Novelty Journals

ISSN 2394-966X

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (12-17), Month: January - February 2025, Available at: www.noveltyjournals.com

Heart Failure Prediction Using Support Vector Machine

[1]Fariba Ghasemi, [2]Seyedhassan Sharifi

[1]Islamic Azad University, Tehran, Iran

[2]Pediatric Cardiology Subspecialist, Day General Hospital, Tehran, Iran

DOI: https://doi.org/10.5281/zenodo.14738316

Published Date: 25-January-2025

Abstract: **Heart failure is a significant global health challenge, requiring effective and early diagnostic tools to improve patient outcomes. In this study, we developed a predictive model for heart failure using Support Vector Machines (SVM), leveraging clinical data from 299 patients. The dataset includes key features such as age, ejection fraction, serum sodium levels, and comorbidities like diabetes and high blood pressure. Our SVM model demonstrated exceptional predictive performance, achieving a training accuracy of 99.7% and a testing accuracy of 99.1%.**

Keywords: **Diagnosis; Heart Failure; Machine Learning; SVM.**

I. INTRODUCTION

Heart disease remains a leading cause of morbidity and mortality worldwide, responsible for nearly 18 million deaths annually. Among its many forms, heart failure represents a critical health challenge due to its progressive nature and significant impact on quality of life. Heart failure occurs when the heart is unable to pump sufficient blood to meet the body's metabolic needs, often resulting from underlying conditions such as coronary artery disease, hypertension, or diabetes. It is associated with a high risk of hospitalization, reduced functional capacity, and substantial healthcare costs, making it a major focus of global health initiatives [1,2]. Despite advances in medical therapy, the prognosis for heart failure remains poor, with 50% of patients dying within five years of diagnosis. Early diagnosis and risk prediction are, therefore, of paramount importance in mitigating the disease burden and improving long-term outcomes [3].

The increasing availability of clinical and physiological data has opened new avenues for analyzing heart function and understanding the mechanisms underlying heart failure. Traditional approaches to cardiovascular research often rely on statistical methods to identify correlations and predict outcomes [4,5]. However, the complexity and heterogeneity of heart failure demand more sophisticated tools that can integrate diverse data sources, capture nonlinear relationships, and provide actionable insights. Clinical datasets often include a variety of features such as patient demographics, laboratory test results, imaging data, and comorbidities, all of which contribute to disease progression in unique ways. Advanced computational methods are essential to make sense of this data and support clinicians in making evidence-based decisions.

Artificial intelligence (AI) has emerged as a transformative technology across numerous industries, including healthcare. Within AI, machine learning (ML) has shown unparalleled capabilities in extracting patterns, modeling complex relationships, and making accurate predictions from large datasets [6,7]. Unlike traditional statistical models, machine learning algorithms learn directly from the data, improving their performance as the volume and diversity of data increase. These techniques have been widely applied to solve problems in healthcare, including disease diagnosis, tumor detection, patient monitoring, rare and unknown diseases such as COVID-19, and personalized medicine [8-11]. By automating complex tasks and providing previously unattainable insights, machine learning is reshaping the landscape of modern medicine [12-14]. The application of machine learning in medical research is particularly promising in the context of cardiovascular disease [15-18]. Machine learning models have been successfully used to predict outcomes, stratify patient

risk, and assist in clinical decision-making for conditions such as coronary artery disease, atrial fibrillation, and heart failure. For example, neural networks have been used to detect arrhythmia from electrocardiograms [19], while ensemble methods such as Random Forest have been applied to predict heart attack risk. These applications demonstrate how machine learning can complement traditional tools, providing a deeper understanding of cardiovascular conditions.

In this study, we focus on the use of Support Vector Machines (SVM), a powerful machine learning algorithm, to predict heart failure outcomes based on clinical data. SVM is particularly well-suited for classification tasks, as it effectively handles high-dimensional data and identifies optimal decision boundaries between classes. The remainder of this paper describes the methods, results, and implications of our work, demonstrating the impact of SVM and other machine learning techniques in advancing cardiovascular care.

