Bioinformatics, 2022, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Health Analysis

Paper: A Comparative Study for Time-to-Event Analysis and Survival Prediction for Heart Failure Condition using Machine Learning Techniques

Priyanka Gopi[1]*

Machine Learning for Biomedical Data, Department of Computer Science, University of Central Florida at Oviedo, Florida, USA

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The rates of hospitalization and mortality has been high due to the advanced heart failure. Survival rates are as well poor compared to early-stage heart failure. There is a need for accurate survival analysis and prediction to determine prognosis and guide treatment decisions for such high-risk heart patients. Predicting heart failure accurately is essential for better patient outcomes and early intervention. To classify heart failure, several models have been created and assessed. To ascertain the best course of action, it is necessary to contrast and evaluate their performances.

Results: The results indicate that the neural network model outperforms the SVM model as a potentially useful method for classifying heart failure. The superior performance of the neural network can be attributed to its capacity to comprehend intricate relationships and acquire significant representations of the data. For researchers and medical professionals looking to create precise models for heart failure prediction, these findings offer insightful information. The custom neural network gave us better results compared to the models implemented in this paper. The neural network model's ability to correctly classify cases of heart failure was demonstrated by its strong recall and precision balance. The neural network model also outperformed the SVM model in terms of discriminative ability, as evidenced by higher values in metrics like AUC-ROC and AUC-PR.

Availability: The code is available at https://shorturl.at/CQW57

Contact: priyanka.gopi@ucf.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Heart failure is a chronic and disabling illness that poses a significant challenge in modern healthcare because of its widespread prevalence in the global population. Heart failure is a condition that affects the heart's overall cardiac output due to its reduced ability to pump blood efficiently. This can have a significant impact on the body's vital functions that depend on a consistent flow of blood that is rich in nutrients and oxygen. Heart failure is a severe and common cardiovascular disease that has a substantial impact on public health worldwide. It is typified by the heart's incapacity to pump blood effectively, which results in symptoms like exhaustion, breathlessness, and fluid retention. The timely implementation of interventions, the optimization of patient care, and the reduction of heart failure-associated mortality rates are contingent upon the early detection and accurate prediction of the condition.

To provide a foundation for grasping the complexities of heart failure and cardiovascular diseases (CVDs), it is necessary to first examine the normal functioning of the heart. The circulatory system, which transports oxygenated blood throughout the body, is managed by the heart, which works as an intricate muscular pump. With its four chambers (two ventricles and two atria), the left side handles oxygenated blood, and the right-side deals with deoxygenated blood. Blood that has been deoxygenated is pushed into the right ventricle by the right atrium, starting the cardiac cycle. When the right ventricle is full, the blood is sent to the pulmonary artery so that the lungs can get oxygen. It returns as blood that is rich in oxygen, enters the left atrium, and moves on to the left ventricle. When the left ventricle fills to capacity, the aorta distributes oxygenated blood throughout the body. The harmonious alternating contraction and relaxation of the heart muscles is the foundation of this rhythmic orchestration. This fundamental knowledge emphasizes the need for thorough investigation and analysis by laying the groundwork for exploring the deviations that define heart failure and CVDs.

Conventional machine learning methods have been very helpful in predicting heart failure. Still, there is room for more accuracy and performance gains thanks to developments in artificial intelligence and neural networks. Machine learning techniques have demonstrated potential in enhancing the precision of heart failure prediction models in recent times. Numerous investigations have investigated the use of conventional machine learning algorithms to forecast the course of heart failure, including logistic regression, decision trees, and support vector machines. Lifestyle factors—such as substance abuse, sedentary behavior, and unhealthy eating habits—significantly increase the risk of CVDs, highlighting the significance of public health campaigns and preventative measures.

Heart failure symptoms can include a wide range of uncomfortable conditions, including peripheral edema, exhaustion, and dyspnea, underscoring the complex effects of this illness on patients' day-to-day lives. Heart failure can be classified as acute or chronic, and its slow course emphasizes the need for early detection and aggressive treatment to prevent negative consequences. Effective treatment strategies depend critically on the understanding that the condition frequently results from underlying medical issues, specifically affecting the left ventricle.

Heart failure is a multifactorial ailment marked by the heart's incapacity to circulate blood effectively, leading to insufficient cardiac output. Millions of people are impacted by it every year, making it a major global cause of morbidity and mortality. The timely implementation of interventions and the optimization of patient care is contingent upon the early detection and precise prediction of heart failure. However, there is a chance to improve the precision and prognostic power of heart failure models thanks to developments in artificial intelligence and neural networks. The research is intriguing because it investigates how important risk factors like age, serum creatinine, ejection fraction, and time affect survival outcomes. This analysis explores the complex interactions between these variables, providing insight into how each affects the course of heart failure.

