

CARDIOCARE: MACHINE LEARNING FOR HEALTH CARE



22UIT404- A MINI PROJECT I REPORT

Submitted by

MADHU MITHA S	(710723205037)
PREETHA J	(710723205045)
SUJIKA S	(710723205058)
THIRUNAAVUKKARASU S	(710723205060)
BAARGAVAN A	(710723205302)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

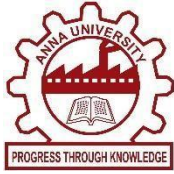
INFORMATION TECHNOLOGY

Dr. N.G.P. INSTITUTE OF TECHNOLOGY COIMBATORE- 641 048

(AN AUTONOMOUS INSTITUTION)

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2025



CARDIOCARE: MACHINE LEARNING FOR HEALTH CARE



22UIT404- A MINI PROJECT I REPORT

Submitted by

MADHU MITHA S	(710723205037)
PREETHA J	(710723205045)
SUJIKA S	(710723205058)
THIRUNAAVUKKARASU S	(710723205060)
BAARGAVAN A	(710723205302)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

Dr. N.G.P. INSTITUTE OF TECHNOLOGY COIMBATORE- 641 048

(AN AUTONOMOUS INSTITUTION)

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2025

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this literature survey report “**CardioCare: Machine Learning for Health Care**” is the bonafide work of “**MADHU MITHA S (710723205037), PREETHA J (710723205045), SUJICA S (710723205058), THIRUNAAVUKKARASU S (710723205060), BAARGAVAN A (710723205302)**” who carried out the literature work under my supervision.

SIGNATURE

**Dr.B.HEMALATHA M.E.,Ph.D.,
SUPERVISOR**

Associate Professor
Department of Information Technology,
Dr. N. G. P Institute of Technology,
Kalapatti Road,
Coimbatore-641 048.

SIGNATURE

**Dr.M. KRISHNAMOORTHY M.E., Ph.D.,
HEAD OF THE DEPARTMENT**

Professor & Head
Department of Information Technology,
Dr. N. G. P Institute of Technology,
Kalapatti Road,
Coimbatore-641 048.

Submitted for the End Semester Mini Project Viva-Voce held on _____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We extend our heartiest thanks to **Dr. NALLA G. PALANISWAMI**, Chairman, KMCH & Dr. N.G.P. Educational Institutions for providing us the necessary infrastructure to do our project work.

We express our gratitude to **Dr. THAVAMANI D. PALANISWAMI**, Secretary, Dr.N.G.P. Institute of Technology, for providing us the facilities to do our project work.

We would like to express our hearty thanks and gratitude to **Dr. S.U. PRABHA M.E.,Ph.D.**, Principal, Dr. N.G.P. Institute of Technology, for her earnest encouragement.

We extend our deep sense of gratitude to **Dr. M. KRISHNAMOORTHY M.E., Ph.D.**, Head of the Department, Department of Information Technology, for his valuable guidance and constructive suggestion at all stages of our project from inception to completion.

We express our hearty thanks to our project guide, **Dr.B.HEMALATHA M.E.,Ph.D.**, Associate Professor, Department of Information Technology, for her valuable guidance and timely help for completing our project.

We express our sincere thanks to our project coordinators **Mr.T.VIJAYA KUMAR** Assistant Professor, Department of Information Technology, for their support in developing our project.

We would also like to express our gratitude to the faculty members of Department of Information Technology and also to our family for their kind patronage.

ABSTRACT

Heart disease remains one of the leading causes of death in the world, and hence, there is a need for dependable and timely prediction methods. The project was commissioned to compare and evaluate the performance of some machine learning algorithms including Support Vector Machines (SVM) , Gradient Boost , KNN and a hybrid SVM optimized with the Jellyfish Optimization Algorithm (SVM-JFO) in predicting heart disease. With a validated clinical dataset, extensive preprocessing was carried out, and model training and evaluation were implemented using key metrics such as accuracy, precision, recall, and F1-score. The final outcomes confirmed that the hybrid SVM-JFO model scored much higher than traditional models, with the highest accuracy and showing superior classification capability in all the metrics explored. These findings illustrate the importance of incorporating nature-inspired optimization algorithms like JFO to enhance machine learning models in the healthcare sector. The hybrid approach of SVM with JFO delivers impressive results, achieving accuracy 95%, recall 98%, F1-score 96%, precision 93%. The improvement is attributed to JFO optimizing hyperparameters like C and gamma, enabling the SVM to find the best decision boundary.

KEYWORDS: Heart Disease Prediction, Support Vector Machine, Gradient Boosting, Jellyfish Optimization, Machine Learning.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	LIST OF ABBREVIATIONS	v
	LIST OF FIGURES	vi
	LIST OF TABLES	vii
1	INTRODUCTION	1
	1.1 About The Project	1
	1.2 Objectives	1
	1.3 Problem Statement	1
	1.4 Scope and Limitations of the Study	1
	1.5 Organization of the Report	2
2	LITERATURE SURVEY	3
	2.1 Introduction	3
	2.2 Review Of Literature Survey	3
	2.2.1 Heart disease prediction using hybrid machine learning: a brief review	3
	2.2.2 Effective feature engineering technique for heart disease prediction with machine learning	4
	2.2.3 Deep learning-based method for detecting parkinson using 1d convolutional neural networks and improved jellyfish algorithms	4
	2.2.4 Machine learning algorithms for predicting hospital readmission and mortality rates in patients with heart failure.	4

2.2.5 Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis.	5
2.2.6 AI-assisted medical imaging and heart disease diagnosis: a deep learning approach for automated analysis and enhanced prediction using ensemble classifiers.	5
2.2.7 Prediction of mortality in young adults with cardiovascular disease using artificial intelligence.	6
2.2.8 Advancing heart disease prediction through machine learning: techniques and insights for improved cardiovascular health.	6
2.2.9 Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison	6
2.2.10 A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis.	7
2.2.11 Forecasting the risk of coronary heart diseases using machine learning algorithms.	7
2.2.12 Study and analysis on detection, classification, and prediction of cardiac arrhythmia using soft computing tool.	7
2.2.13 Hbwo-js: jellyfish search boosted hybrid beluga whale optimization algorithm for engineering applications.	8
2.2.14 Machine learning technology-based heart disease detection models.	8

	2.2.15 Predictive analysis of heart diseases with machine learning approaches	8
	2.4 Summary	18
3	SYSTEM SPECIFICATION	19
	3.1 Introduction	19
	3.2 Hardware Requirements	19
	3.3 Software Requirements	19
	3.4 Software Description	19
4	PROPOSED SYSTEM	21
	4.1 Existing System	21
	4.2 Drawbacks of Existing System	21
	4.3 Proposed Model	22
	4.4 Module Description	24
	4.4.1 Dataset Description	24
	4.4.2 Data Preprocessing	24
	4.4.3 Feature Selection	25
	4.4.4 Support Vector Machine	25
	4.4.5 Jellyfish Optimization	26
	4.4.6 JFO Optimization	26
	4.5 Advantages of Proposed Model	27
5	IMPLEMENTATION AND RESULT	28
	5.1 Overview	28
	5.2 Model Performance and Output	28
6	CONCLUSION AND FUTURE WORK	31
	APPENDIX	32
	SOURCE CODE	32
	REFERENCES	35

