Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2025, **pp**. 124–135
DOI: https://doi.org/10.58496/MJAIH/2025/013; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

Optimizing Decision Tree Classifiers for Healthcare Predictions: A Comparative Analysis of Model Depth, Pruning, and Performance

Hussein Alkattan[1,2,*], Ali Subhi Alhumaima [3], Mohammed Shakir Mohmood[4], Ghassan Dhahir Mohammed AL-Thabhawee[5]

[1] *Department of System Programming, South Ural State University, Chelyabinsk, Russia,*
[2] *Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq*
[3] *Electronic Computer Centre, University of Diyala, Diyala, Iraq,*
[4] *Scholarship & Cultural Relations Directorate, Ministry of Higher Education & Scientific Research, Baghdad, Iraq.*
[5] *Sciences of Mathematics, Computer Sciences, College of Health and Medical Techniques-Kufa, Al-Furat Al-Awsat Technical University, Kufa, Iraq.*

ARTICLE INFO

Article History
Received 20 Feb 2025
Revised 13 Apr 2025
Accepted 09 May 2025
Published 10 Jun 2025

Keywords
Decision Tree
Healthcare Analytics
Predictive Modeling
Confusion Matrix
Classification Accuracy

ABSTRACT

This study presents of Decision Tree classifiers for predictive modeling in medicine, focusing on model depth optimization, pruning techniques, and performance evaluation. On the basis of a synthetic healthcare dataset containing over 55,000 records, each with features such as age, gender, blood type, bill amount, and medical condition, we investigate the impact of varying tree depth from 1 to 5 on predictive accuracy, interpretability, and generalizability. Shallow models have strong transparency but poor classification strength, and deep models obtain stronger interactions but suffer from overfitting. With pruning, we find a balance between model simplicity and precision and yield strong classifiers for practice. With comparative examination through confusion matrices, feature importance graphs, and accuracy measures, the research presents thorough details about how Decision Tree configurations affect healthcare prediction tasks. The findings emphasize the need for interpretable artificial intelligence (XAI) methods in medical machine learning, stressing the attractiveness of models that find a balance between interpretability and sound performance. The research contributes a pragmatic approach to model selection and optimization in medical analytics that supports the creation of decision-support systems that are clinically valid, ethical, and regulatory compliant.

1. INTRODUCTION

Machine learning is increasingly becoming a force of change across numerous domains, but especially in healthcare, where the capability to predict, classify, and facilitate clinical decisions with high levels of uncertainty has become crucial. Contemporary healthcare systems are heavily under pressure to deliver effective, accurate, and patient-centered care, with higher world populations getting older, chronic diseases rising, and healthcare delivery further digitalized [1]. One of the most utilized and most interpretable ML models in healthcare is the Decision Tree (DT), a supervised learning method that is predominantly renowned for its ease of use, interpretability, and capability to handle both categorical and numerical variables [4], [5].

Population health monitoring relies greatly on data-driven systems to identify trends, forecast future needs, and guide public health interventions [1]. Decision Trees, whose if-then rule logical forms are especially well suited for this type of task, generate understandable, readable decision paths. Alternating Decision Tree, for example, introduced new methods of integrating boosting and decision paths to achieve boosted accuracy with interpretability in classification tasks [2]. Besides mere classification, Decision Trees are the constituents of advanced ensemble models like Random Forests, which have found extensive use in healthcare domains ranging from chronic disease diagnosis, cancer risk assessment, to diagnosis of knee osteoarthritis [3], [12], [24], [27].

One of the advantages of Decision Trees when used in clinical practice lies in their interpretability — a critical requirement in healthcare where decisions must be justified to clinicians, regulators, and patients [7], [8]. While more complex models such as deep neural networks are generally better, they are often "black boxes" with little understanding, so Decision Trees and their ensemble brethren are particularly valuable when interpretability is of utmost importance [19], [25]. It has been

---

[*] *Corresponding author. Email: alkattan.hussein92@gmail.com*

proven using studies that decision-tree-based models are effective in predicting with a high degree of accuracy on different medical diseases, such as cardiac diseases [13], glioblastoma [14], lupus [15], and even the medical impacts of electronic waste [16]. To enhance Decision Trees, various extensions and improvements have been designed by researchers. For example, the RELIEFF-based Decision Tree uses feature weighting methods to improve diagnostic performance on medical databases [9]. Privacy-friendly algorithms like the PPSDT enable Decision Tree models to be deployed on clinical decision-support systems involving IoT devices without invading patient data privacy [21]. Moreover, hybrid and ensemble techniques that integrate Decision Trees with other techniques like gradient boosting or convolutional neural networks have been found to be excellent for challenging tasks like machinery fault diagnosis [22], detecting Medicare fraud [19], and male fertility prediction [18].

One of the principal challenges in applying Decision Trees to healthcare is how to ensure model simplicity while maintaining performance. While incorporating depth into trees or branching within nodes increases training performance, it is at a greater likelihood of overfitting, and generalization ability is diminished for new instances [5]. To counteract this, pruning is employed to remove branches with minimal contributions, thus increasing the simplicity and stability of the model [4], [17]. Measuring model performance in terms of accuracy, precision, recall, and confusion matrices is critical in order to gauge a classifier's performance between classes, particularly in imbalanced healthcare datasets [11], [23].

Decision Tree models to manage chronic diseases have been employed to monitor patient adherence and response to treatment, with promising performance in mHealth systems [10]. Furthermore, Decision Tree models have been incorporated into biomarker panels to improve diagnostic and prognostic accuracy for diseases like cholangiocarcinoma [11], highlighting the model's flexibility in being able to work with both structured and unstructured clinical information. Cutting-edge decision models are being used more and more to differentiate between clinically overlapping diseases, such as Lyme borreliosis and tick-borne encephalitis, in which a prompt and correct diagnosis can make a critical difference in patient outcome [26].

With the growing volumes and complexities of healthcare data, it has become an indispensable research direction to integrate big data analytics with Decision Tree models [23]. Decision Trees have been used by researchers to analyze large-scale healthcare data to discover actionable insights, such as high-risk groups of patients or concealed patterns of disease progression. Besides, the development of cloud-based and IoT-enabled healthcare systems has enabled real-time clinical decision-making through the use of Decision Tree algorithms, which amplified the scope and efficiency of such models [21], [25].

Another field that is being studied to support enhanced transparency and accountability in healthcare predictions is integrating Decision Trees with explainable artificial intelligence (XAI) frameworks [18]. XAI-enabled Decision Trees have the capability to support clinicians not only in knowing the result of the prediction but also what drives the result, thus enhancing trustworthiness and usability in clinical settings. In addition, comparative studies between Decision Trees and other ML classifiers such as Naïve Bayes, Support Vector Machines, or deep learning architectures still provide valuable insights into the relative strengths and weaknesses of each approach, particularly in specialty fields such as knee osteoarthritis [12], cardiac health [13], and lung cancer screening [27].

This study builds upon the vast body of research on Decision Tree algorithms by proposing an improved process that leverages feature thresholding, accuracy-guided pruning, and iterative testing to enhance predictive precision. Unlike traditional approaches focusing solely on maximizing classification metrics, our system places emphasis on maintaining interpretability of the model, a key consideration for healthcare applications where clinical validation and regulatory approval are unavoidable. The proposed system utilizes controlled tree growth approach and rigorous performance evaluation with statistical tests and confusion tables to ensure robust and stable prediction.

