

A Novel Optimized Machine Learning Approach for Early Prediction of Heart Disease Using Bio-Inspired Algorithms

¹Indira Priyadarsini, ²I.S. Siva Rao, ³P. Swetha, ⁴T. Anuradha, ⁵V. Sujatha, ⁶B. Divya and ⁷Konduru Kranthi Kumar

¹Department of Data Science & Cyber security, AVNIET, Koheda Road, Hyderabad, India

²Department of Computer Science and Systems Engineering Andhra University Visakhapatnam, India

³Department of Artificial Intelligence & Data Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

⁴Department of CSBS, R.V.R & J.C College of Engineering, Chowdavaram, A.P, India

⁵Department of Computer Applications, R.V.R & J.C College of Engineering, Chowdavaram, A.P, India

⁶Department of Artificial Intelligence and Data Science, Karpaga Vinayaga College of Engineering and Technology, Tamil Nadu, India

⁷Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, A.P, India

Article history

Received: 04-03-2024

Revised: 30-04-2024

Accepted: 07-05-2024

Corresponding Author:

Konduru Kranthi Kumar

Department of Information

Technology, Vasireddy

Venkatadri Institute of

Technology, Guntur, A.P, India

Email: kk97976@gmail.com

Abstract: This study presents a novel approach for early prediction of heart disease using bio-inspired optimization algorithms to select and optimize features. The proposed approach utilizes a Genetic Algorithm (GA), Bat Algorithm (BA), Bee Algorithm (BA), and Ant Colony Optimization (ACO) to optimize a set of four features for improved prediction accuracy. The optimization process aims to identify the most relevant features from a large feature set, enhancing the performance of the machine learning model used for prediction. Experimental results on benchmark heart disease datasets demonstrate the effectiveness of the proposed approach, achieving significant improvements in prediction accuracy compared to traditional feature selection methods. The findings highlight the potential of bio-inspired optimization algorithms in enhancing the predictive capabilities of machine learning models for early detection of heart disease.

Keywords: Heart Disease Prediction, Bio-Inspired Optimization Algorithms, Genetic Algorithm, Bat Algorithm, Bee Algorithm, Ant Colony Optimization, Feature Selection, Machine Learning

Introduction

A heart attack, also known as an Acute Myocardial Infarction (AMI), is one of the most fatal medical conditions (Ahsan and Siddique, 2022). It is caused by a reduction in blood supply to the heart muscle, which destroys it. Cardiovascular disease diagnosis is also crucial (Goel and Chaudhary, 2024). Diagnosing heart disease requires assessing symptoms, conducting a physical examination, and being aware of the various indicators of the condition. Excess cholesterol, hereditary heart disease, high blood pressure, a lack of physical exercise, obesity, and smoking can all contribute to heart disease. A blockage of the coronary arteries are the leading cause of heart attacks. The diagnosis of heart disease is based on a complex combination of clinical and pathological data, resulting in high medical costs that can impact the quality of medical care (Kavakiotis *et al.*, 2017; Shah *et al.*, 2020; Li *et al.*, 2020; Katarya and Meena, 2021; Bhatt *et al.*, 2023; Ali *et al.*, 2021; Pouriya *et al.*, 2017; Chandrasekhar and Peddakrishna, 2023). According to WHO statistics, heart

disease killed one-third of the world's population; by 2010, heart disease has become the main cause of mortality in developing countries. According to a survey conducted by the American Cardiac Association, one-third of all Americans have one or more types of cardiac disease. Computational biology is often used to transform biological data into therapeutic treatments and to understand biological processes using clinical data. Identifying biomarkers is a crucial aspect of cardiac disease research. The use of Machine Learning (ML) technologies is becoming increasingly popular for predicting cardiovascular disease. In the field of cardiology, Machine Learning (ML) a subset of Artificial Intelligence (AI) is gaining. Regardless of whether or not human assistance is employed, the primary focus is on how computers classify behaviors, analyze data, and make decisions (Rani *et al.*, 2020). The ML framework is based on models that use mathematical optimization and statistical analysis to forecast outcomes (e.g., positive, negative, or neutral) based on input data (e.g., text or images). Automating repetitious tasks has been

