ECORFAN

9th International Interdisciplinary Congress on Renewable Energies, Industrial Maintenance, Mechatronics and Informatics

Booklets

RENIECYT
Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas
1702902 CONAHCYT

RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - Google Scholar DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID - V|LEX

Title: Performance comparison in optimization algorithms for heart disease detection model

Authors: Rojas, Rafael, Seseña, Hiram, Zuñiga, Mariana and Martínez, Moises

Universidad Autónoma de Querétaro KWU-5496-2024 0009-0006-0745-9887 2045778
Universidad Autónoma de Querétaro KWU-5498-2024 0009-0005-9599-8954 2002041
Universidad Autónoma de Querétaro KWU-2739-2024 0009-0004-8894-060X 2002165
Universidad Autónoma de Querétaro KWU-9562-2024 0000-0001-5495-5185 370689

Editorial label ECORFAN: 607-8695
BCIERMMI Control Number: 2024-01
BCIERMMI Classification (2024): 241024-0001
RNA: 03-2010-032610115700-14
Pages: 12

CONAHCYT classification:
Area: Engineering
Field: Engineering
Discipline: Systems engineer
Subdiscipline: Computer Sciences

ECORFAN-México, S.C.
Park Pedregal Business. 3580, Anillo Perif., San Jerónimo Aculco, Álvaro Obregón, 01900 Ciudad de México, CDMX,
Phone: +52 1 55 6159 2296
Skype: ecorfan-mexico.s.c.
E-mail: contacto@ecorfan.org
Facebook: ECORFAN-México S. C.
Twitter: @EcorfanC

www.ecorfan.org

Holdings
Mexico Colombia Guatemala
Bolivia Cameroon Democratic Republic of Congo
Spain El Salvador
Ecuador Taiwan
Peru Paraguay Nicaragua