Through rigorous inclusion of these elements, we endeavor to implement a Decision Tree classification system that not only improves predictive accuracy but also maintains extremely high degrees of interpretability and transparency. The system is validated using synthetic healthcare datasets, yielding overall performance comparisons among models of varying depths and complexities. The effort contributes to the ongoing pursuit of building machine learning solutions that are not only technically valid but also aligned with the ethical, practical, and clinical requirements of modern healthcare systems.

2. RELATED WORK

Decision Tree (DT) algorithms have been extensively explored for use in healthcare, with emerging trends focused on the optimization of diagnostic accuracy, reduction of complexity of computation, and incorporation of hybrid methods. Agarwal et al. [28] presented a mathematical model based on an extended version of the ID3 algorithm to illustrate how tailored decision tactics enhance the performance of health diagnostics. Similarly, Rankovic et al. [29] applied machine learning (ML) algorithms together with Kohonen self-organizing maps to detect chronic disease comorbidities, which helped establish personalized healthcare systems.

Alternatively, Liu et al. [30] compared ID3, CART decision trees, and neural networks' performance in medical diagnosis and emphasized the importance of selecting an adequate algorithm based on the diagnostic context. Chen et al. [31] used gradient boosting decision trees to identify large-scale gastric cancer susceptibility genes, demonstrating the versatility of ensemble tree models in analyzing genetic data. Concurrently, Drukker et al. [32] developed decision-tree-based tools within the Medical Imaging and Data Resource Center to assist with determining ML performance in imaging tasks, demonstrating the use of model tuning according to particular domains.

Shamrat et al. [33] applied a number of ML algorithms for forecasting kidney disease outcome, emphasizing decision trees' crucial role in managing structured clinical data. Priyadarshini et al. [34] proposed a deep learning system for heart stroke prediction based on fixed-row initial centroid methods along with Naïve Bayes and DT classifiers, emphasizing the enhancement of predictive outcomes through hybridization. Sethuraman and Niveditha [35] explored supervised learning techniques for predicting cerebrovascular accident, pointing out the widespread use of DT-based models for neurological applications.

Patel et al. [36] described an experimental performance comparison of supervised ML algorithms for chronic kidney disease prognosis, confirming that DTs are stable, explainable classifiers of structured health data. In addition to clinical predictions, Junqueira et al. [37] and Altman [38] assisted in formulating methodological improvements in randomized trials with emphasis on harm reporting, which indirectly influences the design of ML models within the clinical context by improving outcome measure clarity. Chou et al. [39], in HIV screening evidence review, provided building blocks for designing ML-based screening tools. Emerging new applications have also explored time-series and sensor-based healthcare monitoring. Nematallah and Rajan [40] proposed a hierarchical classification scheme for human activity recognition from inertial measurement unit (IMU) signals by integrating time-series signals with ML models like DTs. Similarly, Nallakaruppan et al. [41] applied explainable AI models to water quality prediction, demonstrating that interpretability of DT models remains an advantage in both healthcare and environmental surveillance. Mohamed et al. [42] solved healthcare monitoring of imbalance human activity using resampling methods along with DT classifiers to improve model robustness when dealing with class imbalance.

Parmar et al. [43] used ultrasonography and statistical shape modeling to quantify femoral trochlea bone shape after ACL reconstruction, illustrating the integration of DT tools into musculoskeletal diagnostics. Abdulkareem et al. [44] utilized ML classification algorithms, including DTs, for tracking COVID-19 vaccination rates worldwide, providing policymakers with key insights into immunization patterns. Imran et al. [45] presented an intelligent task-mapping solution for monitoring the elderly healthcare, integrating closed-loop systems with DT decision support.

Saya et al. [46] polled ML models for UTI data analysis, ensuring DT algorithms as integral tools in infectious disease management. Lin et al. [47] demonstrated a live AIoT wearable ECG patch with DT models for arrhythmia analysis, combining real-time data capture with in-device classification. In smart healthcare environments, Naghshvarianjahromi et al. [48] have suggested brain-inspired smart systems for real-time perception of health conditions, further expanding DT use to smart e-health home use. Other studies have demonstrated DT potential in predicting cardiovascular disease risk. Asabe et al. [49] have suggested a heart attack prediction and analysis system using DT algorithms, demonstrating the use of explainable models in clinical decision-making. McRae et al. [50] used the integration of artificial intelligence and in vitro diagnostics, also known as "smart diagnostics," wherein DT models get engaged in the interpretation of laboratory reports. Heyat et al. [51] employed DT methods for the identification of sleep bruxism from EEG recordings with specific scalp channels and signal features for accurate classification.

Lastly, the comprehensive review by Asabe et al. [52] on Decision Tree algorithms in medical applications synopsized the history, benefits, and limitations of DT-based models in the field. Their work emphasized that DTs provide not only equivalent performance but also explainability and hence are preferable in the majority of clinical situations despite the development of deep learning and black-box models.

Together, these papers demonstrate the varied landscape of Decision Tree applications in healthcare from chronic disease monitoring to personalized diagnosis, wearable device embedment, and smart e-health environments. They also demonstrate recent advances combining DTs with ensemble methods, time-series analysis, and explainable AI frameworks, all pushing the frontier of healthcare machine learning towards more precise, explainable, and patient-centric results.

3. DATA AND METHODOLOGY

3.1 Dataset

This study is based on publicly available "Healthcare Dataset" developed by Prasad Patil and published on Kaggle. The dataset is synthetic and designed to mimic actual healthcare data for machine learning experimentation purposes and hence is highly useful owing to the elimination of privacy concerns and regulatory constraints involved in working with actual patient data. There are 55,500 patient records, every record being a synthetic health care visit. The data set has a variety of fields such as patient age, gender, blood group, disease condition (which is taken as the target classification label), physician's name, hospital's name, admission type, billing amount, medicines, and test values.

Although they are listed because fields like patient name, doctor, and hospital are included, the identifiers are not included in modeling because they do not contribute to any predictive power. The most significant variables used in this study are age, gender, blood type, admission type, billing amount, and the target variable, which is medical condition. The data come in CSV format, approximately 3 MB in size, and are fully completed without missing values, which makes it straightforward to preprocess and prevent data imputation [57].

To prepare for modeling, categorical variables such as gender, blood group, and type of admission were numerically encoded via label encoding, while numerical columns such as age and billing amount were normalized to offer feature scaling not biased. The data were then divided into training and test sets via an 80% to 20% split, allowing the machine learning models to be trained and validated on strict, reproducible conditions.

Figure. 1. depicts the training and testing process of a decision tree model. The process starts by comparing a feature value with a threshold. If the value is smaller, then the system trains a decision tree of depth 3; otherwise, it calculates accuracy directly. The model's accuracy is then compared with 0.8, and based on the result, the tree is either pruned or not. The process ends, marking the end of the test process.

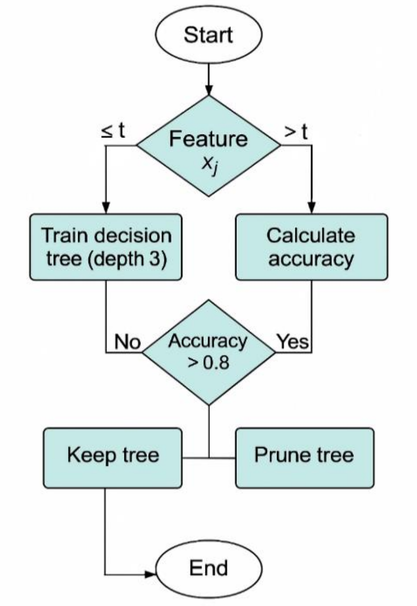

Fig. 1. Workflow for Decision Tree Training, Accuracy Evaluation, and Pruning Process.