LIST OF ABBREVIATIONS

Best of Abbreviations

AI	Artificial Intelligence
SVM	Support Vector Machine
JFO	Jellyfish Optimization Algorithm
SMOTE	Synthetic Minority Over-sampling Technique
ML	Machine Learning
HER	Electronic Health Record
ECG	Electrocardiogram
CHD	Coronary Heart Disease
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic - Area Under Curve
PCA	Principal Component Analysis
DWT	Discrete Wavelet Transform
KNN	K-Nearest Neighbors
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
ML Model	Machine Learning Model

LIST OF FIGURES

FIGURE NO	LIST OF FIGURES	PAGE NO
4.1	Flow Chart for Proposed Model	23
5.1	Performance Result Graph	30

LIST OF TABLES

TABLE NO	TABLE NAME	PAGE NO
2.1	Inferences From Literature Survey	9
5.1	Selected Feature and Cross Validation Scores	29
5.2	Performance metrics of the Proposed SVM with JFO model	30

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Cardiovascular disease (CVDs), and particularly heart disease, are two of the most frequent causes of mortality worldwide. Precise prediction and early detection of heart disease are critical in effective treatment and improving patient outcome. This project focuses on predicting heart disease based on clinical and physiological data using machine learning (ML) models. Traditional ML algorithms such as Support Vector Machines (SVM), Naive Bayes, and Decision Trees were contrasted with a hybrid approach: Support Vector Machine optimized by the Jellyfish Optimization Algorithm (SVM-JFO). The study explores the application of optimization techniques to enhance the predictive power of machine learning models, with the medical data complexity of patterns of heart disease in mind.

1.2 PROBLEM STATEMENT

Heart disease is usually diagnosed late, and this results in unfavorable outcomes. Conventional approaches are labor-intensive and rely on clinical knowledge. Machine learning has promise but may have problems with accuracy and overfitting. This project solves these problems by testing models and employing optimization to enhance predictions.

1.3 OBJECTIVES

- Compare SVM, Gradient Boost, KNN and SVM-JFO models
- Use preprocessing to improve data quality
- Evaluate models using accuracy, precision, recall, and F1-score
- Show how optimization enhances ML performance in heart disease prediction

1.4 SCOPE AND LIMITATIONS OF THE STUDY

Scope

- Uses clinical heart disease dataset

- Implements and compares multiple ML models
- Introduces JFO for optimizing SVM

Limitations

- Limited to structured data
- Results depend on dataset quality
- Higher complexity in hybrid models
- Needs testing on more diverse datasets

1.5 ORGANIZATION OF THE REPORT

This report is organized into several chapters to present a clear and logical understanding of the project. Chapter 1 introduces the project, outlining the background, problem statement, objectives, scope, and organization of the report itself. Chapter 2 presents a literature survey, reviewing related research works and identifying the existing gaps. Chapter 3 explains the system requirements, including both hardware and software components, and their detailed descriptions. Chapter 4 discusses the proposed methodology, system architecture, and module-wise functionality. Chapter 5 focuses on the system implementation and how different components interact to monitor sewage conditions in real time. Chapter 6 provides the results and a discussion of the system's performance and accuracy. Finally, Chapter 7 concludes the report with a summary of outcomes and suggests possible future enhancements. Additional sections include references and appendices that provide supporting documents and source code.

CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

This literature survey explores the application of machine learning techniques in heart disease prediction, focusing on algorithms like Support Vector Machines (SVM), Naive Bayes, and Decision Trees. It examines how these models, along with optimization techniques like Jellyfish Optimization (JFO), have been used to improve prediction accuracy. Previous research highlights the strengths and weaknesses of these models, such as their performance, interpretability, and computational efficiency. Hybrid approaches, like combining SVM with JFO, aim to optimize hyperparameters and enhance predictions. The review also identifies gaps in current studies, such as the need for more interpretable models and better handling of imbalanced data, which will inform the methodology of this study.

2.2 REVIEW OF LITERATURE SURVEY

2.2.1 Heart disease prediction using hybrid machine learning: a brief review

Ahmed et.al [2] explores various hybrid machine learning techniques used to predict heart disease, combining algorithms such as decision trees, ensemble methods (like gradient boosting and random forests), and clustering techniques like K-means. These methods are integrated to capitalize on their unique strengths, improving prediction accuracy and robustness. The study focuses on analyzing extensive datasets, including patient demographics, medical history, and diagnostic test results, to develop effective predictive models. While hybrid approaches offer advantages such as enhanced accuracy, better handling of complex datasets, and improved risk stratification for personalized treatment, the paper also notes challenges, including the potential for overfitting, the black-box nature of some algorithms limiting interpretability, and the need for large, high-quality datasets. Overall, the paper underscores the significance of these advanced techniques in improving cardiovascular health outcomes and supporting informed clinical decision-making.

2.2.2 Effective feature engineering technique for heart disease prediction with machine learning

Babu et.al [4] The study evaluates algorithms including Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), and Logistic Regression (LR) in order to enhance heart disease prediction utilizing patient health data. Principal Component Heart Failure (PCHF), a unique feature engineering technique, was presented in order to improve model accuracy and optimize the feature set. Preprocessing, standardization, and hyperparameter tuning were all included in the study, and measures like sensitivity, specificity, F-measure, and classification accuracy were used to assess performance. 86% accuracy was attained with the Random Forest model, and robustness was guaranteed by K-fold cross-validation. Although the methodology has enhanced predictive capabilities and high accuracy, some practitioners may find it inaccessible due to its computing demands and complexity. Overall, the study shows how machine learning can be used for early heart disease prediction.

2.2.3 Deep learning-based method for detecting parkinson using 1d convolutional neural networks and improved jellyfish algorithms

Shankar et.al [13] study presents a novel approach to Parkinson's disease classification using voice recordings. It is based on 62 samples from both healthy controls and Parkinson's patients. The method makes use of a 1D Convolutional Neural Network (1D-CNN), which has been tuned for feature selection using Maximum Relevance Minimum Redundancy (mRMR) and hyperparameter tuning using the Improved Jellyfish Algorithm (IJFA). This outperforms current models (87.27%-95.09%) with an outstanding accuracy of 88.6%. Robust hyper-parameter optimization, efficient feature selection, and excellent accuracy are among the main benefits. A short dataset, the possibility of overfitting, and the difficulty of combining deep learning with optimization methods are some drawbacks, though. Notwithstanding these difficulties, the approach has potential for enhancing the diagnosis of Parkinson's disease.

2.2.4 Machine learning algorithms for predicting hospital readmission and mortality rates in patients with heart failure.

Rizinde et.al[12] used a quantitative meta-analysis technique in their systematic review to evaluate the bias and quality of the research using the CHARMS checklist. The study assessed a number of machine learning techniques, such as support vector machines (SVM), logistic regression, decision trees, random forests, and K-nearest neighbors (KNN). The results showed that accuracy ranged from 0.71 to 0.93, while precision, recall, and F1-scores ranged from 0.75 to

0.96. With random forests reaching an AUC of up to 0.92, these models demonstrated potential in predicting hospital readmissions and mortality for patients with heart failure. Class disparity, non-standardized health data, and distrust from medical practitioners in LMICs are some of the obstacles that could prevent these predictive technologies from being widely used.

2.2.5 Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis.

Liu et.al [8] evaluated machine learning (ML) algorithms for predicting the risk of cardiovascular disease (CVD), particularly in asymptomatic people. With hyperparameter tuning via grid search, random search, and Bayesian optimization, the study used both data-driven feature selection methods, such as Lasso and random forests, and manual feature selection based on expert knowledge. For model discrimination, the c-statistic/AUROC was the main performance indicator; however, comparisons were made more difficult by the variety of calibration techniques. Beyond traditional risk indicators, machine learning models have the benefit of combining a variety of data sources, which could improve predicted accuracy. Nevertheless, issues with low clinical applicability, methodological variability, significant bias risk, and lack of external validation contrast with well-known risk scores like QRISK and PCE, raising questions about their practicality and the need for better transparency and validation.