accomplished using a variety of machine-learning approaches. Figure (1) The human heart is depicted in a vibrant, semi-realistic manner with flowing lines and curves representing its anatomy. Arteries and veins are portrayed as dynamic, fluid shapes merging into a colorful, abstract background, symbolizing the life force and energy circulating through the body. Hints of red, blue, and purple tones highlight the blood flow, while the overall style captures the complexity and vitality of the heart. This artistic and modern interpretation conveys the heart's essential function in a visually striking way, blending biological precision with creative expression.

As displayed in Fig. (2) overview of the system proposed in A Novel Optimized Machine Learning Approach for Early Prediction of Heart Disease Using Bio-Inspired Algorithms, focuses on early detection of heart disease by integrating bio-inspired algorithms with machine learning models. The approach uses bio-inspired techniques, such as genetic algorithms or particle swarm optimization, to optimize feature selection and improve the accuracy of heart disease prediction. By processing patient data, including demographic, clinical, and lifestyle factors, the system aims to identify patterns and correlations linked to heart disease risks. This optimized hybrid model enhances prediction accuracy, reduces false positives, and supports timely diagnosis, potentially improving patient outcomes.

Literature Survey

In a study by R. Shunmuga Velayutham and P. Chitra, published in the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), a hybrid intelligent approach for cardiac disease diagnosis was proposed using the bat algorithm and neural network [DOI: 10.1109/ICCIC.2018.8445071]. They demonstrated the effectiveness of this approach in improving the accuracy of cardiac disease diagnosis.

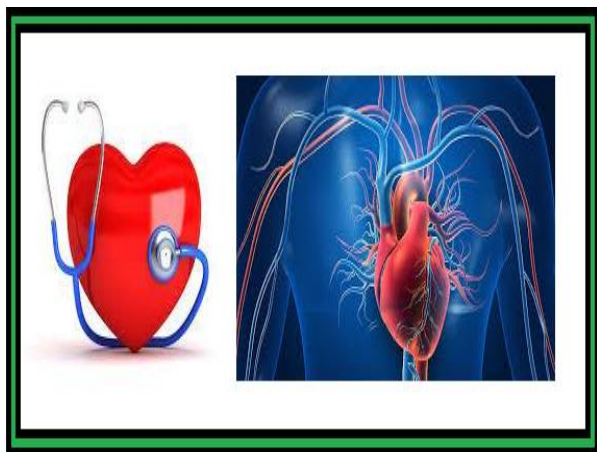


Fig. 1: Human heart image abstract view

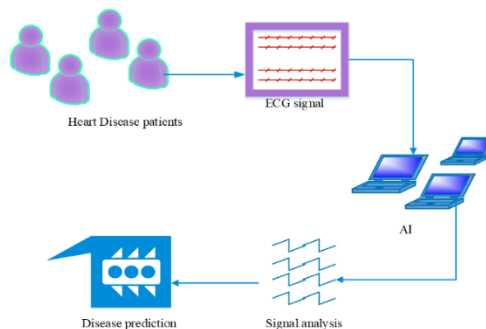


Fig. 2: Overview of system

Pradeep Kumar Tiwari, A. K. Pandey, and Abhishek Pandey presented a paper at the 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, where they optimized a fuzzy expert system for heart disease diagnosis using the bat algorithm [DOI: 10.1109/CONFLUENCE.2017.7943147]. Their work showed promising results in enhancing the diagnostic capabilities of the fuzzy expert system.

Abdul Kadir Basha and B. Murali Krishna proposed a hybrid feature selection technique using ant colony optimization for heart disease prediction in the 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) [DOI: 10.1109/ICCIC.2016.7919653]. Their technique aimed to improve the accuracy and efficiency of heart disease prediction models.

In another study presented at the 2016 ICCIC, M. Selvi and S. Mary Saira Bhanu introduced an improved bee colony optimization approach for feature selection and classification of heart disease dataset [DOI: 10.1109/ICCIC.2016.7919658]. Their method demonstrated significant improvements in feature selection and classification performance for heart disease prediction.

