Electronic Integrated Computer Algorithm Journal
Vol. 3, No. 1, October 2025, pp 11~16

ISSN: 3031-0350, DOI: 10.62123/enigma.v3i1.98

Heart Attack Risk Prediction Using Machine Learning: A Comparative Study of Decision Tree and K-Nearest Neighbors

Fauzi Hizbullah[1*], Alif Noorachmad Muttaqin[1], Rahardian Andiharsa Sih Setiarto[1], Rizki Aulia Hakim[1], Sahidan Abdulmana[2]

[1] *Master of Information System Study Program, School of Industrial Engineering, Telkom University, Main Campus (Bandung Campus), Jl. Telekomunikasi no. 1, Bandung 40257,* West Java, Indonesia.
[2] *Department of Data Science and Analytics, Faculty of Science and Technology, Fatoni University, หมู่ 3 135/8 Khao Tum, Yarang District, Pattani 94160,* Thailand.

DOI : 10.62123/enigma.v3i1.98

Received : August 08, 2025
Revised : September 28, 2025
Accepted : September 30, 2025

Keywords:
Heart Attack Prediction, Machine Learning, Decision Tree, K-Nearest Neighbors, Orange Data Mining

ABSTRACT

Heart disease, particularly heart attacks, is a leading cause of death worldwide, highlighting the importance of early detection and risk prediction. This study develops and evaluates machine learning models to predict heart attack risk using seven health-related attributes: age, marital status, gender, body weight category, cholesterol level, participation in stress management training, and stress level. The dataset, processed with the Orange Data Mining platform, was divided into training (66%) and testing (34%) sets. Two supervised algorithms, Decision Tree and K-Nearest Neighbors (K-NN), were implemented without extensive hyperparameter tuning. Model performance was evaluated using accuracy, precision, recall, and F1 score. The Decision Tree achieved the best results with 84.78% accuracy, 88.52% precision, 79.41% recall, and 83.72% F1 score, indicating its effectiveness in identifying at-risk individuals. Key predictors included age, stress level, and cholesterol, aligning with established medical findings. While the results are promising, limitations include a small dataset and limited algorithm scope. Future research should expand the dataset, include additional clinical features, and explore advanced algorithms to improve accuracy and reduce false negatives, enhancing applicability in preventive healthcare.

1. INTRODUCTION

Heart disease has been causing deaths globally for a very long time, and heart attack is the most frequent sudden manifestation of the disease. Most cardiovascular disease-related deaths annually would have been avoided when discovered at the initial stages, states the World Health Organization, thus focused interventions [1][2]. The task is to detect high-risk individuals before symptoms arise, especially when risk factors like stress, cholesterol level, and lifestyle are typically unavailable early. Predictive modeling emerged as a key innovation area as healthcare systems become more subjected to pressure to deliver more preventive and personalized care [3][4]. To achieve this, machine learning (ML) has proven to be a useful tool for interrogation of large and complex health databases. ML algorithms can detect non-linear relationships and variable interaction which cannot be detected using standard statistical methods. Such algorithms have been widely applied across several biomedical applications including the detection of cancer, diabetes risk assessment, and more recently, heart disease [5][6][7]. By feeding clinical and behavioral health data into classification models, researchers can develop models that predict the likelihood of future instances of health events, which would in turn enable doctors and health clinicians to intervene sooner and more effectively. A series of recent publications have demonstrated the potential of utilizing algorithms such as Decision Tree, Support Vector Machine, and K-Nearest Neighbors to predict cardiovascular risk with promising results [8][9].

This study developed a heart attack risk prediction model using machine learning techniques for example Decision Tree and K-Nearest Neighbors (K-NN) [10]. The data set utilized here includes seven significant health indicators: age, marital status, gender, body weight group, cholesterol level, training in stress management, and stress level. These variables have been used using prior studies and expert knowledge since they are significant to cardiac health. Data was analyzed with Orange, a free data mining computer package with which visual inspection is easy and made for. Through measurement of model performance based on the likes of accuracy, precision, recall, and F1-score, this paper seeks to establish the application of machine learning models in heart attack risk prediction, in line with advancing the whole undertaking of integrating data-intensive approaches in preventive medicine.

