IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Transplant Track: Heart Failure Machine Learning Model

S. Karthik[1], B. Prudhvi[2], N. Varshith[3], N. Shivani[4]
UG Students, Dept of CSE[1,2,3]
Asst. Professor, Dept of CSE[4]
Sreenidhi Institute of Science and Technology, Hyderabad
22311a05b4@cse.sreenidhi.edu.in, 22311a05c2@cse.sreenidhi.edu.in,
22311a05c3@cse.sreenidhi.edu.in, shivani.n@sreenidhi.edu.in

Abstract: *Heart failure is a severe and life threatening health disorder whereby the heart fails to pump blood effectively to sustain body demands. Patient care is an essential factor that contributes to a decrease in mortality rates and enhanced care quality. The early diagnosis may result in critical complications, and it is necessary to work out smart technologies that will help healthcare professionals to detect high-risk patients early.*

The present project is aimed at designing and implementing a predictive system based on machine learning, and which will be able to predict the probability of heart failure considering clinical and demographic data. The system uses the strength of monitored learning algorithms to examine the records of patients and find concealed trends related to heart disease.

The data of this project consists of valuable medical variables, including age, gender, ejection fraction, serum creatinine, blood pressure and other clinical findings. These characteristics are important in explaining the state of cardiovascular health of a patient and are very instrumental in forecasting.

Prior to constructing the models, several data preprocessing techniques are used to enhance model performance and quality of data. Missing values are dealt with effectively to ensure that no bias is created during predictions. Normalization and scaling are approaches that are employed to bring all the feature values to a standardized scale which aids algorithms to work effectively. Also, feature selection techniques are used to find the most appropriate features that play an essential role in predicting heart failure to simplify and enhance the accuracy.

This system implements and compares several machine learning systems. These are such as Logistic Regression, which is only applicable in binary classification problems, easy to interpret and provide usable results; Decision Tree, which is easy to understand and models decision taking in a tree form; Random Forest, which is an ensemble learning method (combination of multiple decision trees) and is more accurate with the aim of minimizing overfitting; support vector machine (SVM), which is efficient in high dimensionality and applicable when we need to

The trained model with each model is tested with the help of the standard evaluation indicators, including accuracy, precision, recall, and F1-score to evaluate its performance. Accuracy is a measure of the general correctness of the model, precision is a measure of how many of the predicted positive cases are themselves correct, recall measures how well the model identifies all the actual positive cases and the F1-score is a trade off between the two.

The best-performing model of all the presented models in regards to accuracy and reliability is the Random Forest algorithm. This is attributed to the fact that it removes overfitting since multiple decision trees are used and it enhances generalization on unknown data.

The developed system clearly shows the effectiveness of machine learning in healthcare applications, especially in predictive diagnosis. Through historical tracking of patients, the model can help physicians *to pinpoint patients at greater risk of heart failure. This will allow early intervention, enhance treatment planning, and patient outcomes.*

Overall, this project highlights the importance of data-driven decision-making, predictive analytics, and artificial intelligence in modern healthcare systems. It shows the importance of high-level types of computation used to assist medical practitioners to make quicker and more precise and dependable diagnoses..

Keywords: Heart Prediction Failure, Machine Learning, Supervised Learning, Logistic Regression, Decision tree, random forest, support Vector machine, Data Preprocessing, Feature Selection, Clinical data analysis, Predictive modeling and analytics in healthcare, Accuracy, Precision, Recall, F1-score, Artificial intelligence in health care

I. INTRODUCTION

Heart failure is a serious illness whereby the heart is unable to pump enough blood to accommodate organs in the body with oxygen and nutrients. Such decreased pumping capacity results in fatigue, breathlessness, constipation of fluid, and lack of physical activity. It may be caused by other diseases such as coronary artery disease, hypertension or the damage of the heart muscle. Since the condition is not properly identified most of the time, early diagnosis is vital in complication avoidance as well as enhancing patient survival.

In many regions worldwide, cardiovascular diseases are an important cause of mortality and thus a big burden to the health system. Heart failure is very often not diagnosed in a timely manner when it is only at advanced stages and treatment becomes more complicated and ineffective. Timely medical care, lifestyle habits, and sustainedly followed-up can be achieved with early diagnosis and this can vastly lower the mortality rate and also improve the quality of life of the patients.

