



Optimized heart disease prediction model using a meta-heuristic feature selection with improved binary salp swarm algorithm and stacking classifier

M. Sowmiya^{a,*}, B. Banu Rekha^b, E. Malar^c

^a Department of ECE, PSG Institute of Technology and Applied Research, Coimbatore, 641062, India

^b Department of Biomedical Engineering, PSG College of Technology, Coimbatore, 641004, India

^c Department of EEE, PSG Institute of Technology and Applied Research, Coimbatore, 641062, India

ARTICLE INFO

Keywords:

Heart disease
Stacking classifier
Meta-heuristic feature selection
Salp swarm algorithm
Predictive modelling

ABSTRACT

Despite technological advancements, heart disease continues to be a major global health challenge, emphasizing the importance of developing accurate predictive models for early detection and timely intervention. This study proposes a heart disease prediction model integrating a stacking classifier with a nature-inspired meta-heuristic algorithm. It employs an improved Binary Salp Swarm Algorithm (BSSA) by incorporating a wolf optimizer and opposition-based learning for optimal feature selection. The proposed Stacking Classifier (SC) architecture features a two-tier ensemble: heterogeneous base classifiers at level 0 and a meta-learner at level 1. The BSSA is used to identify optimal features, which are then utilized to construct the stacking classifier. Experimental results demonstrate superior performance, achieving 95 % accuracy, 0.92 sensitivity, 0.97 specificity, 0.96 precision, and an F1 score of 0.95, with notably low false positive and false negative rates. Further, validation on larger datasets yielded an accuracy of 87.46 %. The feature selection process adopts a multi-objective strategy which enhances the classification accuracy and outperforms conventional techniques. The proposed method demonstrates significant potential for improving the predictive modelling in clinical settings for diagnosing heart diseases.

1. Introduction

Heart disease continues to be a leading cause of mortality worldwide, with projections indicating a significant increase in prevalence by 2030 [1]. According to the European Society of Cardiology, 3.6 million people globally received a diagnosis of a cardiovascular ailment in the last year. Clinical studies identify the following main factors that contribute to heart failure: Modifiable factors that have a significant impact are cholesterol, obesity, smoking, and physical inactivity and factors like sex, age, and family history are uncontrollable. A healthy lifestyle can help to reduce the chance of cardiovascular problems [2]. A precise diagnosis is essential due to the rising rates of heart disease incidence and mortality. Reducing death rates and the progression of the disease are critical to the clinical prognosis. Clinical prediction systems are assisted by computerized methods that help with early risk assessment and diagnosis [3]. The healthcare industry leverages the use of artificial intelligence and data mining to support clinicians. [4]. Specifically,

quantum computing is being used by researchers to reduce training time and improve forecast accuracy [5,6]. Generative Adversarial Networks are deployed to increase the data samples before classification [7].

The prediction process involves primary steps like data collection, feature selection, and categorization. Feature selection plays a vital role in identifying the relevant feature subsets and thereby minimizing the classifier performance degradation and reducing model complexity [8]. It is a key step that improves the classifier's performance and shortens the training time. Various techniques exist for feature selection, primarily filter and wrapper-based methods. Filter methods like analysis of variance, chi-square estimation, and correlation analysis employ statistical procedures to identify decision variable interdependencies [9, 10]. Wrapper techniques utilize classification algorithms to score feature subsets by selecting the subset with the highest classification score for subsequent decisions [11]. Heuristic and random search-based feature selection methods have recently gained prominence [12]. Heuristic search aims to find the optimal feature set, maximizing accuracy or

* Corresponding author.

E-mail addresses: sowmiya@psgitech.ac.in (M. Sowmiya), bbr.bme@psgitech.ac.in (B. Banu Rekha), emr@psgitech.ac.in (E. Malar).

<https://doi.org/10.1016/j.combiomed.2025.110171>

Received 4 December 2024; Received in revised form 4 April 2025; Accepted 7 April 2025

Available online 14 April 2025

0010-4825/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

metaheuristic algorithms, such as PSO, GA, and GWO, which involve additional operations like mutation, crossover, and velocity updates. Moreover, BSSA requires fewer tunable parameters compared to PSO and GWO, simplifying implementation and optimization. While the stacking approach increases computational complexity, it enhances performance and robustness by integrating multiple diverse classifiers. Although this added complexity may demand more resources, it often leads to superior predictive accuracy of 95 %.

