Research article
DOI: https://doi.org/10.18721/JCSTCS.19102
UDC 004.85

ENHANCING BOUNDARY STABILITY IN DECISION TREES AND RANDOM FORESTS: A WEIGHTED SAMPLE DUPLICATION APPROACH

A.V. Konstantinov[1], *A.P. Elizarova*[2], *L.V. Utkin*[1] ✉

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;

[2] National Research University Higher School of Economics, St. Petersburg branch, St. Petersburg, Russian Federation

✉ lev.utkin@gmail.com

Abstract. Decision trees and their ensemble extensions, such as random forests, are widely used as classification models due to their simplicity and interpretability. However, in many real-world tasks where class labels overlap in the feature space, standard decision trees rely on hard splits that create fragile decision boundaries. In these regions, small perturbations in the input values can lead to misclassification, reducing the reliability of the model. To address this issue, we propose a localized data duplication mechanism that modifies the standard CART algorithm by duplicating samples located near the chosen split threshold into both child nodes. To prevent these duplicated samples from overpowering the nodes, they are assigned a reduced weight based on a smoothly decaying function relative to their distance from the threshold. This approach allows both child nodes to learn from ambiguous regions, preserving information about uncertainty while maintaining the axis-aligned deterministic structure of classical decision trees. When applied within a random forest framework, the duplication process also increases ensemble diversity. Experimental evaluation on 11 real-world datasets with varying degrees of class overlap demonstrates that the proposed modification consistently improves ROC-AUC scores and boundary stability while keeping computational costs low.

Keywords: machine learning, decision trees, random forest, classification, data duplication

Acknowledgements: The research was supported by the Russian Science Foundation grant No. 25-21-00103 “New methods and models of multivariate learning in medical diagnostics”. Available online: https://rscf.ru/project/25-21-00103/.

Citation: Konstantinov A.V., Elizarova A.P., Utkin L.V. Enhancing Boundary Stability in Decision Trees and Random Forests: A Weighted Sample Duplication Approach. Computing, Telecommunications and Control, 2026, Vol. 19, No. 1, Pp. 16–25. DOI: 10.18721/JCSTCS.19102

© Konstantinov A.V., Elizarova A.P., Utkin L.V., 2026. Published by Peter the Great St. Petersburg Polytechnic University

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.19102
УДК 004.85

ПОВЫШЕНИЕ УСТОЙЧИВОСТИ ГРАНИЦ В ДЕРЕВЬЯХ РЕШЕНИЙ И СЛУЧАЙНЫХ ЛЕСАХ: ПОДХОД С ИСПОЛЬЗОВАНИЕМ ВЗВЕШЕННОГО ДУБЛИРОВАНИЯ ВЫБОРКИ

А.В. Константинов[1], *А.П. Елизарова*[2], *Л.В. Уткин*[1] ✉

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация;

[2] НИУ ВШЭ, Санкт-Петербургский филиал, Санкт-Петербург, Российская Федерация

✉ lev.utkin@gmail.com

Аннотация. Деревья решений и их ансамблевые расширения, такие как случайные леса, широко используются в качестве моделей классификации благодаря своей простоте и интерпретируемости. Однако во многих реальных задачах, где метки классов перекрываются в пространстве признаков, стандартные деревья решений полагаются на жесткие разбиения, которые создают слабые границы принятия решений. В этих областях небольшие возмущения входных значений могут привести к неправильной классификации, снижая надежность модели. Для решения этой проблемы мы предлагаем механизм локализованного дублирования данных, который модифицирует стандартный алгоритм CART (Classification and Regression Tree) путем дублирования образцов, расположенных вблизи выбранного порога разбиения, в оба дочерних узла. Чтобы предотвратить перегрузку узлов этими дублированными образцами, им присваивается уменьшенный вес на основе плавно убывающей функции относительно их расстояния от порога. Такой подход позволяет обоим дочерним узлам обучаться на неоднозначных областях, сохраняя информацию о неопределенности, одновременно поддерживая выровненную по осям детерминированную структуру классических деревьев решений. При применении в рамках случайного леса процесс дублирования также увеличивает разнообразие ансамбля. Экспериментальная оценка на 11 реальных наборах данных с различной степенью перекрытия классов показывает, что предложенная модификация последовательно улучшает показатели ROC-AUC и устойчивость границ, сохраняя при этом низкие вычислительные затраты.