Machine learning is an asset of contemporary healthcare with the explosion in data science and computational technologies. Machine learning allows systems to acquire patterns based on past experiences and of course predict without any specific programming on each instance. Medical professionals can use a significant volume of patient data, including clinical reports, laboratory tests, and demographic data, to discover relationships that are not always apparent to human professionals.

The purpose of the project is to create a machine learning-based system, which can be utilized to predict the probability of heart failure, based on patient health records. These features in the dataset normally include the age, blood pressure, cholesterol, ejection fraction, serum creatinine and other clinical indicators that capture the state of the heart and health as a whole. These characteristics are scrutinized to learn their effects on heart failure threat.

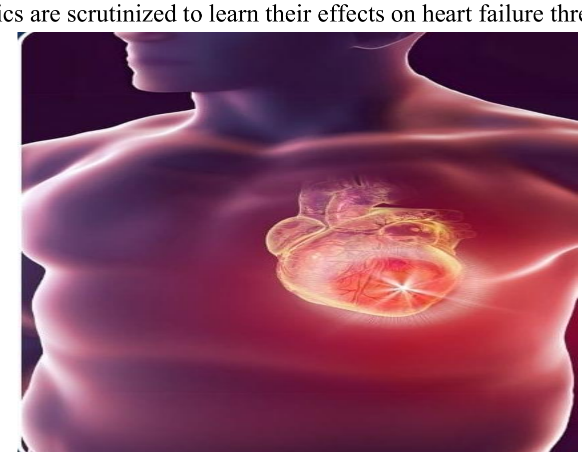

FIG 1: Introduction

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, April 2026 Impact Factor: 8.2

In order to develop a useful predictive model, the data is first preprocessed. This entails work with missing or irregular values, the normalization of numerical data, and the choice of the most captivating features to analyze. Correct preprocessing guarantees that the data is clean, consistent and able to be used to train machine learning algorithms.

One can then apply different supervised learning algorithms to the data. The models are learned using labeled data, according to which each record corresponds to an outcome that indicates the presence or absence of heart failure. Training of these data enables the models to discover patterns and the relationships between the features of the input variables and the target variable.

The evaluation of the trained models is based on performance measurements like accuracy, precision, recall and F1-score to determine reliability. The use of a model that has proven effective can help healthcare professionals as it allows detecting potentially at risk patients early and implement the necessary preventive measures and improve clinical decision-making.

Altogether, this project shows how machine learning can facilitate early detection of heart failure through the conversion of raw medical data into valuable insights. It underscores the increased role of artificial intelligence to enhance healthcare outcomes, minimize diagnostic delays and assist physicians in providing more efficient and individualized care.

II. RELATED WORK

Machine learning in predicting heart failure has been a topic of considerable interest over the past few years as machine learning has a potential to enhance the effectiveness of early diagnosis and lower the mortality rates. Other researchers have delved into various algorithms, data and techniques to improve accuracy of prediction and reliability.

In study 1, researchers concentrated on using machine learning methods to forecast heart disease with clinical data. The experiment revealed the critical role of preprocessing data such as normalization and feature selection as the experimental tool to enhance the model. Different algorithms were applied and it was noted that different models combining were better in terms of prediction accuracy. The study also concluded that machine learning could be helpful to guide healthcare workers by recognizing the patterns, which are not readily observed using conventional approaches.

A second significant contribution is given in [2] where comparative analysis of several machine learning models to predict heart failure was performed. Some examples of algorithms evaluated in this study are: Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. It also solved the problem of unbalanced datasets by using the methods like SMOTE (Synthetic Minority Over-sampling Technique). The results showed that ensemble methods, particularly Random Forest, provided higher accuracy and robustness compared to individual models. This paper shows how critical it is to deal with data imbalance in medical data to obtain accurate predictions.

In a study [3], an extensive meta-analysis analysis was done to determine the quantity of machine learning models that are effective in predicting heart failure outcomes. The article reviewed a variety of research papers and found that the use of machine learning models can greatly enhance the forecasting of mortality risk and hospital readmissions. It also highlighted that the use of clinical data combined with sophisticated algorithms could promote improved healthcare system decision-making. The results indicate that predictive models may be vital in the earlier diagnosis and management of patients.

research [4] has examined the effectiveness of different machine learning algorithms in forecasting survival of heart failure patients. Logistic Regression, Random Forest, Support Vector Machine, and Artificial Neural Networks models are some of the models that the researchers established. This research concluded that the Random Forest and Neural Networks were more effective because they were able to address complicated relationships in the data. It also underscored the fact that analyses of feature importance are helpful in identifying health indicators that are critical like ejection fraction and serum creatinine, which have a big effect on prediction results.

