International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 3, September 2026, pp. 1029~1037
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i3.pp1029-1037

Comparative analysis of decision tree-based machine learning algorithms: a trade-off between accuracy and interpretability

Alwatben Batoul Rashed[1], Aeshah Almutairi[1], Mansir Abubakar[2], Mardiyya Lawal Bagiwa[3]
[1]Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Malaysia
[3]Department of Computer Science and Information Technology, Al-Qalam University, Katsina, Nigeria

Article Info

Article history:

Received Apr 3, 2025
Revised May 24, 2026
Accepted Jun 17, 2026

Keywords:

Classification
Classifiers
Decision trees
JRip
Machine learning
Projective and recurring trees

ABSTRACT

This paper presents a comparative analysis of decision tree (DT)-based machine learning (ML) algorithms using 11 publicly available datasets from the knowledge extraction based on evolutionary learning (KEEL) repository. The study focuses on two critical aspects of DT-based algorithms: interpretability and accuracy. These measures often conflict, making it challenging to optimize both simultaneously. The research examines the trade-offs between accuracy and interpretability in three algorithms: Java implementation of the repeated incremental pruning to produce error reduction algorithm (JRip), projective and recurring trees (PART), and decision table. Experimental results demonstrate that JRip achieves the highest accuracy among the three algorithms across the datasets. However, decision table exhibits the highest interpretability, followed by PART, while JRip scores the lowest due to its pruning mechanism. These findings provide valuable insights into the balance between accuracy and interpretability, assisting researchers and practitioners in selecting the most suitable algorithm for classification tasks. The findings show a definite trade-off between accuracy and interpretability. Although JRip generates more complicated rule sets, it consistently delivers better accuracy across the majority of datasets. Decision tables offer simpler and easier-to-understand models, despite their lower accuracy.

Corresponding Author:

Mansir Abubakar
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA
Shah Alam, Selangor 4000, Malaysia
Email: mansir@uitm.edu.my

1. INTRODUCTION

Huge amounts of information are commonly processed by experts using statistical algorithms. However, these statistical algorithms are not capable enough to extract the required knowledge from the available data and information. This necessitated the exploration of alternative algorithms for knowledge discovery (KD) aimed at extracting useful information. The process of KD is carried out using several algorithms, which include classification, clustering, regression, and summarization [1]. Classification is a type of data mining task which includes discovering rules that divide data into distinct categories. It is essentially an algorithm of developing a model that explains and differentiates data classes in a test dataset based on a set of training data. Decision tree (DT)-based algorithms, such as decision table, projective and recurring trees (PART), and repeated incremental pruning to produce error reduction (JRip), have received a lot of attention because of their efficacy and interpretability in handling classification tasks. Rules-based classification algorithms focus more on improving accuracy without considering the rules generated.

The operational procedures of the system have not only considered the accuracy of the system but also the interpretability that indicates the capability to describe efficiently the operational procedure of a model.

The interplay between interpretability and accuracy becomes evident through the literature. The interpretability of machine learning (ML) models is frequently required [2], despite the fact that the subjective character of interpretability makes such a measurement difficult to define [3]. While accuracy remains a fundamental benchmark for model performance, the call for interpretability reflects a broader understanding of model utility. Models that are highly accurate yet lacking in interpretability may hinder their adoption in contexts where understanding the underlying decision-making process is critical. This is particularly evident in domains like healthcare and finance, where model predictions directly impact human lives and substantial resources [4]. Numerous points have been raised to emphasize the importance of incorporating interpretability in conjunction with accuracy.

Recent research has examined the use of various algorithms on databases, illuminating their advantages, drawbacks, and relative performance. This section aims to offer a concise examination of ongoing research focused on classification models. In a study conducted [5], a comparative analysis was performed to forecast bitcoin prices using both logistic regression (LR) and DT. The findings revealed that LR exhibited superior performance compared to DT based on accuracy, achieving 97.59%. Demir *et al*. [6] examined the effectiveness of ML algorithms in their study, including long-short term memory (LSTM), artificial neural networks (ANN), naive Bayes (NB), DT, and support vector machines (SVM). Qomariyah *et al.* [7] used the Waikato environment for knowledge analysis (WEKA) data mining tool to implement pairwise comparison and build a DT for learning user preferences. The performance of various DT algorithms, namely J48, ID3, random forest (RF), and random tree, was evaluated to assess their ability to estimate user preferences accurately based on the pairwise comparison dataset. The results demonstrated the remarkable capability of J48 in accurately estimating user preferences. Balabanova *et al.* [8] presented a comparative study focusing on identifying root mean square (RMS) noise levels of tones with different frequencies using ML algorithms. It employed k-nearest neighbors (KNN), DT, and NB, where KNN achieved an impressive accuracy. Both DT and KNN exhibited admirable efficiency and performance.