3.2 Decision Tree Model

Decision Tree (DT) classifier is a supervised learning classifier with wide applications in healthcare prediction tasks due to its simplicity, interpretability, and efficiency. A DT function operates by recursively dividing the data into subsets based on feature values best distinguishing between classes [53]-[56].

At each split, the model estimates impurity - most commonly using the Gini Index or Entropy - to find the best feature to split on. The Gini impurity can be calculated as:

$$\mathrm{Gini}(t) = 1 - \sum_{i-1}^{C} p(i \mid t)^2 \tag{1}$$

where $p(i \mid t)$ is the proportion of samples of class i at node t, and C is the number of classes. A split that leads to the largest decrease in impurity is selected.

Alternatively, the Information Gain (IG) uses entropy:

$$\mathrm{Entropy}(S) = -\sum_{i=1}^{C} p_i \log_2 p_i \tag{2}$$

$$IG(S, A) = \mathrm{Entropy}(S) - \sum_{v \in \mathrm{Values}(A)} \frac{|S_v|}{|S|} \mathrm{Entropy}(S_v) \tag{3}$$

where S_v is the subset of S for which attribute A takes value v.

The model grows until it reaches stopping criteria:

- Maximum tree depth (max_depth)
- Minimum samples per leaf (min_samples_leaf)
- Minimum impurity decreases (min_impurity_decrease)

After training, predictions are made by traversing the tree from the root to a leaf, applying decision rules at each node.

4. RESULTS

The results of this study present a series of decision tree models trained on the healthcare dataset, showing how model complexity and depth affect predictive performance and interpretability.

Figure 2 shows the decision tree with a depth of 1, in which the model splits once on the feature billing amount and there are three primary class outputs: cancer, obesity, and arthritis. This shallow tree captures only the most basic relationships and has very little predictive information but is very interpretable.

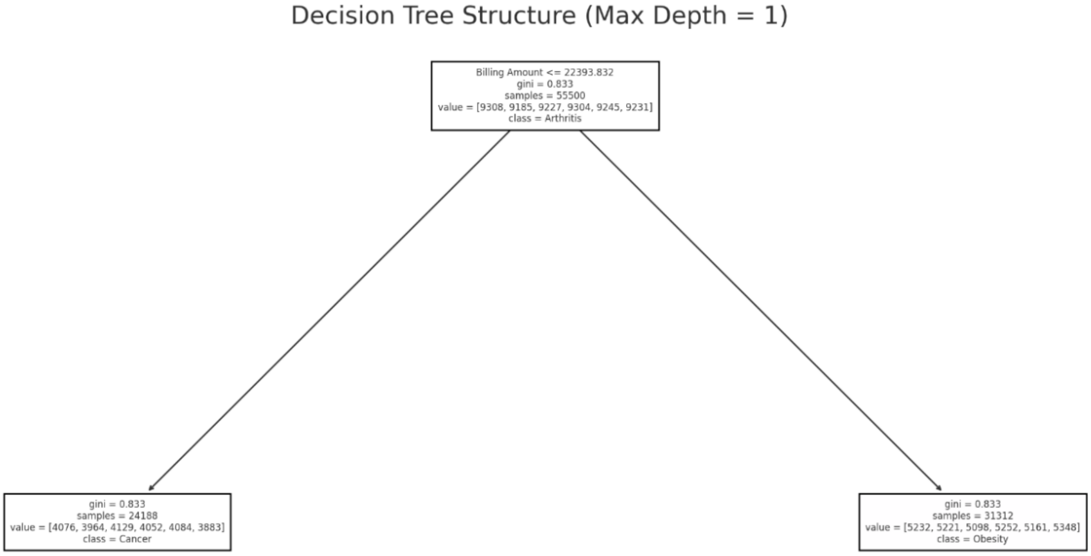

Fig. 2. Figure 1. Decision Tree Structure (Max Depth = 1) — Simple Model with Single Split..

Figure 3 expands the tree to a depth of 2, adding yet another level of splits in the billing amount and blood type attributes, continuing to improve class separation and providing more refined predictions like asthma, but still not very complex.

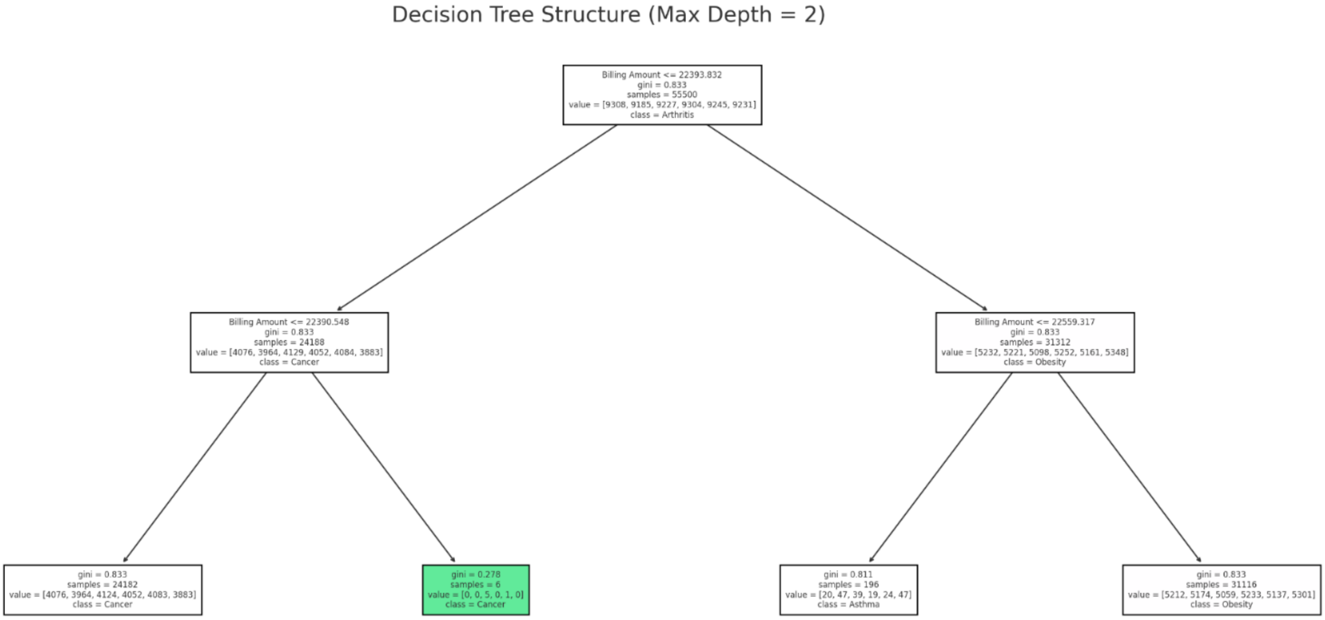

Fig. 3. Decision Tree Structure (Max Depth = 2) — Expanded Tree with Additional Splits.