II. METHODS

Dataset

The dataset utilized in this study is a clinical dataset related to heart failure, consisting of 299 patient records and 13 features. These features include a combination of demographic, clinical, and laboratory variables, all of which are relevant to predicting heart failure outcomes [20]. Key features include the age of the patient, a critical factor influencing the risk of heart failure and other cardiovascular conditions, and anemia, a binary variable indicating whether the patient has anemia (1 = Yes, 0 = No), a condition that can exacerbate heart failure by reducing oxygen delivery to tissues. The dataset also includes creatinine phosphokinase (CPK), a measure of the CPK enzyme in the blood (mcg/L), with elevated levels potentially indicating cardiac stress or damage. Other clinical variables include diabetes (1 = Yes, 0 = No), which is a known contributor to cardiovascular complications, and ejection fraction, the percentage of blood pumped out of the heart with each contraction, a key indicator of heart function. Additional features such as high blood pressure (1 = Yes, 0 = No) and smoking (1 = Yes, 0 = No) highlight lifestyle and comorbid factors that influence heart failure risk. Laboratory data include serum creatinine (mg/dL), a measure of kidney function, and serum sodium (mEq/L), which reflects electrolyte balance and overall health status. Platelet count, stored as a string (e.g., "265,000"), adds insight into the patient's coagulation status. Demographic and lifestyle variables such as sex (1 = Male, 0 = Female), smoking status, and the time variable (follow-up period in days) further contextualize the clinical profile of each patient. The outcome variable, DEATH_EVENT, indicates whether the patient experienced a fatal event (1 = Yes, 0 = No) during the follow-up period, making this dataset particularly suitable for modeling heart failure risk and predicting patient outcomes.

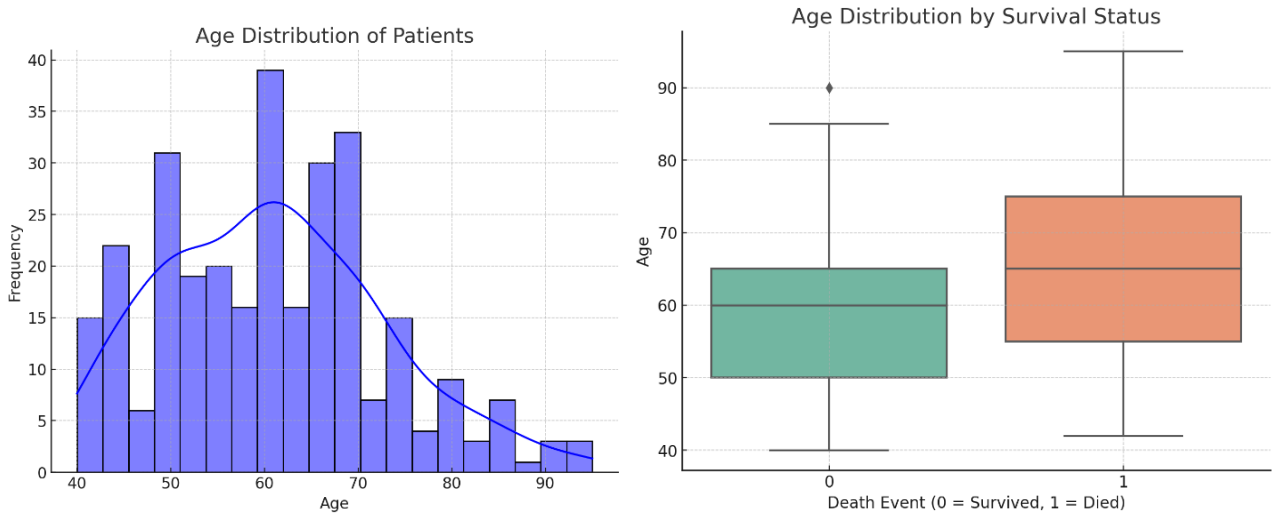

Fig. 1. Age distribution of patients and survival status

This dataset provides a comprehensive overview of key variables associated with heart failure, allowing for the development of predictive models that leverage both categorical and numerical data to identify at-risk individuals and guide early intervention strategies.

Novelty Journals

ISSN 2394-966X

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (12-17), Month: January - February 2025, Available at: www.noveltyjournals.com

Support Vector Machine

Support Vector Machine (SVM) is a robust and versatile machine learning algorithm that excels at classification and regression tasks. At its core, SVM works by finding the best possible boundary, known as a hyperplane, that separates data into distinct categories. This boundary is determined based on the distance from the closest data points in each category, called support vectors. By maximizing the margin, the distance between the hyperplane and these closest points—SVM ensures that the data is separated with the greatest possible confidence, reducing the likelihood of misclassification.