PRESENTATION CONTENT

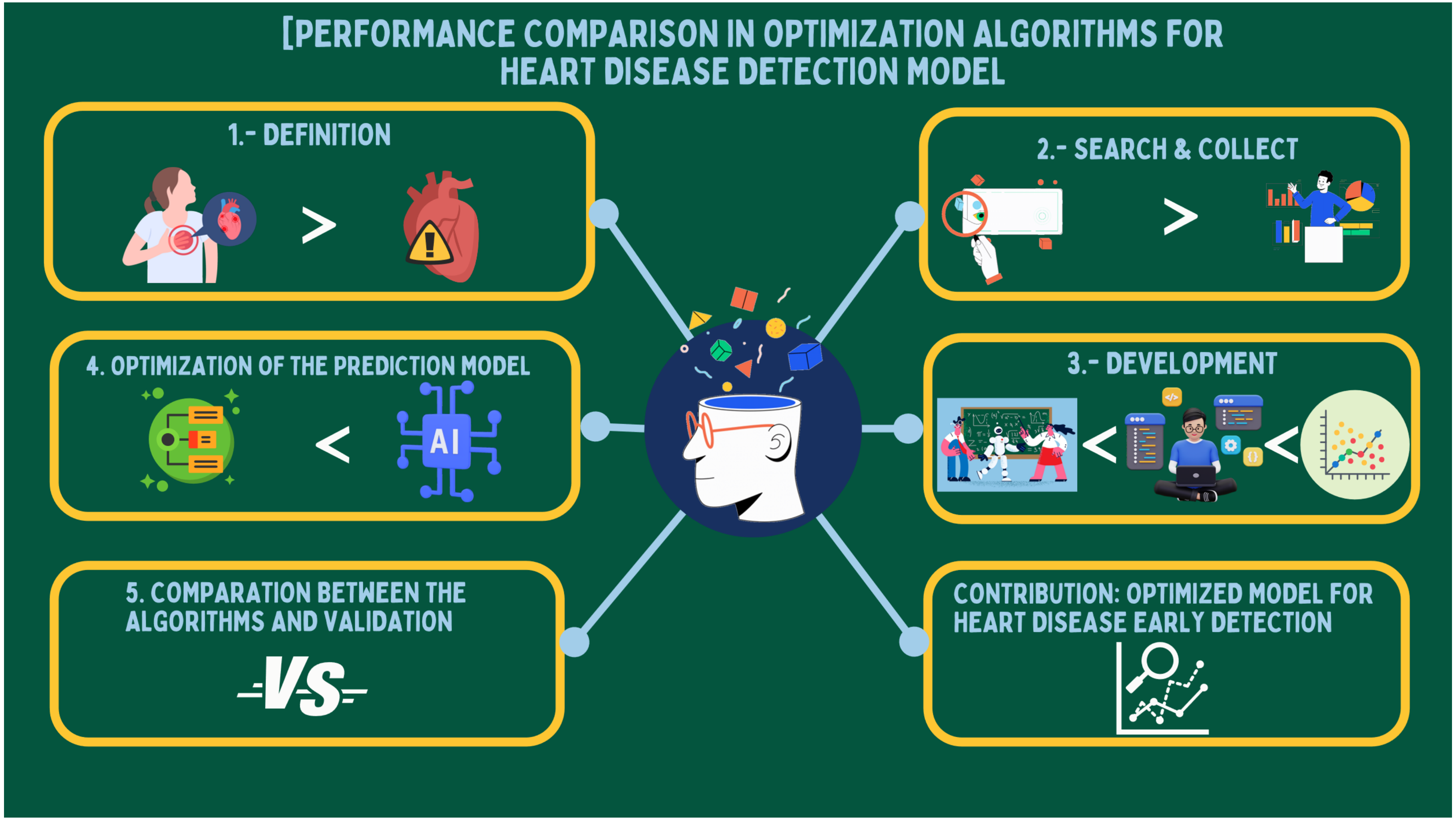
INTRODUCTION

- Global Context: Cardiovascular diseases (CVD) are the leading cause of global mortality. The risk factors include chest pain, resting blood pressure, cholesterol levels, and fasting blood sugar levels. Resting ECG, maximum heart rate, exercise-induced angina, and the pqrst segment also play significant roles in cardiovascular risk assessment. These data highlight the need for a comprehensive approach to evaluating and managing cardiovascular risk.

- The objective of this work is to compare the performance of a SVM model enhanced with optimization algorithms (PSO and GA) for early detection of heart diseases.

- The performance metrics are cross-validation, test precision, F1-Score, etc

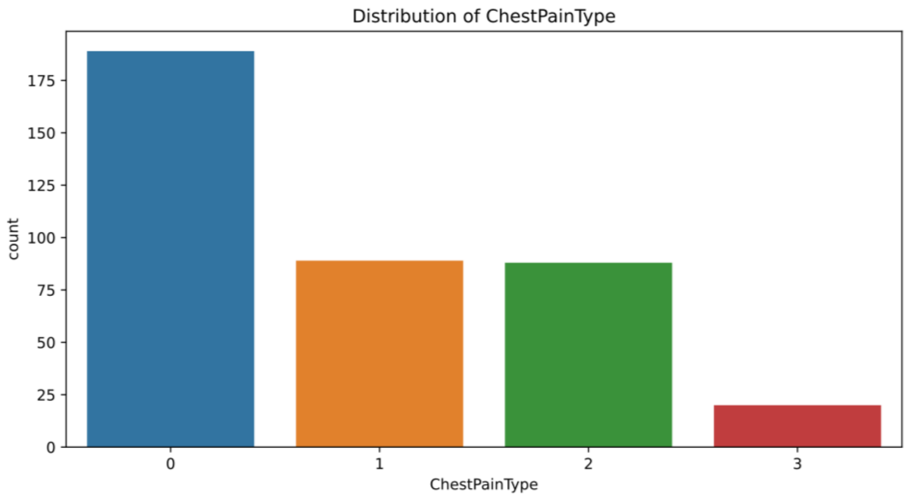
Visual abstract



Analysis EDA

For this study, a public heart disease database was utilized (Kaggle, 2024), encompassing various variables. Variables like age, sex, chest pain type, resting BP, cholesterol, FastingBS, RestingECG, Max HR, Exercise Angina, Oldpeak, ST_Slope, and the target variable HeartDisease. The dataset was balanced before use to ensure an equal number of male and female participants.