2.2.6 AI-assisted medical imaging and heart disease diagnosis: a deep learning approach for automated analysis and enhanced prediction using ensemble classifiers.

Vara et.al[14] described machine learning (ML) approaches for predicting the risk of cardiovascular disease (CVD), emphasize deep learning techniques like CNNs and RNNs as well as algorithms like logistic regression, support vector machines (SVM), and random forests. These algorithms frequently exceed conventional risk assessments by analyzing patient data from wearable technology and electronic health records (EHRs) to reveal complex connections. However, there are obstacles that can restrict clinical trust and generalizability, such as lack of external validation, poor model interpretability, and problems with data quality. Comparing models is made more difficult by variations in feature engineering and selection between research. To guarantee the clinical efficacy of machine learning in CVD risk assessment, the paper recommends better techniques and validation.

2.2.7 Prediction of mortality in young adults with cardiovascular disease using artificial intelligence.

Ahmad et.al[1] developed a predictive model for mortality in young adults (ages 18-45) with cardiovascular disease (CVD) using a CHAID algorithm, which achieved an accuracy of 88.27% and an AUC of 0.79. The CHAID model identified key predictors for mortality, starting with pulse pressure, followed by diastolic blood pressure, ischemic heart disease (IHD), and geographical location. Patients with pulse pressure above 73 had the highest death rate (8.64%). The study highlighted the importance of using AI for early risk prediction in younger populations and emphasized CHAID's effectiveness in analyzing critical factors related to mortality outcomes in CVD patients.

2.2.8 Advancing heart disease prediction through machine learning: techniques and insights for improved cardiovascular health.

Bhowmik et.al[5] used the Cleveland dataset from the UCI Machine Learning Repository, containing 70,000 patient records and 12 features, to predict heart disease using Logistic Regression, Random Forest, and Support Vector Machines (SVM). Logistic Regression was the most effective model overall, achieving the highest ROC-AUC score of 0.7810, which balanced true positives and false positives better than the others. While SVM showed the highest accuracy (72.33%), its slightly lower ROC-AUC (0.7789) placed it second. Random Forest performed the least effectively for this dataset. The study emphasizes the potential of simple models like Logistic Regression for early heart disease detection, offering reliable predictions while balancing precision and recall.

2.2.9 Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison.

Hossain et al[6] used data from hospitals in Bangladesh to predict heart disease with machine learning methods like Logistic Regression, Naïve Bayes, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). Random Forest performed the best, with 80% accuracy, 80.91% precision, 90% recall, and an F1-score of 80.91%, especially when using selected important features. The study emphasized that choosing the right features improves prediction accuracy. However, some methods, like Naïve Bayes, didn't perform as well, partly due to the small dataset and overlapping features. This research aims to help doctors detect heart disease early.

2.2.10 A Systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis.

Alkayyali et.al[3] In a Systematic Literature Review of Deep and Machine Learning Algorithms in Cardiovascular Illness Diagnosis by Zakaria K.D. Alkayyali et al. analyzes 37 publications to evaluate the contribution of deep learning (DL) and machine learning (ML) to the diagnosis of cardiovascular illnesses (CVD). The study uses specifications like accuracy, sensitivity, and precision to evaluate the effectiveness of important supervised algorithms for CVD prediction, including CNN, SVM, KNN, and RNN. It highlights the necessity of improved ensemble methodologies, bigger datasets, and modern DL algorithms like ResNet and Xception. The study notes the lack of research employing modern techniques and actual clinical datasets for accurate CVD prediction, while also highlighting the future potential of ML and DL to enhance healthcare diagnostics.

2.2.11 Forecasting the risk of coronary heart diseases using machine learning algorithms.

Lakshmi et.al[7] in his literature review used the Framingham Heart Study dataset to predict heart disease using a variety of machine learning methods, such as Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN). According to the study, KNN and Random Forest came in second and third, with accuracies of 80.57% and 82.56%, respectively, while Gradient Boosting had the greatest accuracy of 88.08%. These algorithms' strengths include their capacity to handle big datasets and produce predictions with a high degree of accuracy, both of which are critical for the early identification of cardiac conditions. A notable drawback, still, was the possibility of overfitting, especially with complex models like Gradient Boosting, which can make it difficult to apply the outcomes to fresh, untested data.

2.2.12 Study and analysis on detection, classification, and prediction of cardiac arrhythmia using soft computing tool.

Pratima et.al[10] paper examines a number of techniques and algorithms for the detection and classification of cardiac arrhythmias. These include deep learning techniques like Depth-wise Separable Convolution Neural Network with Focal Loss (DSC-FL-CNN) and machine learning models like K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and K-NN. These models use efficient pre-processing methods like Discrete Wavelet Transform (DWT) to lower noise in ECG signals and exhibit great accuracy, with SVM and RF reaching up to 99.51% accuracy. These methods have the advantages of automatic

detection, great dependability, and enhanced performance due to advanced feature extraction. However, issues including data imbalance, deep learning models' complexity, and their dependence on high-quality training data can make them less useful in practical settings.

2.2.13 Hbwo-js: jellyfish search boosted hybrid beluga whale optimization algorithm for engineering applications.

Yuan et.al[15] developed the HBWO-JS algorithm using a hybrid methodology that combines the Beluga Whale Optimization (BWO) algorithm with the Jellyfish Search (JS) optimizer. By improving the original BWO's exploration and exploitation capabilities, this method improves the combined accuracy of optimization problems across a range of dimensions. The approach makes use of a combination of dynamic parameter changes and creative search techniques, both of which are carried out by means of computational instruments aimed for comparison with well-known methods. The HBWO-JS algorithm's strengths include its stability across various test functions, especially in high-dimensional environments, and its exceptional performance in locating near-optimal solutions. Increased computational complexity and the requirement for meticulous parameter optimization to obtain the best outcomes in particular applications, however, could be drawbacks.

2.2.14 Machine learning technology-based heart disease detection models.

Nagavelli et.al[9] In the study of machine learning technologies for the detection of cardiac disease, highlight the value of early heart failure diagnosis. In order to improve prediction accuracy, the study will review a number of machine learning models, such as XGBoost, support vector machines (SVM), and Naïve Bayes. The methodology shows that the diagnostic capabilities and patient outcomes are significantly improved by including these models into clinical decision support systems. Important results show that XGBoost performs better than other models in terms of F1-measure, accuracy, precision, and recall. By demonstrating how machine learning may transform the detection and treatment of heart disease, the study advances healthcare services with the ultimate goal of improving patient care and lowering the death rates linked to cardiovascular disorders.

2.2.15 Predictive analysis of heart diseases with machine learning approaches.

Ramesh TR et.al[11] paper "Predictive Analysis of Heart Diseases with Machine Learning Approaches" discusses the vital role that machine learning plays in the early detection and management of heart conditions. It describes the goals of improving prediction accuracy and locating important characteristics in datasets related to heart disease. The methodology involves

applying different classification algorithms to datasets such as the UCI heart disease dataset, feature selection, and data preprocessing. Major findings show improved prediction performance by comparing several machine learning models. The study's contributions to healthcare are highlighted in its conclusion, along with the possibility of better patient outcomes and suggestions for further research in the medical area using predictive analytics.