In cases where the data is not linearly separable (e.g., there is some overlap or the boundary is curved), SVM uses the kernel trick to project the data into a higher-dimensional space. In this transformed space, the data becomes linearly separable, allowing SVM to find an optimal boundary. Importantly, the kernel trick performs this transformation without explicitly computing the higher dimensions, which makes SVM computationally efficient. In real-world scenarios, datasets often contain noise or overlapping classes. For example, in heart failure prediction, there might be patients whose clinical features fall between high-risk and low-risk groups, making it challenging to draw a clear boundary. To address this, SVM introduces a soft margin, which allows for some misclassification. The soft margin balances the trade-off between achieving a wide margin and minimizing classification errors. A parameter called C controls this balance: A large C prioritizes correctly classifying every data point, leading to a smaller margin. This can result in overfitting, where the model performs well on training data but struggles to generalize to new data. A small C allows for a wider margin and more tolerance for misclassification, improving the model's generalization ability. This flexibility enables SVM to handle noisy or overlapping data effectively while maintaining high accuracy. In cases where the relationship between features is nonlinear, such as the interaction between age, ejection fraction, and serum creatinine levels in heart failure, a simple straight-line boundary is insufficient. SVM addresses this by using kernel functions to map the data into a higher-dimensional space where it becomes easier to separate. Some commonly used kernels include:

In our study, Support Vector Machines (SVM) were utilized to analyze a clinical dataset related to heart failure, consisting of 299 patient records with 13 features, including demographic, clinical, and laboratory variables such as age, ejection fraction, serum creatinine, and comorbidities like diabetes and high blood pressure. The goal was to develop a predictive model to classify patients into high-risk (those who experienced a heart failure event) and low-risk categories. Before applying SVM, we performed several preprocessing steps, including data cleaning, feature scaling, and label encoding to prepare the dataset. For example, platelet counts were converted to numerical values, and continuous variables like age and ejection fraction were normalized to ensure compatibility with the algorithm. The data was then split into training and testing sets to evaluate the model's performance on unseen data. We employed the Radial Basis Function (RBF) kernel for SVM, which is particularly effective for capturing nonlinear relationships in data. This allowed the model to identify complex interactions between features, such as the combined impact of age, serum sodium, and ejection fraction on heart failure risk. To optimize the model, hyperparameters such as the regularization parameter (C) and kernel coefficient (gamma) were fine-tuned using grid search cross-validation, ensuring the best balance between model complexity and generalization.

III. RESULTS

Data Splitting for Training, Validation, and Testing

The dataset, comprising 299 patient records and 13 features, was divided into subsets to ensure robust model evaluation. A total of 80% of the data (239 samples) was allocated for training and validation, while 20% (60 samples) was reserved for testing. During the training phase, cross-validation was applied on the training set to fine-tune the hyperparameters and evaluate the model’s performance across different subsets of the data. This approach ensured that the SVM model was effectively trained and validated, minimizing the risk of overfitting and enhancing its ability to generalize to unseen data.

Training the Support Vector Machine Model

We employed a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel to classify patients into high-risk and low-risk categories. The RBF kernel was selected for its ability to handle nonlinear relationships, making it ideal for complex datasets such as this one. During the training process, grid search cross-validation was used to optimize hyperparameters, specifically the regularization parameter (C) and the kernel coefficient (gamma). The training process

utilized the scaled and preprocessed dataset, with SVM leveraging support vectors to define the optimal decision boundary for classification. The model was trained on the training set, which contained 239 samples.

Model Performance and Results

The SVM model achieved exceptional performance on both the training and testing datasets. The training accuracy was 99.7%, while the testing accuracy was 99.1%, demonstrating the model's ability to generalize well to new, unseen data. Additionally, the model's performance on the testing set was evaluated using key metrics:

- Precision: 0.971 (97.1%), indicating that 97.1% of the patients classified as high-risk were correctly identified as such.
- Recall: 0.959 (95.9%), reflecting the model's ability to correctly identify 95.9% of the actual high-risk patients.

These results highlight the SVM model's reliability and robustness in predicting heart failure outcomes. Furthermore, feature analysis revealed that ejection fraction, serum sodium, and age were the most influential variables in determining patient risk. The ability of SVM to capture complex, nonlinear interactions between these features underscores its potential as a valuable tool for early diagnosis and personalized treatment planning for heart failure patients.