The main objective of an analysis EDA is to study the dataset and find relations between the variables. Some relations are shown in Figure 2 and Figure 4.



The Figure 2 shows the distribution of the "ChestPainType" variable with categories coded as 0 (asymptomatic), 1 (atypical angina), 2 (non-anginal pain), and 3 (typical angina). The bar chart reveals that category 0 is the most common, while category 3 is the least frequent. Categories 1 and 2 have moderate occurrences.

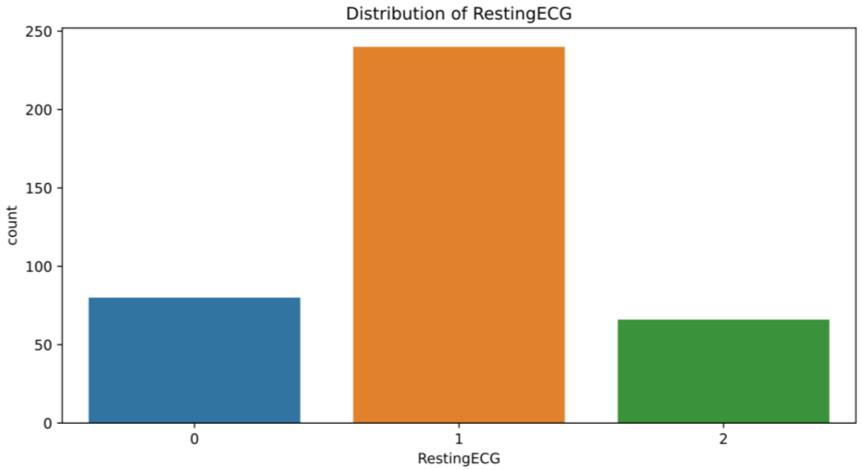
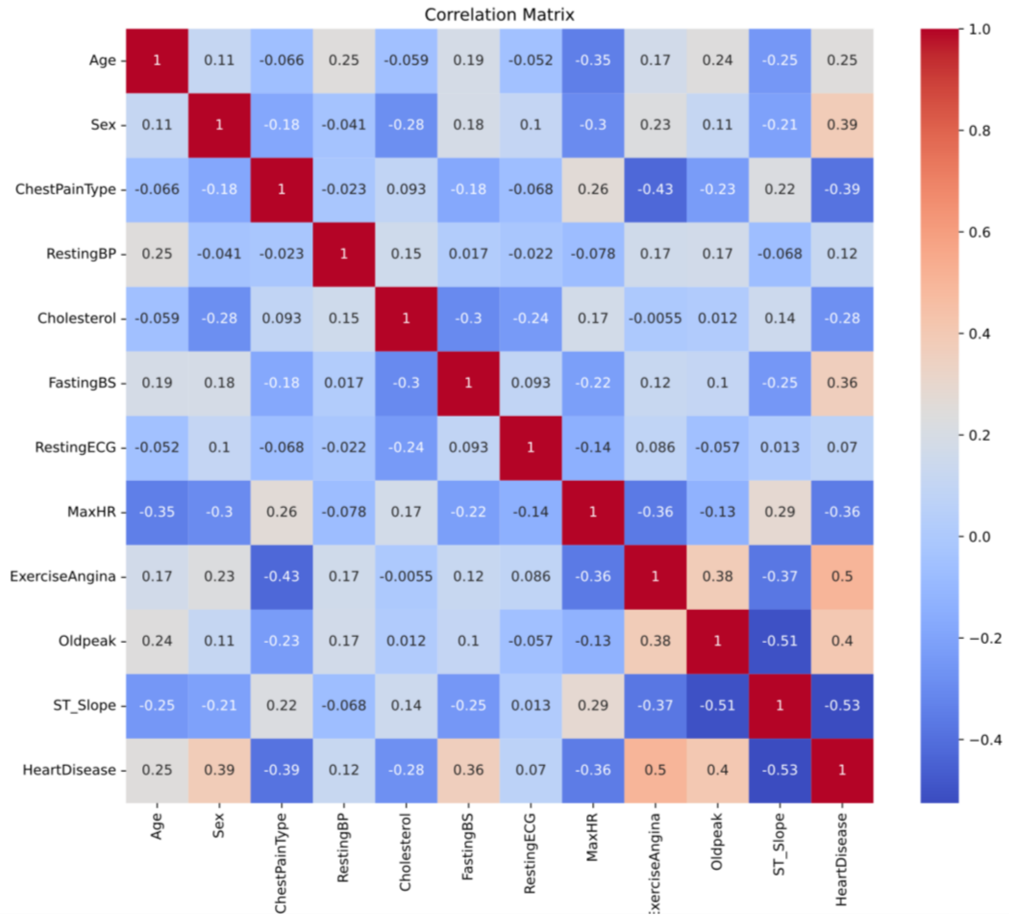