A study was conducted that compared various ML algorithms for classification applications [9]. The research utilized the WEKA tool to analyze and compare the performance of different classifiers, including C4.5 DT, NB, ZeroR, decision table, and OneR classification algorithms. Multiple performance indicators such as true negative, true positive, false negative, false positive, margin curve, and model building time were employed to evaluate the efficacy of the models. This investigation's conclusions shed light on the relative strengths and shortcomings of several ML algorithms in classification tasks. Kumar *et al.* [10] compared ML algorithms for detecting organic and non-organic cotton illnesses. The aim of the study was to evaluate and compare the efficacy of various ML algorithms in the identification of illnesses affecting cotton crops. The researchers tested the usefulness of several classifiers in discriminating between organic and non-organic cotton illnesses.

The investigation included performance criteria such as precision, accuracy, F1-score, and recall. The study gives insights into the appropriateness and effectiveness of several ML algorithms for identifying illnesses in cotton crops through their comparative examination. A study on establishing a knowledge-based system for newborn illness diagnosis and therapy suggestions [11]. To extract insights from a clinical dataset, the study applies a design science research strategy and a hybrid data mining process model. The paper considers three classification algorithms in WEKA tools, including PART, J48, and JRip. Among these methods, the PART algorithm performs best, with an accuracy of 98.06% under 10-fold cross-validation. Yavuz [12] investigated the variables connected to the COVID-19 epidemic utilizing data mining tools and approaches in the United States. The author develops rules and insights and examines the efficacy of data mining algorithms through the study. The findings show that the JRip algorithm has the greatest accurate classification rate of 53.29% and the highest accuracy of 0.491. Almutairi and Khan [13] presented a variety of ML techniques: Bayesian network (BN), NB, SVM, KNN, DT (J48), and RF. The dataset was acquired from Kaggle. The experimental results showed that RF outperformed the other methods, achieving an impressive accuracy. Pandian [14] provided an overview of big data analytics and ML in handling substantial volumes of data for valuable information extraction. Vijayakumar [15] proposed a capsule neural network (CapsNet) based classification system, trained on a limited dataset, for identifying cancerous brain tumors.

The analysis involved various ML techniques including gradient boosted trees (GBT), KNN, neural networks (NN), and ensemble learning, using data from coinmarketcap.com. Notably, the ensemble learning method outperformed state-of-the-art models [16]. Samaddar *et al.* [17] employed the WEKA data mining tool to build DT models through pairwise comparison for learning user preferences. J48 performed best when splitting data into 65% training and 35% test sets. A comparative study on ML approaches for detecting RMS noise levels of tones with different frequencies [18], [19] evaluated DT, KNN, and NB, where KNN exhibits high efficiency.

The necessity for accurate and comprehensible models has increased due to the growing reliance on ML systems in high-stakes industries including healthcare, finance, and education. Even though contemporary algorithms like RFs and gradient boosting machines [20]–[22] offer excellent prediction performance, they frequently operate as black-box models, which restricts openness and confidence. Interpretable machine learning (IML) and explainable artificial intelligence (XAI) are becoming more popular as a result [23]. An appealing mix between interpretability and accuracy is offered by DT-based algorithms, especially rule-based classifiers like JRip [24], PART, and decision tables [25], [26]. A basic problem still exists, though: increasing predictive performance frequently results in a more complex model, which makes it harder to interpret [27], [28].