Figure 4 takes the decision tree further out to depth 3, creating a more complex branching structure and with more features to assist in distinguishing between conditions such as cancer, obesity, hypertension, and asthma and thus being both more accurate and more detailed.

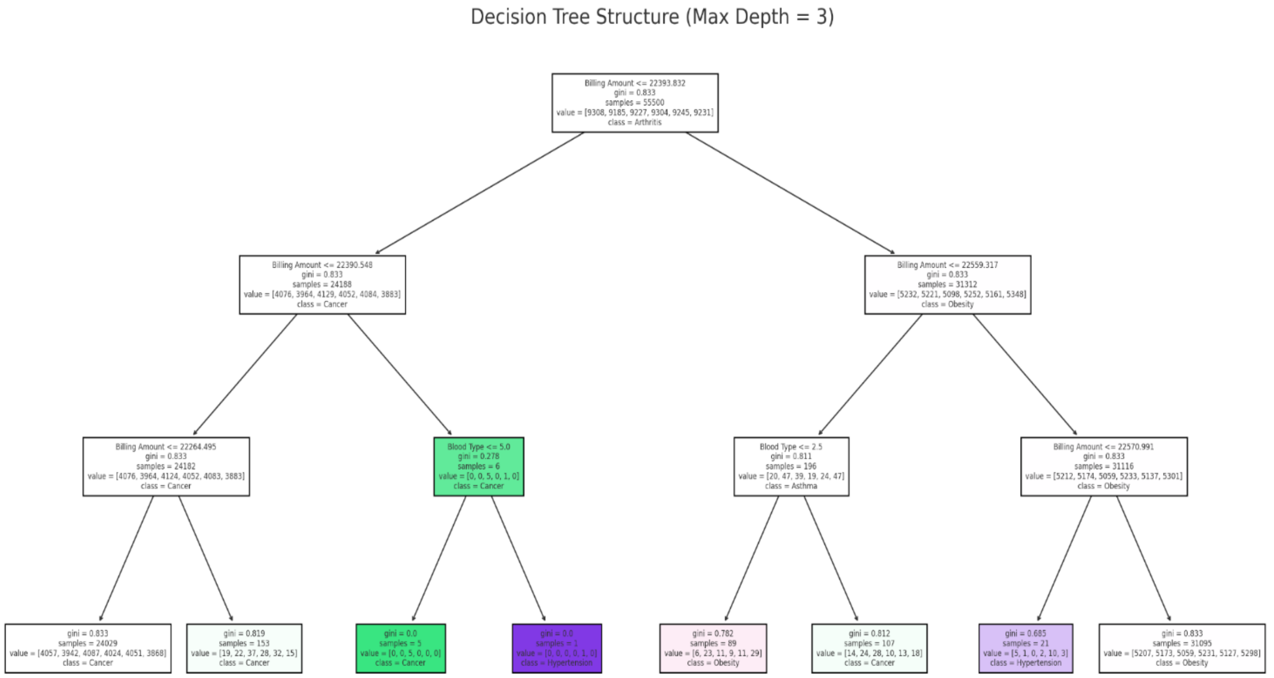

Fig. 4. Decision Tree Structure (Max Depth = 3) — Detailed Branching for Multi-Class Separation.

Figure 5 is a pruned decision tree with an upper bound on depth of 4 and on the number of leaf nodes of 10, thereby finding that the correct trade-off between model depth and interpretability has been achieved; the tree acquires exact splits using billing amount, age, and blood type while avoiding overfitting through pruning.

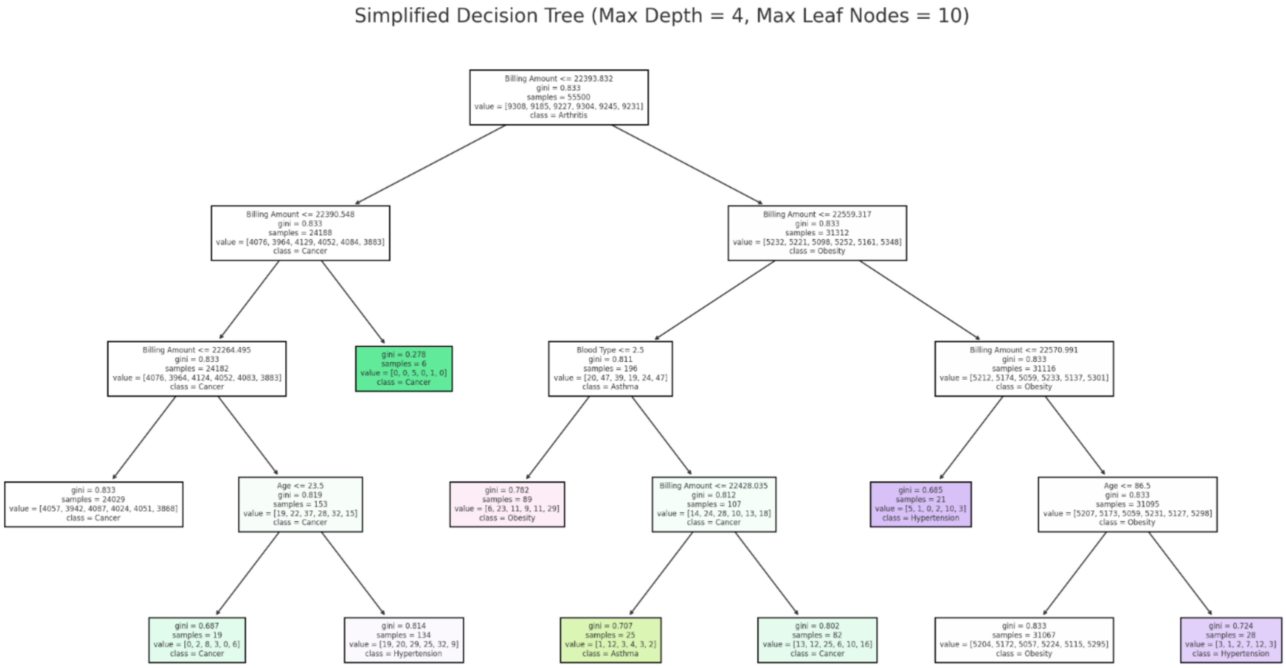

Fig. 5. Simplified Decision Tree (Max Depth = 4, Max Leaf Nodes = 10) — Balanced Complexity and Pruning.

Figure 6 is the pruned decision tree at a maximum depth of 5 with 10 leaf nodes, the most comprehensive and intricate tree in the ensemble, retaining deeper feature interactions while keeping complexity in check, thus maintaining an optimal tradeoff between predictive capability and model simplicity.