2. LITERATURE REVIEW

2.1 Machine Learning for Heart Disease Prediction

*Corresponding Author Email: fhizbullah@student.telkomuniversity.ac.id

Machine learning is a new methodology used in cardiovascular disease prediction with its unmatched capability in risk estimation and early detection. The application of computational intelligence techniques in the diagnosis of heart disease is paradigmatic with respect to conventional clinical assessment methods to data-driven predictive models. State-of-the-art research now conclusively demonstrates that machine learning software can outperform more conventional risk assessment computer programs by identifying leads to cardiovascular pathology by exploiting complex patterns in large-scale healthcare data sets. By combining heterogenous algorithmic approaches ranging from classic supervised learning methods to cutting-edge architectures in deep learning, a compelling foundation was thus set in the development of accurate, scalable, and clinically meaningful models for predicting the prospects of improving patient outcomes through early intervention [11][12][13].

The groundwork of machine learning algorithms used in cardiovascular disease prediction consists of a wide range of algorithmic approaches, each with their respective computational efficacies and merits. Random Forest is the most widely used algorithm, utilized in approximately 25% of all research works published thus far and which has a habit of yielding high performance measures on diverse datasets [14]. Traditional supervised learning methods such as Support Vector Machines, Logistic Regression, and Decision Trees form the basis of predictive modeling, while ensemble methods such as Gradient Boosting, AdaBoost, and XGBoost have demonstrated higher capability in detecting complex non-linear relationships among cardiovascular risk factors. The shift towards deep learning architecture introduced powerful neural network models like Convolutional Neural Networks for processing ECG signals, Long Short-Term Memory networks for identifying temporal patterns, and CNN-LSTM hybrid architectures that achieve over 98% accuracy in certain areas. Feature engineering techniques, such as Recursive Feature Elimination with Principal Component Analysis, played a crucial role in optimizing the model through the identification of the most informative cardiovascular biomarkers without contributing to computational complexity [15][16]. Current advances in federated learning and privacy-preserving machine learning techniques provided significant groundwork in protecting patient data while facilitating joint model development across institutions [17].

2.2 Overview of Selected Algorithms

Decision Trees and K-Nearest Neighbors (KNN) are two fundamental supervised learning paradigms which have gained significant significance in medical diagnosis applications. Decision Trees are hierarchical rule-based classifiers that partition the data by a series of binary tests on feature thresholds, creating an easy-to-understand tree structure with every internal node a feature test, branches as test results, and leaf nodes as class predictions. The working of the algorithm is recursive binary partitioning of the data using metrics such as Gini impurity or entropy to discover optimal decision boundaries that will lead to maximum information gain at a node. K-Nearest Neighbors, by contrast, is an instance-based learning algorithm that classifies new instances by determining the most similar training instances in the feature space and then classifying most frequent class among these neighbors. The KNN algorithm relies on distance measures, typically Euclidean distance, to determine similarity between cases, making predictions based on the premise that similar cases should have similar outcomes, which fits clinical judgment where patients with similar symptoms will have similar diagnoses [18][19][20].

Both algorithms possess special advantages that make them particularly suitable for medical applications, while each respond to differing clinical needs and constraints. Decision Trees possess better clinical interpretability in the sense that they are clear, rule-based decision flow diagrams that trace clinical decision-making processes and are easily understandable by health care practitioners without advanced machine learning expertise. Their ability to handle both categorical and continuous variables, address missing values effectively, and generate actionable clinical rules makes them indispensable for the development of clinical decision support systems. K-Nearest Neighbors demonstrates improved performance in applications to personalized medicine where treatment recommendations must be customized based on individual patient characteristics because it was used in "patients-like-me" algorithms that identify the closest previous cases for the purpose of treatment in a bid to direct. The non-parametric nature of the algorithm enables it to pick up complex, non-linear patterns in medical information without assuming distributions in the underlying data, making the algorithm ideal for rare diseases or heterogeneous presentations of conditions where standard statistical models will not work. However, KNN's computational expense during prediction and its sensitivity to k-value and distance metric choice necessitate accurate optimization, while Decision Trees are susceptible to overfitting and instability, which can be alleviated with ensemble methods and proper pruning [21][22][23].