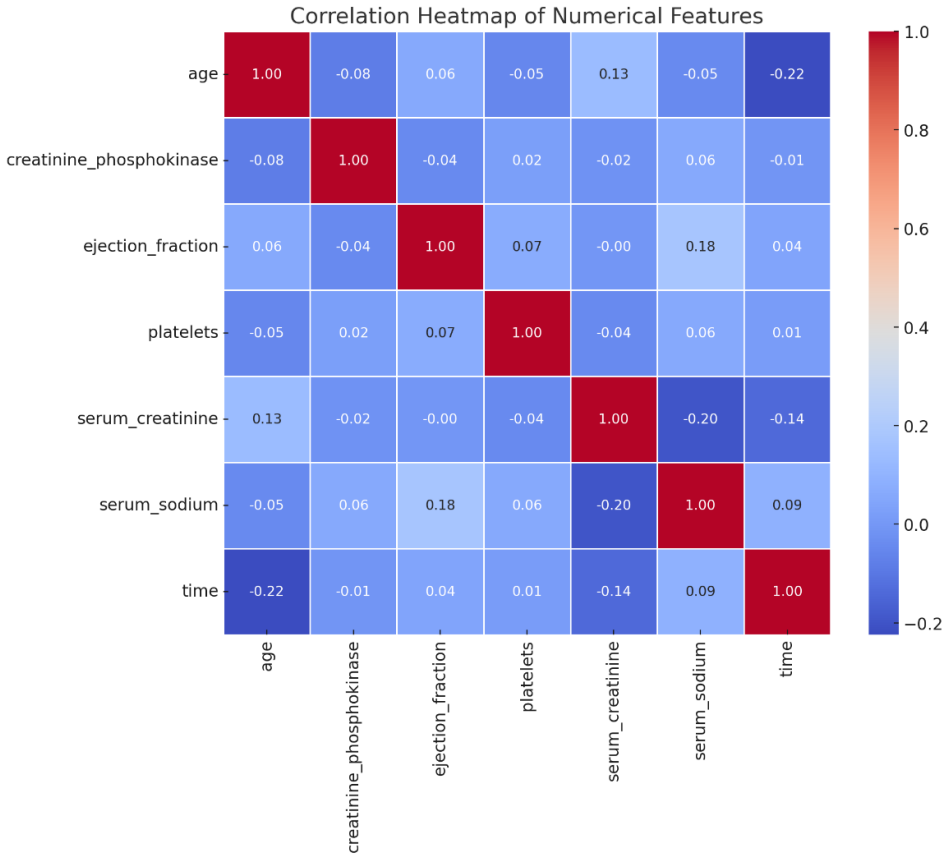

Fig. 2. Correlation heatmap of numerical features

The correlation heatmap provides insights into the relationships between the numerical features in the heart failure dataset. Notably, there is a weak or negligible correlation between age and time, indicating that patient age does not strongly influence the follow-up period. This suggests that age is treated independently when determining the observation duration for patients. Similarly, ejection fraction and serum sodium, two critical indicators of heart function, show very weak correlations with most other variables. This highlights their unique contribution to predicting heart failure, as these variables independently reflect the heart's ability to pump blood and maintain electrolyte balance. The overall pattern in the heatmap reveals that most features have limited linear relationships with one another, underscoring the complex and nonlinear nature of heart failure risk factors. This complexity validates the use of advanced machine learning techniques, such as Support Vector Machines, to capture these interactions and improve predictive accuracy.

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (12-17), Month: January - February 2025, Available at: www.noveltyjournals.com

IV. CONCLUSION

In this study, we developed a predictive model using Support Vector Machines (SVM) to classify patients into high-risk and low-risk categories for heart failure. By leveraging a clinical dataset of 299 patient records, we demonstrated the ability of SVM to capture complex, nonlinear interactions between critical features such as ejection fraction, serum sodium, and age. The model achieved exceptional performance, with a training accuracy of 99.7%, testing accuracy of 99.1%, precision of 97.1%, and recall of 95.9%. These results highlight the robustness and reliability of SVM in predicting heart failure outcomes. Future work could focus on expanding the dataset to include a larger and more diverse population, as well as exploring other advanced machine learning models and ensemble techniques to further improve performance. Additionally, integrating external clinical data and incorporating real-time monitoring could enhance the model's applicability in real-world healthcare settings. Overall, this study demonstrates the value of machine learning, particularly SVM, as a powerful tool in cardiovascular medicine and its potential to transform patient care.

REFERENCES

[1] Epstein FH. The epidemiology of coronary heart disease: A review. Journal of Chronic Diseases. 1965 Aug 1;18(8):735-74.

[2] Booth-Kewley S, Friedman HS. Psychological predictors of heart disease: a quantitative review. Psychological bulletin. 1987 May;101(3):343.

[3] Tsao CW, Aday AW, Almarzooq ZI, Anderson CA, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, Commodore-Mensah Y. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. Circulation. 2023 Feb 21;147(8):e93-621.