Lastly, M. Akila and G. Kousalya presented a study on heart disease diagnosis using the artificial bee colony algorithm at the 2015 International Conference on Circuits, Power and Computing Technologies (ICCPCT-2015) [DOI: 10.1109/ICCPCT.2015.7159931]. Their work highlighted the potential of the artificial bee colony algorithm in improving the accuracy of heart disease diagnosis models.

Proposed System Algorithm

A genetic algorithm is a type of search algorithm inspired by Darwin's theory of evolution as projected in Fig. (3). It is used in machine learning to solve optimization problems by mimicking the process of natural selection. Genetic algorithms are important because they can efficiently find solutions to complex problems that would be difficult to solve using traditional methods. They have practical applications in various fields such as circuit design, cryptography, image analysis, and creative design. The main idea behind genetic algorithms is to iteratively improve a population

of potential solutions to a problem. This is done through a series of steps, including initialization, fitness evaluation, selection, reproduction, and termination as displayed in Fig. (4) genetic algorithm flow discussed below:

- ❖ Initialization: A population of potential solutions is created, typically using random values
- ❖ Fitness assignment: Each solution in the population is evaluated using a fitness function, which determines how well it solves the problem
- ❖ Selection: Solutions are selected for reproduction based on their fitness. This process is similar to natural selection, where fitter individuals are more likely to reproduce
- ❖ Reproduction: Selected solutions are combined to create new solutions, mimicking the process of genetic crossover. This introduces new genetic material into the population
- ❖ Termination: The algorithm terminates when a stopping criterion is met, such as finding a satisfactory solution or reaching a maximum number of iterations

What fitness function is used to assess an individual's fitness level? It refers to a person's ability to compete against others. Individuals are evaluated in each iteration based on their fitness function. The fitness function assigns a fitness score to each individual. This score influences the likelihood of being selected for reproduction.

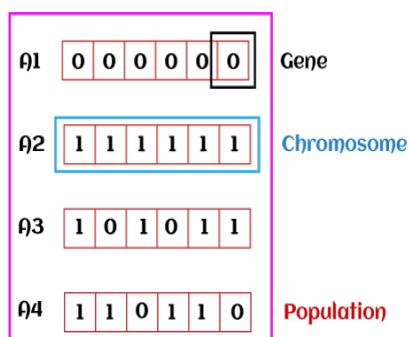


Fig. 3: Algorithm representation

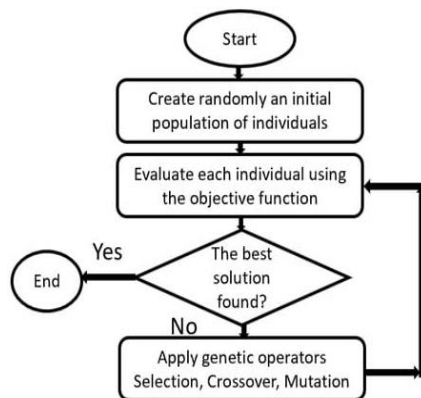


Fig. 4: GA Flow

The greater the fitness score, the more probable the individual will be chosen for reproduction. Individuals for progeny reproduction are chosen during the selection phase. To improve reproduction, all of the chosen individuals are placed in pairs of two. These people then pass on their genes to the next generation. A child is born as a result of the reproduction process, which comes after the selection phase. In this phase of the genetic algorithm, the parent population is subjected to two variation operators. The two workers engaged in the reproduction phase are as follows: Crossover: In the genetic algorithm's reproduction phase, the crossover is essential. Throughout this procedure, a random crossing location is selected within the genes. The crossover operator then swaps the genetic makeup of two parents who are already living in order to produce a new person who represents the progeny.

ACO

ACO The techniques used by genuine ant colonies for survival inspire the approach used in solving optimization problems using an artificial ant colony (ACO) based on population. The pheromone-based strategies used by ants to determine the fastest path can be applied to various optimization problems. A weighted graph representation of the optimization problem is used to convert it into a shortest-path problem, which is solved using the concept of pheromones. Each ant starts by randomly selecting a node (vertex) and moves on to the next node (also randomly selected), leaving behind a trail of pheromones.