2.3 Tools and Platforms in Health Data Analysis

Orange Data Mining is a health data analysis paradigm change that introduces a component-based visual programming environment that demystifies the machine learning and data mining capabilities for healthcare professionals without requiring extensive programming knowledge [24][25]. Orange stands out due to its visually intuitive drag-and-drop interface where complex analytical processes are constructed by connecting pre-specified widgets that have specialized data processing, analysis, and visualization capabilities. The widget-based architecture of the platform provides end-to-end support for clinical applications like data import and preprocessing, statistical modeling, application of machine learning algorithms, and visualization using cutting-edge techniques, all via a graphical user interface eliminating the usual barriers of script-based programming [26][27]. The performance of Orange in healthcare applications has been demonstrated through numerous applications, from analyzing the COVID-19 data that classified with 82% accuracy using Naive Bayes algorithms to cardiovascular disease prediction research being performed using k-means, hierarchical, and density-based clustering. Platform semantics that are clear when it comes to well-recognized graphical icons and double coding ideas coupled with automatic textual marking offer an intellectually transparent system that is not only better for pedagogical use but professional clinical research [28].

The selection of Orange as a healthcare data analysis tool is strategically justified by its integration of access, comprehensiveness, and clinical applicability similar to the specific needs of healthcare data science. Compared to the more

formally programming-focused systems such as R or Python, Orange has immediate graphical feedback and interpretable outcomes with minimal effort, making it highly suited to interdisciplinary healthcare teams where clinical subject-matter experts are not necessarily computationally adept. Open-source platform availability ensures cost-effectiveness as well as transparency, which are needed in cost-limited healthcare environments that require explainability regulatory compliance for algorithms. The versatile set of Orange algorithms support all phases of healthcare data mining tasks, including disease prediction with classification, patient stratification with clustering, outcome prediction with regression, and clinical pattern discovery with association rule mining. The tool's established record of success in medical use, evidenced by successful uses across subject domains from the prediction of nitinol alloy performance (such as extremely successful R^2 outcomes via k-NN algorithms) to health-oriented social media data sentiment analysis, establishes it as an established and overall healthcare informatics instrument. Additionally, the educational nature of Orange makes it a great tool for teaching data science methodology to clinicians, instilling a culture of data in clinics while maintaining the interpretability required for evidence-based medicine decision-making [29][30].

3. RESEARCH METHODS

There was a sequential process followed by this study with data harvesting, preprocessing, model training, and performance evaluation. The overall aim was to develop a prediction model that would determine individuals at risk of heart attack based on selected health predictors using machine learning methods. The overall procedure is illustrated in Figure 1.