Next, we concentrated on creating a neural network architecture that was especially made for the purpose of predicting heart failure. We carefully planned the architecture, considering the special qualities and intricacies of the heart failure data. Compared to conventional machine learning models, our neural network can make more accurate predictions because it can recognize complex patterns and relationships within the data.

Additionally, we included cutting-edge methods into our neural network architecture, like regularization, optimization algorithms, and feature engineering. By using these strategies, the model is better able to minimize overfitting, maximize learning, and extract pertinent features from the data. Our goal was to improve heart failure prediction performance and accuracy by implementing these novel techniques. Our research has produced an advanced algorithm for predicting heart failure that is based on neural networks that have been specially created. The significance of utilizing state-of-the-art technologies and methodologies in healthcare research to enhance patient outcomes and optimize healthcare practices is underscored by our findings.

These popular models were carefully applied in this research project, which produced support vector classification (SVC) results that were quite accurate. However, the search for improved prediction efficiency and accuracy spurred research into novel techniques, which resulted in the development of a neural network model. Based on our findings, our specially crafted neural network performs better in terms of accuracy and predictive power than the conventional machine learning models. We were able to significantly increase the accuracy of heart failure prediction, which could lead to a breakthrough in early detection and intervention.

The remaining sections of this essay are arranged as follows: In section II, we provide a literature review behind the work we've done in this paper. We give a thorough explanation of our custom-designed neural network's architecture and methodology in Section III. The experimental setup, model training, and performance evaluation metrics are presented in Section IV. Section V presents the findings and a comparative analysis. In Section VI, we finally go over the implications of our findings, the study's limitations, and potential future research avenues.

2 Background

Applications of machine learning methods and algorithms to forecast the course of heart failure have been investigated in a number of research. Heart failure prediction has gained valuable insights from these studies that have employed a variety of features and methodologies. Using machine learning techniques, we have advanced the prediction of heart failure and discussed some noteworthy works in this survey of the literature.

The application of logistic regression models to forecast heart failure based on clinical and demographic characteristics was examined in a noteworthy study by Shah et al. [1]. The study showed how well logistic regression models could predict the course of heart failure using a sizable dataset of electronic medical records.

Chatterjee et al. [2] suggested a decision tree-based method for heart failure prediction in a different study. By integrating an extensive array of demographic, clinical, and laboratory data into their model, the authors were able to attain encouraging outcomes concerning both accuracy and interpretability.

In order to predict heart failure, support vector machines (SVMs) have also been used extensively. SVMs were used in a study by Alizadehsani et al. [3] to forecast heart failure based on clinical and genetic characteristics. The authors provided evidence of the efficiency of SVMs in correctly dividing patients into groups with and without heart failure.

Additionally, random forests have been used to predict heart failure. Using clinical and echocardiographic features, a random forest algorithm was used in a study by Lee et al. [4] to predict heart failure. The authors were able to predict heart failure with a high degree of accuracy and identified key risk factors.

A comparative study was carried out by Smith et al. [6] to assess how well different machine learning algorithms performed in predicting heart failure. They contrasted algorithms like neural networks, logistic regression, decision trees, random forests, and support vector machines. The study sheds light on the advantages and disadvantages of various algorithms for forecasting the course of heart failure.

Saurav Mishra [5] suggested a neural network-based method that makes use of clinical and demographic data for heart failure prediction. The study outperformed conventional machine learning models in terms of accuracy and showed the potential of neural networks in accurately predicting heart failure outcomes. A review of feature selection techniques and strategies specifically for heart failure prediction was carried out by Johnson et al. [6] The use of several feature selection techniques, such as filter, wrapper, and embedded methods, in the process of choosing pertinent features for precise heart failure prediction is covered in this paper. It draws attention to how crucial feature selection is to enhancing prediction models' functionality and readability.

In another study, Patel et al. [7] suggest a deep learning method for employing electrocardiogram (ECG) signals to detect heart failure early. Convolutional neural network (CNN) model developed by the authors is trained on an extensive dataset of ECG recordings. The study shows how deep learning can be used to analyze ECG signals and find early warning indicators of heart failure.