Table 2.1 Inferences from Literature Survey

S.No	Title of the paper	Author Details	Publication & Year	Remarks
1	Heart disease prediction using hybrid machine learning:a brief review	Ahmed, Mohammed, and Idress Husien	Journal of Robotics and Control (JRC) (2024)	<ul style="list-style-type: none"> The hybrid approach improves prediction accuracy by combining multiple algorithms, leading to more reliable heart disease risk assessments. The complexity of the models can reduce interpretability, making it challenging for clinicians to understand and trust the predictions.
2	Effective feature engineering technique for heart disease prediction with machine learning	Babu, Ch Kiran, M. Iswarya,R. Manikanta Kumar, And M. Pavan Sai	Journal of Nonlinear Analysis and Optimization (2024)	<ul style="list-style-type: none"> The study achieves 86% accuracy with the Random Forest model and improves prediction accuracy through innovative

				<p>feature engineering (PCHF) and robust K-fold cross-validation.</p> <ul style="list-style-type: none"> • The methodology's high computational demands and complexity may make it inaccessible or difficult to implement for practitioners without advanced technical resources.
3	<p>Deep learning-based method for detecting parkinson using 1d convolutional neural networks and improved jellyfish algorithms</p>	<p>Shankar, Sharmila</p>	<p>International journal of electrical and computer engineering systems (2024)</p>	<ul style="list-style-type: none"> • The study achieves 88.6% accuracy in Parkinson's disease classification, utilizing robust feature selection (mRMR) and hyperparameter tuning with the Improved Jellyfish Algorithm (IJFA). • The small dataset and the risk of overfitting, along with the complexity of integrating deep learning with optimization methods, pose

				challenges for broader applicability.
4	Machine learning algorithms for predicting hospital readmission and mortality rates in patients with heart failure.	Rizinde, T., I. Ngaruye, and N. D. Cahill	African Journal of Applied Research (2024)	<ul style="list-style-type: none"> The study shows strong predictive potential with machine learning models, especially random forests, achieving an AUC of 0.92 for predicting heart failure readmissions and mortality. Challenges like class disparity, non-standardized health data, and distrust from medical practitioners in LMICs may hinder the widespread adoption of these models.
5	Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis.	Liu, Tianyi, Andrew Krentz, Lei Lu, and Vasa Curcin	European Heart Journal-Digital Health (2024)	<ul style="list-style-type: none"> The study demonstrates that machine learning models, by combining diverse data sources, can improve the accuracy of predicting cardiovascular

				<p>disease risk beyond traditional risk factors.</p> <ul style="list-style-type: none"> Issues like methodological variability, bias, and lack of external validation raise concerns about the clinical applicability and reliability of these models compared to established risk scores.
6	<p>AI-assisted medical imaging and heart disease diagnosis: a deep learning approach for automated analysis and enhanced prediction using ensemble classifiers.</p>	<p>Nidadavolu Durga Venkat Sai Siva Vara Prasad Raju</p>	<p>Journal of Artificial Intelligence General science (JAIGS) (2024)</p>	<ul style="list-style-type: none"> The study highlights that deep learning (CNNs, RNNs) and other ML algorithms can enhance cardiovascular disease risk prediction by analyzing complex data from wearables and electronic health records. Challenges such as lack of external validation, poor interpretability, and data quality issues limit the clinical

				trust and generalizability of these models.
7	Prediction of mortality in young adults with cardiovascular disease using artificial intelligence	Vara, Nidadavolu Venkat Durga Sai Siva, Prasad Raju, and Penmetsa Naveena Devi	Electronic Journal of General Medicine (2024)	<ul style="list-style-type: none"> The CHAID algorithm accurately predicts mortality in young adults with CVD, achieving 88.27% accuracy and identifying key predictors like pulse pressure and ischemic heart disease. The model's focus on young adults (ages 18-45) may limit its generalizability to older populations or those with different demographic characteristics.
8	Advancing heart disease prediction through machine learning: techniques and insights for improved cardiovascular health.	Ahmad, M., S. Bani Hani, M. Abu-Abeeleh, and I. Aljarah	British Journal of Nursing Studies (2024)	<ul style="list-style-type: none"> Logistic Regression proved highly effective for early heart disease detection, achieving the highest ROC-AUC score of 0.7810 and balancing precision and recall.

				<ul style="list-style-type: none"> Random Forest performed the least effectively, and while SVM had the highest accuracy (72.33%), its lower ROC-AUC score showed variability in model performance.
9	Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison.	Hossain, Md Imam, Mehadi Hasan Maruf, Md Ashikur Rahman Khan, Farida Siddiqi Prity, Sharmin Fatema, Md Sabbir Ejaz, and Md Ahnaf Sad Khan	Iran Journal of Computer Science (2023)	<ul style="list-style-type: none"> Logistic Regression achieved the highest ROC-AUC score (0.7810), proving to be an effective and reliable model for early heart disease detection. Random Forest underperformed, and although SVM had the highest accuracy (72.33%), its slightly lower ROC-AUC showed variability in model effectiveness.
10	A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis.	Alkayyali, Z. K., S. Anuar Bin Idris, and Samy S. Abu-Naser.	Journal of Theoretical and Applied Information Technology (2023)	<ul style="list-style-type: none"> Gradient Boosting achieved the highest accuracy (88.08%) and is effective in handling large datasets for early heart disease

				<p>detection.</p> <ul style="list-style-type: none"> • The complexity of Gradient Boosting and other models may lead to overfitting, making it difficult to apply the results to new, unseen data.
11	Forecasting the risk of coronary heart diseases using machine learning algorithms.	Lakshmi, J. V. N., and Anirban Das	Journal Press India (2023)	<ul style="list-style-type: none"> • Gradient Boosting achieved the highest accuracy (88.08%) and effectively handles large datasets, making it highly suitable for early heart disease prediction. • The complexity of models like Gradient Boosting increases the risk of overfitting, which may limit their ability to generalize to new, unseen data.
12	Study and analysis on detection, classification and prediction of cardiac arrhythmia using soft computing tool.	Pratima, A., K. GopalaKrishna, and S. N. Prasad	Journal of Physics: Conference Series (2023)	<ul style="list-style-type: none"> • SVM and Random Forest achieved up to 99.51% accuracy in detecting cardiac arrhythmias, benefiting from advanced feature extraction and pre-

				<p>processing techniques like Discrete Wavelet Transform (DWT).</p> <ul style="list-style-type: none"> • Issues like data imbalance, the complexity of deep learning models, and dependence on high-quality training data can hinder their practical use in real-world applications.
13	<p>Hbwo-js: jellyfish search boosted hybrid beluga whale optimization algorithm for engineering applications.</p>	<p>Yuan, Xinguang, Gang Hu, Jingyu Zhong, and Guo Wei</p>	<p>Journal of Computational Design and Engineering (2023)</p>	<ul style="list-style-type: none"> • The HBWO-JS algorithm offers high stability and exceptional performance in high-dimensional optimization problems by combining Beluga Whale Optimization and Jellyfish Search. • Its increased computational complexity and the need for careful parameter optimization may limit its practical

				application and ease of use.
14	Machine learning technology-based heart disease detection models.	Nagavelli, Umarani, Debabrata Samanta, and Partha Chakraborty.	Journal of Healthcare Engineering (2022)	<ul style="list-style-type: none"> • XGBoost outperforms models like SVM and Naïve Bayes in accuracy, precision, and recall, enhancing heart disease diagnosis and patient outcomes. • Integrating these machine learning models into clinical systems may face challenges like data quality issues, interpretability, and the need for healthcare expertise.
15	Predictive analysis of heart diseases with machine learning approaches.	Ramesh, T. R., Umesh Kumar Lilhore, M. Poongodi, Sarita Simaiya	Malaysian Journal of Computer Science (2022)	<ul style="list-style-type: none"> • Machine learning models, combined with feature selection and data preprocessing, enhance heart disease prediction accuracy, leading to improved patient outcomes. • The study's reliance on specific datasets

				like the UCI heart disease dataset generalizability, requiring further research for broader applicability.
--	--	--	--	------------------------------------------------------------------------------------------------------------