Researchers investigated algorithms that balance these variables, such as joint optimization algorithms that take interpretability and accuracy into account. Contreras and Bocklitz [29] perform an empirical investigation of the interpretability-accuracy trade-off in explainable models. The study's goal is to look at how interpretability affects the performance of these models. The authors conduct a series of experiments to compare several ML algorithms, such as ANN, SVM, and DT, across numerous datasets [30]. While the findings indicate that interpretability often comes at the expense of accuracy, the study underlines the significance of achieving a balance between interpretability and accuracy. They investigate the efficacy of various models, such as SVM, DT, and rule-based classifiers, in terms of accuracy and interpretability in identifying attacks on networks. Contreras and Bocklitz [29] compare IML algorithms for predicting real estate prices using linear regression, DT, and rule-based models. They evaluate the interpretability of these algorithms using feature analysis and rule extraction. An interpretable DT algorithm is proposed in [23] with sparsity and accuracy optimization. The findings show that the suggested algorithm achieves a superior balance of sparsity and accuracy, exceeding the baseline algorithms in terms of interpretability as well as prediction accuracy. The difficulties in creating precise and comprehensible ML models when training data is scarce or diverse have been further brought to light by recent research. The choice of suitable classification methods is made even more crucial by the fact that small datasets frequently raise the danger of overfitting and decrease model generalizability [30]. Furthermore, explainable ML has drawn a lot of interest in high-stakes applications, proving that transparent models enhance user confidence and enable well-informed decision-making without significantly sacrificing predictive accuracy [31]. However, acknowledging and balancing the trade-off between interpretability and accuracy in ML algorithms remain a serious concern.

2. METHOD

According to Ortiz *et al.* [24], data mining is defined as the extraction of implicit, previously unknown, and potentially valuable information from data. The major steps involved in data mining are extract, transform, and load (ETL), followed by the analysis of this data using various algorithms to generate knowledge. Classification is a widely utilized supervised data mining technique that aims to find a model describing the data classes or concepts. Its primary purpose is to predict the class of objects using the model for unknown class labels. The classification model is built using training datasets and can be represented in different forms such as tables, trees, or rules. This paper focuses on rule-based classification algorithms, specifically PART, JRip, and decision tables.

Decision table: decision tables, as described by [25], are a collection of conditions and actions that organize data and knowledge. They are highly useful due to their transparent content presentation, minimizing redundancy and excessive complexity. Decision tables employ algorithmic techniques that involve a set of decision rules explaining the arrangement of circumstances to determine the appropriate actions to take.

JRip algorithm: JRip [26] is a popular algorithm used to analyze classes of different sizes and generate an initial set of rules for each class. It uses an incrementally reduced error method to classify all instances of a given class in the training data. The method then derives a set of rules that apply to all members of that class. This algorithm works through the following steps, as shown in Algorithm 1.

Algorithm 1. JRip algorithm

1: Building stage: repeat steps 2 and 3 until the description length (DL) of the ruleset is reached.
2: Develop phase: construct a rule by iteratively adding antecedents to it in a greedy manner until it becomes perfect.
3: Pruning phase: sequentially prune each rule, including any final sequences of antecedents.
4: Optimization stage: after generating the initial ruleset, produce and prune two variations of each rule using randomized data, following steps 2 and 3.
5: Finally, after examining the rules, if residual positives remain, generate additional rules based on those residual positives.

PART algorithm: PART is an algorithm derived from the C4.5 and RIPPER/JRip algorithms, designed to construct a partial DT rather than a fully expanded one. This algorithm combines the building and pruning algorithms to generate a stable subtree that cannot be further simplified [16]. Algorithm 2 describes how the tree-building algorithm works.

Algorithm 2. PART algorithm
1: Recursively divide a set of inputs into a partial tree.
2: In the initial step, select a test and partition the set into subsets.
3: Extend the subsets in decreasing order of average entropy, starting from the smallest subset. This recursive process continues until a subset becomes a leaf, followed by backtracking.
4: Pruning takes place when an internal node with its offspring extended into leaves is encountered. The algorithm evaluates whether this node should be replaced with a single leaf.
5: If replacement occurs, the algorithm proceeds as usual by examining the siblings of the newly added node.

2.1. Classification evaluation datasets

In this study, ten publicly available datasets were selected from knowledge extraction based on evolutionary learning (KEEL) repository, a globally well-known benchmark dataset, for evaluation. The datasets have been popularly used in several studies to evaluate the performance of algorithms. Table 1 provides an overview of the key features of these datasets. The datasets present several desirable features, such as having a dissimilar number of variables from 3 to 18, and diverse sizes ranging from 155 to 846 patterns.

Table 1. The selected datasets are characterized by their specific properties

Dataset	# Variables	# Classes	# Instances
BUPA	7	1	345
New thyroid	6	2	213
Pima	9	3	767
Cleveland	13	5	303
SAheart	9	2	462
Haberman	3	2	306
Hert	13	2	270
Win	13	3	178

2.2. Experiment setup

In terms of evaluating algorithm performance, the fundamental measure of accuracy, denoted as ACC(X), is calculated as the ratio of correctly classified instances (Nc) to the total number of instances (Nt). On the other hand, interpretability is assessed by considering the complexity, which is typically determined by the number of rules and should be minimized. To mitigate potential biases associated with randomly splitting a dataset into learning and test sets, the main experiment employed the cross-validation method. The entire dataset was randomly divided into distinct subsets that were mutually exclusive, containing roughly equal numbers of records and exhibiting similar class distributions. The learning process was repeated k times, where each time a different subset was utilized as the test set while the remaining subsets served as the training set. The average results across k runs were then computed. For consistency, the study employed a similar 10-fold cross-validation approach for both the algorithms and baseline systems. Each dataset was randomly partitioned into one-fold for the testing phase and nine folds for the training phase.