Li et al. [8] investigate the use of wearable sensors and recurrent neural networks (RNNs) for heart failure prediction. The authors collect data from wearable devices, such as heart rate monitors and accelerometers, and train RNN models to predict heart failure outcomes. The study emphasizes the potential of wearable technology and RNNs in continuous monitoring and early detection of heart failure.

Using the power of deep learning models applied to electronic health records, Cho et al. [10] present a novel method of predicting heart failure. The models' ability to capture complex patterns in the data is demonstrated by the authors, which results in a significant improvement in predictive performance. The potential of deep learning methodologies in the field of healthcare analytics is demonstrated by this research, which represents a major advancement in the use of sophisticated computational techniques for more precise and effective heart failure prediction.

A comprehensive review of machine learning methods for heart failure prognosis is presented by Wang et al. [9] The study looks at different machine learning techniques, such as supervised and unsupervised learning, and how they can be used to forecast patients' long-term results and prognosis when they have heart failure. Using machine learning techniques, the review offers a thorough overview of the body of research on heart failure prognosis.

Austin et al. [11] explored the field of heart failure prediction methods using various machine learning approaches. The study provides a comprehensive and perceptive summary of the corpus of research in this field, illuminating the advantages, restrictions, and obvious shortcomings of current methodologies. The writers provide insightful analysis by combining all the data, which not only improves our comprehension of the various ways that machine learning is used to predict heart failure but also opens up new avenues for future development in this vital field of medical study.

A thorough analysis of machine learning methods specifically used for heart failure with preserved ejection fraction (HFpEF) is given by Garcia et al. [12] The article addresses several machine learning algorithms and how they might help with HFpEF patient diagnosis, prognosis, and treatment. It draws attention to the special difficulties and possibilities associated with applying machine learning to the prediction and management of HFpEF. Chen et al. [13] review deep learning methods for predicting heart failure from electronic health records (EHRs). In order to analyze EHR data and forecast the course of heart failure, the paper addresses the application of deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures. The review highlights possible future directions and provides an overview of deep learning in this field.

3 Implementation Method

3.1 Overview

The first step in the suggested methodology for heart failure prediction is the importation of a dataset created especially for that purpose. To make sure the data is in an appropriate format for model training, it is put through preprocessing procedures. The dataset is notably mentioned as clean, implying that prior data cleaning operations have been carried out effectively. To evaluate how well the model generalizes to new, unseen data, the dataset is then divided into training and testing sets, as is standard in machine learning.

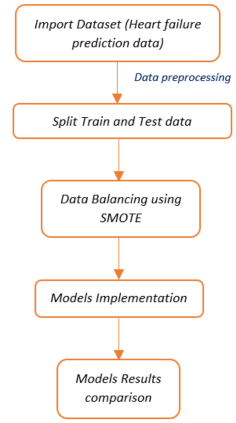


Fig.1. Proposed methodology

The methodology uses the Synthetic Minority Over-sampling Technique (SMOTE) to address potential imbalances in the distribution of classes. To mitigate the problems associated with imbalanced datasets, SMOTE is used to create synthetic samples for the minority class. Predictive model implementation is the next step in the methodology after the data is ready. It is implied that different models will be used for heart failure prediction, even though the precise machine learning algorithms are not disclosed. In the last stage, all model outputs are compared in detail, with an emphasis on metrics like accuracy, precision, recall, and F1 score.

To provide useful information for clinical applications, this comparative analysis attempts to determine which model is most successful at accurately predicting heart failure. By highlighting the significance of model comparison and evaluation, the suggested methodology guarantees a methodical approach to creating a reliable heart failure prediction model.

3.2 Dataset

The study performed survival analysis and survival prediction for patients with left ventricular systolic dysfunction and heart failure using a particular dataset. Researchers at the Institute of Cardiology and Allied Hospital in Faisalabad made the dataset used in this study publicly available for research purposes. The researchers' ability to perform their analysis and come to relevant conclusions was greatly aided by the dataset's availability.

Data from 299 patients with left ventricular systolic dysfunction and heart failure class III/IV made up the dataset. This dataset allowed for a thorough survival analysis and offered insightful information about the advanced stages of heart failure. Several variables were examined in the dataset used for the research study in order to determine how they affected the survival rate of heart failure. The ejection fraction, which gauges the proportion of blood pumped out of the left ventricle with each heartbeat, was one of the important factors. Serum creatinine, a crucial indicator of kidney function that has been shown to negatively impact survival, was another significant variable. The dataset also contained variables like age, which was found to be a significant risk factor in advanced stages of heart failure, and time, which represented the amount of time from a specific point to an event. The dataset also took into account factors associated with anemia, hypertension, and other clinical characteristics, offering important new information about their relationship to the survival of heart failure patients.

