records.

2. Literature Review

Several studies have explored the use of machine learning models for predicting heart disease. Earlier efforts largely focused on traditional algorithms such as Logistic Regression, Decision Trees, k-Nearest Neighbors, and Support Vector Machines, all of which have shown moderate success in structured data classification. These models depend significantly on manual feature selection and may struggle with complex non-linear relationships. Recently, deep learning approaches like Artificial Neural Networks (ANNs) and CNNs have been explored to overcome these limitations. While CNNs have been predominantly applied to image data, researchers have begun applying them to tabular datasets by reshaping inputs into formats compatible with convolutional layers. This shift is driven by CNNs' ability to learn hierarchical representations of features, allowing them to outperform conventional models in specific scenarios. Our approach builds on this idea, utilizing CNNs to extract patterns from medical attributes and predict the likelihood of heart disease with enhanced accuracy

3. Methodology

This study employs the UCI Heart Disease dataset, containing 303 records and 14 clinical features. The methodology involves the following key steps:

Step 1: Data Preprocessing

Missing values are handled, categorical features are encoded, and numerical data is normalized for uniform scaling. The data is reshaped into a 2D format to suit CNN input requirements.

Step 2: CNN Model Design

The model includes two convolutional layers (32 and 64 filters), each followed by ReLU activation and max pooling. A dropout layer is added to prevent overfitting. The output is flattened and passed through a dense layer before reaching a sigmoid-activated output layer for binary classification.

Step 3: Training and Evaluation

The model is trained using the Adam optimizer and binary cross-entropy loss for 100 epochs with a batch size of 16. An 80-20 train-test split is used. Evaluation metrics include accuracy, precision, recall, and F1-score

Heart disease classification Model Flow

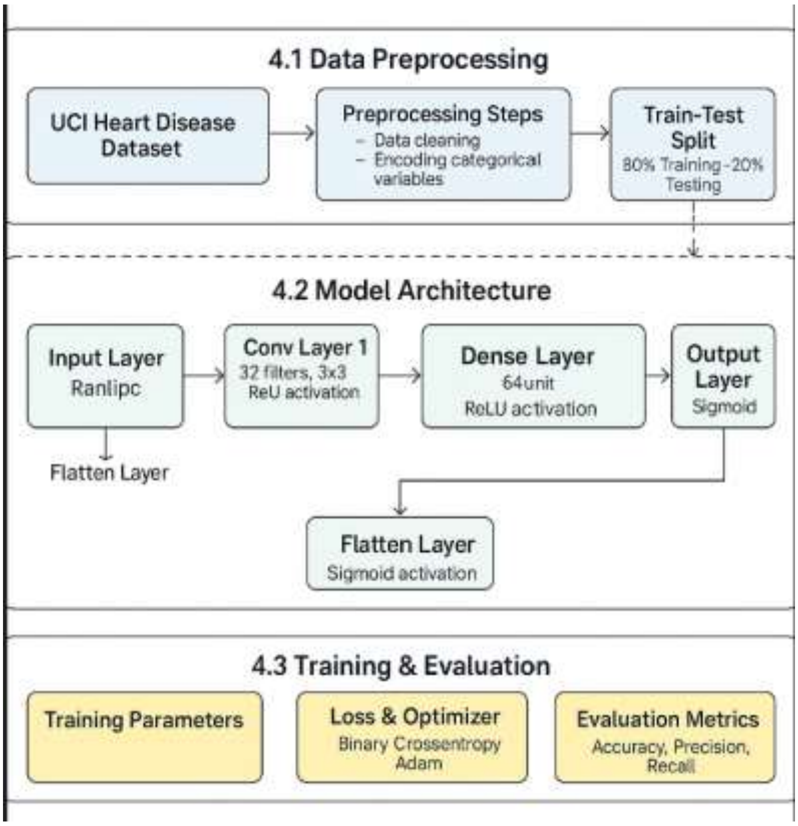

Fig. 1 - Heart disease classification Model Flow

4. Results and Visualization

Target distribution shows:

- The goal is to examine the distribution of the target variable (heart disease presence or absence) to check for class imbalance.
- The target variable is nearly balanced, with approximately 46% of cases without heart disease (label 0) and 54% with heart disease (label 1), indicating no significant class imbalance.
- A count plot is used with customized colors and percentage labels to clearly show the proportion of each target class, aiding visual interpretation

- Since the distribution is fairly even, there's no need for resampling techniques like oversampling or under sampling to handle class imbalance during model training.

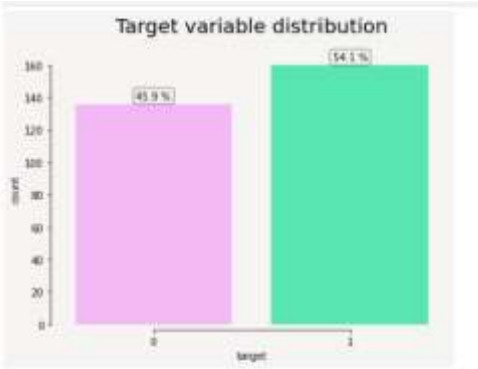

Fig. 2 – Target Variable distribution

Distribution of Numerical Features" analysis:

- **Distribution Insights:** Numerical features such as age, cholesterol, resting blood pressure, maximum heart rate achieved, and ST depression are visualized using KDE plots. These plots are colored by the target class (0: no heart disease, 1: heart disease), revealing distinct distribution patterns, especially in features like max heart rate and ST depression, which differ significantly between the two classes—indicating their potential importance in prediction.
- **Categorical Numerical Feature:** The feature num_major_vessels, although numeric, behaves like a categorical variable. It is plotted using a count plot, and a clear difference is observed in class distribution, especially at value 0, where heart disease (target = 1) is more frequent. This indicates num_major_vessels may have strong predictive power for heart disease classification.