3. RESULTS AND DISCUSSION

In this research, the experiment is conducted using the WEKA tool, which is a software tool that encompasses a range of ML algorithms for tasks such as data classification, clustering, regression, and visualization. A significant advantage of utilizing WEKA is its customization flexibility to meet specific requirements. The primary objective of this study is to conduct a comparative analysis, evaluating the performance of these algorithms based on metrics of accuracy and interpretability. In many classification rules, algorithms are generally used to generate the classification rules. In the present section, a comparison of three classification algorithms: PART decision, J48, and RepTree using the WEKA tool are presented in Table 2.

The results show that for accuracy and interpretability, JRip outperformed PART and decision table in most eleven datasets. For PIMA datasets, JRip outperformed the PART and decision table algorithms in terms of interpretability and became competitive in accuracy. The results show that the algorithm recorded

74.48% as average accuracy and four rules as interpretability value. For the BUPA dataset, decision table outperformed the PART and JRip algorithms in terms of interpretability, and JRip outperformed the PART and decision table algorithms in accuracy. For the new thyroid dataset, JRip outperformed the PART and decision table algorithms in terms of interpretability and became competitive in accuracy. The results show that the algorithms recorded 93.02% as average accuracy and four rules as interpretability value. For the Haberman dataset, JRip outperformed the PART and decision table algorithms in terms of interpretability and became competitive in accuracy. The results show that the algorithm recorded 73.53% as average accuracy and two rules as interpretability value. For Wisconsin datasets, JRip outperformed PART and decision table. The algorithms were evaluated based on their interpretability and accuracy.

The results also demonstrated that the algorithm recorded 95.57% as average accuracy and five rules as interpretability value, as shown in Figure 1. For the breast cancer dataset, decision table outperformed the PART and JRip algorithms in terms of accuracy. The results show that the decision table algorithm recorded 72.38% as average accuracy and two rules as interpretability value for JRip. For the Wisconsin diagnostic breast cancer (WDBS) dataset, JRip outperformed the PART and decision table. The algorithm recorded 71.5% as average accuracy and 17 rules as interpretability value for JRip. For the Cleveland dataset, JRip outperformed the PART and decision table. The results show that the algorithm recorded 54.78% as average accuracy and one rule as interpretability value.

Table 2. The performance of the classification algorithms

Dataset	Decision table	JRip	PART
PIMA	Rules: 63 Accuracy: 74.48%	Rules: 4 Accuracy: 74.48%	Rules: 13 Accuracy: 74.78%
BUPA	Rules: 3 Accuracy: 59.71 %	Rules: 5 Accuracy:67.8%	Rules: 5 Accuracy: 64.1%
New thyroid	Rules: 21 Accuracy: 93.4884%	Rules: 4 Accuracy: 93%	Rules: 43 Accuracy: 93.95%
Haberman	Rules: 2 Accuracy: 72.5%	Rules: 2 Accuracy: 73.5%	Rules: 4 Accuracy: 73.8%
Wisconsin	Rules: 20 Accuracy: 93.8%	Rules: 5 Accuracy: 95.6%	Rules: 11 Accuracy: 94.7%
Breast Cancer	Rules: 11 Accuracy: 72.3776%	Rules: 2 Accuracy: 70.2%	Rules: 24 Accuracy: 68.5315%
WDBS	Rules: 21 Accuracy: 93.6731%	Rules: 6 Accuracy: 93.8489%	Rules: 9 Accuracy: 92.4429%
WINE	Rules: 33 Accuracy: 87.0787 %	Rules: 5 Accuracy: 93.2584 %	Rules: 5 Accuracy: 91.573%
SAheart	Rules: 29 Accuracy: 68.8312%	Rules: :3 Accuracy: 67.7489%	Rules: 8 Accuracy: 71.2121%
Vehicle	Rules: 62 Accuracy: 67.2577%	Rules: 17 Accuracy: 69.0307%	Rules: 29 Accuracy: 71.513%
Cleveland	Rules: 8 Accuracy: 54.5%	Rules: 1 Accuracy: 54.7%	Rules: 48 Accuracy: 52.4%