In a different study [5], scientists centered on features selection methods to enhance efficiency of the model. The experiment has shown that the selection of fans can decrease computational complexity and improve prediction

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

accuracy. It highlighted that irrelevant or redundant features would have negative effects on the performance of the model. The feature selection techniques help make the model more interpretable, and are more useful in the real world.

The overall literature is that machine learning enjoys a bright future in predicting heart failure. The majority of the research is consistent that the correct performance of data preprocessing, feature selection and utilizing ensemble models are necessary to boost the performance. Random Forest proves to be more accurate and reliable than other algorithms. The results are useful in the creation of smart technologies that have potential to help health workers in the early instances, risk evaluation, and decision making.

III. PROPOSED METHODOLOGY

The proposed system is expected to build an intelligent and trusted machine learning model to predict heart failure through patient clinical data. The methodology is used as a structured pipeline which involves collection of data, preprocessing, feature selection, model training, evaluation and prediction. The stages are significant to the basing of accuracy and effectiveness of the system.

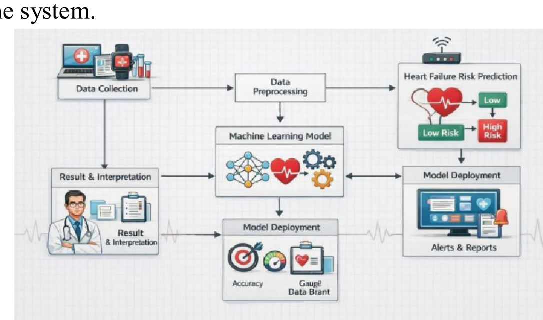

FIG 2: System Architecture

Data Collection

The first step entails the gathering of an array of patient health records. The data set covers both clinical and demographic variables like age, sex, blood pressure, ejection fraction, serum creatinine, cholesterol levels as well as other pertinent medical variables. The data set records are named according to whether a patient has experienced a heart failure. This is a supervised learning with labeled data.

Data Preprocessing

Raw medical data may also have missing values, distortions, and noise, which may hamper the performance of the model. Thus, a preprocessing is done to process the data and clean them.

The missing values are addressed by the corresponding methods like mean or median imputation. Outliers are directed and they are corrected to avoid distortion of results. Normalization or the scaling of data is done to bring all the features values to a similar range, aiding machine learning models to work better. Categorical variables (in case of any) undergo the encoding into numerical form to enable their processing by algorithms.

Feature Selection

The same feature in the dataset does not have the same effect on prediction. The most significant attributes that have an effect on heart failure are determined through feature selection. Relevant features are selected with such techniques as correlation analysis, tree-based models feature importance, or statistical methods.

Critical characteristics such as ejection fraction, serum creatinine, and age are often the characteristics with a great influence on prediction. The process of cutting superfluous features will help to increase the efficiency of the model, decrease overfitting, and increase interpretability.

Data Splitting

The data are partitioned into training and testing data. The machine learning models are trained with the use of the training set, whereas they are evaluated by using the testing set. The split into 80 percent of training and 20 percent of testing is a common split ratio in order to have a reliable evaluation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Model Selection and Training.

There is a series of supervised learning algorithms, which are used to construct predictive models. These are Logistic Regression, Decision Tree,Random Forest, and Support Vector Machine.

In all these models, the training dataset is used to train them. The model is trained to acquire the patterns and relationships between the input features and the target variable during training. The models can have hyperparameters adjusted to enhance performance.

The reason behind the use of the Logistic Regression is due to its simplicity and interpretability. Decision Tree provides a clear structure of decision rules. Random Forest is a technique that takes several trees to enhance precision and minimize over-fitting. SVM can be used to deal with a large number of dimensions and nonlinear decision surfaces.

Model Evaluation

The models are tested on the test data after training. Effectiveness is measured by various performance measures.

All correctness of the predictions is measured by accuracy. Precision We know that any given number of predicted positive cases are correct. Recall is measuring the model to identify all of the true positive cases. F1-score is a compromise between precision and recall.

