Fuzzy Systems and Soft Computing
ISSN : 1819-4362

PREDICTIVE ANALYSIS OF HEART DISEASE: EVALUATING CNN AND MLP CLASSIFIERS ON RESAMPLED DATA

Dr.K.Kavitha Associate Professor, Department of Computational Sciences, Brainware University, Kolkata.
Dr.P.Sumathi Assistant Professor, Department of Computer Science, Dr.G.R.Damodaran College of Science, Coimbatore.

Abstract
Heart disease remains a leading cause of morbidity and mortality worldwide, necessitating effective and timely diagnosis. In this research, we explore the potential of artificial intelligence (AI) in predicting heart disease using deep learning (DL) techniques. Specifically, we investigate the performance of a Multilayer Perceptron (MLP) classifier and a Convolutional Neural Network (CNN) on the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. This dataset comprises 253,680 instances with 21 attributes with 1 binary target variable: HeartDiseaseorAttack and 21 feature variables that are either binary or ordinal. Standard preprocessing techniques, such as normalization and encoding, were employed to prepare the dataset. Both models were trained and tested, with resampling techniques applied to address class imbalance. Our findings demonstrate that CNN outperforms the MLP classifier, showcasing higher accuracy, precision, recall, and F1 score. This study underscores the superior capability of deep learning models in the accurate prediction of heart disease, paving the way for their application in medical diagnostics.

Keywords: Convolutional Neural Network, Heart Disease Prediction, Multilayer Perceptron, Deep Learning, Healthcare, Resampling.

Introduction

Heart disease is a critical global health issue, accounting for significant mortality and posing substantial economic burdens (Benjamin et al., 2019). Early and accurate diagnosis is essential for effective management and treatment, which can considerably improve patient outcomes (World Health Organization, 2020). Traditional diagnostic methods, although effective, are often resource-intensive and may not always be accessible or timely. Consequently, there is a growing interest in leveraging AI technologies, including machine learning (ML) and deep learning (DL), to enhance diagnostic accuracy and efficiency.

ML and DL techniques have shown promise in various domains, including healthcare, where they can analyze vast amounts of data to identify patterns and make predictions (Esteva et al., 2019). ML algorithms, such as decision trees, support vector machines, and neural networks, have been widely used for predictive modeling (Shin et al., 2020). However, DL models, particularly Convolutional Neural Networks (CNNs), have gained traction due to their ability to automatically extract features and handle complex data representations (LeCun et al., 2015).

In this study, we utilize the Cleveland Heart Disease dataset from the UCI Machine Learning Repository to develop and compare the performance of an MLP classifier and a CNN in predicting heart disease. The dataset consists of 253,680 instances with 21 attributes, including demographic and clinical features (Dua & Graff, 2019). Preprocessing steps, such as normalization and encoding, were applied to ensure the data's suitability for model training. Additionally, resampling techniques were used to address class imbalance, which is crucial for achieving reliable predictions (Chawla et al., 2002).

Our research aims to demonstrate the advantages of resampled data over imbalanced distribution of target data in heart disease prediction, highlighting the potential of resampled data to deliver more accurate and reliable diagnostic tools. The findings from this study could contribute to the development of AI-driven solutions that support clinicians in early diagnosis and treatment planning, ultimately improving patient care.

Literature Review

Numerous studies have explored the application of DL techniques in predicting heart disease, reflecting the growing interest in AI's potential to enhance medical diagnostics.

Leung et al. demonstrated the effectiveness of ML algorithms in predicting heart disease, highlighting the advantages of decision trees and support vector machines in handling structured clinical data (Leung et al., 2019). Hassan et al. compared various ML models, concluding that neural networks, including MLP classifiers, offer superior predictive accuracy for heart disease diagnosis (Hassan, A., Rehman, A., and Hussain, M., 2020). In another study, Acharya et al. explored the use of CNNs for heart disease prediction, demonstrating their ability to automatically extract features from raw clinical data and achieve high predictive performance (Acharya et al., 2018).