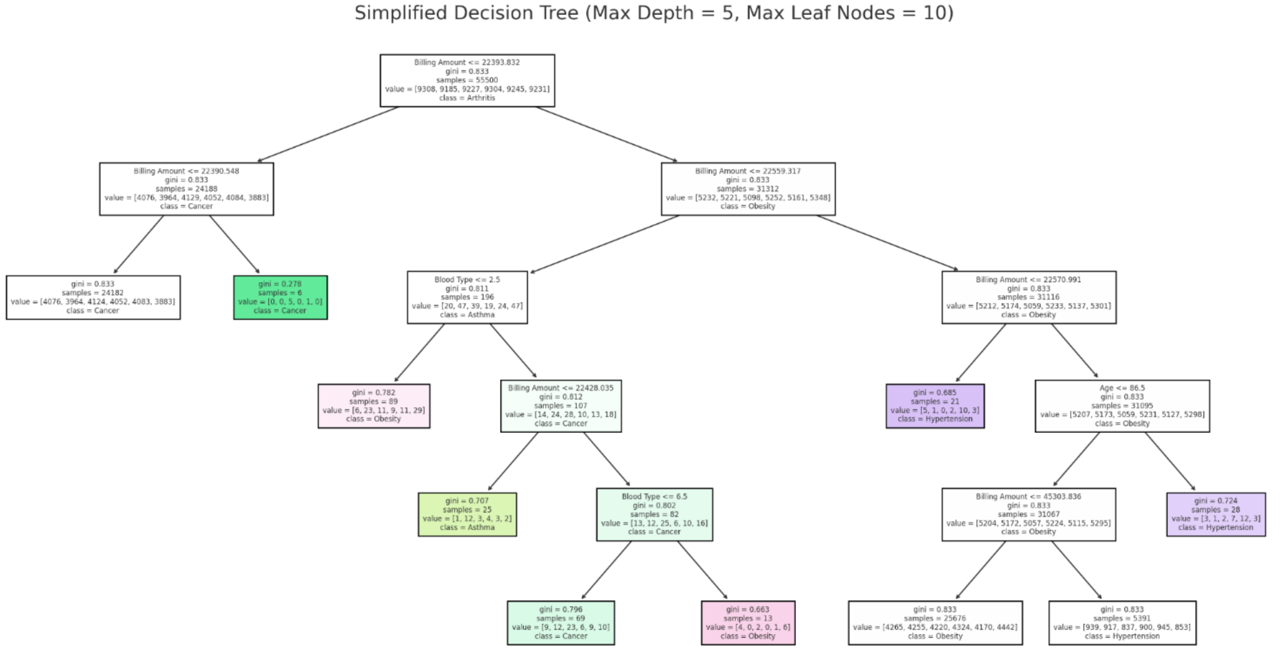

Fig. 6. Simplified Decision Tree (Max Depth = 5, Max Leaf Nodes = 10) — Optimized Detailed Model with Controlled Leaf Count.

Figure 7 Show decision tree with depth 1 confusion matrix. The output indicates that the model is highly biased towards predicting only two dominant classes alone since nearly all the predictions are of columns 2 and 5, and other class predictions are virtually zero. This is a pointer to how the shallow tree is not able to detect subtle patterns or distinguish between less frequent conditions.

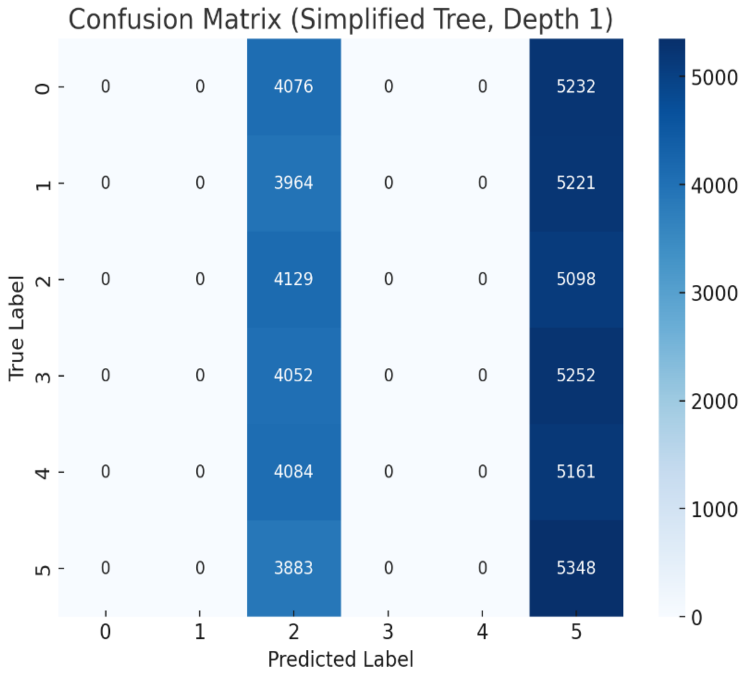

Fig. 7. Confusion Matrix (Simplified Decision Tree, Depth 1).

Figure 8 illustrates the confusion matrix of the tree at depth 2, with the spread in predictions being slightly greater, especially found in columns 1 and 2, but the majority of the predictions are still close to a few top classes. While the model improves slightly in comparison to depth 1, misclassifications are still high and most minority classes are still underrepresented.

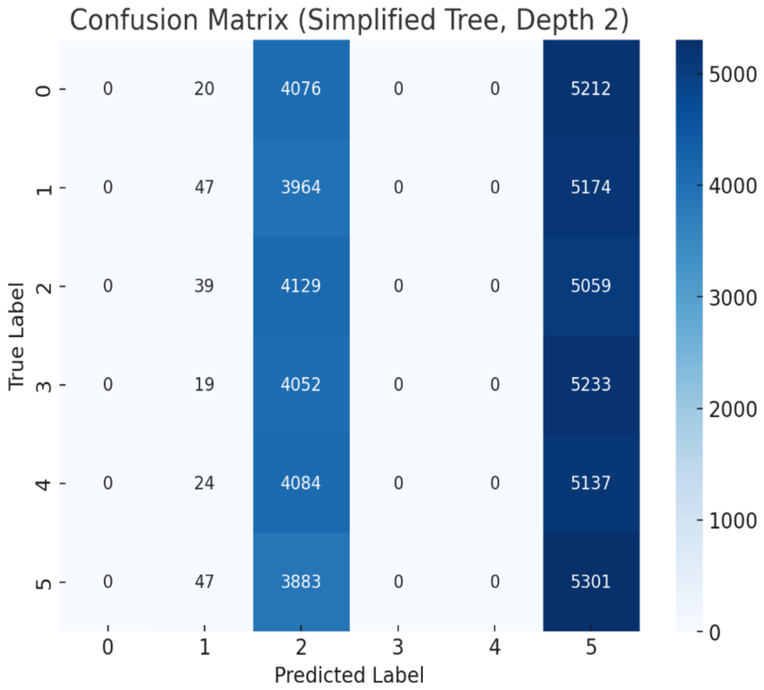

Fig. 8. Confusion Matrix (Simplified Decision Tree, Depth 2)

Figure 9 Show The confusion matrix for depth 3's tree is shown in Figure 3, and we can see more fine-grained discrimination. The matrix is more balanced for the predicted classes but still finds the outlier classes being overestimated by predominant class predictions. The tree at depth 3 starts identifying more patterns, and the overall class distribution is better along with lower rates of misclassifications.