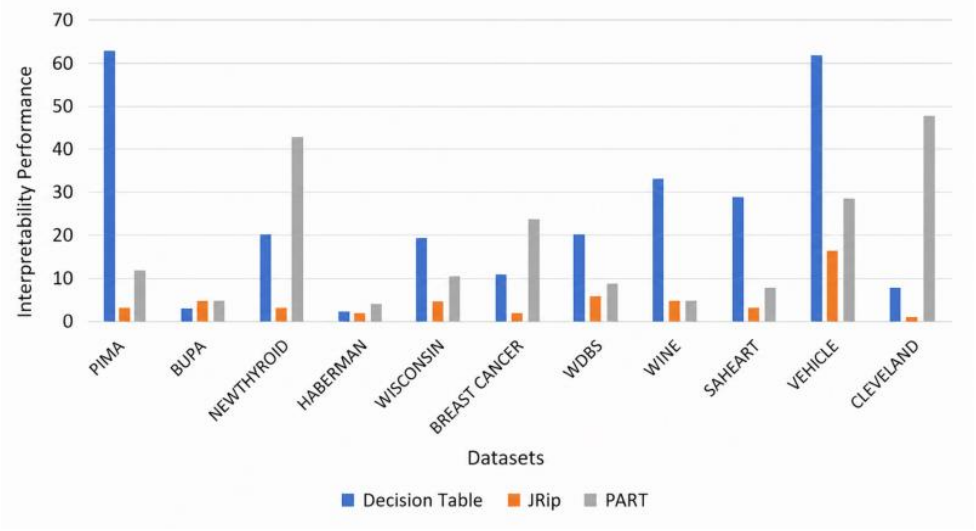


Figure 1. Comparison of different algorithms in interpretability results

The results show that the algorithm recorded 93.89% as average accuracy and six rules as interpretability value. For the Wine dataset, JRip outperformed the PART and decision table. The results show that the algorithm recorded 93.3% as average accuracy and five rules as interpretability value. For the SAheart dataset, Rip outperformed the PART and decision table algorithms in terms of interpretability, and PART outperformed the JRip and decision table algorithms in accuracy. The results show the PART algorithm recorded 71.2% as average accuracy and three rules as interpretability value for JRip. For the vehicle dataset, Rip outperformed the PART and decision table algorithms in terms of interpretability, and PART outperformed the JRip and decision table algorithms in accuracy. The results show the PART in terms of interpretability, the results shown in Figure 1, JRip had significantly lesser interpretability because of its pruning method, whereas decision table attained the maximum interpretability and PART came in second [22].

Figure 2 illustrates the classification accuracy of three ML algorithms: decision table, JRip, and PART across eleven datasets. The comparison shows that the decision table algorithm consistently achieves a comparable accuracy to JRip and PART, making it the most effective model in most cases. Datasets such as new thyroid, Wisconsin breast cancer, and WDBS exhibit the highest accuracy levels across all three algorithms, exceeding 90%, indicating that these datasets are well-suited for classification. Conversely, datasets such as BUPA, Haberman, SAheart, vehicle, and Cleveland demonstrate lower accuracy, suggesting that they pose more challenges for classification due to potential data complexities or imbalances [30].

When comparing the three algorithms, decision table generally outperforms both JRip and PART, reinforcing its reliability for classification tasks [22]. However, JRip and PART display similar performance trends, with JRip slightly outperforming PART on certain datasets. This suggests that while decision table provides the best accuracy overall, JRip may still be a competitive choice depending on the dataset. The results highlight the importance of dataset characteristics in influencing algorithm effectiveness and suggest that selecting the right classification model requires careful consideration of both accuracy and interpretability for depending on specific applications [26]. The findings show a definite trade-off between accuracy and interpretability [18]. Although JRip generates more complicated rule sets, it consistently delivers better accuracy across the majority of datasets. Decision tables offer simpler and easier-to-understand models, despite their lower accuracy.