2.4 SUMMARY

This project explores the use of machine learning and hybrid models for predicting cardiovascular and neurological diseases, showing high accuracy with methods like Random Forest, Gradient Boosting, and Logistic Regression. Advanced techniques such as feature engineering, cross-validation, and optimization algorithms (e.g., IJFA, HBWO-JS) significantly improve prediction accuracy. Models achieved up to 90.51% accuracy in arrhythmia detection and 88.6% in Parkinson's classification. While effective, these models often suffer from complexity, overfitting, and low interpretability. Issues like small datasets, data imbalance, and lack of external validation limit generalizability. Integration into clinical settings is challenged by computational demands and practitioner distrust. Nevertheless, machine learning enhances early diagnosis and has strong potential to improve healthcare outcomes when properly implemented.

CHAPTER 3

SYSTEM SPECIFICATION

3.1 INTRODUCTION

The system requirements necessary to develop and execute the heart disease prediction project. It includes both hardware and software specifications, as well as a brief description of the software components used in the implementation.

3.2 HARDWARE SPECIFICATION

- **Processor:** Intel Core i5 or AMD Ryzen 5 (or higher)
- **RAM:** 8 GB minimum
- **Storage:** 256 GB SSD or HDD
- **Operating System:** Windows 10/11, Ubuntu 18.04+, or macOS

3.3 SOFTWARE SPECIFICATION

- **Platform:** Google Colab
- **Programming Language:** Python
- **Libraries:** NumPy, Pandas, Scikit-learn, Imbalanced-learn, Matplotlib, Seaborn
- **Algorithms:** Support Vector Machine (SVM), Jellyfish Optimization Algorithm (JFO), SMOTE (Synthetic Minority Over-sampling Technique), SelectKBest with Chi-squared Test, Cross-Validation

3.4 SOFTWARE DESCRIPTION

Google Colab

Google Colab (Colaboratory) is a cloud-based interactive coding environment developed by Google, designed especially for machine learning, data science, and artificial intelligence projects. It is built on *Jupyter Notebooks*, allowing users to write and execute

Python code in cells, view outputs instantly, and include formatted text, images, and visualizations—all in the same document.

One of the major advantages of Google Colab is that it requires no installation or setup. All necessary libraries such as NumPy, pandas, matplotlib, scikit-learn, TensorFlow, and PyTorch come pre-installed, saving time and effort. Colab provides free access to powerful hardware, including GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units), which significantly speed up training for deep learning models.

Colab supports loading data from local systems, Drive, or cloud sources, making it easy to use real-world datasets. It is especially beneficial for students, researchers, and developers who may not have access to high-end computers but want to work on high-performance computing tasks.

Python

Python is a high-level, interpreted programming language created by Guido van Rossum and first released in 1991. It is known for its simple, readable syntax that makes it easy to learn and use, especially for beginners. Python supports multiple programming styles, including procedural, object-oriented, and functional programming. It is widely used in various fields such as web development, data science, artificial intelligence, automation, and more. With a vast ecosystem of libraries and tools, Python enables rapid development and problem-solving. It runs on multiple operating systems including Windows, macOS, and Linux. Python's strong community contributes to its continuous growth and support. Its flexibility and power make it a favorite among both beginners and professionals in the programming world.

CHAPTER 4

PROPOSED SYSTEM

GAP IDENTIFIED FROM LITERATURE SURVEY

- The scope of several models is limited by overfitting and small datasets.
- Comparison and consistency of the model suffers by various feature engineering techniques.
- Practical implementation in healthcare is limited by a lack of scientific validation.
- Few research includes a variety of real-world datasets or focus on young adult populations.
- Development of an innovative model that combines SVM and Jellyfish Optimization to outperform traditional techniques.
- The project uses the Jellyfish Optimization Algorithm (JFO) for feature selection and hyperparameter optimization, which improves the performance of Support Vector Machines (SVM) and increases the accuracy of heart disease prediction.
- Comparison of advanced and traditional approaches, providing a comprehensive understanding of their advantages and disadvantages.
- The focus is on creating flexible, comprehensible, and useful cardiac disease prediction solutions.

4.1 EXISTING SYSTEM

Existing heart disease prediction systems use traditional machine learning models like SVM, Naïve Bayes, Decision Trees, and Random Forest. These models rely on clinical data to predict the presence of heart disease.

4.2 DRAWBACKS OF EXISTING SYSTEM

- Limited accuracy and recall
- Struggle with complex, high-dimensional data
- Risk of overfitting
- Lack of optimization reduces prediction reliability

4.3 PROPOSED MODEL

Develop and compare heart disease prediction models using Jellyfish Optimization Algorithm (JFO) integrated with SVM, alongside standard SVM, Gradient Boosting, and KNN

Approach:

- Analyze datasets related to heart disease using exploratory data analysis (EDA) and preprocessing.
- Use JFO to choose features and adjust hyperparameters for SVM and decision trees.
- Develop and verify models.
- Analyze performance indicators including ROC-AUC, F1-score, recall, accuracy, and precision..

Data source :

- Heart disease datasets from UCI or similar sources.
- Features include age, gender, chest pain type, cholesterol levels, and others.

Tools:

Python (NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn).

Algorithms:

- Jellyfish Optimization Algorithm (JFO): Hyperparameter optimization.
- Support Vector Machine (SVM): Kernel-based classification.
- Gradient Boosting Probabilistic model for baseline comparison.
- KNN : Probabilistic model for baseline comparison

Hybrid optimization framework:

JFO-SVM to enhance feature hyperparameter tuning.

Baseline models:

Standard SVM, Gradient Boosting, and KNN for comparison.

Justification of chosen approach :

- JFO for Optimization: Enhances feature selection and parameter tuning efficiency, improving model performance (AlMahadin et al.[3], Abualigah et al.[13]).
- SVM Strength: Handles non-linear data effectively (Bhowmik et al.[8]).

- Gradient Boosting : Serves as a probabilistic baseline model for comparison.
- KNN : Provide interpretable classification models with the ability to handle mixed data types.