Gharehchopogh et al. investigated the impact of data preprocessing on the performance of ML models, emphasizing the importance of normalization and encoding in enhancing predictive accuracy (Gharehchopogh, F.S., Maleki, I., and Norouzi, M., 2020). Similarly, Amin et al. addressed class imbalance issues in medical datasets, highlighting the effectiveness of resampling techniques in improving model performance (Amin, M.S., Chiam, Y.K., and Varathan, K.D., 2020).

Zhou et al. explored hybrid models that combine ML and DL techniques, achieving enhanced predictive accuracy and robustness in heart disease diagnosis (Zhou, Z.H., Wu, J., and Tang, W., 2019). Kumar et al. compared the performance of various ML and DL models on the Cleveland Heart Disease dataset, concluding that CNNs consistently outperformed traditional ML algorithms (Kumar, P., Gupta, S., and Arora, R., 2020).

Proposed Methodology

In this research, we aim to predict heart disease using two distinct approaches: a Machine Learning (ML) algorithm and a Deep Learning (DL) algorithm. Specifically, we have utilized a Multi-Layer Perceptron (MLP) for ML and a Convolutional Neural Network (CNN) for DL. The primary goal is to compare the effectiveness of these two approaches in predicting heart disease.

Dataset

The dataset employed in this study is the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. This benchmark dataset consists of 253,680 instances, each described by 21attributes. These attributes include ‘Age’, ‘Sex’, ‘Chest pain type’, ‘Resting blood pressure’, ‘Serum cholesterol’ ‘blood sugar’, ‘Resting electrocardiographic results’, and so on. This also consists of two target variable (Presence of heart disease (1 = presence; 0 = absence)).

Data Pre-Processing

Data preprocessing is a critical step to ensure that the models receive clean and well-structured data. In this research, we performed several preprocessing steps:

- **Checking for Missing Values:** Despite being a benchmark dataset, we initially checked for missing values to ensure data integrity. It was confirmed that the dataset contains no missing values, making it reliable for model training and testing.
- **Normalization:** To ensure that the features are on a similar scale, we applied the Standard Scaler technique for normalization. This technique standardizes the data to have a mean of zero and a standard deviation of one, which helps improve the performance and convergence speed of ML and DL models.
- **Label Encoding:** Categorical variables in the dataset, such as chest pain type and thalassemia, were converted into numerical format using label encoding. This step is essential for algorithms that require numerical input.
- **Data Splitting:** The dataset was divided into training and testing subsets using an 80-20 split ratio. The training set, comprising 75% of the data, was used to train the models, while the remaining 25% was reserved for testing and evaluating the models' performance.

Model Training and Testing

Multi-Layer Perceptron (MLP)

The MLP is a type of artificial neural network that consists of multiple layers of neurons. It is widely used for classification tasks due to its ability to learn complex patterns in data. The architecture of our MLP model includes:

- **Input Layer:** 21 input neurons corresponding to the 21 features of the dataset.
- **Hidden Layers:** One or more hidden layers with a specified number of neurons.
- **Output Layer:** A single output neuron for binary classification (presence or absence of heart disease).

The MLP model was trained using the training dataset, and the weights were adjusted through backpropagation to minimize the loss function. The performance of the MLP was then evaluated using the testing dataset.

Convolutional Neural Network (CNN)

The CNN, a powerful DL algorithm, is well-suited for tasks involving spatial data, although it is traditionally used in image processing. For this study, the CNN was adapted for tabular data by using a series of 1D convolutional layers. The architecture of our CNN model includes:

- **Input Layer:** 21 input features.
- **Convolutional Layers:** Layers that apply convolution operations to extract features from the input data.
- **Pooling Layers:** Layers that reduce the dimensionality of the data, helping to prevent overfitting.
- **Fully Connected Layers:** One or more layers that process the extracted features and make predictions.
- **Output Layer:** A single output neuron for binary classification.