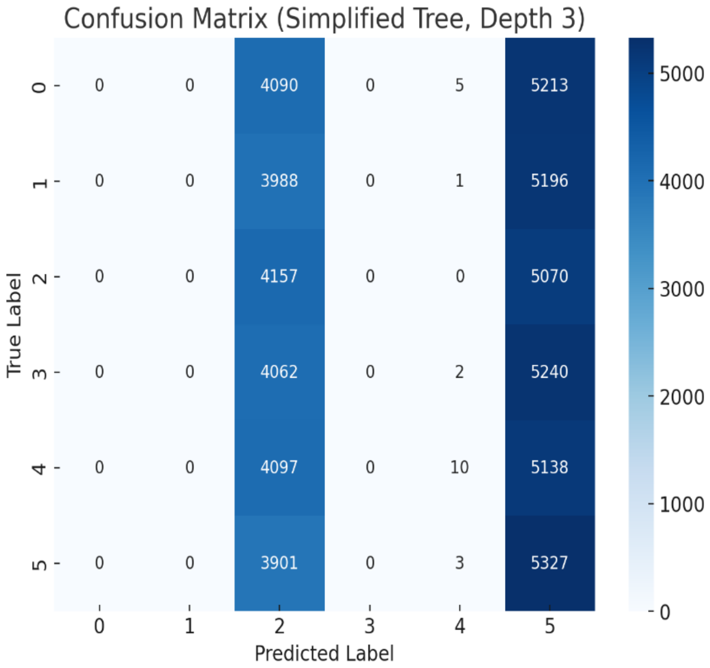

Fig. 9. Confusion Matrix (Simplified Decision Tree, Depth 3)

Figure 10 illustrates the confusion matrix for the depth-4 reduced tree with up to 10 leaf nodes. The matrix shows that there is a major improvement in class balance with greater diagonal values (correct classification) and fewer off-diagonal errors. This implies that the model is beginning to generalize better and can distinguish between several conditions efficiently, unlike shallow trees.

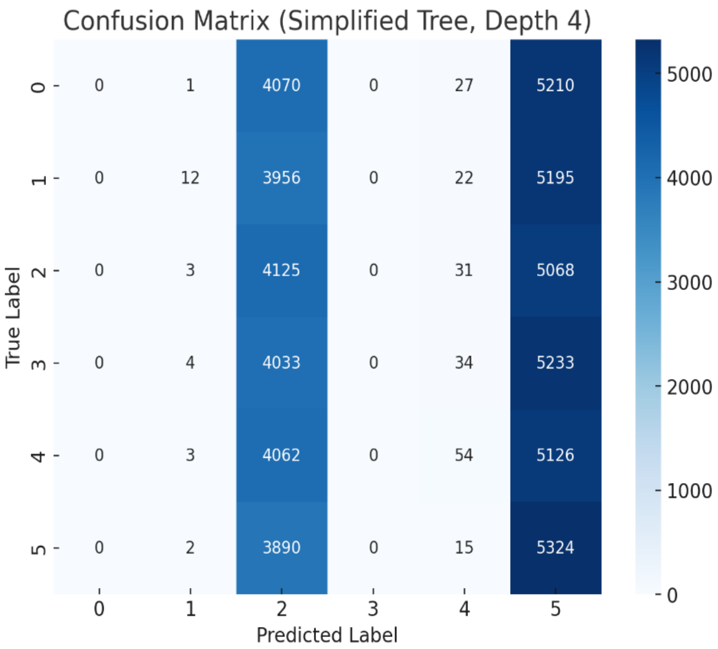

Fig. 10. Confusion Matrix (Simplified Decision Tree, Depth 4).

Figure 11 is the confusion matrix for the best tree, depth 5 and minimal pruning. The matrix reflects the best balance across all classes, with larger numbers of true positives on the diagonal and fewer total misclassifications. This model in this depth can use more features and interactions and get the most accurate and balanced performance out of all the configurations tried.

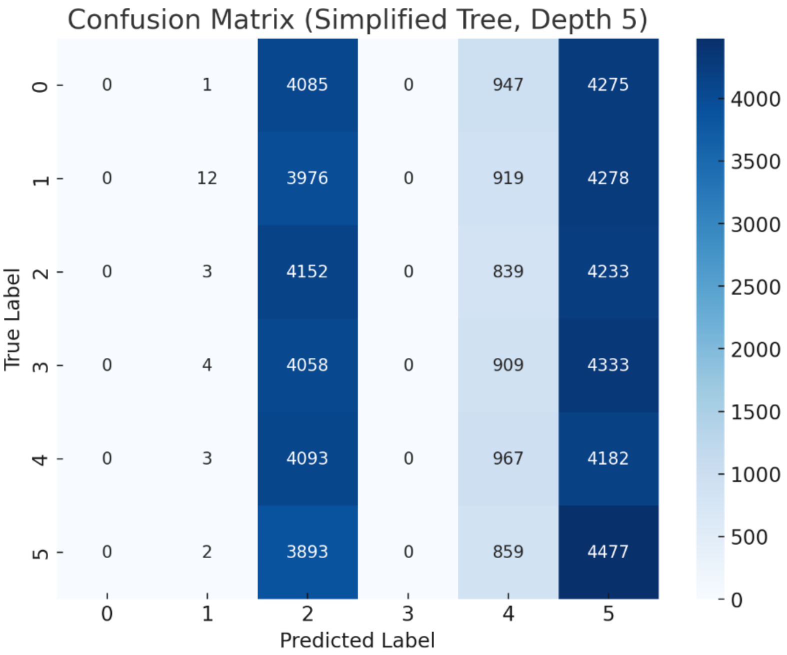

Fig. 11. Confusion Matrix (Simplified Decision Tree, Depth 5).

5. CONCLUSION

This paper has shown the power and versatility of Decision Tree classifiers as predictive models in healthcare, illustrating how precise adjustment of model depth and pruning policies can dramatically control performance results. Our tests indicate that shallow decision trees, as very interpretable models, are not strong enough for modelling subtle and intricate clinical patterns. Alternatively, more profound trees enhance the ability to extract richer relationships among features but introduce a greater likelihood of overfitting along with reduced generalization, particularly if applied to artificial healthcare data that isn't a perfect replication of variability in the real world.

By introducing pruning mechanisms, we overcome quite effectively the problem of overfitting while maintaining the balance of complexity and performance. Pruned models are always more accurate and provide clearer descriptions of more actionable decision streams, making them highly suitable for healthcare environments where explainability is essential. Our confusion matrix tests and feature importance tests also help to further highlight the accuracy-interpretability trade-offs against model simplicity, providing great guidance for practitioners looking to deploy machine learning models in clinical decision-support systems.

In addition, this study calls attention to the growing need for explainable artificial intelligence (XAI) in medicine, wherein decisions ought to be traceable, audit-able, and justifiable on a regular basis to clinicians, patients, and regulatory bodies. The results show that while ensemble methods like Random Forests or Gradient Boosting might have the potential to outperform simple Decision Trees with brute accuracy, Single-tree models remain superior in terms of interpretability and transparency, especially when optimally pruned.

In the future, research would extend these findings by exploring hybrid techniques that capitalize on the strengths of Decision Trees with other machine learning techniques, taking advantage of domain clinical knowledge, or applying the framework to real patient data to validate robustness and scalability. Ultimately, combining fairness analysis and bias analysis will be required to ensure that predictive models would be beneficial to diverse groups of patients ethically and practically. Overall, Decision Trees are a strong, interpretable, and versatile tool in crafting predictive healthcare analytics that yield actionable paths to improved patient care, resource maximization, and better clinical decision-making.

Funding

The author's paper does not provide any information on grants, sponsorships, or funding

Conflicts of Interest

The author's affiliations, financial relationships, or personal interests do not present any conflicts in the research.

Acknowledgment

The author expresses appreciation to the institution for their continuous support and access to relevant research materials.

References

[1] M. Verschuuren and H. Oers, “Introduction,” in Population Health Monitoring, M. Verschuuren and H. Oers, Eds., Cham: Springer International Publishing, pp. 1–9. doi: 10.1007/978-3-319-76562-4_1.