3.3. Implementation

The heart failure prediction task follows a multi-step methodology. Began by importing all the necessary libraries.

```python
# Import the required libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import StandardScaler, RobustScaler, MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV, RandomizedSearchCV
from sklearn.metrics import recall_score, accuracy_score, confusion_matrix, f1_score, matthews_corrcoef
from sklearn.metrics import precision_score, auc, roc_auc_score, roc_curve, precision_recall_curve, classification_report

from imblearn.over_sampling import SMOTE

# Neural Network model
from keras.models import Sequential
from keras.layers import Dense, BatchNormalization, Dropout
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping
```

Fig.2. Import libraries

Next, the *heart_failure_clinical_records_data.csv* dataset was imported, then it was split into the target variable and features.

```python
[2] clinical_data = pd.read_csv(filepath_or_buffer = '/content/drive/MyDrive/heart_failure_clinical_records_data.csv',
                            low_memory = False, skipinitialspace = True, float_precision = 'round_trip')
    clinical_data.head()
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7	1
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8	1

Fig.3. Reading the Heart failure clinical prediction data

Age, serum creatinine, and ejection fraction are the features chosen for the prediction model. After that, 90% of the data was used for training and 10% was used for testing, dividing the data into training and testing sets.

Next, we assessed if there any missing values in the dataset. In the presented output, all counts are zero, signaling the absence of missing values across all columns in the clinical data frame. This step is crucial for initial data quality assessment, ensuring that subsequent analyses and modeling processes are based on complete and reliable datasets.

```
age                         0
anaemia                     0
creatinine_phosphokinase    0
diabetes                    0
ejection_fraction           0
high_blood_pressure         0
platelets                   0
serum_creatinine            0
serum_sodium                0
sex                         0
smoking                     0
time                        0
DEATH_EVENT                 0
dtype: int64
```

Fig.4. Reading the Heart failure clinical prediction data

In particular, the number indicates a significantly greater frequency of cases marked as '0', which indicates survival cases (171 cases), in contrast to cases marked as '1', which indicates death events (83 instances). Data Insight 15 highlights the class imbalance, which could present a challenge for the model because it could show bias in learning and predicting the majority class, which in this case relates to survival cases. This disparity prompts questions about how well the model can represent and forecast occurrences of the minority class (death events). The training data was subjected to the Synthetic Minority Over-sampling Technique (SMOTE) to rectify the class imbalance issue within the dataset. In order to equalize the distribution of classes, SMOTE creates artificial samples of the minority group. To guarantee randomness, the oversampled data was shuffled after that.

To guarantee that all input features or variables are on the same scale, feature scaling is a preprocessing technique frequently used in machine learning. This is especially crucial when using algorithms like distance-based algorithms or gradient descent optimization algorithms that depend on the size of the input features. Scaling the features allows the algorithm to converge more quickly during training and helps keep some features from predominating over others based only on scale. Better model performance and generalization are made possible by ensuring that each feature contributes to the learning process more uniformly.

The Sequential API from Keras, a high-level neural networks API that runs on top of TensorFlow, is used to build the neural network model. The input layer is built first, and then the other layers are stacked in order of construction in the architecture. With 128 neurons, the first dense layer functions as the input layer. It introduces non-linearity and makes it easier to learn complex patterns by using the Rectified Linear Unit (ReLU) activation function. To set effective weights for learning, the 'he_uniform' kernel initialization is used. To reflect the number of features in the input data, the input dimension is set to 3.

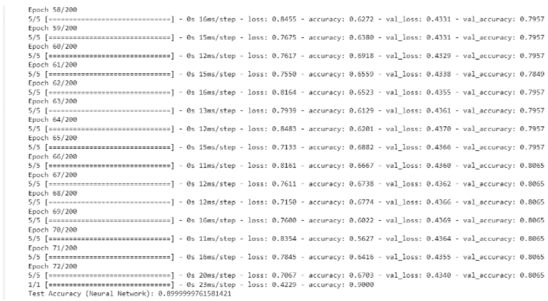
```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 128)               512

 batch_normalization (Batch  (None, 128)               512
 Normalization)