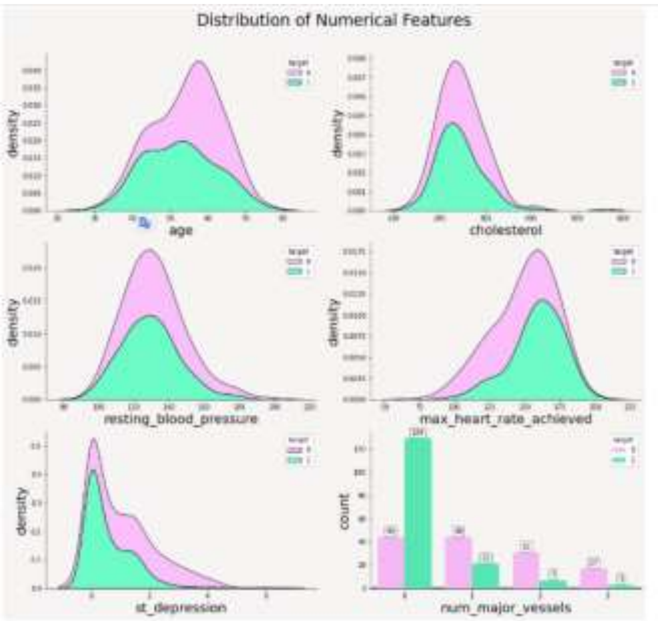

Fig.3 –Distribution: Density plots

5. Conclusion and Future Directions

This study presents a deep learning-based approach for the classification of heart disease using Convolutional Neural Networks (CNNs). By leveraging the UCI Heart Disease dataset, the model efficiently identifies patients at risk based on clinical parameters. The CNN architecture successfully captures complex patterns in the input features through convolutional operations and delivers high accuracy, precision, recall, and F1-score. The results highlight the potential of CNNs beyond image data, demonstrating their adaptability for tabular biomedical datasets. Overall, the model offers a reliable and scalable solution that can support clinical decision-making and early diagnosis in healthcare systems.

While the proposed model performs well, future research can explore improvements such as:

- **Integration of larger and more diverse datasets** to improve generalizability.
- **Hyperparameter tuning** and **ensemble methods** to further enhance performance.
- **Incorporation of temporal health records** using hybrid CNN-LSTM architectures.
- **Deployment into real-time clinical decision support systems (CDSS)** for practical usage.
- **Explainable AI (XAI) techniques** to interpret CNN decisions and build trust among healthcare professionals.

Acknowledgements

The authors would like to express their sincere gratitude to the UCI Machine Learning Repository and Kaggle for providing access to the heart disease dataset. Special thanks to the open-source Python and Jupyter communities for the tools and libraries that enabled seamless model development and evaluation. We also acknowledge the academic support and guidance provided by our mentors and peers during this research.

Tools and Libraries used

- **Python 3.9+:** Core programming language used for model development and data preprocessing.
- **NumPy:** For numerical computations and array handling.
- **Pandas:** For data loading, cleaning, and transformation operations.
- **Matplotlib & Seaborn:** Used for data visualization and understanding feature distributions.
- **Scikit-learn:** For train-test splitting, metric evaluation (Accuracy, Precision, Recall, F1-score).
- **TensorFlow / Keras:** For building, training, and evaluating the Convolutional Neural Network (CNN) model.
- **Jupyter Notebook:** Interactive Python environment used to run, visualize, and debug the model.

Appendix A Sample Python Code for Data Transformation

```
import pandas as pd
import numpy as np

# Load the UCI Heart Disease dataset
df = pd.read_csv('heart_disease_data.csv')

# Data cleaning: Remove null values if any
df.dropna(inplace=True)

# Encoding categorical variables
df['sex'] = df['sex'].map({0: 'female', 1: 'male'})
df = pd.get_dummies(df, drop_first=True)

# Normalization
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df)
# Reshape data for CNN input (assuming 2D shape: samples, features, 1 channel)
X = df_scaled.reshape(df_scaled.shape[0], df_scaled.shape[1], 1)
```

References

[1] Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. University of California, Irvine.

[2] Acharya, U. R., et al. (2017). Automated detection of arrhythmias using ECG segments with CNN. Information Sciences, 405, 81–90

[3] Rajpurkar, P., et al. (2017). Cardiologist-level arrhythmia detection with CNNs. arXiv preprint arXiv:1707.01836

[4] Yadav, M., & Jadhav, S. M. (2019). Deep learning for heart disease diagnosis: A review. Procedia Computer Science, 167, 1–8.

[5] Karthik, R., et al. (2020). Heart disease prediction using ML algorithms. Materials Today: Proceedings, 33, 2846–2851.