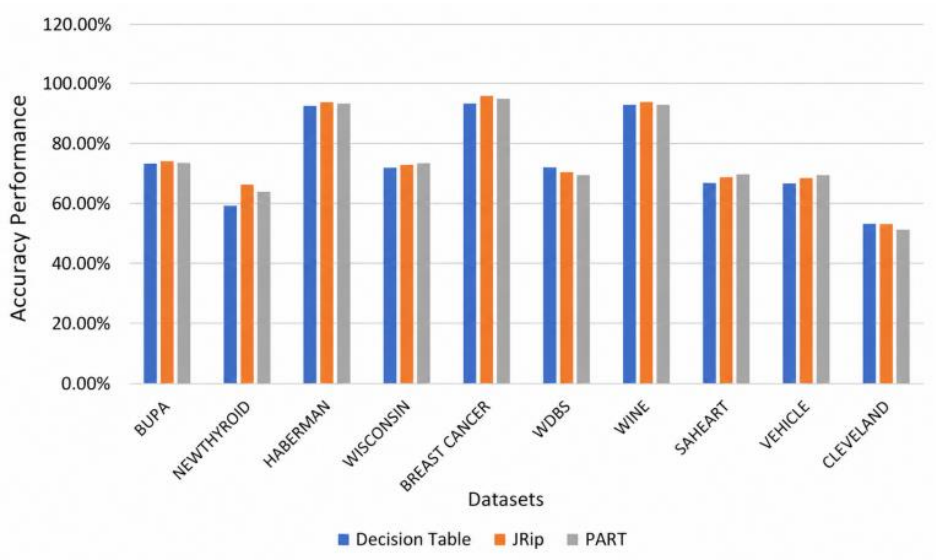


Figure 2. Comparison of different algorithms in accuracy

4. CONCLUSION

This research paper presents a comparative analysis to evaluate the performance of DT-based algorithms in terms of accuracy and interpretability over 11 datasets. On the evaluation of the performance of DT-based algorithms, the basic accuracy measure used was ACC(X), which indicates the number of correctly classified instances and the number of total instances. When the classification was performed using the attributes in individual X, the accuracy of the classifier (X) was maximized. The basic measure of DT-based algorithms' interpretability is INT(X). On complexity at the rule-based level, interpretability is based on number of rules. Interpretability was minimized. The result recorded significant improvement in both

accuracy and interpretability with JRip compared to the DT-based algorithms in the result table in most of the eleven datasets. In terms of interpretability JRip had significantly outperformed because of its pruning method, whereas decision table attained the maximum interpretability and PART came in second. In terms of accuracy, the PART and JRip are the most accurate then the decision table. This research demonstrates that no single model maximizes interpretability and accuracy. While decision table provides better interpretability, JRip performs exceptionally well in terms of prediction. Future research will investigate SHAP-based analysis and hybrid explainable models.

ACKNOWLEDGMENTS

The authors would like to thank the Deanship of Scientific Research, Qassim University, Buraidah, Saudi Arabia, and College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Malaysia, for all necessary support.

FUNDING INFORMATION

This work is funded by the Deanship of Scientific Research, Qassim University, Buraidah, Saudi Arabia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Alwatben Batoul Rashed	✓	✓	✓	✓	✓	✓		✓		✓	✓		✓	✓
Aeshah Almutairi		✓				✓		✓	✓	✓	✓	✓	✓	
Mansir Abubakar	✓		✓	✓			✓		✓	✓	✓			
Mardiyya Lawal Bagiwa		✓	✓	✓		✓				✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest in this research.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

All datasets used in this study are open access, secondary data which can be found at the KEEL machine learning repository at https://sci2s.ugr.es/keel/datasets.php.

REFERENCES

[1] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.

[2] Y. Kodratoff, "Guest editor's introduction," *AI Communications*, vol. 7, no. 2, pp. 83–85, 1994, doi: 10.3233/AIC-1994-7201.

[3] S. Rüping, "Learning interpretable models." Ph.D. dissertation, Universität Dortmund, Dortmund, Germany, 2006, doi: 10.17877/DE290R-8863.

[4] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures," *Information Sciences*, vol. 181, no. 20, pp. 4340–4360, Oct. 2011, doi: 10.1016/j.ins.2011.02.021.

[5] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-currency price prediction using decision tree and regression techniques," in *2019 3rd International Conference on Trends in Electronics and Informatics*, Apr. 2019, pp. 190–194, doi: 10.1109/ICOEI.2019.8862585.

[6] A. Demir, B. N. Akilotu, Z. Kadiroglu, and A. Sengur, "Bitcoin price prediction using machine learning methods," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, Nov. 2019, pp. 1–4 doi: 10.1109/UBMYK48245.2019.8965445.