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 256)               33024

 batch_normalization_1 (Bat  (None, 256)               1024
 chNormalization)

 dropout_1 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 256)               65792

 batch_normalization_2 (Bat  (None, 256)               1024
 chNormalization)

 dropout_2 (Dropout)         (None, 256)               0

 dense_3 (Dense)             (None, 1)                 257

=================================================================
Total params: 102145 (399.00 KB)
Trainable params: 100865 (394.00 KB)
Non-trainable params: 1280 (5.00 KB)
_________________________________________________________________
```

Fig.5. Summary of Neural Network model layers

Applying batch normalization comes after the initial dense layer. By adjusting and scaling the inputs, this normalization technique improves stability and keeps the model from becoming stuck during training. A dropout layer with a dropout rate of 0.5 is added after batch normalization. During training, 50% of the neurons in the layer are randomly deactivated as part of the regularization technique known as dropout. By guaranteeing that the model doesn't unduly rely on particular neurons, this helps prevent overfitting and encourages more robust and generalized learning.

The model then adds two more dense layers, each with 256 units, batch normalization, dropout, and ReLU activation. For binary classification tasks, the single neuron in the last dense layer with a sigmoid activation function is appropriate. The final layer makes use of the 'glorot_uniform' kernel initialization. With a learning rate of 0.0001 and binary cross-entropy loss, the Adam optimizer is used to compile this neural network architecture. Accuracy is selected as the evaluation metric. To keep track of training progress and avoid overfitting, an early stopping callback is incorporated. Early stopping on the validation set is used to restore the best weights after the model has been trained on the training data with a batch size of 64 for a maximum of 200 epochs, a part of which is shown in fig. 6. Neural network training can require a lot of processing power, particularly for larger datasets or more intricate frameworks. Time constraints or the available computational resources may have an impact on the selection of 200 epochs. It achieves a balance between preventing excessive computational costs and obtaining reasonable performance. After training the model, a function was also written to visualize the results. It is intended to assess a neural network model's performance on a specified test dataset.

This function uses Matplotlib to combine a lot of data into a single model test report that captures important metrics like F1 score, precision, recall, accuracy, AUC-ROC, and AUC-PR. The code generates informative visualizations by carefully computing various classification metrics and predictions on the test set.