Figure 4 shows the distribution of the “RestingECG” variable, with categories coded as 0 (left ventricular hypertrophy), 1 (normal), and 2 (ST-T wave abnormality). The bar chart indicates that category 1 (normal) is the most common among participants. Category 0 (LVH) has a moderate number of occurrences, while category 2 (ST-T wave abnormality) is the least frequent.

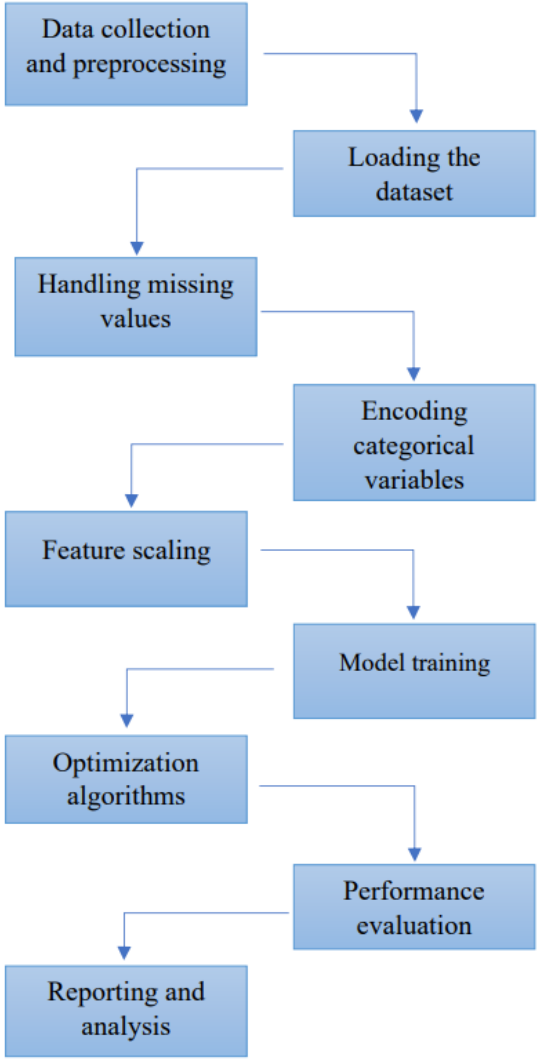


Correlational matrix

- HeartDisease and Sex: There is a strong positive correlation between HeartDisease and Sex. This suggests that the likelihood of heart disease is higher in one gender compared to the other. Typically, a higher correlation with Sex often indicates that males (coded as 1) are more likely to have heart disease.
- MaxHR and HeartDisease: Maximum Heart Rate (MaxHR) has a moderate negative correlation with HeartDisease. This suggests that individuals with a higher maximum heart rate are less likely to have heart disease.
- RestingBP and Age: Resting Blood Pressure (RestingBP) shows a positive correlation with Age, suggesting that older individuals tend to have higher resting blood pressure.
- RestingECG and HeartDisease: The correlation between RestingECG and HeartDisease is relatively weak, indicating that resting ECG results may not be a strong predictor of heart disease in this dataset.