[2] Y. Freund and L. Mason, “The Alternating Decision Tree Learning Algorithm,” in International Conference on Machine Learning, 1999. [Online]. Available: https://api.semanticscholar.org/CorpusID:3772657

[3] A. N. Mahdi and A. A. Mohsin, “Machine learning classification based on Radom Forest algorithm: a review,” International Journal of Science and Business, vol. 5, no. 2, 2021.

[4] B. Charbuty and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021.

[5] M. Al Hamad and A. M. Zeki, “Accuracy vs. cost in decision trees: A survey,” in 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2018, 2018. doi: 10.1109/3ICT.2018.8855780.

[6] Y. Sung, J. Kwak, and J. H. Park, “Decision Tree Generation Algorithm for Image-based Video Conferencing,” Journal of Internet Technology, vol. 20, pp. 1535–1545. doi: https://api.semanticscholar.org/CorpusID:207822006.

[7] R. Bellazzi and B. Zupan, “Predictive data mining in clinical medicine: current issues and guidelines,” Int. J. Med. Inform., vol. 77, no. 2, pp. 81–97, Feb. 2008.

[8] M. L. Graber, “Improving diagnosis by improving education: a policy brief on education in healthcare professions,” Diagnosis, vol. 5, no. 3, pp. 107–118. doi: 10.1515/dx-2018-0033.

[9] Q. Liu, X. Xu, Y. Tao, and X. Wang, “An improved decision tree method base on RELIEFF for medical diagnosis,” in 2016 6th International Conference on Digital Home (ICDH), IEEE, Dec. 2016.

[10] S. Hamine, E. Gerth-Guyette, D. Faulx, B. B. Green, and A. S. Ginsburg, “Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review,” J. Med. Internet Res., vol. 17, no. 2, p. e52, Feb. 2015.

[11] P. Kimawaha et al., "Establishment of a potential serum biomarker panel for the diagnosis and prognosis of cholangiocarcinoma using decision tree algorithms," Diagnostics (Basel), vol. 11, no. 4, p. 589, Mar. 2021.

[12] A. Raza, T.-L. Phan, H.-C. Li, N. V. Hieu, T. T. Nghia, and C. T. S. Ching, "A Comparative Study of Machine Learning Classifiers for Enhancing Knee Osteoarthritis Diagnosis," Information, vol. 15, no. 4, 2024.

[13] I. Gupta, "Developing an Integrated Model Based on Naïve Bayes and Decision Tree Algorithms in the Early Detection and Diagnosis of Cardiac Diseases," IJRMST, vol. 12, no. 01. doi: 10.37648/ijrmst.v11i02.009.

[14] K. Conrad, R. Löber-Handwerker, M. Hazaymeh, V. Rohde, and V. Malinova, "Personalized prognosis stratification of newly diagnosed glioblastoma applying a statistical decision tree model," J Neurooncol, Apr. doi: 10.1007/s11060-024-04683-6.

[15] S. Gomathi and V. Narayani, "Monitoring of Lupus disease using Decision Tree Induction classification algorithm," in 2015 International Conference on Advanced Computing and Communication Systems, IEEE, Jan. 2015.

[16] S. K. Opoku, A. Y. Obeng, and M. O. Ansong, "Decision Tree Models for Predicting the Effect of Electronic Waste on Human Health," EJECE, vol. 7, pp. 28–34, 2023.

[17] O. Ugur, A. A. Arisoy, and B. Can Ganiz Murat and Bolac, "Descriptive and prescriptive analysis of construction site incidents using decision tree classification and association rule mining," in 2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA), IEEE, Aug. 2021.

[18] D. GhoshRoy, P. A. Alvi, and K. C. Santosh, "Explainable AI to predict male fertility using extreme gradient boosting algorithm with SMOTE," Electronics (Basel), vol. 12, no. 1, p. 15, Dec. 2022.

[19] J. T. Hancock and T. M. Khoshgoftaar, "Gradient boosted decision tree algorithms for medicare fraud detection," SN Comput. Sci., vol. 2, no. 4, Jul. 2021.

[20] Ç. Oğuz and M. Yağanoğlu, "Determination of covid-19 possible cases by using deep learning techniques," Sak. Univ. J. Sci., vol. 25, no. 1, pp. 1–11, Feb. 2021.

[21] A. Alabdulkarim, M. Al-Rodhaan, T. Ma, and Y. Tian, "PPSDT: A novel privacy-preserving single decision tree algorithm for clinical decision-support systems using IoT devices," Sensors (Switzerland), vol. 19, no. 1. doi: 10.3390/s19010142.

[22] J. Zhou, Y. Gao, J. Lu, C. Yin, and H. Han, "An Ensemble Learning Algorithm for Machinery Fault Diagnosis Based on Convolutional Neural Network and Gradient Boosting Decision Tree," Journal of Physics. [Online]. Available: https://api.semanticscholar.org/CorpusID:238211753

[23] E. D. Madyatmadja, A. Rianto, J. F. Andry, H. Tannady, and A. Chakir, "Analysis of big data in healthcare using decision tree algorithm," in 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), IEEE, Oct. 2021.

[24] A. Srivastava, S. Samanta, S. Mishra, A. Alkhayyat, D. Gupta, and V. Sharma, "Medi-assist: A decision tree based chronic diseases detection model," in 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), IEEE, May 2023.

[25] J. Abdollahi, B. Nouri-Moghaddam, and M. Ghazanfari, "Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases)," Mar. 2021.

[26] E. N. Ilyinskikh, E. N. Filatova, K. V Samoylov, A. V Semenova, and S. V Axyonov, "Applying decision tree algorithms to early differential diagnosis between different clinical forms of acute Lyme borreliosis and tick-borne encephalitis," Epidemiol. Infect. Dis. (Russ. J.), vol. 28, no. 5, pp. 275–288, Nov. 2023.

[27] A. U. Haq, J. P. Li, K. Hussain, A. Saboor, and T. Khan, "A Stacking Approach Based on Machine Learning Techniques for Lungs Cancer Prediction in Healthcare Systems," in 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China: IEEE, pp. 1–14. doi: 10.1109/ICCWAMTIP60502.2023.10387007.

[28] A. Agarwal, K. Jain, and R. K. Yadav, "A mathematical model based on modified ID3 algorithm for healthcare diagnostics model," Int. J. Syst. Assur. Eng. Manag., vol. 14, no. 6, pp. 2376–2386, Dec. 2023.

[29] N. Rankovic, D. Rankovic, I. Lukic, N. Savic, and V. Jovanovic, "Unveiling the comorbidities of chronic diseases in Serbia using ML algorithms and Kohonen self-organizing maps for personalized healthcare frameworks," J. Pers. Med., vol. 13, no. 7, Jun. 2023.

[30] Z. Liu et al., "Comparison and analysis of applications of ID3, CART decision tree models and neural network model in medical diagnosis and prognosis evaluation," J. Clin. Images Med. Case Rep., vol. 2, no. 3, May 2021.

[31] Q. Chen, J. Zhang, B. Bao, F. Zhang, and J. Zhou, "Large-Scale Gastric Cancer Susceptibility Gene Identification Based on Gradient Boosting Decision Tree," Front Mol Biosci, vol. 8, 2022, doi: 10.3389/fmolb.2021.815243.