Ключевые слова: машинное обучение, деревья решений, случайный лес, классификация, дублирование данных

Финансирование: Исследование выполнено за счет гранта Российского научного фонда в рамках реализации проекта «Новые методы и модели многовариантного обучения в задачах медицинской диагностики» (Соглашение №25-21-00103, https://rscf.ru/project/25-21-00103/).

Для цитирования: Konstantinov A.V., Elizarova A.P., Utkin L.V. Enhancing Boundary Stability in Decision Trees and Random Forests: A Weighted Sample Duplication Approach // Computing, Telecommunications and Control. 2026. Т. 19, № 1. С. 16–25. DOI: 10.18721/JCSTCS.19102

Introduction

Decision trees and their ensemble extensions, such as random forests, are popular because they are simple, often perform well in practice, and offer interpretability. Each decision tree splits the feature space in a hierarchical way, which allows users to trace exactly how a prediction is made. Ensembles average over many trees, which tends to reduce variance and improve generalization.

© Константинов А.В., Елизарова А.П., Уткин Л.В., 2026. Издатель: Санкт-Петербургский политехнический университет Петра Великого

In many real-world classification tasks, especially in medical or environmental settings, class-conditional distributions overlap in feature space. That means that for certain regions, observations from more than one class lie very close to each other in the features. In these overlapping regions, small perturbations of input values or small changes in the split thresholds can lead to different class predictions. This reduces predictive stability and increases variance near class boundaries. A standard decision tree uses hard splits: at each internal node it chooses a feature and a threshold that minimize impurity (such as Gini impurity or entropy). When classes overlap, small perturbations can lead to misclassification due to fragile decision boundaries. As trees grow deeper, the number of samples in each node tends to shrink, increasing sensitivity to small changes and decreasing robustness near class boundaries.

This work proposes a modification of tree construction that introduces a stochastic or softening element in handling samples near decision boundaries. Specifically, after a split is determined, samples whose values are close to the threshold (i.e., near the hyperplane defined by the feature and threshold) are duplicated into both child nodes. The influence of each duplicated sample decreases as its distance from the threshold increases. This duplication has two key effects: it preserves information about ambiguous or uncertain regions, and it increases diversity among trees when used in a random forest ensemble. The goal is to make decision boundaries more precise, reduce errors arising in overlapping class regions, and improve learning in difficult parts of feature space, while retaining much of the interpretability and simplicity of standard CART-style trees.

Related works

The problem of class overlap, uncertainty, and decision boundary instability has been extensively addressed in recent literature. Existing approaches can broadly be categorized into probabilistic tree modifications, fuzzy logic parameterizations, and data-level boundary interventions. These approaches help illustrate why the proposed method is needed.

The concept of soft or probabilistic splits replaces rigid decision boundaries with weighted contributions from both child branches. While foundational work demonstrated that soft thresholds reduce generalization error, soft decision trees (SDTs) have recently seen a massive resurgence as a means to balance deep learning's predictive power with interpretability [1]. For instance, the fusion of SDTs with concept-based convolutional models was used to create transparent classification systems in complex, overlapping spaces [3]. Recent advancements also include embedding spatial visual attention directly into the inner nodes of SDTs to ensure path-based interpretability [2]. Despite these improvements, SDTs fundamentally alter the standard tree routing mechanism by relying on fully differentiable gradient descent and sigmoid-based probability branching rather than classical hard splits [1, 2].

Fuzzy decision trees (FDTs) explicitly model uncertainty by mapping features to fuzzy membership functions [5, 7]. Recent advancements in this domain emphasize adaptability and scalability. Rabcan et al. (2025) proposed a dynamic FDT that uses cumulative mutual information to incrementally adapt to evolving, overlapping data streams [4]. FDT optimization has also been applied successfully in complex environments, such as personalized hybrid learning systems, utilizing cost-complexity pruning to prevent boundary overfitting [6]. Furthermore, FDTs are proven highly effective for regression and classification in uncertain physical environments, such as climate factor modeling [5]. However, these approaches introduce significant mathematical overhead: they require predefined triangular or trapezoidal membership functions and linguistic parameters that must be manually tuned, deviating from the algorithmic simplicity of axis-aligned crisp splits [4, 6].