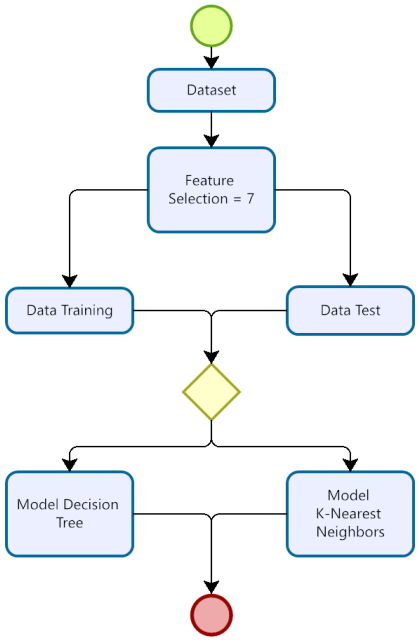

Figure 1. Research Stages

Data is made up of structured health-related data obtained from diverse individuals. It contains a set of predictors pre-specified in clinical research to be predictive of cardiovascular health, i.e., age, marital status, sex, body weight category, cholesterol level, presence or absence at stress management training, and stress level. They are physiological, psychological, and lifestyle predictors, which have been established to be important risk factors for heart attacks. Table 1 kept all seven attributes available from the dataset. Preprocessing of the data and machine learning modeling were both done with Orange. Orange is an integrative visual programming software also involved in data mining and other machine learning operations.

Table 1. Parameters

Parameters	Type
Age	Numeric
Marital Status	Numeric
Gender	Categorical
Weight Category	Numeric
Cholesterol	Numeric
Stress Management Training	Categorical
Stress Level	Numeric

The simplicity of Orange makes it a favorite among an enormous number of users in schools and universities as it possesses data transformation facilities, training of models, and visualization of results. Preprocessing of data was done prior to model training to achieve consistency and readiness for classification. The gender categorical and stress management participation variables were

converted to numerical variables through one-hot encoding with Orange's preprocessing widgets. The parameters in the dataset were not dropped since all of them were assumed to be potentially valuable to the task of prediction. The dataset consisted of 120 individual records. The data were randomly split into training (66%, 79 records) and testing (34%, 41 records) subsets. In this way, most of the data would be used in training the models and losing not too large a part to independent testing. The two supervised machine learning algorithms used to solve this problem were Decision Tree and K-Nearest Neighbors. The two algorithms were chosen because they are interpretable, can handle classification problems, and have been shown previously to be successful in solving medical prediction problems. Both models were constructed with Orange visual widgets without aggressively hyperparameter tuning, relying on the defaults to simulate a standard-use scenario. For evaluating the performance of the predictive models, four standard classification metrics were utilized: accuracy, precision, recall, and F1 score. Accuracy returns the proportion of overall correct predictions, precision returns the proportion of correct positive predictions out of all positive predictions, recall is the ability of the model to predict all positive instances with correctness, and F1 score is the harmonic mean between recall and precision. These indicators are a good balanced measure of the quality of performance of the model, particularly in medical diagnosis applications where false positives and negatives are catastrophic. With this approach, the study aims to be able to discern the possibility and effect of machine learning solutions in providing early detection of risk of heart attack, and by doing so add to data-driven studies of health solutions.

4. RESULTS AND DISCUSSIONS

4.1 Model Performance Evaluation

The models that were trained, namely Decision Tree and K-Nearest Neighbors (K-NN), were compared against the test dataset. Data was split using a 66:34 test/train ratio. Accuracy, precision, recall, and F1-score are the criteria for measurement. These were achieved from confusion matrix results in Orange.

Table 2. Model Evaluation Metrics

Matrix	Score
Accuracy	84,78%
Precision	88,52%
Recall	79,41%
F1 Score	83,72%

4.2 Confusion Matrix Analysis

The confusion matrix provides us with more information about the classification outcome. For this specific example, the Decision Tree model produced:

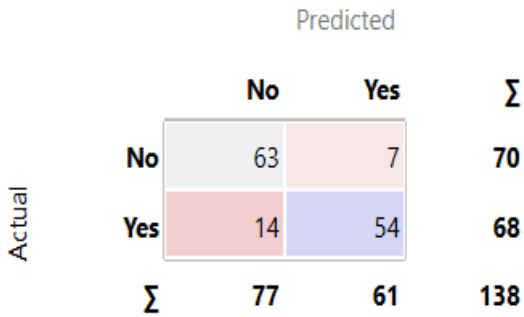

Figure 2. Confusion Matrix

The false negatives are of particular concern in healthcare scenarios, as they are patients who are at risk but classified as low risk. False negatives must be kept to a minimum to avoid missed diagnoses.