These visualizations provide valuable insights into the dataset, highlighting the distribution of key categorical variables and the relationships between various features. This information is crucial for understanding the data and preparing it for further analysis, such as training a SVM model for predicting cardiovascular health outcomes.

METHODOLOGY



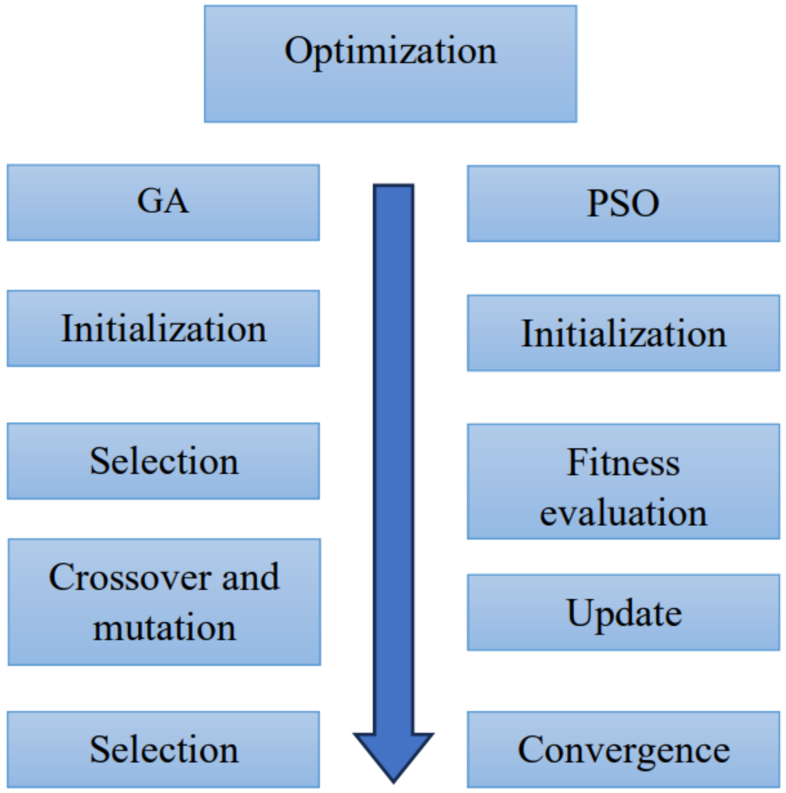
- Dataset: Kaggle heart health dataset.
- Preprocessing: Handling of missing values, coding of categorical variables, scaling.
- Models: Standard SVM, GA-optimized SVM, PSO-optimized SVM.
- Algorithms: GA uses crossover and mutation; PSO adjusts the position and velocity of particles.

Optimization algorithms

Genetic Algorithm (GA): Involves initialization, selection, crossover and mutation, and further selection to evolve better solutions.

Particle Swarm Optimization (PSO): Involves initialization, fitness evaluation, updating positions, and convergence to optimize solutions.

Both aim to improve model performance through iterative processes.