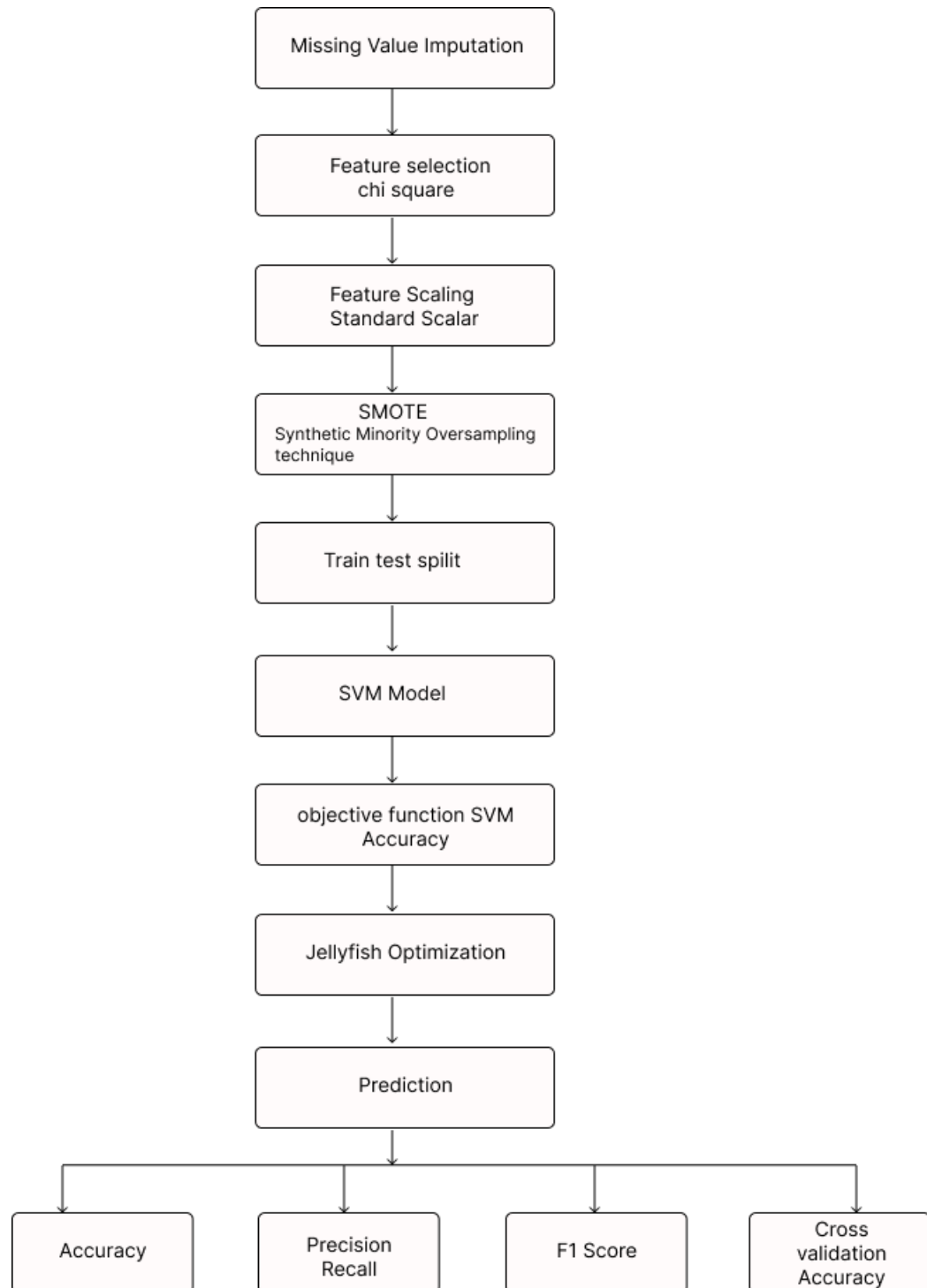


Figure 4.1 Flow Chart for Proposed Model

The flowchart illustrates the machine learning pipeline for heart disease prediction using an SVM model optimized with the Jellyfish Optimization Algorithm. It begins with missing value

imputation to handle incomplete data, followed by feature selection using the chi-square test to retain significant features. Standard Scaler is then applied to scale the features uniformly. To address class imbalance, the SMOTE technique generates synthetic samples for the minority class. The data is split into training and testing sets, and an initial SVM model is built. The SVM's accuracy serves as the objective function for optimization, and the Jellyfish Optimization Algorithm is used to fine-tune its hyperparameters. The optimized model makes predictions, which are evaluated using metrics such as accuracy, precision, recall, F1 score, and cross-validation accuracy to ensure robust performance from the Figure 4.1.

4.4 MODULE DESCRIPTION

4.4.1 Dataset description

The study's dataset, which included patient records with cardiovascular characteristics like age, gender, blood pressure, cholesterol, fasting blood sugar, ECG readings, and ST depression levels, was acquired via Kaggle. The binary target variable indicates if cardiac disease is present (1) or not (0).

4.4.2 Data preprocessing

The process of transforming raw data into a format suitable for model training is known as data preprocessing. By handling missing values, normalizing features, eliminating outliers, and choosing pertinent attributes, it improves the quality of data. Proper preprocessing ensures that the model learns correctly, potentially enhancing accuracy, generalization, and overall performance. 1) Handling Missing Values: The mean imputation method is used to impute missing values:

1) Handling Missing Values: The mean imputation method is used to impute missing values:

Where:

$$X_{\text{new}} = \frac{\sum_{i=1}^n X_i}{n}$$

Where:

- X_{new} is the replaced value.
- X_i represents the observed values.
- n is the count of available data points.

2) Z-Score Standardization

To normalize data and ensure all features have a standard deviation of 1 and mean of 0, Z score standardization is applied:

$$Z = \frac{y - \mu}{\sigma}$$

Where:

- y = Original value
- μ = Mean of the variables
- σ = Standard deviation

4.4.3 Feature selection

Feature selection removes irrelevant attributes, enhancing model accuracy and reducing complexity.

Feature Selection Using Chi-Square Test: The ChiSquare Test assesses feature importance by measuring dependence between independent variables and the target variable:

$$\chi^2 = \sum \frac{(A_i - B_i)^2}{B_i}$$

Where:

- A_i = Observed frequency
- B_i = Expected frequency

4.4.4 Support Vector Machine

This algorithm is widely used in Machine Learning techniques due to its insensitivity to noise, and it performs well with large datasets. It characterizes each feature in the n-dimensional Euclidean space and attempts to find the corresponding hyperplane that distinguishes between the various classes. For data that does not split, Support Vector Machines (SVM) use the kernel trick. It renders the problem of class separation in the original space simpler by mapping the data into a higher-dimensional space. Because of this method, SVM are applicable to more complex datasets which have complex fracture surfaces. Common kernels like polynomial kernels and Radial Basis Function (RBF) take advantage of features like distance, similarity, and inner product. It can lead to improving accuracy and generalization of SVM.

$$f(y) = w \cdot p + b$$

Where:

- w = Weight vector
- p = Input feature vector
- b = Bias term

The RBF kernel used in this model is defined as:

$$K(x, x') = \exp(-\gamma ||x - x'||^2)$$

Where:

- γ = Kernel coefficient, controlling the flexibility of the decision boundary

4.4.5 Jellyfish optimization

Jellyfish Optimization Algorithm (JFO) is a novel method derived from nature, specifically the motion of swimming by jellyfish, which aims to solve complex problems. The major objective of JFO is to explore the solution space effectively by modifying behavioral approaches in a pattern similar to a jellyfish. JFO operates based on two main movements:

• **Random movement:** This movement reflects the way jellyfish glide smoothly in the water, enabling the program to move around the solution space in various directions. This aspect enables exploration in all directions so that the algorithm does not become stuck in one position.

$$X^{t+1} = X^t + \alpha \cdot (X^r - X^t)$$

Where:

- X^r = Randomly selected solution
- α = Random step size

• **Directed movement:** This replicates how jellyfish swim towards food, guiding the jellyfish towards better solutions. This enables the algorithm to concentrate on the best solution by limiting the potential answers.

$$X^{t+1} = X^t + \beta \cdot (X^g - X^t)$$

Where:

- X^g = Best solution found so far
- β = Attraction coefficient

4.4.6 JFO Parameters:

- P = Population size (set of candidate solutions)
- t = Current iteration
- T_{\max} = Maximum number of iterations
- X^t = Position of jellyfish t
- F = Objective function (accuracy or F1-score)