4.3 Parameters Importance and Model Interpretability

A good thing about the Decision Tree model is that it is easy to interpret. The tree model identifies which variables most heavily affected predictions. From the tree plot, age, stress level, and cholesterol were the splitting variables most frequently utilized.

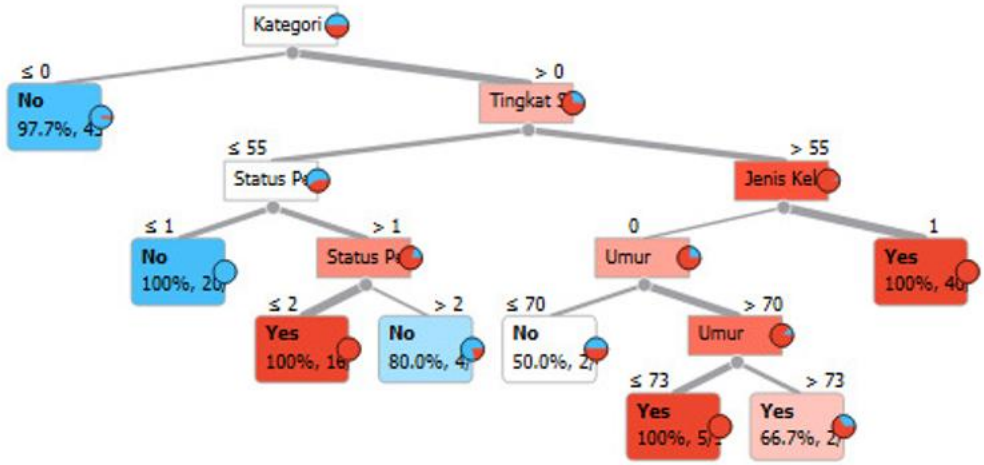

Figure 3. Decision Tree Result

4.4 Tool Utilization and Practical Application

Orange was used as the frontend for data preprocessing, model training, and result visualization. Graphical, interactive frontend of Orange allowed easy model implementation without programming. Pre-built widgets for classification, test & score, confusion matrix, and visualization provided quick model verification. Orange also allows easy comparison of models. Decision Tree was marginally better than K-NN on all the measures in this research work. However, K-NN was still competitive and may be even more optimized with refinement of the K value or application of parameters scaling techniques.

4.5 Limitations and Consideration

Some things are worth noting for the limitations. The dataset used was small and may not represent the full spectrum of patient populations. Furthermore, only two algorithms of machine learning were tried out with default settings of parameters. Stronger algorithms like Random Forest or Gradient Boosting, and hyperparameter optimization, might further improve results. Further, the parameters used here do not contain clinical variables like blood pressure, glucose, or family history. Adding more medically specific information might improve prediction power and model generalization.

5. CONCLUSSION

This study developed and evaluated Decision Tree and K-Nearest Neighbors (K-NN) models to predict heart attack risk based on seven health-related attributes using the Orange Data Mining platform. The Decision Tree achieved the best performance, with 84.78% accuracy, 88.52% precision, 79.41% recall, and 83.72% F1 score, demonstrating the feasibility of applying machine learning for early risk detection using simple health indicators. Key predictors such as age, stress level, and cholesterol were consistent with established medical findings, reinforcing their importance in cardiovascular risk assessment. Nevertheless, the research is limited by its small dataset, lack of hyperparameter optimization, and restricted algorithm scope. Future work should involve larger and more diverse datasets, additional clinical parameters, advanced algorithms such as Random Forest or Gradient Boosting, and optimization techniques to improve accuracy and reduce false negatives, which are critical in medical diagnostics.