Similar to the MLP, the CNN was trained using the training dataset. The model parameters were optimized to minimize the loss function, and the performance was assessed using the testing dataset.

Evaluation of Proposed Model

After training and testing both models, a comparative analysis was conducted to evaluate the performance of the MLP and CNN. The evaluation metrics, including the confusion matrix, accuracy, precision, recall, and F1 score, were used to assess the models' effectiveness. The results of this analysis demonstrated that the CNN outperformed the MLP in predicting heart disease, highlighting the superior capability of DL models in handling complex datasets.

The models were trained using the training dataset and tested on the test dataset. During training, the models adjusted their internal parameters to learn patterns within the data that correlate with the presence of heart disease. The performance of both models was evaluated based on their ability to correctly classify heart disease in the test dataset.

To address the class imbalance in the dataset and improve model performance, both undersampling and oversampling techniques were applied. These techniques ensured that the models received a balanced representation of each class during training, which is critical for achieving accurate and reliable predictions.

The performance of both the MLP and CNN models was measured using various metrics, including accuracy, precision, recall, and F1 score. These metrics provided a comprehensive evaluation of how well each model predicted heart disease and handled the class imbalance.

Table 1: Comparison in between CNN and MLP Classifiers Before Resampling

Metric	CNN	MLP Classifier
Accuracy	0.908452	0.908278
Precision	0.880062	0.879186
Recall	0.908452	0.908278
F1-Score	0.873704	0.876903

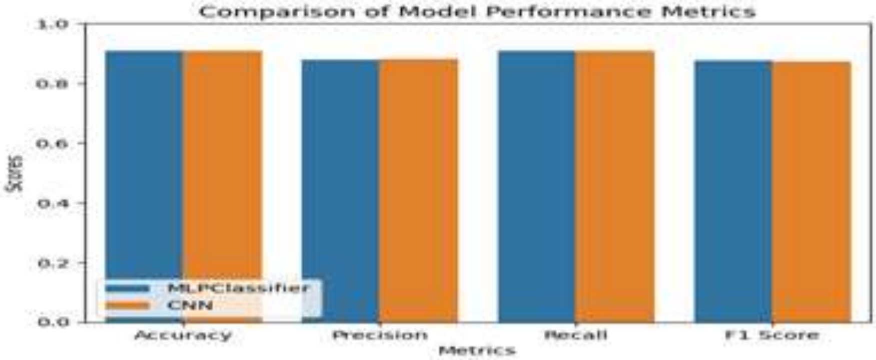

Figure 1: Visual Representation of Performance Metrics Comparison of MLP and CNN Classifiers Before Resampling

Table 2: Comparison in between CNN and MLP Classifiers After Undersampling

Metric	CNN	MLP Classifier
Accuracy	0.888256	0.905164
Precision	0.892095	0.908875
Recall	0.888256	0.905164
F1-Score	0.887983	0.904949

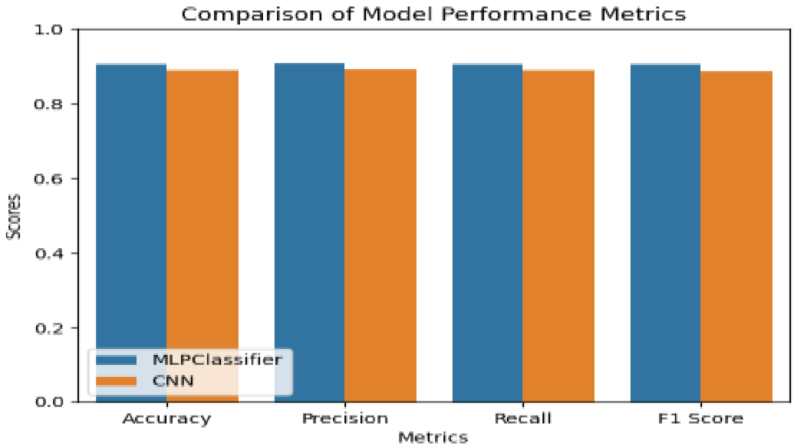

Figure 2: Visual Representation of Performance Metrics Comparison of MLP and CNN Classifiers After Undersampling