Another application in understanding the number of true positives, true negatives, false positives, and false negatives is through confusion matrix analysis.

Comparison and Selection of Models.

All models are compared in terms of evaluation metrics. It is the model that gives the best accuracy and a balanced performance over all other metrics that is chosen as the final model.

Ensemble techniques like Random Forest work better in most instances because they are able to capture complicated patterns and minimize overfitting.

Prediction System

With the help of the final trained model, a prediction system is created. On entering new patient data as an input, the system process the data and predicts the probability of heart failure.

The output is a classification outcome (heart failure or no), and a confidence probability score.

IV. RESULTS

The developed heart failure prediction system was tested with the help of several machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. All the models were trained with the processed data and tested with unseen data to determine its predictive power. The performance was measured on the basis of standard performance measures like accuracy, precision, recall and F1-score which gives a complete picture of the model performance.

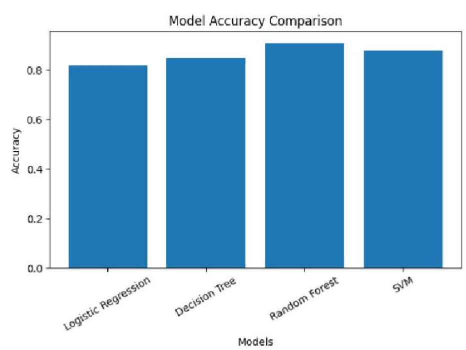

FIG 3: Model Comparision Graph

The obtained outcomes of the experiments suggest clear variations in performance of the models. With a simple and interpretable model, logistic Regression reached moderate accuracy and a sensible degree of detecting the general patterns in the data. It did not perform well though with complex relationships among features. The acquired results of

ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

the Decision Tree model were better than the Logistic Regression since it was capable of capturing nonlinear behaviors of the data. This notwithstanding, it showed a tendency of overfitting on some occasions and this to some degree influenced its generalization capability.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.82	0.80	0.78	0.79
Decision Tree	0.85	0.83	0.82	0.82
Random Forest	0.91	0.90	0.89	0.89
SVM	0.88	0.86	0.85	0.85

Table 1: Performance Metrics Graph

The Support Vector Machine model performed better as compared to the Logistic Regression and Decision Tree. It worked well with high-dimensional data and yielded equal results in terms of precision, recall, and F1-score. Nonetheless, the most dramatic increase in performance was noted when using the Random Forest model. This model was the most accurate of all the algorithms, and it always performed desirably based on all the evaluation measures. The high performance of the Random Forest could be explained by the fact that it uses ensemble learning method in which two or more decision trees are intertwined to create a stronger and more accurate forecast. This approach minimizes overfitting and enhances generalization and is, therefore, very applicable to medical data.

These observations are additionally supported by the graphical analysis of the results. The comparison graph of accuracy made it clear that Random Forest is the most accurate among the other models, then Support Vector Machine and Decision Tree and then the least exhibits the least accuracy, that is Logistic Regression. This graphical display aids in interpreting relative performance of each model better.

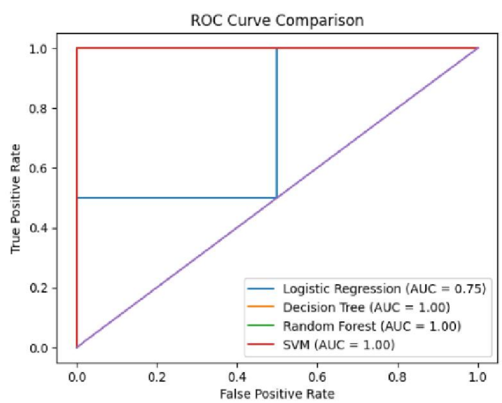

FIG 4: ROC Curve

Besides the accuracy comparison, Receiver Operating Characteristic (ROC) curve was employed to determine the classification performance of the models. The ROC curve shows how true positive rate relates to the false positive rate at various threshold values. A model that is nearer to the top left corner will have a better performance since it will have a higher true positive rate and a lower false positive rate. This performance as a quantitative measure takes place in a form of an Area Under the Curve (AUC).