Despite the promising results, this study has several limitations. First, the dataset used was relatively small and therefore may not fully represent diverse patient populations. Second, only seven parameters were included, excluding crucial clinical indicators such as blood pressure, glucose level, smoking history, and family history of cardiovascular disease, which may limit the clinical relevance of the findings. Finally, ROC-AUC analysis, which is a standard evaluation metric in medical machine learning studies, was not applied in this work due to data and scope limitations. Future research should address these limitations by employing larger and more diverse datasets, incorporating additional clinical parameters, and including ROC-AUC analysis to provide more comprehensive performance evaluation.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Dr. Romi Satria Wahono, M.Eng., Ph.D., for providing the dataset used in this research and for his valuable insights into the field of data mining and machine learning. His contribution played a significant role in enabling the development and evaluation of the predictive models presented in this study.

REFERENCES

[1] S. Rajagopalan and P. J. Landrigan, "Pollution and the Heart," *New England Journal of Medicine*, vol. 385, no. 20, pp. 1881–1892, Nov. 2021, doi: 10.1056/NEJMra2030281.

[2] M. Di Cesare *et al.*, "The Heart of the World," *Glob Heart*, vol. 19, no. 1, Jan. 2024, doi: 10.5334/gh.1288.

[3] L. C. E. Huberts, R. J. M. M. Does, B. Ravesteijn, and J. Lokkerbol, "Predictive monitoring using machine learning algorithms and a real-life example on schizophrenia," *Qual Reliab Eng Int*, vol. 38, no. 3, pp. 1302–1317, Apr. 2022, doi: 10.1002/qre.2957.

[4] D. M. Lloyd-Jones *et al.*, "Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association," *Circulation*, vol. 146, no. 5, Aug. 2022, doi: 10.1161/CIR.0000000000001078.

[5] S. S. Dhanda *et al.*, "Advancement in public health through machine learning: a narrative review of opportunities and ethical considerations," *J Big Data*, vol. 12, no. 1, p. 154, Jul. 2025, doi: 10.1186/s40537-025-01201-x.

[6] A. A. Armoundas *et al.*, "Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association," *Circulation*, vol. 149, no. 14, Apr. 2024, doi: 10.1161/CIR.0000000000001201.

[7] C. Sirocchi, A. Bogliolo, and S. Montagna, "Medical-informed machine learning: integrating prior knowledge into medical decision systems," *BMC Med Inform Decis Mak*, vol. 24, no. S4, p. 186, Jun. 2024, doi: 10.1186/s12911-024-02582-4.

[8] M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovasc Disord*, vol. 25, no. 1, p. 212, Mar. 2025, doi: 10.1186/s12872-025-04627-6.

[9] S. A. Alowais *et al.*, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC Med Educ*, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.

[10] J. L. B. Munthe, K. Sinaga, and Santi Prayudani, "Sentiment Analysis of Hate Speech against DPR-RI on Twitter Using Naive Bayes and KNN Algorithms," *Electronic Integrated Computer Algorithm Journal*, vol. 2, no. 1, pp. 52–60, Oct. 2024, doi: 10.62123/enigma.v2i1.39.

[11] R. Kumar *et al.*, "A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions," *Front Artif Intell*, vol. 8, May 2025, doi: 10.3389/frai.2025.1583459.

[12] Kayam Saikumar, "Heart Disease Prediction using Machine Learning and Deep Learning Approaches: A Systematic Survey," *Panamerican Mathematical Journal*, vol. 35, no. 2s, pp. 72–81, Nov. 2024, doi: 10.52783/pmj.v35.i2s.2398.

[13] S. Baral, S. Satpathy, D. P. Pati, P. Mishra, and L. Pattnaik, "A Literature Review for Detection and Projection of Cardiovascular Disease Using Machine Learning," *EAI Endorsed Transactions on Internet of Things*, vol. 10, Mar. 2024, doi: 10.4108/eetiot.5326.

[14] S. D. Prakoso, A. E. Permanasari, and A. R. Pratama, "Heart Disease Prediction Using Machine Learning: A Systematic Literature Review," in *2023 10th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, IEEE, Aug. 2023, pp. 155–159. doi: 10.1109/ICITACEE58587.2023.10277209.