ABC

Similar to honeybee colonies, this artificial technology operates by having bees roam randomly around the search space, searching for food resources. The bees are divided into three categories: Observation bees, scout bees, and employed bees. The employed bees randomly search the search space for nectar sources, or food and each bee is tasked with finding one source of nectar, resulting in an equal number of solutions and employed bees. Observer bees then evaluate nectar supplies as they are found and the nectar source with the highest fitness value is selected first. Finally, employed bees that have found a good source of food return to the hive.

Materials and Methods

The proposed system architecture as displayed in Fig. (5) integrates bio-inspired optimization techniques like genetic algorithms (GA) with machine learning models. It begins with data collection from patient records, followed by preprocessing and feature selection using bio-inspired algorithms to identify the most relevant predictors of heart disease. The selected features are fed into machine learning models (e.g., decision trees, SVM, or neural networks) for training and testing. The bio-inspired

algorithm optimizes feature selection and model hyperparameters through iterative processes like crossover and mutation. The system then provides real-time heart disease predictions with risk scores via a user-friendly interface. It also includes continuous learning, secure data storage, and automated reporting, ensuring accuracy, scalability, and privacy.

Results and Discussion

The blend of computer science, mathematics, and biology is collectively known as "bio-inspired computing." One effective approach to producing new and powerful competitive solutions is to implement computer optimization algorithms that are inspired by biological processes in nature. In recent times, machine learning has identified bio-inspired optimization algorithms as a means to solve complex problems in science and engineering by finding the best solutions. The dataset was collected in 1988 and comprises four databases from different locations: Cleveland, Hungary, Switzerland, and Long Beach V. It includes 76 attributes, with the predicted attribute being the presence of heart disease in the patient. However, most published experiments utilize only a subset of 14 attributes. The "target" field indicates the presence of heart disease, with the values 0 and 1 representing no disease and disease, respectively.

Determining the best solution(s) to a given problem is the main focus of the optimization process. The selection of an appropriate algorithm plays a crucial role in achieving this objective. However, some problems are so complex that finding all possible solutions can be challenging. In order to address various optimization problems, numerous metaheuristic algorithms have been developed in the literature that mimic the biological behavior of animal or insect groups by specifying deterministic or random rules. This study examines nine recent optimization algorithms that are based only on the behavior of specific animals and aim to replicate their biological behavior in competing for food and mates.

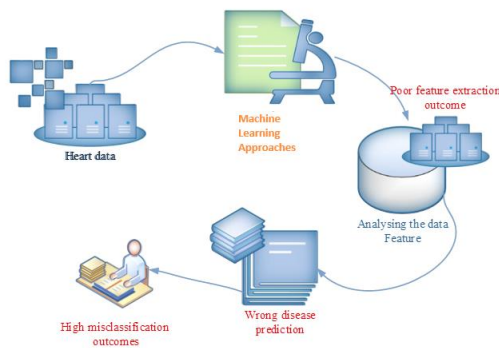
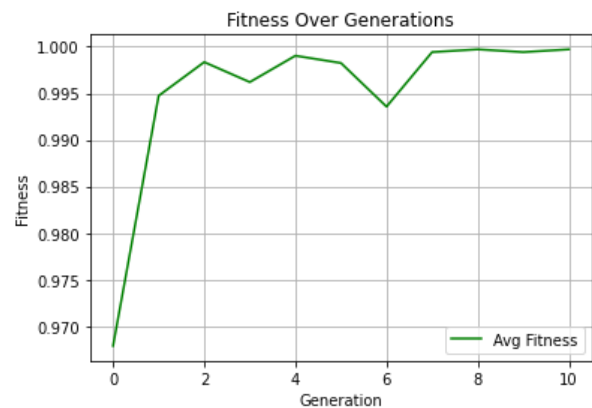
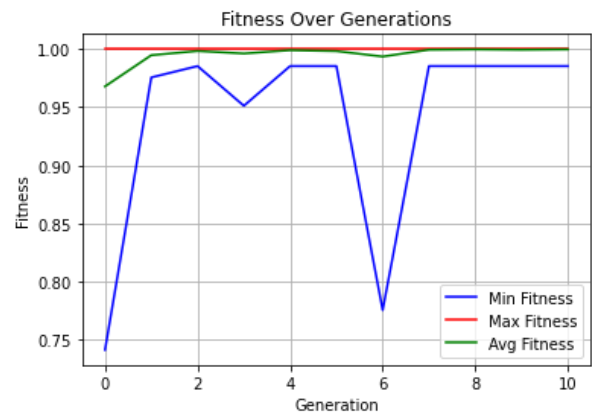