Instead of modifying the tree structure probabilistically, several studies focus directly on the geometric boundaries and data overlap. Class overlap – where instances from different categories share the exact same feature space – is proven to degrade the performance and interpretability of predictive models significantly [8]. To tackle this, novel decision rules based on boundary mixed attribute dependency have been introduced, allowing trees to factor the boundary region explicitly as a measure of knowledge uncertainty [9]. Additionally, when dealing with noisy class variables at the boundaries, random forest algorithms exhibit measurably higher robustness compared to standalone C4.5 or CART trees [10].

When individual tree boundaries fail to resolve class overlap, ensemble and distributed techniques provide an alternative. Coalition-based decision trees now use decision template fusion to manage conflicting or overlapping data sources across distributed networks [11]. In complex medical domains characterized by overlapping tumor classes, multiclass tree ensembles combining perfusion and spectroscopy data have achieved high accuracy by extracting precise, overlapping rule combinations [12]. Nevertheless, these techniques typically operate as global or post-hoc ensembles, not fundamentally altering how the individual underlying decision tree processes the ambiguous boundary during the split search.

Despite these advances, a critical gap remains. Current state-of-the-art methods either mandate computationally expensive gradient routing [1, 3], require complex fuzzy membership tuning [4, 6], or rely on global data-level interventions [8]. There is a distinct need for methods that preserve the clean CART structure, do not require extensive global parameter tuning, and maintain interpretability while dynamically improving robustness at class boundaries.

Method

The proposed method modifies the construction of standard decision trees in order to handle uncertain or overlapping class regions more effectively, as well as small sample size, while keeping the simplicity and interpretability of classical CART. The main idea is to duplicate samples that lie close to a chosen split threshold, so that both resulting child nodes can learn from data in ambiguous regions. This adjustment helps the model to better capture local structure near class boundaries and to become less sensitive to small variations in input data or threshold placement. Furthermore, it allows for a more precise determination of split thresholds if they are shared among child nodes with the same parent.

In a traditional decision tree, each node contains a subset of the training data and selects one feature, defined as x_j, and a threshold value, called t. The goal of this step is to find the split that best reduces impurity, for example measured by Gini impurity or entropy. Once the threshold is determined, all samples with feature values less than or equal to t are sent to the left child node, and the remaining points, i.e., those with values greater than t are sent to the right child node. The algorithm then continues recursively on both child nodes until a stopping criterion is met, such as a minimum node size or maximum depth.

Let D denote the current node dataset. At the root node, D is the whole dataset:

$$D = \left\{ \left(x^{(i)}, y^{(i)} \right) \right\}_{i=1}^{N},$$

where each feature vector $x = (x_1, \ldots, x_f)$ is of size f. In classical CART algorithm, at the current node, when the split is found, D_L and D_R are subsets to which D is partitioned after splitting by feature x_j and threshold t:

$$D_L = \left\{ (x, y) \in D \middle| x_j \leq t \right\}, \quad D_R = \left\{ (x, y) \in D \middle| x_j \rangle t \right\}.$$

The subset D_L is considered in the left child node, and D_R – in the right.

This deterministic splitting process works well when the classes are clearly separated, but it can be unreliable when the data contain regions of overlap, or the sample size within the node becomes insufficient for reliable splitting threshold estimation. In such areas, some samples from different classes have very similar feature values, and even a small change in t or in the input values can send them to opposite sides of the split. As the tree grows deeper, each node contains fewer samples, which increases the instability of these boundaries, leading to possibly higher error rate.

To address this issue, we propose a localized weighted sample duplication mechanism. After the best split feature and threshold are found, the samples that lie close to the chosen threshold along the selected feature are selected (near-boundary samples), where closeness is defined by a distance between the feature value and the threshold and the number of samples that are nearest to the threshold. Once the near-boundary samples are identified, they are duplicated so that they appear in both child nodes. In other words, each of these samples contributes to learning in both branches of the split. This ensures that the information about uncertain regions is not lost. To prevent the duplication from overpowering the data in either node, each duplicated sample receives a reduced influence, which depends on how far it is from the threshold. Samples that are exactly at or very close to the threshold have the highest influence, and those further away have gradually smaller influence.