[7] N. N. Qomariyah, E. Heriyanni, A. N. Fajar, and D. Kazakov, "Comparative analysis of decision tree algorithm for learning ordinal data expressed as pairwise comparisons," in *2020 8th International Conference on Information and Communication Technology*, Jun. 2020, pp. 1–4, doi: 10.1109/ICoICT49345.2020.9166341.

[8] I. S. Balabanova, V. I. Markova, S. S. Kostadinova, and G. I. Georgiev, "Comparative analysis between machine learning methods in tones classification," in *2020 28th National Conference with International Participation (TELECOM)*, Oct. 2020, pp. 45–48, doi: 10.1109/TELECOM50385.2020.9299535.

[9] V. Sheth, U. Tripathi, and A. Sharma, "A comparative analysis of machine learning algorithms for classification purpose," *Procedia Computer Science*, vol. 215, pp. 422–431, 2022, doi: 10.1016/j.procs.2022.12.044.

[10] S. Kumar *et al.*, "A comparative analysis of machine learning algorithms for detection of organic and nonorganic cotton diseases," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–18, Jun. 2021, doi: 10.1155/2021/1790171.

[11] D. Wendimu and K. Biredagn, "Developing a knowledge-based system for diagnosis and treatment recommendation of neonatal diseases," *Cogent Engineering*, vol. 10, no. 1, Dec. 2023, doi: 10.1080/23311916.2022.2153567.

[12] Ö. Yavuz, "A data mining analysis of COVID-19 cases in states of United States of America," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 1754–1758, Apr. 2022, doi: 10.11591/ijece.v12i2.pp1754-1758.

[13] A. Almutairi and R. U. Khan, "Image-based classical features and machine learning analysis of skin cancer instances," *Applied Sciences*, vol. 13, no. 13, Jun. 2023, doi: 10.3390/app13137712.

[14] A. P. Pandian, "Review of machine learning techniques for voluminous information management," *Journal of Soft Computing Paradigm*, vol. 1, no. 2, pp. 103–112, Dec. 2019, doi: 10.36548/jscp.2019.2.005.

[15] T. Vijayakumar, "Classification of brain cancer type using machine learning," *Journal of Artificial Intelligence and Capsule Networks*, vol. 1, no. 2, pp. 105–113, Dec. 2019, doi: 10.36548/jaicn.2019.2.006.

[16] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," *Physica A: Statistical Mechanics and its Applications*, vol. 551, Aug. 2020, doi: 10.1016/j.physa.2020.124569.

[17] M. Samaddar, R. Roy, S. De, and R. Karmakar, "A comparative study of different machine learning algorithms on bitcoin value prediction," in *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies*, Feb. 2021, pp. 1–7, doi: 10.1109/ICAECT49130.2021.9392629.

[18] N. Jo, S. Aghaei, J. Benson, A. Gomez, and P. Vayanos, "Learning optimal fair decision trees: trade-offs between interpretability, fairness, and accuracy," in *2023 AAAI/ACM Conference on AI, Ethics, and Society*, Aug. 2023, pp. 181–192, doi: 10.1145/3600211.3604664.

[19] M. Czajkowski, K. Jurczuk, and M. Kretowski, "Steering the interpretability of decision trees using lasso regression - an evolutionary perspective," *Information Sciences*, vol. 638, Aug. 2023, doi: 10.1016/j.ins.2023.118944.

[20] W. J. Murdoch, C. Singh, K. Kumbier, R. A.-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.

[21] J. Alòs, C. Ansótegui, and E. Torres, "Interpretable decision trees through MaxSAT," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8303–8323, Aug. 2023, doi: 10.1007/s10462-022-10377-0.

[22] R. Kohavi, "The power of decision tables," in *Machine Learning: ECML-95*, Berlin, Heidelberg: Springer, 1995, pp. 174–189, doi: 10.1007/3-540-59286-5_57.

[23] M. Razavi *et al.*, "Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: scoping review," *JMIR Mental Health*, vol. 11, Aug. 2024, doi: 10.2196/53714.

[24] B. L. Ortiz *et al.*, "Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care," *JMIR mHealth and uHealth*, vol. 12, Sep. 2024, doi: 10.2196/59587.

[25] S. Golazad, A. Mohammadi, A. Rashidi, and M. Ilbeigi, "From raw to refined: data preprocessing for construction machine learning (ML), deep learning (DL), and reinforcement learning (RL) models," *Automation in Construction*, vol. 168, Dec. 2024, doi: 10.1016/j.autcon.2024.105844.