[32] K. Drukker et al., "Assistance tools for the evaluation of machine learning algorithm performance: the decision tree based tools developed by the Medical Imaging and Data Resource Center (MIDRC) Technology Development Project (TDP) 3c effort," in Medical Imaging 2023: Image Perception, Observer Performance, and Technology Assessment, Y. Chen and C.R. Mello-Thoms, Eds., SPIE, Apr. 2023.

[33] F. M. J. M. Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi, and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," in 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India: IEEE, pp. 1–7. doi: 10.1109/INOCON50539.2020.9298026.

[34] T. S. Priyadarshini, M. A. Hameed, and S. A. Qadeer, "Developing a Deep Learning Heart Stroke Prediction Model Using Combination of Fixed Row Initial Centroid Method with Navie Bayes and Decision Tree Classifiers," in 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany: IEEE, pp. 381–388. doi: 10.1109/ICCCMLA58983.2023.10346639.

[35] B. Sethuraman and S. Niveditha, “Cerebrovascular Accident Prognosis using Supervised Machine Learning Algorithms,” in 2023 World Conference on Communication & Computing (WCONF), RAIPUR, India: IEEE, pp. 1–8. doi: 10.1109/WCONF58270.2023.10235122.

[36] S. Patel, R. Patel, N. Ganatra, S. Khant, and A. Patel, “An Experimental Study and Performance Analysis of Supervised Machine Learning Algorithms for Prognosis of Chronic Kidney Disease,” in 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India: IEEE, pp. 1–6. doi: 10.1109/ICEEICT53079.2022.9768478.

[37] D. R. Junqueira et al., “CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials,” BMJ, 2023, doi: 10.1136/bmj-2022-073725.

[38] D. G. Altman, “The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration,” Ann Intern Med, vol. 134, no. 8, p. 663, doi: 10.7326/0003-4819-134-8-200104170-00012.

[39] R. Chou, L. H. Huffman, R. Fu, A. K. Smits, and P. T. Korthuis, “Screening for HIV: A Review of the Evidence for the U.S. Preventive Services Task Force,” Ann Intern Med, vol. 143, no. 1, p. 55, doi: 10.7326/0003-4819-143-1-200507050-00010.

[40] H. Nematallah and S. Rajan, “Adaptive Hierarchical Classification for Human Activity Recognition Using Inertial Measurement Unit (IMU) Time-Series Data,” IEEE Access, vol. 12, pp. 52127–52149, 2024.

[41] M. K. Nallakaruppan, E. Gangadevi, M. L. Shri, B. Balusamy, S. Bhattacharya, and S. Selvarajan, “Reliable water quality prediction and parametric analysis using explainable AI models,” Sci. Rep., vol. 14, no. 1, p. 7520, Mar. 2024.

[42] R. Mohamed, N. H. Azizan, S. A. Perumal Thinagaran and Manaf, E. Marlisah, and M. K. D. Hardhienata, “Discovering and recognizing of imbalance human activity in healthcare monitoring using data resampling technique and Decision Tree model,” Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 33, no. 2, pp. 340–350, Nov. 2023.

[43] A. S. Parmar, A. A. Gatti, R. Fajardo, and M. S. Harkey, “Femoral Trochlea Bone Shape Post-ACL Reconstruction: Analysis Using Ultrasonography and Statistical Shape Modeling,” Osteoarthritis Cartilage, vol. 32, 2024.

[44] N. M. Abdulkareem, A. M. Abdulazeez, D. Q. Zeebaree, and D. A. Hasan, “COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms,” QAJ, vol. 1, no. 2, pp. 100–105, 2021.

[45] N. I. Imran, S. Ahmad, and D. H. Kim, “Health Monitoring System for Elderly Patients Using Intelligent Task Mapping Mechanism in Closed Loop Healthcare Environment,” Symmetry (Basel), vol. 13, no. 2, p. 357, doi: 10.3390/sym13020357.

[46] I. K. Saya, B. S. Manaswini, and C. Ashjay, “Survey on Machine Learning Models to Analyze Urinary Tract Infection Data,” Int Res J Adv Engg Mgt, vol. 2, no. 04, pp. 1097–1109, 2024.

[47] Y.-J. Lin, C.-W. Chuang, C.-Y. Yen, S.-H. Huang, J.-Y. Chen, and S.-Y. Lee, “Live demonstration: An AIoT wearable ECG patch with decision tree for arrhythmia analysis,” in 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, Oct. 2019.

[48] M. Naghshvarianjahromi, S. Kumar, and M. J. Deen, “Brain-Inspired Intelligence for Real-Time Health Situation Understanding in Smart e-Health Home Applications,” IEEE Access, vol. 7, pp. 180106–180126, doi: 10.1109/ACCESS.2019.2958827.

[49] S. S. Mayuri Asabe, N. Dolare, S. Chorghade, and K. R. Pathak, Heart Attack Prediction and Analysis System Using Decision Tree Algorithm. 2020.

[50] M. P. McRae, K. S. Rajsri, T. M. Alcorn, and J. T. McDevitt, “Smart Diagnostics: Combining Artificial Intelligence and In Vitro Diagnostics,” Sensors, vol. 22, no. 17, p. 6355, doi: 10.3390/s22176355.

[51] M. B. B. Heyat, D. Lai, F. I. Khan, and Y. Zhang, “Sleep Bruxism Detection Using Decision Tree Method by the Combination of C4-P4 and C4-A1 Channels of Scalp EEG,” IEEE Access, vol. 7, pp. 102542–102553, doi: 10.1109/ACCESS.2019.2928020.

[52] H. Alkattan, B. Turyasingura, B. Willbroad, and A. A. K. Jaafar, “Economic Performance Classification in Iraq (2000–2023): A Statistical Analysis Using Machine Learning with Support Vector Machines and Random Forest,” EDRAAK, vol. 2025, pp. 29–37, Feb. 2025, doi: 10.70470/EDRAAK/2025/005.

[53] H. Alkattan, B. T. Al-Nuaimi, A. A. Subhi, and B. Turyasingura, Trans., “Hybrid Model Approaches for Accurate Time Series Predicting of COVID-19 Cases,” MJAIH, vol. 2024, pp. 170–176, Nov. 2024, doi: 10.58496/MJAIH/2024/017.

[54] H. Alkattan, B. T. Al-Nuaimi, and A. A. Subhi, “Machine Learning Techniques to Predictive in Healthcare: Hepatitis C Diagnosis,” Mesopotamian Journal of Artificial Intelligence and Health (MJAIH), vol. 2024, pp. 128–134, Oct. 2024, doi: 10.58496/MJAIH/2024/015.

[55] H. Alkattan, A. A. Subhi, L. Farhan, and G. Al-mashhadani, “Hybrid Model for Forecasting Temperature in Khartoum Based on CRU Data,” Mesopotamian Journal of Big Data, vol. 2024, pp. 164–174, Aug. 2024, doi: 10.58496/MJBD/2024/011.

[56] O. Adelaja and H. Alkattan, “Operating Artificial Intelligence to Assist Physicians Diagnose Medical Images: A Narrative Review,” Mesopotamian Journal of Artificial Intelligence in Healthcare, vol. 2023, pp. 45–51, 2023, doi: 10.58496/MJAIH/2023/009.

[57] P. Patil, “Healthcare Dataset,” Kaggle, [Online]. Available: https://www.kaggle.com/datasets/prasad22/healthcare-dataset. [Accessed: March 10, 2025].