4.10. Limitations of the proposed approach

This study is developed using UCI benchmark datasets to explore the potential of the stacking method. However, the performance of the proposed model needs to be validated on the real-time dataset. A dataset with more clinical biomarkers should also be utilized to test this study. The proposed model is computationally intensive and can be deployed using larger processing environments and cloud-based solutions to manage extensive datasets. With preprocessing provided, the model can handle heterogeneous data such as biomarkers, vital signs, and electronic health records. Feature analysis enhances interpretability in clinical settings by identifying key predictors of heart disease, enabling early detection and monitoring.

5. Conclusion

An optimized ensemble stacking classifier is presented for heart disease classification. The proposed method incorporated the optimal feature selection approach using a modified Salp swarm algorithm. Then, the selected features are classified using a two-level stacking classifier. The stacking classifier employed the traditional machine learning classifiers in the hierarchical structure. Experimentation using the UCI heart disease dataset demonstrated the proposed BSSA-SC which achieved an accuracy of 95 % and showed better performance over SVM, RF, MLP, XGB, and ADB classifiers. Incorporating an improved salp optimizer enhanced the prediction process, resulting in a 5.5 % improvement in accuracy. The suggested approach performed better than the previous methods, with accuracy gains of 49.37 %, 14.45 %, 5.55 %, 4.39 %, and 11.76 % against PSO + KNN, Dragonfly + hybrid classifier, stacked SVM, DUMF + Ensemble Classifier, and MI + MLP. In the future, hybrid optimization and deep learning classifiers might further improve the effectiveness of the proposed method. Despite the contributions, this study can be investigated for the real-time biomarkers from the affected patients.

CRedit authorship contribution statement

M. Sowmiya: Writing – original draft, Methodology, Conceptualization. **B. Banu Rekha:** Writing – review & editing, Validation, Supervision, Formal analysis, Conceptualization. **E. Malar:** Writing – review & editing, Supervision, Investigation.

Data availability

The data presented in this study are publicly available.

Ethics statement

This research utilized publicly available open-source data, and no human or animal subjects were involved in the study. All data used in this research are freely accessible from reputable sources and have been used in compliance with the terms and conditions set forth by the respective data providers.

The authors declare that there are no conflicts of interest related to this work. The manuscript has not been previously published and is not being considered for publication elsewhere.