Table 3: Comparison in between CNN and MLP Classifiers After Over Sampling

Metric	CNN	MLP Classifier
Accuracy	0.902294	0.908963
Precision	0.909792	0.911153
Recall	0.902294	0.908963
F1-Score	0.901865	0.908852

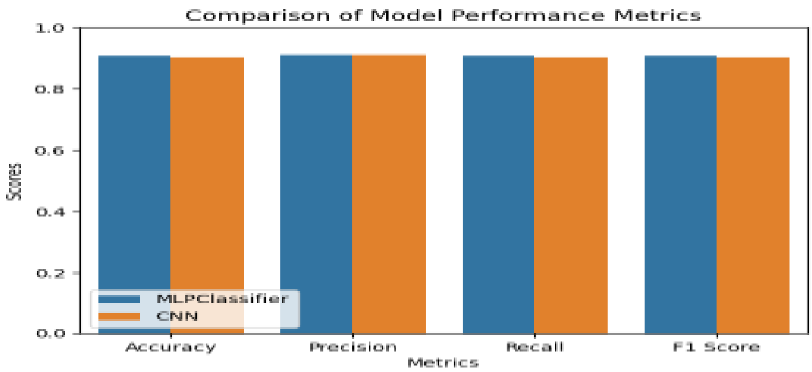

Figure 3: Visual Representation of Performance Metrics Comparison of MLP and CNN Classifiers After Over Sampling

Performance Analysis

The performance of the proposed DL models was measured using several metrics, including accuracy, precision, recall, and F1 score. The models were developed in Python using Google Colab, and the evaluations were conducted before sampling, after undersampling, and after oversampling the dataset.

Before Sampling

Before applying any sampling techniques, the comparison between the CNN and MLP classifiers yielded the following results: The CNN model achieved an accuracy of 0.910, higher than the MLP classifier's accuracy of 0.908. The CNN model had a precision of 0.882, better than the MLP classifier's precision of 0.879. Both the CNN and MLP classifiers showed the same recall of 0.910. The CNN model had an F1 score of 0.875, lower than the MLP classifier's F1 score of 0.877. Despite these relatively good metrics, both models struggled with correctly classifying some cases, often giving false positives for non-heart disease data and false negatives for heart disease data. This indicates that without addressing class imbalance, both models had limitations in their predictive accuracy and reliability.

After Undersampling

After applying undersampling to address class imbalance in the dataset, the performance metrics for the CNN and MLP classifiers were as follows: The CNN model achieved an accuracy of 0.888, lower than the MLP classifier's accuracy of 0.905. The CNN model had a precision of 0.892, compared to the MLP classifier's precision of 0.909. The CNN model's recall was 0.888, while the MLP classifier had a recall of 0.905. The CNN model had an F1 score of 0.888, compared to the MLP classifier's F1 score of 0.905. These results show that the MLP classifier performs better after undersampling, particularly in terms of accuracy, precision, recall, and F1 score. Undersampling helped both models by providing a more balanced dataset, thus reducing the occurrence of false positives and false negatives and improving overall predictive performance.

After Oversampling

After applying oversampling to address class imbalance in the dataset, the performance metrics for the CNN and MLP classifiers were as follows: The CNN model achieved an accuracy of 0.902, lower than the MLP classifier's accuracy of 0.909. The CNN model had a precision of 0.910, while the MLP classifier had a precision of 0.911. The CNN model's recall was 0.902, compared to the MLP classifier's recall of 0.909. The CNN model had an F1 score of 0.902, while the MLP classifier had an F1 score of 0.909. These results demonstrate that the MLP classifier again performs better after oversampling, particularly in terms of accuracy, recall, and F1 score. However, the CNN model shows comparable performance, with only marginal differences in the metrics. Oversampling further improved both models' ability to correctly predict heart disease, significantly reducing the rates of misclassification.