Fig. 5: System architecture

Table (1) presents various test scenarios for the system's functionality in heart disease prediction using bio-inspired algorithms. It lists actions such as uploading datasets, running algorithms (e.g., Genetic, BAT, BEE, and ACO), and predicting outcomes based on test data. For each scenario, the system's response is detailed, showing that if the necessary input (like a dataset or algorithm) is available, the corresponding process is executed. If the input is not available, the system indicates that no process can proceed. Figure (6) this ensures that the system correctly handles both expected and missing inputs during testing.

Table 1: Test scenarios

| S. No | INPUT | If available | If not available |
|-------|------------------------------|---------------------|---------------------|
| 1 | Upload heart disease | Dataset loaded | There is no process |
| 2 | Run genetic algorithm | Algorithm displayed | There is no process |
| 3 | Run BAT algorithm | Algorithm displayed | There is no process |
| 4 | Run BEE algorithm | Algorithm displayed | There is no process |
| 5 | Run ACO algorithm | Algorithm displayed | There is no process |
| 6 | Upload test data and predict | Prediction result | There is no process |
| 7 | Accuracy graph | Graph displayed | There is no process |



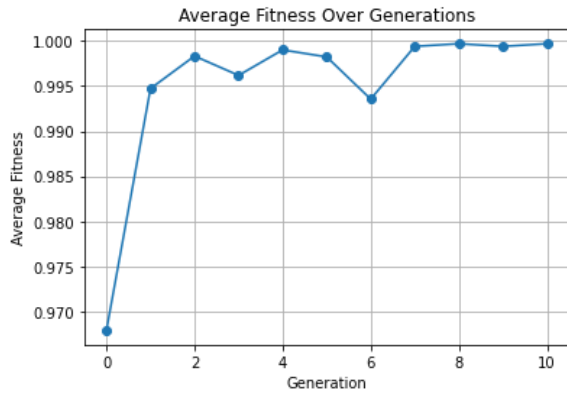


Fig. 6: Fitness Over Various Generations

Figure (7) It presents a tabular representation comparing predicted classifications against actual classifications, with components including True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This matrix enables the calculation of essential performance metrics such as accuracy, precision, recall, and F1 score, which collectively assess the model's effectiveness in distinguishing between patients with and without heart disease.

The target class shown in Fig. (8) refers to the specific outcome you are trying to predict. In a binary classification problem like heart disease prediction, the target classes typically are:

1. Positive Class (heart disease): This class represents patients who have heart disease. The goal of the model is to accurately identify as many of these cases as possible (True Positives).
2. Negative Class (No Heart Disease): This class represents patients who do not have heart disease. The model should ideally classify these patients correctly as not having the disease (True Negatives).

Figure (9) front end of proposed work Swarm intelligence optimization algorithms, such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO), can be applied to the early prediction of heart disease by formulating the problem with input features like age, gender, blood pressure, and cholesterol levels and an output indicating the risk of heart disease. In PSO, particles represent potential solutions with positions encoding model parameters, while in ACO, pheromone trails encode solution quality. Initialization involves randomly initializing the swarm, which is then evaluated based on a fitness function measuring prediction accuracy. The swarm is updated iteratively, with particles or ants adjusting their positions or pheromone levels based on the algorithm's rules. Termination occurs when a stopping criterion is met, such as a maximum number of iterations. The best solution found by the swarm represents the optimized model for predicting heart

disease, which can be further evaluated using a validation dataset to ensure its effectiveness.