Hyperparameters of the optimization algorithms

Box 9

Table 1

Hyperparameters of the GA

Parameter	Value
Population	50
Generations	20
Crossing	0,6
Mutation	0,3
Tournament	3

Box 10

Table 2

Mutation parameters

Parameter	Value
mu	0
Sigma	1
Probability of mutating each attribute	0,2

Box 11

Table 3

Hyperparameters of the PSO

Parameter	Value
No. of particles	50
Dimensions	2
Cognitive Coefficient (c1)	0,5
Social coefficient	0,3
Lower limit	[0.1, 0.0001]
Upper limit	[10, 1]
Iterations	20

RESULTS

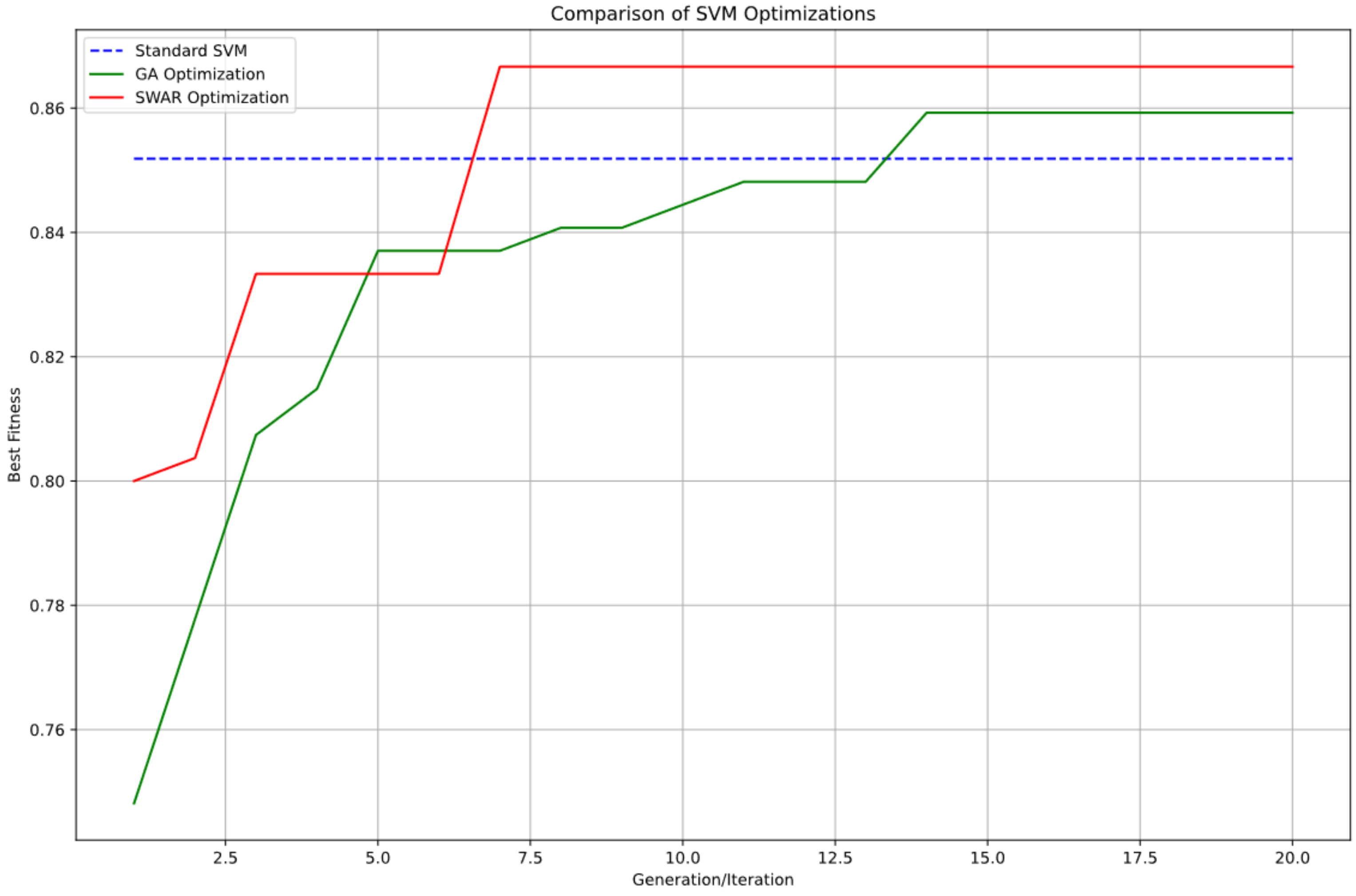
- Standard SVM: Cross-validation accuracy of 0.85.
- GA-optimized SVM: Cross-validation accuracy improved to 0.86.
- PSO-optimized SVM: Best performance with accuracy of 0.87 and F1 Score of 0.87.