4.5 ADVANTAGES OF PROPOSED MODEL

- Improves model accuracy by optimizing SVM hyperparameters and feature selection using JFO.
- Enhances model reliability through comparison with multiple baseline classifiers (SVM, Gradient Boosting, KNN).
- Handles non-linear and high-dimensional medical data effectively using SVM with kernel functions.
- Provides a robust hybrid optimization framework (JFO-SVM) for efficient tuning and selection.
- Offers comprehensive evaluation using multiple metrics such as ROC-AUC, F1-score, recall, accuracy, and precision.
- Ensures model interpretability and simplicity through inclusion of KNN and Gradient Boosting as baseline models.
- Utilizes real-world medical datasets (e.g., KAGGLE) ensuring practical relevance and applicability.
- Demonstrates the effectiveness of nature-inspired optimization algorithms (JFO) in medical diagnostics.
- Aids early detection of heart disease, potentially reducing healthcare costs and improving patient outcomes.
- Leverages widely-used and open-source tools (Python, Scikit-learn) for reproducibility and ease of development.

CHAPTER 5

IMPLEMENTATION AND RESULT

5.1 OVERVIEW

This chapter provides an overview of the system developed for heart disease prediction. The system is designed to process clinical data, apply machine learning models, and predict the presence of heart disease with improved accuracy using optimized algorithms. The workflow involves data preprocessing, feature selection, handling imbalanced data, model training, and evaluation using performance metrics. The project's performance comparison of machine learning models analyzes different classifiers, including SVM, KNN, Gradient Boosting, and SVM with JFO, using key metrics such as accuracy, precision, recall, and F1-score. The results clearly demonstrate that the hybrid model (SVM with JFO) outperforms the other models in all evaluation criteria. Support Vector Machine (SVM) with Jellyfish Optimization (JFO) improves heart disease prediction by combining JFO's feature selection with SVM's classification power. JFO mimics jellyfish behavior, identifying key patient information (e.g., age, blood pressure, cholesterol). SVM then classifies the optimized data, reducing noise, overfitting, and training time. JFO improves accuracy to 98% compared to SVM's 90%

5.2 MODEL PERFORMANCE AND OUTPUT

Performance analysis of models :

Here are the formulas for accuracy, F1 score, precision, and recall which are commonly used to evaluate machine learning models.

1) Accuracy: Measures the correctness of the model.

$$\text{Accuracy} = \frac{T_{RP} + T_{RN}}{T_{RP} + T_{RN} + F_{AP} + F_{AN}}$$

2) Precision (Positive Predictive Value): Measures how many of the predicted positive cases are actually positive.

$$P_R = \frac{T_{RP}}{T_{RP} + F_{AN}}$$

3) Recall (True Positive Rate or Sensitivity): Measures how many actual positive cases were correctly identified.

$$R_E = \frac{T_{RP}}{T_{RP} + F_{AN}}$$

4) F1-score: The mean of Precision and Recall, balancing both metrics.

$$F1\text{-score} = 2 \times \frac{PR \times RE}{PR + RE}$$

Where:

- T_{RP} = True Positive
- T_{RN} = True Negative
- F_{AP} = False Positive
- F_{AN} = False Negative

Table 5.1 Selected Feature and Cross Validation Scores

Selected Feature	Crossvalidation Scores
Cp	0.92
Thalach	0.92
Exang	0.97
Oldpeak	0.95
Ca	0.94
Mean	0.94

Cross-validation evaluates a model's performance by splitting the data into multiple subsets. The model is trained on some of the subsets and tested on the remaining subsets. The displayed results show five cross-validation scores, averaging 94.41% accuracy, indicating high model reliability and generalization across different data splits, reducing overfitting risks. In this method, the model used for 70% of values for Training, 20% for validation, and 10% for Testing.

Table 5.2 Performance metrics of the Proposed SVM with JFO model

Performance Metrics	Classifier Model			
	SVM	KNN	Gradient Boosting	SVM With JFO
Accuracy	0.79	0.83	0.88	0.95
Precision	0.85	0.87	0.90	0.93
Recall	0.76	0.82	0.89	0.98
F1-Score	0.80	0.85	0.89	0.96

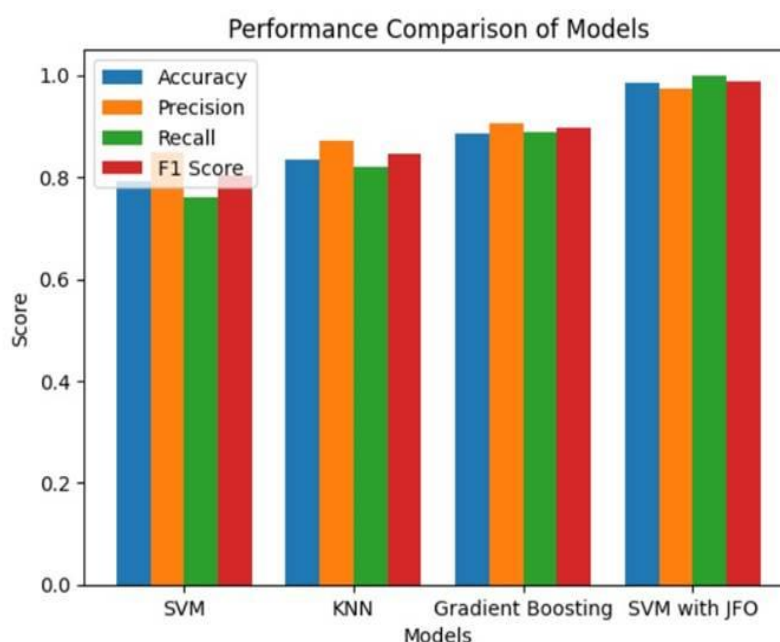


Figure 5.1 Performance Result Graph

The Table 5.2 and Figure 5.1 describes the hybrid approach of SVM with JFO delivers impressive results, achieving 95% accuracy and a recall of 0.98, indicating that a significant majority of actual positive cases were classified correctly. While the F1-score of 0.96 indicates a well-balanced performance between precision and recall, the precision of 0.93 indicates a high ability to limit false positives. The improvement is attributed to JFO optimizing hyperparameters like C and gamma, enabling the SVM to find the best decision boundary.

CHAPTER 6

CONCLUSION AND FUTURE WORK

CONCLUSION

The hybrid SVM-JFO model significantly outperforms traditional machine learning models in heart disease prediction. It ensures higher accuracy and recall, making it highly reliable for real-world applications. This demonstrates the effectiveness of optimization techniques like Jellyfish Optimization Algorithm (JFO) in fine-tuning model parameters for better performance. Additionally, the model reduces overfitting by selecting optimal hyperparameters, leading to improved generalization on unseen data. Its robustness and adaptability make it a promising approach for medical diagnostics and decision support systems

FUTURE WORK

- **Integration with Real-Time Data:**

Incorporate data from wearable devices or IoT-based health monitors for real-time heart disease prediction.

- **Deep Learning Models:**

Explore advanced deep learning techniques such as CNNs or RNNs for improved prediction accuracy and handling of more complex datasets.

- **Larger and diverse datasets:**

Use larger, multi-source datasets to improve the generalizability and robustness of the model.

- **Mobile or web application:**

Develop a user-friendly application for patients and doctors to easily access predictions and monitor heart health.