Funding

The authors received no financial support for the research work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Ankur Kalra, Arun Pulikkottil Jose, Poornima Prabhakaran, Ashish Kumar, Anurag Agrawal, Ambuj Roy, Balram Bhargava, Nikhil Tandon, Dorairaj Prabhakaran, The burgeoning cardiovascular disease epidemic in Indians – perspectives on contextual factors and potential solutions, *The Lancet Regional Health - Southeast Asia* 12 (2023).
- [2] J.P. Li, A.U. Haq, S.U. Din, J. Khan, A. Khan, A. Saboor, Heart disease identification method using machine learning classification in e-healthcare, *IEEE Access* 8 (2020) 107562–107582.
- [3] N.L. Fitriyani, M. Syafrudin, G. Alfian, J. Rhee, HDPM: an effective heart disease prediction model for a clinical decision support system, *IEEE Access* 8 (2020) 133034–133050.
- [4] Mohd Javaid, Abid Haleem, Ravi Singh, Rajiv Suman, Shanay Rab, Significance of machine learning in healthcare: features, pillars, and applications, *International Journal of Intelligent Networks* 3 (2022), <https://doi.org/10.1016/j.ijin.2022.05.002>.
- [5] Saud Alotaibi, Hanan Mengash, Sami Dhabhi, Sana Alazwari, Radwa Marzouk, Mimouna Alkhonaini, Abdullah Mohamed, Anwer Hilal, Quantum-enhanced machine learning algorithms for heart disease prediction, *Human-centric Computing and Information Sciences* (2023), <https://doi.org/10.22967/HGIS.2023.13.041>.
- [6] Huda Enad, Mazin Mohammed, A review on artificial intelligence and quantum machine learning for heart disease diagnosis: current techniques, challenges and issues, recent developments, and, *Future Directions* 11 (2023) 8–25, <https://doi.org/10.54216/FPA.110101>.
- [7] R.R. Sarra, A.M. Dinar, M.A. Mohammed, M.K.A. Ghani, M.A. Albahar, A robust framework for data generative and heart disease prediction based on efficient deep learning models, *Diagnostics* 12 (12) (2022) 2899, <https://doi.org/10.3390/diagnostics12122899>.
- [8] José Sanz, Mikel Galar, Aranzazu Jurio, Antonio Brugos, Miguel Pagola, Humberto Sola, Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system, *Appl. Soft Comput.* 20 (2014) 103–111, <https://doi.org/10.1016/j.asoc.2013.11.009>.
- [9] C. Kuzudisli, B. Bakir-Gungor, N. Bulut, B. Qaqish, M. Yousef, Review of feature selection approaches based on grouping of features, *PeerJ* 11 (2023 Jul 17) e15666, <https://doi.org/10.7717/peerj.15666>. PMID: 37483989; PMCID: PMC10358338.
- [10] G. Chandrashekar, F. Sahin, A survey on feature selection methods, *Comput. Electr. Eng.* 40 (2014) 16–28.
- [11] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1997) 273–324.
- [12] Tansel Dökeröglü, Firdevsi Ayça Deniz, Hakan Kızılöz, A comprehensive survey on recent metaheuristics for feature selection, *Neurocomputing* 494 (2022) 269–296, <https://doi.org/10.1016/j.neucom.2022.04.083>, 2022.
- [13] M. Sharma, P. Kaur, A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem, *Arch. Comput. Methods Eng.* 28 (2021) 1103–1127.
- [14] Dongwan Lu, Yinggao Yue, Zhongyi Hu, Minghai Xu, Yinsheng Tong, Hanjie Ma, Effective detection of Alzheimer's disease by optimizing fuzzy K-nearest neighbors based on salp swarm algorithm, *Comput. Biol. Med.* 159 (2023), <https://doi.org/10.1016/j.combiomed.2023.106930>.
- [15] Nilesh Kunhare, Ritu Tiwari, Joydip Dhar, Intrusion detection system using hybrid classifiers with meta-heuristic algorithms for the optimization and feature selection by genetic algorithm, *Comput. Electr. Eng.* 103 (2022), <https://doi.org/10.1016/j.compeleceng.2022.108383>.
- [16] P. Swetha, D.S. Rao, Effective feature selection-based meta-heuristics optimization approach for spam detection, *SN COMPUT. SCI.* 4 (2023) 681, <https://doi.org/10.1007/s42979-023-02126-z>.
- [17] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (2003) 1157–1182, <https://doi.org/10.1162/1532443032753616>.
- [18] Majdi Mafarja, Seyedali Mirjalili, Whale optimization approaches for wrapper feature selection, *Appl. Soft Comput.* 62 (2018) 441–453, <https://doi.org/10.1016/j.asoc.2017.11.006>.
- [19] M. Rostami, K. Berahmand, E. Nasiri, S. Forouzandeh, Review of swarm intelligence-based feature selection methods, *Eng. Appl. Artif. Intell.* 100 (2021) 104210.
- [20] Yuaning Liu, Gang Wang, Huiling Chen, Hao Dong, Xiaodong Zhu, Sujing Wang, An improved particle swarm optimization for feature selection, *JBE* 8 (2011) 191–200, [https://doi.org/10.1016/s1672-6529\(11\)60020-6](https://doi.org/10.1016/s1672-6529(11)60020-6).
- [21] Hossam Zawbaa, Eid Emary, Aboul Ella Hassanien, Binary grey wolf optimization approaches for feature selection, *Neurocomputing* 172 (2016) 371–381, <https://doi.org/10.1016/j.neucom.2015.06.083>.