[26] M. Atzmueller, J. Fürnkranz, T. Kliegr, and U. Schmid, "Explainable and interpretable machine learning and data mining," *Data Mining and Knowledge Discovery*, vol. 38, no. 5, pp. 2571–2595, Sep. 2024, doi: 10.1007/s10618-024-01041-y.

[27] A. Tursunalieva, D. L. J. Alexander, R. Dunne, J. Li, L. Riera, and Y. Zhao, "Making sense of machine learning: a review of interpretation techniques and their applications," *Applied Sciences*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020496.

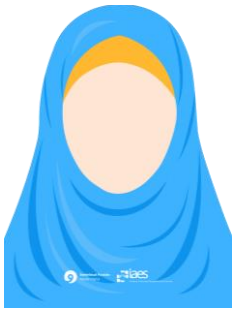
[28] N. Mehdiyev, M. Majlatow, and P. Fettke, "Interpretable and explainable machine learning methods for predictive process monitoring: a systematic literature review," *Artificial Intelligence Review*, vol. 58, no. 12, Oct. 2025, doi: 10.1007/s10462-025-11399-0.

[29] J. Contreras and T. Bocklitz, "Explainable artificial intelligence for spectroscopy data: a review," *Pflügers Archiv - European Journal of Physiology*, vol. 477, no. 4, pp. 603–615, Apr. 2025, doi: 10.1007/s00424-024-02997-y.

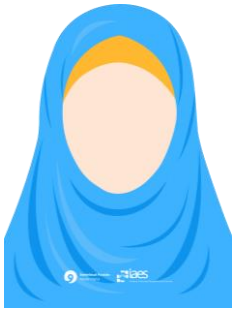
[30] M. Z. Naser, "A review of machine learning with small and limited data," *Journal of Big Data*, vol. 13, no. 1, Jan. 2026, doi: 10.1186/s40537-025-01346-9.

[31] A. Trelles, T. F. Ruiz, and A. P. Rojo, "Systematic review and meta-analysis of explainable machine learning models for clinical depression detection," *Behavioral Sciences*, vol. 15, no. 11, Oct. 2025, doi: 10.3390/bs15111476.

BIOGRAPHIES OF AUTHORS

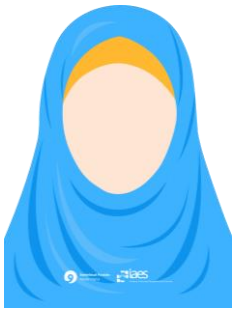


Alwatben Batoul Rashed holds a Ph.D. in Computer Science from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, specializing in intelligent computing. She is currently an assistant professor at the College of Computer, Qassim University, Saudi Arabia. Her research focuses on intelligent computing, data mining, and algorithm optimization, with an emphasis on developing efficient computational models for complex problem-solving. She has published and contributed to numerous research projects, advancing the fields of machine learning, predictive analytics, and data-driven decision-making. She can be contacted at email: bwtban@qu.edu.sa.



Aeshah Almutairi received her Ph.D. degree in Computer Science from Durham University, United Kingdom. She is currently working as an assistant professor at College of Computer, Qassim University, Saudi Arabia. Her research interests are in the area of computer graphics. In particular, she aims to employ machine learning techniques for better recognition and identification. She can be contacted at email: aeshah.almutairi@qu.edu.sa.

Mansir Abubakar is a senior lecturer at Universiti Teknologi MARA. He earned his Ph.D. degree from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, with specialization in intelligent computing. He is a certified LMS facilitator by the New Creation Foundation, Netherlands. He has been a member of Volunteer Expert Group for the development of National AI Policy (VEG-NAIP) in Nigeria. He felicitated the National Adopted Village for Smart Agriculture (NAVSA) program at AUK in collaboration with the National Information Technology Development Agency (NITDA), Nigeria. He has been an IEEE member since 2016. He has attended many conferences, seminars, and workshops in the field of computer science and emerging technologies. He has published many articles in reputable journals. His primary research interests include intelligent computing, semantic web technology, NLP, data mining, and algorithm optimization. He can be contacted at email: mansir@uitm.edu.my.



Mardiyya Lawal Bagiwa is a lecturer in the Department of Computer Science and Information Technology at Al-Qalam University, Katsina, Nigeria. She holds a bachelor of science (B.Sc.) in Computer Science from Al-Qalam University, Katsina, and a master of science (M.Sc.) in Information Technology from Bayero University, Kano. With a strong passion for technology and education, she is dedicated to advancing research and knowledge in the field of computer science and emerging technologies. She can be contacted at email: mardiyyalawalbagiwa@auk.edu.ng.