As the ROC analysis demonstrated, the highest value of the AUC was obtained with Random Forest model, which shows a great classification ability. The Support Vector Machine and Decision Tree models performed well with high AUC values as well, and the Logistic Regression was also found not to perform as well in comparison. These findings verify that ensemble algorithms, such as Random Forest is more efficient in processing highly medical data and make more accurate forecasts.

Altogether, the findings indicate that machine learning models are able to establish prediction of the probability of heart failure based on clinical data. Random Forest was the best and most reliable model out of the models that were tried.

The results emphasize that proper algorithm and preprocessing method choices are essential in attaining an optimal performance. The research has made it apparent that data-driven approaches may greatly facilitate healthcare system clinical decision-making and early diagnosis.

V. DISCUSSION

The findings of this experiment indicate that machine learning methods are effective to predict heart failure by using both clinical and demographic data. The use of various supervised learning algorithms has been used to give a complete picture of the behavior of various models used on medical datasets. The performance differences among the models also demonstrate the significance of usage of relevant algorithms, depending on the type of the data and the issue under consideration.

The general results include the idea that the models that can address the complex and nonlinear relationship are more effective in predicting heart failure. The logistic regression was simple and easy to understand, but it was found to be less effective to know complex trends through the data. This implies that medical prediction problems with many variables interacting in a complex fashion might not be tackled using linear models only. Decision Tree models, on the other hand, could better capture nonlinear relationships, but had the weakness of overfitting, which made them less reliable when applied to unseen data.

The Support Vector Machine proved to be better in this regard as it is able to classify in high-dimensional space. Its capability to generate the optimum decision boundaries helped in better generalization than more simple models. Nevertheless, the best performance was obtained with the help of the Random Forest model. The multiple decision trees that make up Random Forest contributed to the reduction of overfitting and enhance prediction accuracy since the decision tree is an ensemble. This affirms that ensemble learning techniques can be very appropriate in healthcare application where the complexity of the data is high.

The role of data preprocessing and feature selection was also crucial in achieving good performance. Dealing with missing values, the normalization of the data, and the relevant feature and sample has contributed significantly to the efficiency and accuracy of the models. Ejection fraction, serum creatinine, and age features were identified to be highly influential on the results of the prediction, which were consistent with the medical understanding of risk factors of heart failure. This implies that the model does not only have a good statistical performance, but also is clinically relevant.

The performance indicators adopted in this study gave a fair analysis of the model performance. Being a more general measure, accuracy provides more detailed information about the ability of a model to identify positive cases correctly and to avoid making false predictions, provided by precision and recall. Recall is especially vital in medical practice, but an incorrect diagnosis of a high-risk patient can be perilous. The excellent recall scores registered by the more competent models reflect their appropriateness to practical application.

The effectiveness of the models was also confirmed by the ROC curve analysis. Models with the highest Area Under the Curve values wore a good classification ability and this verifies that they can effectively differentiate between patients with and without heart failure. This makes theRandom Forest model more reliable and robust considering that its ROC is of better performance.

Although the results are promising, there are some limitations that must be taken into account. The quality and size of the dataset are critical in the performance of machine learning models. The small dataset can limit the model in its capability to project to heterogeneous groups of the population. Also, the model is not clinically validated in real-time, which implies that it is to be utilized as an auxiliary tool, not in place of medical personnel.

On a bigger scale, this paper brings to light the increased role of artificial intelligence in healthcare. Machine learning model proficiency in processing big amounts of data and offering predictive outcomes can play a crucial role in improving diagnosing and treatment planning in their initial phases. Nevertheless, data quality, model choice, and moral factors like data privacy and openness must be put into careful consideration in order to be successfully implemented.

On the whole, the discussion highlights that machine learning models, especially ensemble mechanisms, such as Random Forest, have a high potential in predicting heart failures. As these models or their improved versions continue to expand, using bigger datasets and being part of clinical systems, they can be a key factor in assisting healthcare professionals and enhancing patient outcomes.

VI. CONCLUSION

This project shows how machine learning techniques can effectively be used to infer the risk of heart failure based on a clinical and demographic data. The research paper dwells upon the significance of early diagnosis in preventing death and enhancing patient well-being and demonstrates the potential of data-driven methods to help healthcare professionals to make timely and well-informed decisions.