- [22] Abdelaziz Hammouri, Majdi Mafarja, Mohammed Al-Betar, Mohammed Awadallah, Iyad Abu Doush, An improved Dragonfly Algorithm for feature selection, *Knowl. Base Syst.* 203 (2020) 106131, <https://doi.org/10.1016/j.knosys.2020.106131>.
- [23] Majdi Mafarja, Seyedali Mirjalili, Whale optimization approaches for wrapper feature selection, *Appl. Soft Comput.* 62 (2017), <https://doi.org/10.1016/j.asoc.2017.11.006>.
- [24] Ali Asghar Heidari, Hossam Faris, Ibrahim Aljarah, Majdi Mafarja, Huiling Chen, Harris hawks optimization: algorithm and applications, *Future Gener. Comput. Syst.* 97 (2019), <https://doi.org/10.1016/j.future.2019.02.028>. <https://aliasgharheidari.com/HHO.html.849-872>.
- [25] M. Zivkovic, C. Stoean, A. Chhabra, N. Budimirovic, A. Petrovic, N. Bacanin, Novel improved 3. <https://doi.org/10.3390/s22051711>, 2022.
- [26] M. Mijwil, Alaa K. Faeq, Mohammad Aljanabi, Early detection of cardiovascular disease utilizing machine learning techniques: evaluating the predictive capabilities of seven algorithms, *Iraqi Journal for Computer Science and Mathematics* 5 (1) (Feb. 2024) 263–276.
- [27] Ahmed Dinar, Mazin Mohammed, Karrar Abdulkareem, Raniya Sarra, Enhanced heart disease prediction based on machine learning and χ^2 statistical optimal feature selection model, *Design* 6 (2022) 87, <https://doi.org/10.3390/designs6050087>.
- [28] Burak Kolkusa, Burcu Bakir-Gungor, Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis, *Comput. Stand. Interfac.* 84 (2023) 103706, <https://doi.org/10.1016/j.csi.2022.103706>.
- [29] M. Sowmiya, B. Banu Rekha, E. Malar, Ensemble classifiers with a hybrid feature selection approach for the diagnosis of coronary artery disease, *The Scientific Temper* 14 (3) (2023) 726–734, <https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.3.24>.
- [30] Vikas Chaurasia, Aparna Chaurasia, Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique, *Biomedical Materials & Devices* 1 (2023), <https://doi.org/10.1007/s44174-022-00060-x>.
- [31] S. Nandy, M. Adhikari, V. Balasubramanian, et al., An intelligent heart disease prediction system based on a swarm-artificial neural network, *Neural Comput. Appl.* 35 (2023) 14723–14737, <https://doi.org/10.1007/s00521-021-06124-1>.
- [32] C.B. Gokulnath, S.P. Shanharajah, An optimized feature selection based on genetic approach and support vector machine for heart disease, *Clust. Comput.* 22 (Suppl 6) (2019) 14777–14787, <https://doi.org/10.1007/s10586-018-2416-4>.
- [33] J. Nourmohammadi-Khiarak, M.R. Feizi-Derakhshi, K. Behrouzi, et al., New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection, *Health Technol.* 10 (2020) 667–678, <https://doi.org/10.1007/s12553-019-00396-3>.
- [34] S. Balasubramanian, Relief-based feature selection and gradient Squirrel search algorithm enabled deep maxout network for detection of heart disease, *Biomed. Signal Process Control* 87 (2024) 1–8, <https://doi.org/10.1016/j.bspc.2023.105446>.
- [35] T Sheeba Paul, Deepjyoti Roy, Mohammad Haider Syed, A metaheuristic-enabled training system for ensemble classification technique for heart disease prediction, *Adv. Eng. Software* 174 (2022), <https://doi.org/10.1016/j.advengsoft.2022.103297>.
- [36] Sireesha Moturi, S.N. Tirumala Rao, Srikanth Vemuru, Grey wolf assisted dragonfly-based weighted rule generation for predicting heart disease and breast cancer, *Comput. Med. Imag. Graph.* 91 (2021), <https://doi.org/10.1016/j.compmedimag.2021.101936>.
- [37] Renji P. Cherian, Noby Thomas, Sunder Venkitachalam, Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm, *J. Biomed. Inf.* 110 (2020), <https://doi.org/10.1016/j.jbi.2020.103543>.
- [38] Q. Al-Tashi, H. Rais, S. Jadid, Feature selection method based on grey wolf optimization for coronary artery disease classification, in: F. Saeed, N. Gazem, F. Mohammed, A. Busalim (Eds.), *Recent Trends in Data Science and Soft Computing*. IRICT 2018. *Advances in Intelligent Systems and Computing*, vol. 843, Springer, Cham, 2019, https://doi.org/10.1007/978-3-319-99007-1_25.
- [39] Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra, Shubhendu Kumar Sarangi, A hybrid machine learning algorithm for heart and liver disease prediction using modified particle swarm optimization with support vector machine, *Procedia Comput. Sci.* 218 (2023) 818–827, <https://doi.org/10.1016/j.procs.2023.01.062>.