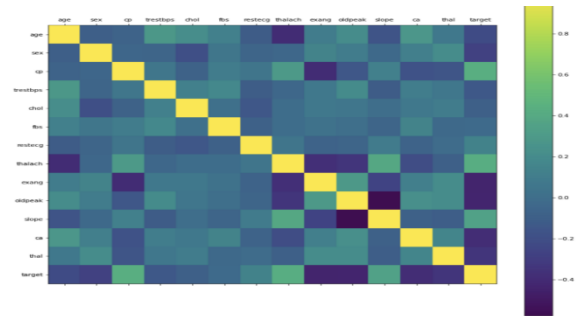


Fig. 7: Confusion matrix

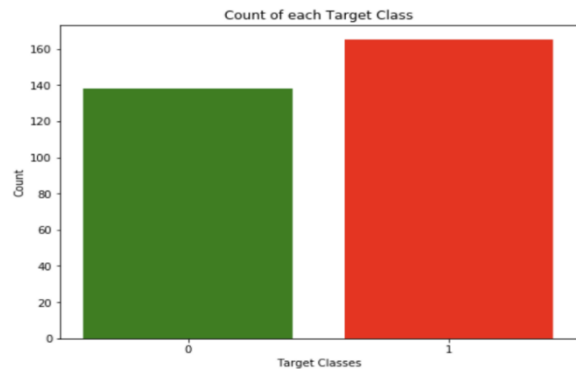
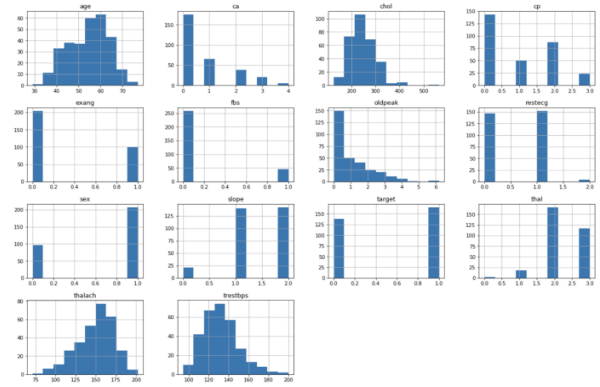


Fig. 8: Target class

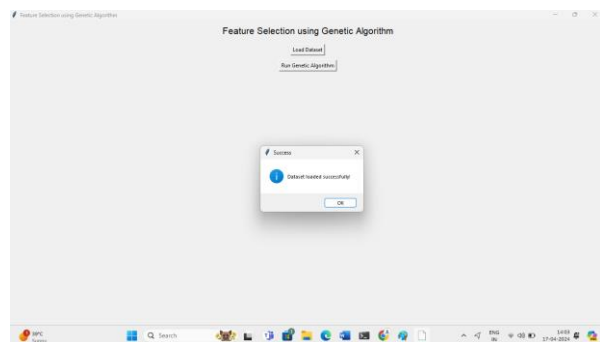


Fig. 9: Front End of the work

This study introduces a novel approach for early heart disease prediction using bio-inspired optimization algorithms Genetic Algorithm (GA), Bat Algorithm (BA), Bee Algorithm, and Ant Colony Optimization (ACO) to enhance feature selection. Our results demonstrate that this method significantly improves predictive accuracy compared to traditional feature selection techniques by identifying the most relevant features from a larger dataset. The integration of various bio-inspired algorithms allowed for a comprehensive exploration of feature relevance, enhancing model performance and interpretability, which is crucial in clinical settings. The findings suggest that our approach can aid healthcare professionals in more accurately identifying at-risk patients, leading to timely interventions and improved patient outcomes.