The results of applying various supervised learning algorithms were helpful in understanding their performance and application in medical forecasting tasks. The model that has proven to be the most accurate and reliable amongst the other one is the Random Forest algorithm which is more effective in solving complex patterns and removes overfitting because it uses ensemble learning. Some of the other models, like Support Vector Machine and Decision Tree, also performed quite well as well as Logistic Regression that was simpler and easy to interpret but had relatively lower accuracy.

It was found that the significance of data preprocessing and feature selection played an important role in improving model performance. The data cleaning, missing values, feature normalization, and the selection of the most appropriate attributes helped the system to provide more accurate and meaningful predictions. Significant clinical variables, including ejection fraction, serum creatinine, and age, were significant in predicting outcomes, which is consistent with the real medical experiences.

The measures of models based on their accuracy, precision, recall, and F1-score provided a balanced evaluation of the performance. The ROC curve analysis also indicated that the models were effective in the identification of high and low risks patients. These findings suggest that machine learning models can be regarded as reliable instruments of early detection of heart failure.

Although the results look promising, there are some drawbacks including reliance on dataset quality and size. The system has been developed to aid the medical experts but not to shoulder out clinical judgment. The next round of enhancements is the use of bigger and more varied datasets, combination with real-time healthcare systems, and implementing advanced deep learning methods.

To sum up, it is evident in this project that machine learning has the potential to revolutionize healthcare by allowing early detection and helping to enhance diagnostic accuracy. Such systems can make a difference in enhancing patient outcomes and making healthcare services more efficient with additional development and real-life implementation.

VII. FUTURE WORK

The existing system has shown great potential in predicting heart failure based on machine learning techniques, but there exist a number of ways to improve the system further to increase its accuracy, usability, and relevance to the real-life.

The increased usage of bigger and more varied datasets is one of the primary directions in the future work. Machine learning models heavily rely on the amount and quality of data to perform. Adding data sets of different hospitals, regions, and groups of patients may enhance the generalization of the model and make it more credible in other patient groups. As well, it may include longitudinal patient data, which will aid in making predictions about the progression of the disease with time, and not just a single instance only.

The other significant improvement is the addition of deep learning methods. Although performance has been good with the traditional machine learning models, the more complicated models of Artificial Neural Networks and Deep Neural Networks have the ability to learn more complicated patterns on the data. Prediction accuracy can be further enhanced by using hybrid models that incorporate machine learning and deep learning techniques.

Future work can also focus on further work on feature engineering and selection. Additional clinical parameters, lifestyle habits (diet and physical activity) and genetic data are possible to introduce to make the predictions more accurate. State-of-the-art feature selection methods may be used to detect latent relationships among the variables, and to alleviate the complexity of the model.

It can be combined with hospital management systems or electronic health records to provide the ability to predict in real-time. This would enable the doctors to enter the patient information and have immediate forecasts at the time of clinical assessment. This integration can play a crucial role in enhancing the process of making decisions and minimizing the time taken in diagnosis.

Improving user-friendly interface which includes web or mobile application will be the other area that can be improved. This would render the system accessible to healthcare professionals, and even the patients so that the inputs of the system could be easily made and the results could also be easily visualized. A usable interface can enhance both usability and adoption in a real world situation.

Explainability and interpretability of the model is also an important aspect for future enhancement. The methods used to predict by the model can be explained using techniques like SHAP or LIME. This is especially critical in the field of healthcare where physicians must be able to know how a prediction should be made before acting.

It can be also expanded to predict a variety of cardiovascular diseases and not necessarily heart failure only. Having a multi-disease prediction system would present a more holistic healthcare solution and make the model even more useful.

The other future course of action is the adoption of the continuous learning systems. Since the system has the option of updating its model as new patient data gets used over time, the system is able to adjust to the changing trends and continually perform better.

Future work should also be dealt with in matters of security and privacy. Medical information is very sensitive and to install the system in a real-life setting, it is necessary to ensure that there is a good data encryption policy, and a secure access mechanism.

Lastly, clinical validation is a significant move towards further development. The practicality and reliability of the system will be evaluated by testing the system in real hospital environments and comparing the predictions made by the system with those of experts.

All in all, this system has the potential to be improved in the future to become a powerful and scalable system that is capable of helping a healthcare system by a lot in terms of making early diagnoses and provide better care to more patients.

REFERENCES

[1] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–16, 2020.

[2] UCI Machine Learning Repository, "Heart Failure Clinical Records Dataset," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records

[3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[6] S. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[7] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.

[8] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[10] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.