- [40] S. Balasubramanian, K. Satheesh Kumar, V. Kavitha, A. Prasanth, T.A. Sivakumar, Feature selection and dwarf mongoose optimization enabled deep learning for heart disease detection, *Comput. Intell. Neurosci.* (2022) 2819378, <https://doi.org/10.1155/2022/2819378>, 2022 Dec 7; PMID: 36531922; PMCID: PMC9750776.
- [41] Q. Al-Tashi, S.J. Abdul Kadir, H.M. Rais, S. Mirjalili, H. Alhussian, Binary optimization using hybrid grey wolf optimization for feature selection, *IEEE Access* 7 (2019) 39496–39508, <https://doi.org/10.1109/ACCESS.2019.2906757>.
- [42] Jaishri Wankhede, S. Palaniappan, S. Magesh, Heart Disease Prediction Using Hybrid Random Forest Model Integrated with Linear Model, 2021, <https://doi.org/10.3233/APC210160>.
- [43] L. Ali, et al., An optimized stacked support vector machines based expert system for the effective prediction of heart failure, *IEEE Access* 7 (2019) 54007–54014, <https://doi.org/10.1109/ACCESS.2019.2909969>.
- [44] Serhat Kılıçarslan, PSO + GWO: a hybrid particle swarm optimization and Grey Wolf optimization based Algorithm for fine-tuning hyper-parameters of convolutional neural networks for Cardiovascular Disease Detection, *J. Ambient Intell. Hum. Comput.* 14 (2022) 1–11, <https://doi.org/10.1007/s12652-022-04433-4>.
- [45] K. Velswamy, R. Velswamy, I.T.J. Swamidason, et al., Classification model for heart disease prediction with feature selection through modified bee algorithm, *Soft Comput.* 26 (2022) 13049–13057, <https://doi.org/10.1007/s00500-021-06330-y>.
- [46] Animesh Dubey, Dr Amit Sinhal, Richa Sharma, Heart disease classification through crow intelligence optimization-based deep learning approach, *Int. J. Inf. Technol.* 16 (2023), <https://doi.org/10.1007/s41870-023-01445-x>.
- [47] Ritesh Sonawane, Hitendra Patil, Automated heart disease prediction model by hybrid heuristic-based feature optimization and enhanced clustering, *Biomed. Signal Process Control* 72 (2022) 103260, <https://doi.org/10.1016/j.bspc.2021.103260>.
- [48] Ehsan Roshan, Michael Kattan, Random forest swarm optimization-based for heart diseases diagnosis, *J. Biomed. Inf.* 115 (2021) 103690, <https://doi.org/10.1016/j.jbi.2021.103690>.
- [49] Dileep Pulugu, Kunjam Rao, Bodapati Prajna, Sitaratnam Gokuruboyina, Revathy Peddi, Amit Grover, Anu Sheetal, An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm, *Neural Comput. Appl.* 35 (2023), <https://doi.org/10.1007/s00521-022-07064-0>.
- [50] A. Janosi, W. Steinbrunn, M. Pfisterer, R. Detrano, Heart disease, UCI Machine Learning Repository (1989). <https://doi.org/10.24432/C52P4X>.
- [51] Statlog (Heart). UCI Machine Learning Repository. <https://doi.org/10.24432/C57303>.
- [52] Seyedali Mirjalili, Amir H. Gandomi, Seyedeh Zahra Mirjalili, Shahrzad Saremi, Hossam Faris, Seyed Mohammad Mirjalili, Salp Swarm Algorithm, A bio-inspired optimizer for engineering design problems, *Adv. Eng. Software* 114 (2017) 163–191, <https://doi.org/10.1016/j.advengsoft.2017.07.002>.
- [53] M. Braveen, S. Nachiyappan, R. Seetha, et al., Retracted article: ALBAE feature extraction based lung pneumonia and cancer classification, *Soft Comput.* 28 (Suppl 2) (2024) 589, <https://doi.org/10.1007/s00500-023-08453-w>.
- [54] R. Subramanian, P. Aruchamy, An effective speech emotion recognition model for multi-regional languages using threshold-based feature selection algorithm, *Circ. Syst. Signal Process.* 43 (2024) 2477–2506, <https://doi.org/10.1007/s00034-023-02571-4>.
- [55] Jabbar Akhil, Prediction of heart disease using k-nearest neighbor and particle swarm optimization, *Biomed. Res.* 18 (9) (2017).
- [56] M.T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Avelaira-Mata, J. A. Benítez-Andrades, Heart disease risk prediction using deep learning techniques with feature augmentation, *Multimed. Tool. Appl.* (2023) 1–15.
- [57] Abdulwahab Almazroi, Eman Aldhahri, Saba Bashir, Sufyan Ashfaq, A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning, 2023, <https://doi.org/10.1109/ACCESS.2023.3285247>. IEEE Access. PP. 1–1.
- [58] J. Mishra, M. Tiwari, IoT-enabled ECG-based heart disease prediction using three-layer deep learning and meta-heuristic approach, *SIVIP* 18 (2024) 361–367, <https://doi.org/10.1007/s11760-023-02743-4>.