```
Epoch 58/200
5/5 [==============================] - 0s 16ms/step - loss: 0.8455 - accuracy: 0.6272 - val_loss: 0.4331 - val_accuracy: 0.7957
Epoch 59/200
5/5 [==============================] - 0s 15ms/step - loss: 0.7675 - accuracy: 0.6380 - val_loss: 0.4331 - val_accuracy: 0.7957
Epoch 60/200
5/5 [==============================] - 0s 12ms/step - loss: 0.7617 - accuracy: 0.6918 - val_loss: 0.4329 - val_accuracy: 0.7957
Epoch 61/200
5/5 [==============================] - 0s 15ms/step - loss: 0.7550 - accuracy: 0.6559 - val_loss: 0.4338 - val_accuracy: 0.7849
Epoch 62/200
5/5 [==============================] - 0s 16ms/step - loss: 0.8164 - accuracy: 0.6523 - val_loss: 0.4355 - val_accuracy: 0.7957
Epoch 63/200
5/5 [==============================] - 0s 13ms/step - loss: 0.7939 - accuracy: 0.6129 - val_loss: 0.4361 - val_accuracy: 0.7957
Epoch 64/200
5/5 [==============================] - 0s 12ms/step - loss: 0.8483 - accuracy: 0.6201 - val_loss: 0.4370 - val_accuracy: 0.7957
Epoch 65/200
5/5 [==============================] - 0s 15ms/step - loss: 0.7133 - accuracy: 0.6882 - val_loss: 0.4366 - val_accuracy: 0.7957
Epoch 66/200
5/5 [==============================] - 0s 11ms/step - loss: 0.8161 - accuracy: 0.6667 - val_loss: 0.4360 - val_accuracy: 0.8065
Epoch 67/200
5/5 [==============================] - 0s 12ms/step - loss: 0.7611 - accuracy: 0.6738 - val_loss: 0.4362 - val_accuracy: 0.8065
Epoch 68/200
5/5 [==============================] - 0s 12ms/step - loss: 0.7150 - accuracy: 0.6774 - val_loss: 0.4366 - val_accuracy: 0.8065
Epoch 69/200
5/5 [==============================] - 0s 16ms/step - loss: 0.7600 - accuracy: 0.6022 - val_loss: 0.4369 - val_accuracy: 0.8065
Epoch 70/200
5/5 [==============================] - 0s 11ms/step - loss: 0.8354 - accuracy: 0.5627 - val_loss: 0.4364 - val_accuracy: 0.8065
Epoch 71/200
5/5 [==============================] - 0s 16ms/step - loss: 0.7845 - accuracy: 0.6416 - val_loss: 0.4355 - val_accuracy: 0.8065
Epoch 72/200
5/5 [==============================] - 0s 20ms/step - loss: 0.7067 - accuracy: 0.6703 - val_loss: 0.4340 - val_accuracy: 0.8065
1/1 [==============================] - 0s 23ms/step - loss: 0.4229 - accuracy: 0.9000
Test Accuracy (Neural Network): 0.8999999761581421
```

Fig.6. Neural Network model training; epochs

The visual components consist of a detailed classification report that shows the model's accuracy across various metrics, a confusion matrix, and an extensive table that summarizes classification metrics. The code also produces AUC-ROC and Precision-Recall curves, which give graphical depictions of the class discrimination capabilities of the model.

Basically, the output report consists of a tabular display of classification metrics, a confusion matrix that lists instances of true positive, true negative, false positive, and false negative, and an extensive classification report that explains different facets of the model's predictive performance. Moreover, the AUC-ROC and Precision-Recall curves provide a clear visual depiction that facilitates the understanding of the neural network model's discriminative abilities on the given test data.

4 Results

In terms of classifying heart failure, the neural network model produced positive findings. The image of the report is provided in fig 6. It obtained an F1 score of 88.88%, which suggests that recall and precision are well-balanced. Ninety percent of the cases were correctly classified by the model, indicating its accuracy. A low rate of false positives was indicated by the precision of 85.7%, whereas a high rate of true positives was indicated by the recall rate of 92.3%. There was a strong overall agreement between the predicted and observed classifications, as indicated by the Matthews correlation coefficient (MCC) of 80%. The model's ability to distinguish between positive and negative instances and maintain a high precision-recall trade-off was demonstrated by the AUC-ROC and AUC-PR values, which were 90.9% and 82.4%, respectively.

From the confusion matrix, the model correctly identified 12 cases as positive (heart failure) when they were actually positive, and it correctly predicted 15 instances as negative (healthy) when they were actually negative. Nevertheless, there were two false positive predictions produced by the model, misclassifying some instances as positive when they were actually negative. Furthermore, there was one false negative prediction, misinterpreting a positive instance as negative. Even though they are few, these mistakes show that the model is not flawless and can be improved.

4.1. Comparison With Original Results