APPENDIX

SOURCE CODE

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.feature_selection import SelectKBest, chi2
from imblearn.over_sampling import SMOTE

data = pd.read_csv("heart.csv")
data.fillna(data.mean(), inplace=True)

X = data.drop(columns=["target"])
y = data["target"]

chi2_selector = SelectKBest(chi2, k=5)
X_selected = chi2_selector.fit_transform(X, y)
selected_features = X.columns[chi2_selector.get_support()].tolist()

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_selected)

smote = SMOTE(sampling_strategy='auto', random_state=42)
X_res, y_res = smote.fit_resample(X_scaled, y)

X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, test_size=0.2, random_state=42)

def objective_function(params):
    C, gamma = params
```

```

model = SVC(C=C, gamma=gamma, kernel='rbf')
scores = cross_val_score(model, X_train, y_train, cv=5, scoring='accuracy')
return np.mean(scores)

```

```

class JellyfishOptimization:

```

```

    def __init__(self, objective_function, bounds, population_size, iterations, alpha=3, beta=0.1):
        self.objective_function = objective_function
        self.bounds = np.array(bounds)
        self.population_size = population_size
        self.iterations = iterations
        self.alpha = alpha
        self.beta = beta
        self.population = None
        self.best_solution = None
        self.best_score = float("-inf")

```

```

    def initialize_population(self):
        self.population = np.random.uniform(self.bounds[:, 0], self.bounds[:, 1], (self.population_size,
len(self.bounds)))

```

```

    def evaluate_population(self):
        fitness = np.array([self.objective_function(ind) for ind in self.population])
        best_index = np.argmax(fitness)
        if fitness[best_index] > self.best_score:
            self.best_score = fitness[best_index]
            self.best_solution = self.population[best_index]
        return fitness

```

```

    def move_jellyfish(self, fitness):
        new_population = np.copy(self.population)
        for i in range(self.population_size):
            if np.random.rand() < 0.5:
                r1, r2 = np.random.choice(self.population_size, 2, replace=False)
                direction = self.population[r1] - self.population[r2]

```

```

        new_position = self.population[i] + self.alpha * np.random.rand() * direction
    else:
        jellyfish_mean = np.mean(self.population, axis=0)
        direction = jellyfish_mean - self.population[i]
        new_position = self.population[i] + self.beta * np.random.rand() * direction
    new_population[i] = np.clip(new_position, self.bounds[:, 0], self.bounds[:, 1])
    return new_population

def optimize(self):
    self.initialize_population()
    for _ in range(self.iterations):
        fitness = self.evaluate_population()
        self.population = self.move_jellyfish(fitness)
    return self.best_solution

bounds = [(0.1, 1000), (0.001, 1)]
jellyfish = JellyfishOptimization(objective_function, bounds, population_size=10, iterations=50)
best_params = jellyfish.optimize()
svm_model = SVC(C=best_params[0], gamma=best_params[1], kernel='rbf')
svm_model.fit(X_train, y_train)
y_pred = svm_model.predict(X_test)
cv_scores = cross_val_score(svm_model, X_train, y_train, cv=5)
jfo_accuracy = accuracy_score(y_test, y_pred)
jfo_precision = precision_score(y_test, y_pred)
jfo_recall = recall_score(y_test, y_pred)
jfo_f1 = f1_score(y_test, y_pred)

print("Selected Features:", selected_features)
print("Accuracy:", jfo_accuracy)
print("Precision:", jfo_precision)
print("Recall:", jfo_recall)
print("F1 Score:", jfo_f1)
print("Cross-Validation Scores:", cv_scores)
print("Mean Cross-Validation Accuracy:", np.mean(cv_scores))

```

REFERENCES

- [1] Ahmad, M., S. Bani Hani, M. Abu-Abeeleh, and I. Aljarah. *Prediction of mortality in young adults with cardiovascular disease using artificial intelligence. Electron J Gen Med.* 2024; 21 (3): em584. Electronic Journal of General Medicine, 2024.
- [2] Ahmed, Mohammed, and Idress Husien. "Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review." *Journal of Robotics and Control (JRC)* 5, no. 3 (2024): 884-892.
- [3] Alkayyali, Z. K., S. Anuar Bin Idris, and Samy S. Abu-Naser. "A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis." *Journal of Theoretical and Applied Information Technology* 101, no. 4 (2023): 1353-1365
- [4] Babu, Ch Kiran, M. Iswarya, R. Manikanta Kumar, And M. Pavan Sai. "Effective feature engineering technique for heart disease prediction with machine learning." *Journal of Nonlinear Analysis and Optimization* 15, no. 1 (2024).
- [5] Bhowmik, Proshanta Kumar, Mohammed Nazmul Islam Miah, Md Kafil Uddin, Mir Mohtasam Hossain Sizan, Laxmi Pant, Md Rafiqul Islam, and Nisha Gurung. "Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health." *British Journal of Nursing Studies* 4, no. 2 (2024): 35-50.
- [6] Hossain, Md Imam, Mehadi Hasan Maruf, Md Ashikur Rahman Khan, Farida Siddiqi Prity, Sharmin Fatema, Md Sabbir Ejaz, and Md Ahnaf Sad Khan. "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison." *Iran Journal of Computer Science* 6, no. 4 (2023): 397-417.
- [7] Lakshmi, J. V. N., and Anirban Das. "Forecasting the Risk of Coronary Heart Diseases using Machine Learning Algorithms." *Journal Press India* 3, no. 2 (2023).
- [8] Liu, Tianyi, Andrew Krentz, Lei Lu, and Vasa Curcin. "Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis." *European Heart Journal-Digital Health* (2024).
- [9] Nagavelli, Umarani, Debabrata Samanta, and Partha Chakraborty. "Machine Learning Technology-Based Heart Disease Detection Models." *Journal of Healthcare Engineering* 2022, no. 1 (2022): 7351061.
- [10] Pratima, A., K. GopalaKrishna, and S. N. Prasad. "Study and analysis on detection, classification, and prediction of cardiac arrhythmia using soft computing tool." In *Journal of Physics: Conference Series*, vol. 2571, no. 1, p. 012010. IOP Publishing, 2023.

- [11] Ramesh, T. R., Umesh Kumar Lilhore, M. Poongodi, Sarita Simaiya, Amandeep Kaur, and Mounir Hamdi. "Predictive analysis of heart diseases with machine learning approaches." *Malaysian Journal of Computer Science* (2022): 132-148.
- [12] Rizinde, T., I. Ngaruye, and N. D. Cahill. "Machine Learning Algorithms for Predicting Hospital Readmission and Mortality Rates in Patients with Heart Failure." *African Journal of Applied Research* 10, no. 1 (2024): 316-338.
- [13] Shankar, Sharmila. "Deep Learning-Based Method for Detecting Parkinson using 1D Convolutional Neural Networks and Improved Jellyfish Algorithms." *International journal of electrical and computer engineering systems* 15, no. 6 (2024): 515-522.
- [14] Vara, Nidadavolu Venkat Durga Sai Siva, Prasad Raju, and Penmetsa Naveena Devi. "AI-Assisted Medical Imaging and Heart Disease Diagnosis: A Deep Learning Approach for Automated Analysis and Enhanced Prediction Using Ensemble Classifiers." *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023* 6, no. 1 (2024): 210-229.
- [15] Yuan, Xinguang, Gang Hu, Jingyu Zhong, and Guo Wei. "HBWO-JS: jellyfish searchboosted hybrid beluga whale optimization algorithm for engineering applications." *Journal of Computational Design and Engineering* 10, no. 4 (2023): 1615-1656.