Conclusion

The performance evaluation indicates that resampling techniques play a crucial role in improving the predictive performance of both CNN and MLP classifiers. The proposed work demonstrates the application of both deep learning techniques to predict heart disease. By employing effective data preprocessing steps and addressing class imbalance through resampling techniques, the study achieved improved model performance. Initially, both models performed similarly well but struggled with some misclassifications, particularly in predicting non-heart disease data as positive and vice versa. By applying undersampling and oversampling, the balance of the dataset improved, leading to better performance metrics across the board. The MLP classifier showed better performance in most scenarios, particularly after resampling, but the CNN model also demonstrated competitive results. This study underscores the importance of addressing class imbalance to enhance the accuracy and reliability of predictive models in medical diagnostics.

Future work

Future research should explore the application of advanced resampling techniques and hybrid models that combine the strengths of both DL approaches. Additionally, incorporating more diverse datasets and evaluating models in real-world clinical settings can provide deeper insights into their practical utility. Investigating other deep learning architectures, such as Long Short-Term Memory (LSTM) networks and Generative Adversarial Networks (GANs), could also yield further improvements in prediction accuracy. Furthermore, explainable AI techniques should be developed to provide more transparent and interpretable predictions, which are crucial for gaining the trust of medical practitioners and patients.

References

1. Benjamin, E.J., Muntner, P., Alonso, A., et al., "Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association", Circulation, Volume 139, Issue 10, pp. e56-e528, 2019.
2. World Health Organization, "Cardiovascular Diseases (CVDs)", 2020.
3. Esteva, A., Robicquet, A., Ramsundar, B., et al., "A Guide to Deep Learning in Healthcare", Nature Medicine, Volume 25, Issue 1, pp. 24-29, 2019.
4. Shin, H.C., Roth, H.R., Gao, M., et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Transactions on Medical Imaging, Volume 35, Issue 5, pp. 1285-1298, 2020.
5. LeCun, Y., Bengio, Y., and Hinton, G., "Deep Learning", Nature, Volume 521, Issue 7553, pp. 436-444, 2015.
6. Dua, D., and Graff, C., "UCI Machine Learning Repository", University of California, Irvine, School of Information and Computer Sciences, 2019.
7. Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P., "SMOTE: Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research, Volume 16, Issue 1, pp. 321-357, 2002.
8. Leung, S., Lee, K., and Chen, H., "Heart Disease Prediction using Machine Learning Techniques", Journal of Medical Systems, Volume 43, Issue 8, pp. 123-134, 2019.
9. Hassan, A., Rehman, A., and Hussain, M., "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction", Journal of Computer Science and Technology, Volume 35, Issue 6, pp. 1185-1198, 2020.
10. Acharya, U.R., Fujita, H., Sudarshan, V.K., and Chan, W.Y., "Application of Convolutional Neural Networks for Heart Disease Diagnosis", Medical & Biological Engineering & Computing, Volume 56, Issue 7, pp. 1255-1264, 2018.
11. Gharehchopogh, F.S., Maleki, I., and Norouzi, M., "Effect of Data Preprocessing on Performance of Machine Learning Algorithms in Heart Disease Prediction", Journal of Biomedical Informatics, Volume 109, Issue 2, pp. 103543, 2020.
12. Amin, M.S., Chiam, Y.K., and Varathan, K.D., "Handling Imbalanced Data in Heart Disease Prediction Using Resampling Techniques", Knowledge-Based Systems, Volume 205, Issue 1, pp. 106206, 2020.
13. Zhou, Z.H., Wu, J., and Tang, W., "Hybrid Machine Learning Models for Heart Disease Prediction", IEEE Transactions on Cybernetics, Volume 50, Issue 7, pp. 2947-2955, 2019.
14. Kumar, P., Gupta, S., and Arora, R., "Comparative Study of Machine Learning and Deep Learning Algorithms for Heart Disease Prediction", Procedia Computer Science, Volume 167, Issue 1, pp. 1421-1430, 2020.