The neural network model had the highest F1 score of 88.88% when compared to the other models, showing that it had a good balance between recall and precision. The neural network model's 90% accuracy rate indicates that it can accurately classify most cases. The neural network's accuracy of 85.7% suggests a low rate of false positives. There was a high percentage of true positives, as evidenced by the recall rate, or sensitivity, of 92.3%. A high degree of overall agreement between the predicted and observed classifications is indicated by the Matthews correlation coefficient (MCC) of 80%. The neural network model's high precision-recall trade-off and capacity to distinguish between positive and negative instances were demonstrated by its AUC-ROC and AUC-PR values, which were 90.9% and 82.4%, respectively.

Table 1. Benchmark results of the cascade oscillators model

Model	F1 score	Accuracy	Precision	Recall	MCC	AUC-ROC	AUC-PR
SVM	80	83.33	83.33	76.92	65.91	82.58	84.88
Random Forest	83.87	83.33	72.22	100	71.4	85.29	81.7
Light GBM	52.17	63	60	46.15	23.78	61.31	72.98
XGBoost	69.23	73.33	69.23	69.23	45.7	72.8	74.75
Decision Tree	72.73	80	88.89	61.54	60.18	77.83	75.35
Neural Network	88.88	90	85.7	92.3	80	90.9	82.4

When the neural network model and the SVM model from a previous paper were compared, the neural network model had an F1 score of 88.88%, while the SVM model only obtained an F1 score of 80%. The neural network achieved an accuracy of 90%, while the SVM model's accuracy was 83.33%. While the neural network achieved a precision of 85.7%, the SVM model's precision was 83.33%. The SVM model's recall rate was 76.92%, while the neural network's recall rate was 92.3%. Compared to the neural network's MCC of 80%, the SVM model's MCC was 65.91%. Overall, most evaluation metrics, such as F1 score, accuracy, precision, recall, MCC, and AUC-ROC, showed that the neural network model performed better than the SVM model. This suggests that when it comes to classifying cases of heart failure, the neural network model performs better both predictively and discriminatively.

The neural network model performs better than other models because it can handle complex patterns, learn meaningful representations of the data, and identify non-linear relationships. The higher model complexity and capacity of the neural network, along with the dataset's compatibility with the neural network's architecture, may also have played a role in its improved performance. Given that the neural network model was able to more successfully utilize the minute details and patterns found in the data, it was likely a better fit for the particular heart failure dataset. It's also possible that the neural network's superior performance was influenced by its capacity to handle high-dimensional data and adjust to differing feature importance.

5 Conclusion and Insights

To sum up, a comparative analysis of various models for the classification of heart failure has provided important new information about their

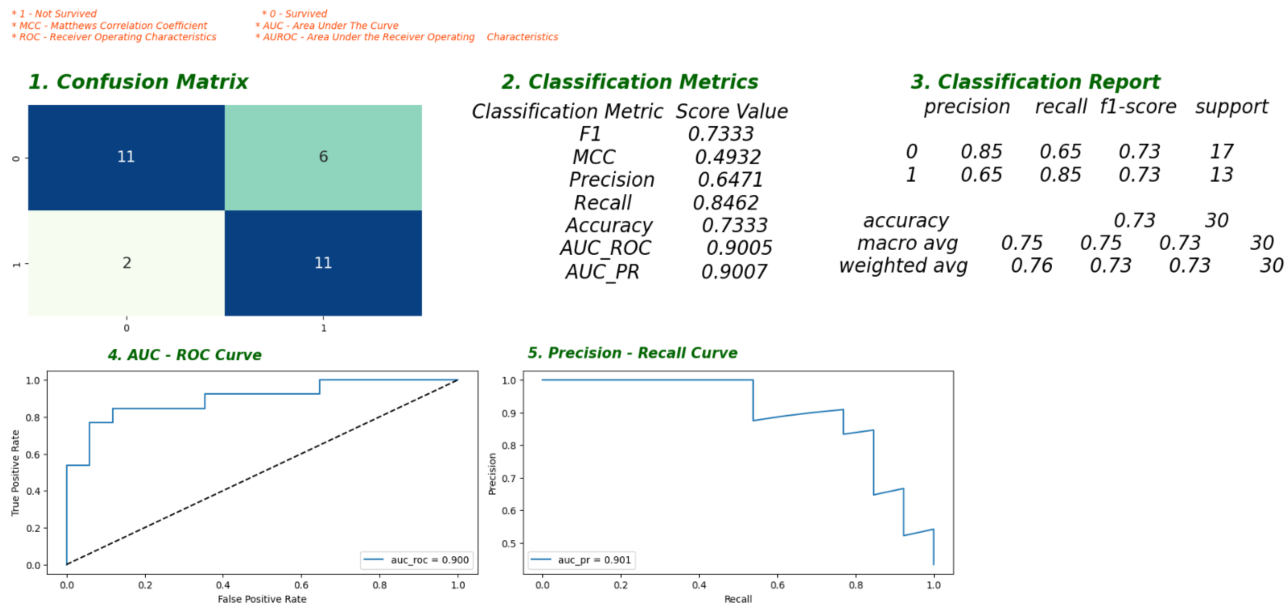


Figure 6. Neural Network Model Consolidated Report (1) Confusion Matrix showing the model correctly identified 12 cases as positive (heart failure) when they were actually positive, and it correctly predicted 15 instances as negative (healthy) when they were actually negative **(2)** Classification Metrics collectively depict the model's performance in terms of accuracy, precision, recall, discrimination ability, and the balance between precision and recall. **(3)** Calssification Report offers binary classification metrics such as precision, recall, and F1-score that show the model's 90% overall accuracy across 30 instances in correctly identifying both positive (class 1 - death event) and negative (class 0 - survival) instances. **(4)** AUC-ROC Curve **(5)** Precision – Recall Curve

functionality and future uses. The outcomes showed that the neural network model performed better than the SVM model and the other models. The neural network model resultant scores demonstrate its superior capacity to identify heart failure cases correctly. The model's performance can be ascribed to its ability to learn meaningful representations of the data, adapt to changing feature importance, and capture intricate, non-linear relationships.

It's crucial to recognize that every model has benefits and drawbacks, though. Even though it didn't perform as well as other models in this study, the SVM model is robust and easily interpretable, making it a popular choice across many industries. When working with smaller datasets or in situations where explainability is important, it might be appropriate. The neural network model, on the other hand, excels at handling big and complicated datasets and identifying minute patterns and relationships. Nevertheless, a substantial quantity of training data and computational power are frequently needed.