This mechanism preserves the deterministic structure of a standard decision tree. The split itself remains hard and axis-aligned, so the interpretability of the resulting model is unchanged. However, by allowing both child nodes to learn from near-boundary samples, the model becomes more robust. It no longer discards information from uncertain regions, and it gains a more stable decision boundary, less sensitive to small variations in the data.

When this method is used within a random forest, the duplication process also introduces additional diversity among trees. Because each tree in the ensemble is built from a bootstrap sample and may select different features and thresholds, the specific boundary samples that are duplicated vary between trees. This increases the variety of the ensemble members, which in turn improves generalization and reduces the correlation of their errors, especially in regions of class overlap.

A few practical considerations are important for implementation. Since duplicating samples changes the number of observations seen by each node, stopping criteria based on sample counts may need to account for the reduced influence or weight of duplicated points rather than counting them as separate full samples. It is also necessary to choose the width of the near-boundary region carefully: if it is too wide, unnecessary duplication may increase computational cost or overfit the boundary; if it is too narrow, the effect may be too small. Similarly, the decay rule for sample influence must ensure a smooth transition between duplicated and non-duplicated samples. Despite these considerations, the method remains simple to integrate into standard tree algorithms because it does not alter impurity calculations, prediction procedures, or the overall recursive structure.

In summary, the method retains the key advantages of CART, including simplicity, interpretability, and computational efficiency, while improving robustness in overlapping class regions. It achieves this by allowing both branches of a split to learn from data that lie close to uncertain boundaries, thereby reducing error sensitivity and enhancing the reliability of both individual trees and ensembles built from them.

This mechanism introduces randomized noise into sampling procedure for child node data, as shown in Fig. 1. By exposing subsequent splits to both sides of the uncertain region, the tree becomes less sensitive to exact threshold placement and gains resilience to perturbations.

Let the absolute distance from the decision boundary for sample i along the splitting feature j be defined as:

$$d_i = \left| x_{ij} - t \right|.$$

Let r be a user-defined parameter controlling the proportion of copied samples. The number of samples duplicated from the opposite branch is:

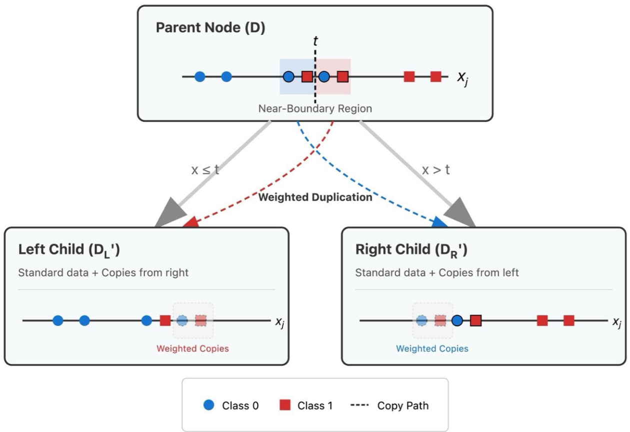

Fig. 1. Sample duplication near decision threshold

$$k = \lfloor r \cdot \min(|D_L|, |D_R|) \rfloor.$$

Select k samples with smallest d_i values from each side and duplicate them into the opposite child subset. Thus, final subsets become:

$$D_{L'} = D_L \cup \mathrm{Copy}_R, \quad D_{R'} = D_R \cup \mathrm{Copy}_L.$$

Each duplicated observation receives a weight reflecting its proximity to the decision boundary. We use a smooth decaying function:

$$w_i = e^{-a|d_i|}, \quad a > 0,$$

where a controls the decay rate. A larger a causes a rapid drop in weight for samples further from the boundary, while a smaller a allows a wider band of samples to influence the child nodes. Alternatively, a polynomial decay function can be utilized:

$$w_i = (1 + |d_i|)^{-a}.$$

This ensures that exact boundary samples have maximum weight (1), while farther ones contribute less.