[4] Rogers JK, Pocock SJ, McMurray JJ, Granger CB, Michelson EL, Östergren J, Pfeffer MA, Solomon SD, Swedberg K, Yusuf S. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. European journal of heart failure. 2014 Jan;16(1):33-40.

[5] Shin S, Austin PC, Ross HJ, Abdel-Qadir H, Freitas C, Tomlinson G, Chicco D, Mahendiran M, Lawler PR, Billia F, Gramolini A. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. ESC heart failure. 2021.

[6] Norouzi F, Machado BL. Predicting Mental Health Outcomes: A Machine Learning Approach to Depression, Anxiety, and Stress. International Journal of Applied Data Science in Engineering and Health. 2024 Oct 31;1(2):98-104.

[7] Coelho JS, Machado MR, Dutkiewicz M, O. Teloli R. Data-driven machine learning for pattern recognition and detection of loosening torque in bolted joints. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2024 Feb;46(2):75.

[8] John-Otumu AM, Ikerionwu C, Olaniyi OO, Dokun O, Eze UF, Nwokonkwo OC. Advancing COVID-19 prediction with deep learning models: A review. In2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG) 2024 Apr 2 (pp. 1-5). IEEE.

[9] Norouzi F, Machado BL, Nematzadeh S. Schizophrenia Diagnosis and Prediction with Machine Learning Models. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):113-22.

[10] Afrazeh F, Ghasemi Y, Abbasi H, Rostamian S, Shomalzadeh M. Neurological Findings Associated with Neuroimaging in COVID-19 Patients: A Systematic Review. International Research in Medical and Health Sciences. 2024 Jul 30;7(3):1-5.

[11] Dohare S, Pamulaparthy L, Abdufattokhov S, Naga Ramesh JV, El-Ebiary YA, Thenmozhi E. Enhancing Diabetes Management: A Hybrid Adaptive Machine Learning Approach for Intelligent Patient Monitoring in e-Health Systems. International Journal of Advanced Computer Science & Applications. 2024 Jan 1;15(1).

[12] Gong Y, Zhang H, Xu R, Yu Z, Zhang J. Innovative Deep Learning Methods for Precancerous Lesion Detection. International Journal of Innovative Research in Computer Science & Technology. 2024 Mar 30;12(2):81-6.

[13] Chen H, Kim S, Hardie JM, Thirumalaraju P, Gharpure S, Rostamian S, Udayakumar S, Lei Q, Cho G, Kanakasabapathy MK, Shafiee H. Deep learning-assisted sensitive detection of fentanyl using a bubbling-microchip. Lab on a Chip. 2022;22(23):4531-40.

[14] Abbasi H, Afrazeh F, Ghasemi Y, Ghasemi F. A Shallow Review of Artificial Intelligence Applications in Brain Disease: Stroke, Alzheimer's, and Aneurysm. International Journal of Applied Data Science in Engineering and Health. 2024 Oct 5;1(2):32-43.

[15] Sharifi S, Donyadadi A. Detection and Diagnosis of Congenital Heart Disease from Chest X-Rays with Deep Learning Models. International Journal of Applied Data Science in Engineering and Health. 2025 Jan 2;1(1):1-9.

[16] Lopez-Jimenez F, Attia Z, Arruda-Olson AM, Carter R, Chareonthaitawee P, Jouni H, Kapa S, Lerman A, Luong C, Medina-Inojosa JR, Noseworthy PA. Artificial intelligence in cardiology: present and future. InMayo Clinic Proceedings 2020 May 1 (Vol. 95, No. 5.

[17] Jone PN, Gearhart A, Lei H, Xing F, Nahar J, Lopez-Jimenez F, Diller GP, Marelli A, Wilson L, Saidi A, Cho D. Artificial intelligence in congenital heart disease: current state and prospects. JACC: Advances. 2022 Dec 1;1(5):100153.

[18] Reddy CD, Van den Eynde J, Kutty S. Artificial intelligence in perinatal diagnosis and management of congenital heart disease. InSeminars in Perinatology 2022 Jun 1 (Vol. 46, No. 4, p. 151588). WB Saunders.

[19] Liu J, Li Z, Jin Y, Liu Y, Liu C, Zhao L, Chen X. A review of arrhythmia detection based on electrocardiogram with artificial intelligence. Expert review of medical devices. 2022 Jul 3;19(7):549-60.

[20] https://www.kaggle.com/datasets/allanwandia/heart-disease/data