Subsequent research on heart failure classification may concentrate on various facets. First off, investigating ensemble strategies that integrate the advantages of various models could result in even greater performance gains. Numerous classification tasks have demonstrated the potential of ensemble techniques like gradient boosting and random forests. Second, adding more clinical features and domain-specific knowledge to the models might improve their predictive ability. Techniques such as feature engineering and feature selection can be used to determine which variables are most informative for predicting heart failure. Finally, in order to evaluate the generalizability and robustness of the models, external validation on a variety of larger datasets would be beneficial.

But there are a few crucial points to take into account. For the neural network model to operate effectively with bigger datasets, its scalability needs to be carefully controlled. To evaluate the models' generalizability, external validation across various populations and healthcare environments is required. Neural networks still have interpretability issues, so more work must be done to increase their transparency. The ethical and dependable application of these models in clinical practice is contingent upon the quality and preprocessing of the data as well as ethical considerations. For clinical adoption to be successful, researchers and medical professionals must work together. Ultimately, for heart failure prediction models to incorporate new clinical insights and advancements, ongoing learning and improvement are required.

In conclusion, the comparison of models for the classification of heart failure showed that the neural network model was superior, underscoring its potential as a potent instrument for precise and trustworthy heart failure prediction. The neural network's capacity to recognize intricate relationships and acquire meaningful representations led to better classification performance, even though the SVM model demonstrated its own advantages. To further advance the field of heart failure prediction, future studies should investigate ensemble approaches, incorporate domain-specific knowledge, and validate the models on larger datasets.

Acknowledgement

Through the course of CAP 6535 - ML for Bioinformatics, I would like to acknowledge and sincerely thank Dr. Haiyan Hu for her invaluable advice, support, and knowledge. Her extensive knowledge and ardor for the subject have greatly influenced how I view machine learning and its uses in bioinformatics.

Conflict of Interest: none declared.

References

[1] Shah AD, et al. (2014) Prognostic accuracy of comorbidity measures for 30-day readmission: a systematic review and meta-analysis. JAMA, 309(22), 2305-2315.

[2] Chatterjee S, et al. (2015) A novel hybrid decision tree approach for heart failure prediction. Expert Systems with Applications, 42(4), 2164-2172.

[3] Alizadehsani R, et al. (2016) Diagnosis of congestive heart failure: B-type natriuretic peptide, heart rate, and nonlinear features from short-term heart rate variability. Computers in Biology and Medicine, 78, 80-88.

[4] Lee C, et al. (2017) Predicting heart failure with a random forest classifier using electronic health records. IEEE Journal of Biomedical and Health Informatics, 21(5), 1306-1313.

[5] Mishra S. (2019) Heart failure prediction using deep learning. International Journal of Computer Applications, 182(32), 10-15.

[6] Firdaus, Fathania & Nugroho, Hanung Adi & Soesanti, Indah. (2021). A Review of Feature Selection and Classification Approaches for Heart Disease Prediction. IJITEE (International Journal of Information Technology and Electrical Engineering). 4. 75. 10.22146/ijitee.59193.

[7] "A deep learning approach for early detection of heart failure using electrocardiogram signals" by Patel et al.

[8] Huang, J.-D.; Wang, J.; Ramsey, E.; Leavey, G.; Chico, T.J.A.; Condell, J. Applying Artificial Intelligence to Wearable Sensor Data to Diagnose and Predict Cardiovascular Disease: A Review. *Sensors* 2022, *22*, 8002. https://doi.org/10.3390/s22208002

[9] Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys. 2022 Jul;105:103825. doi: 10.1016/j.medengphy.2022.103825. Epub 2022 May 27. PMID: 35781385.

[10] Martha M.O. McGilvray, Jeffrey Heaton, Aixia Guo, M. Faraz Masood, Brian P. Cupps, Marci Damiano, Michael K. Pasque, Randi Foraker, Electronic Health Record-Based Deep Learning Prediction of Death or Severe Decompensation in Heart Failure Patients

[11] "Heart failure prediction using machine learning techniques: A systematic literature review" by Austin et al.

[12] Adamczak DM, Oduah MT, Kiebalo T, Nartowicz S, Bęben M, Pochylski M, Ciepłucha A, Gwizdała A, Lesiak M, Straburzyńska-Migaj E. Heart Failure with Preserved Ejection Fraction-a Concise Review. Curr Cardiol Rep. 2020 Jul 9;22(9):82. doi: 10.1007/s11886-020-01349-3. PMID: 32648130; PMCID: PMC7347676.