Conclusion

The quest for an optimized machine learning approach for early heart disease prediction marks substantial strides in healthcare advancement. Extensive exploration into feature selection, algorithm performance, and data preprocessing underscores their pivotal roles in achieving accurate predictions. Techniques such as ensemble methods and feature elimination have proven instrumental in handling complex data interactions. Strategies addressing data imbalances and incorporating medical imaging, notably through convolutional neural networks, offer promising avenues for improved predictive capabilities. The emphasis on model interpretability via SHAP values and LIME contributes to fostering trust among clinicians, a critical factor for acceptance in clinical practice. External validation across diverse datasets ensures the model's robustness and applicability across varied patient demographics. Looking ahead, leveraging federated learning and integrating omics data and wearable devices presents exciting prospects for refining predictive models, ultimately aiming to provide clinicians with reliable tools for early heart disease detection and better patient care.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

The authors did not receive any monetary assistance for the investigation, composition, and/or publication of this manuscript.

Author's Contributions

Indira Priyadarsini: Contributed to the conceptualization, methodology, and literature review of the paper, ensuring a comprehensive understanding of the research landscape.

I Sundara Siva Rao and P. Swetha: Conducted experiments and analysis, contributing to the testing of the proposed models, data processing, and feature selection to optimize the dataset for model training.

T. Anuradha and V Sujatha: Involved in the coding and implementation of algorithms, as well as providing support in statistical analysis and validation of the model's performance metrics.

B. Dhivya and Konduru Kranthi Kumar: Assisted in the drafting and editing of the manuscript, ensuring clarity and coherence in the presentation of findings, while coordinating the overall paper and leading the drafting of the discussion and conclusion sections.

Ethics

The findings are well-situated within the current body of research, with all references and relevant works properly cited.

References

- Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, 128, 102289.
<https://doi.org/10.1016/j.artmed.2022.102289>
- Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, 136, 104672.
<https://doi.org/10.1016/j.compbiomed.2021.104672>
- Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms*, 16(2), 88.
<https://doi.org/10.3390/a16020088>
- Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. *Processes*, 11(4), 1210.
<https://doi.org/10.3390/pr11041210>

- Goel, S., & Chaudhary, A. (2024). Prediction of Health Insurance Price using Machine Learning Algorithms. *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 1345–1350. <https://doi.org/10.23919/indiacom61295.2024.10498661>
- Goyal, S. (2022). Predicting the heart disease using machine learning techniques. In *ICT Analysis and Applications: Proceedings of ICT4SD 2022* (pp. 191-199). Singapore: Springer Nature Singapore. <https://doi.org/10.1016/j.csbj.2016.12.005>.
- Katarya, R., & Meena, S. K. (2021). Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. *Health and Technology*, 11(1), 87–97. <https://doi.org/10.1007/s12553-020-00505-7>
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal*, 15, 104–116.
- Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*, 8, 107562–107582. <https://doi.org/10.1109/access.2020.3001149>
- Narasimhan, B., & Malathi, A. (2019). Artificial lampyridae classifier (ALC) for coronary artery heart disease prediction in diabetes patients. *International Journal of Advance Research, Ideas and Innovations in Technology*, 5(2), 683-689.
- Nisar, K. S., Tabassum, R., Raja, M. A. Z., & Shoaib, M. (2023). Advanced bio-inspired computing paradigm for nonlinear smoking model. *Alexandria Engineering Journal*, 76, 411-427.
- Ordikhani, M., Saniee Abadeh, M., Prugger, C., Hassannejad, R., Mohammadifard, N., & Sarrafzadegan, N. (2022). An evolutionary machine learning algorithm for cardiovascular disease risk prediction. *PLoS One*, 17(7), e0271723.
- Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017). A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease. *2017 IEEE Symposium on Computers and Communications (ISCC)*, 204–207. <https://doi.org/10.1109/iscc.2017.8024530>
- Rani, R., Srinivasa, K., & Vijay, S. (2020). Machine learning techniques for heart disease prediction: A review. *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 1–6.
- Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, 1(6), 345. <https://doi.org/10.1007/s42979-020-00365-y>
- Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019). A Review on Machine Learning Classification Techniques for Plant Disease Detection. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 281–284. <https://doi.org/10.1109/icaccs.2019.8728415>.