Box 14

Table 5

Performance metrics

Model	Cross-Validation Accuracy	Test Accuracy	F1 Score
Standard SVM	0,851852	0,853448	0,841121
Normal SVM	0,859259	0,844828	0,83018
PSO Optimized SVM	0,822222	0,87931	0,86792



CONCLUSION

- PSO optimization shows superior results in test accuracy and F1 Score.
- SVM models can benefit from optimization techniques like GA and PSO.
- Future work: Focus on avoiding overfitting and improving generalization.
- These findings emphasize the importance of evaluating machine learning models using multiple metrics and across different phases of the optimization process to gain a comprehensive understanding of their performance.

REFERENCES

Ahmad, A. A., & Polat, H. (2023). Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. Diagnostics, 13(14). [*https://doi.org/10.3390/diagnostics13142392*](https://doi.org/10.3390/diagnostics13142392)

Immanuel, J. D., Abraham Leo, S. E., Saranya, S., Nithiya, C., Arunkumar, K., & Pradeep, D. (2024). An Intelligent Heart Disease Prediction by Machine Learning Using Optimization Algorithm. Journal of Information Technology Management, 16(1), 167–181. [*https://doi.org/10.22059/jitm.2024.96381*](https://doi.org/10.22059/jitm.2024.96381)

Kaggle. (18 de agosto de 2024). Kaggle. Obtenido de Kaggle: [*https://www.kaggle.com/datasets/ronanazarias/heart-desease-dataset*](https://www.kaggle.com/datasets/ronanazarias/heart-desease-dataset)

Krittanawong, C. Z. (2017). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology. doi: 10.1016/j.jacc.2017.03.571

Mandala, V., Surabhi, S. N. D., Balaji, V. R., Patil, D. R., Waris, S. F., & Shobana, G. (2024). Wild Horse Optimizer and Support Vector Machine (SVM) Classifier Predicts the Heart Disease Converging Nature-Motivated Optimization and Machine Learning. Journal of Angiotherapy, 8(3). [*https://doi.org/10.25163/angiotherapy.839535*](https://doi.org/10.25163/angiotherapy.839535)

Nagavelli, U. &. (2024). Prediction of heart disease and improving classifier performance using particle swarm optimization. Smart Innovation, Systems and Technologies. [*https://doi.org/10.1007/978-981-99-7711-6_19*](https://doi.org/10.1007/978-981-99-7711-6_19)

Nelson, B. B. (2023). Nature-inspired methods and machine learning algorithms for intelligent prediction of heart diseases. AIP Conference Proceedings. [*https://doi.org/10.1063/5.0128873*](https://doi.org/10.1063/5.0128873)

REFERENCES

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. Diagnostics, 14(2) https://doi.org/10.3390/diagnostics14020144

Ramkumar, M. A. (2023). Deep convolutional neural network optimized with hybrid marine predator's and nomadic people optimization for cardiac arrhythmia classification using ECG signals. Biomedical Signal Processing and Control. https://doi.org/10.1016/j.bspc.2023.105157

Sandeep Tomar, D. D. (2024). Analysis and Enhancement of Prediction of Cardiovascular Disease Diagnosis using Machine Learning Models SVM, SGD, and XGBoost. International Journal of Advanced Computer Science and Applications (IJACSA), 4. DOI: 10.14569/IJACSA.2024.0150449

Warsi, S. U. (2024). A hybrid approach for heart disease prediction using genetic algorithm and SVM. 2024 5th International Conference on Advancements in Computational Sciences (ICACS). doi: 10.1109/icacs60934.2024.10473308

ECORFAN®