[15] A. Eleyan, E. AlBoghbaish, A. AlShatti, A. AlSultan, and D. AlDarbi, "RHYTHMI: A Deep Learning-Based Mobile ECG Device for Heart Disease Prediction," *Applied System Innovation*, vol. 7, no. 5, p. 77, Aug. 2024, doi: 10.3390/asi7050077.

[16] R. Garg, P. K. Sarangi, A. K. Sahoo, and J. Jha, "Cardiovascular Disease Prediction: Performance Analysis and Comparison of Various Supervised Machine Learning Algorithms," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/ICICAT57735.2023.10263671.

[17] M. S. Al Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," *IEEE Access*, vol. 11, pp. 121574–121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[18] V. V, R. A C, P. B.D, and A. K. Sharma, "Brain Stroke Prediction using Decision Tree Algorithm," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/EASCT59475.2023.10392706.

[19] R. M. Parry *et al.*, "k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction," *Pharmacogenomics J*, vol. 10, no. 4, pp. 292–309, Aug. 2010, doi: 10.1038/tpj.2010.56.

[20] H. A. Abdulqader and A. M. Abdulazeez, "Review on Decision Tree Algorithm in Healthcare Applications," *Indonesian Journal of Computer Science*, vol. 13, no. 3, Jun. 2024, doi: 10.33022/ijcs.v13i3.4026.

[21] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci Rep*, vol. 12, no. 1, p. 6256, Apr. 2022, doi: 10.1038/s41598-022-10358-x.

[22] S. Dağıstanlı *et al.*, "A novel survival algorithm in COVID-19 intensive care patients: the classification and regression tree (CRT) method," *Afr Health Sci*, vol. 21, no. 3, pp. 1083–1092, Sep. 2021, doi: 10.4314/ahs.v21i3.16.

[23] F. Alemi, H. Erdman, I. Griva, and C. H. Evans, "Improved Statistical Methods are Needed to Advance Personalized Medicine," *Open Transl Med J*, vol. 1, no. 1, pp. 16–20, Nov. 2009, doi: 10.2174/1876399500901010016.

[24] J. Demšar and B. Zupan, "Hands-on training about data clustering with orange data mining toolbox," *PLoS Comput Biol*, vol. 20, no. 12, p. e1012574, Dec. 2024, doi: 10.1371/journal.pcbi.1012574.

[25] J. Saputra, C. Lawrencya, J. M. Saini, and S. Suharjito, "Hyperparameter optimization for cardiovascular disease data-driven prognostic system," *Vis Comput Ind Biomed Art*, vol. 6, no. 1, p. 16, Aug. 2023, doi: 10.1186/s42492-023-00143-6.

[26] Z. Dobesova, "Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics," *Computer Applications in Engineering Education*, vol. 32, no. 4, Jul. 2024, doi: 10.1002/cae.22735.

[27] D. Vaishnav and B. R. Rao, "Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool," in *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Nov. 2018, pp. 603–607. doi: 10.1109/ICICT43934.2018.9034442.

[28] U. Thange, V. K. Shukla, R. Punhani, and W. Grobbelaar, "Analyzing COVID-19 Dataset through Data Mining Tool 'Orange,'" in *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE, Jan. 2021, pp. 198–203. doi: 10.1109/ICCAKM50778.2021.9357754.

[29] J. Wang and P.-R. Cornely, "Data-Driven Decisions: Exploring the Impact of Data Mining in Healthcare," in *2023 6th International Conference on Computing and Big Data (ICCBD)*, IEEE, Oct. 2023, pp. 38–45. doi: 10.1109/ICCBD59843.2023.10607356.

[30] P. Subathra, R. Deepika, K. Yamini, P. Arunprasad, and S. K. Vasudevan, "A Study of Open Source Data Mining Tools and its Applications," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 10, no. 10, pp. 1108–1132, Aug. 2015, doi: 10.19026/rjaset.10.1880.