These weights affect computation of Gini impurity and class assignment in leaves, defined as:

$$G = \frac{1}{W} \sum_c \left(\sum_{i \in C_c} w_i \right) \left(1 - \frac{\sum_{i \in C_c} w_i}{W} \right),$$

where W is the normalizing term representing the sum of all sample weights in the node, and C_c represents the subset of samples belonging to class c.

If a sample is duplicated across multiple splits (tree depth more than one), its influence must gradually diminish. Therefore, each time a sample is re-copied, its weight is updated recursively:

$$w_i^{(k+1)} = w_i^{(k)} \cdot e^{-ad_i}.$$

Duplication is only allowed up to a specified maximum copy depth δ to avoid exponential growth.

The proposed tree construction algorithm consists of the following steps:

1. For each node:
   - o Compute impurity for all possible splits.
   - o Select best feature x_j and threshold t.
2. Identify boundary-near samples using distances d_i.
3. Duplicate k closest samples into opposite nodes.
4. Assign weights via chosen decay function.
5. Recurse until stopping criteria (max depth, min sample size) are met.

The computational complexity of the proposed algorithm remains close to the original one. In a standard CART algorithm, finding the optimal split at a given node requires $O(f\,N\log N)$ operations, where f is the number of features and N is the number of samples in the node. The proposed modification introduces three additional operations per node: calculating the absolute distances d_i from the threshold for all samples ($O(N)$), identifying the k closest samples (which can be achieved in $O(N)$ time using selection algorithms), and copying and weighting these k samples ($O(k)$). Because k is much smaller than N and these additional steps scale linearly, the computational complexity of the node-splitting process remains dominated by the original $O(f\,N\log N)$ split search. While duplicating samples increases the effective N for subsequent child nodes, the strict enforcement of the maximum copy depth δ caps the duplication process, preventing exponential growth of the dataset. Consequently, the overall training time increases only by a bounded constant factor, preserving the efficiency and scalability of the baseline random forest framework.

Numerical experiments

The proposed method was evaluated on 11 publicly available datasets obtained from the UCI Machine Learning Repository and Kaggle. These datasets cover a range of application areas and data characteristics. Most of them are related to medical classification problems such as heart disease, diabetes, and cancer diagnosis. The datasets differ in size, ranging from 306 to 1484 samples, and in dimensionality, from 3 to 34 features. Full description of the used data is provided in Table 1. Basic preprocessing was applied to all datasets: missing values were removed and categorical features were encoded using label encoding.

Table 1

Description of data sets

Title	M	N	C
Breast Cancer Wisconsin (Original)	9	682	2
Haberman's Survival	3	306	2
Titanic	3	714	2
Heart Attack	13	303	2
Diabetes	8	768	2
Hepatitis C virus	14	615	5
Ecoli	7	336	8
Dermatology	34	366	6
Heart Failure	12	918	2
Yeast	8	1484	10
Cirrhosis	20	418	4

For quantitative evaluation, the standard classification metric independent of threshold choice, the area under the receiver operating characteristic curve (ROC-AUC), was used. Each metric value was obtained by averaging results from five-fold stratified cross-validation repeated ten times with different random seeds to ensure statistical reliability.

As a baseline, the standard random forest implementation based on the CART algorithm from the scikit-learn library was employed. The main hyperparameters of this baseline were selected as follows: the number of trees was set to 100, and the maximum depth of individual trees was limited to values between 5 and 10, depending on the dataset (the best parameters were selected via grid search with cross-validation). For the proposed modification, hyperparameters were also tuned using grid search, including the parameter r that determines the width of the near-boundary region, tested in increments between 0 and 0.9, the maximum tree depth at which sample duplication is applied (δ, ranging from 1 to 5), and parameters describing the weighting function. The weight decay parameter was fixed to $a = 1$ across all datasets.

Table 2

Experiment results on real data, ROC-AUC

Data set	Max tree depth	Baseline RF	Proposed RF		
		mean ± std	r	δ	mean ± std
Breast Cancer	8	0.993 ± 0.005	0.1	5	**0.994 ± 0.005**
Haberman's Survival	5	0.683 ± 0.049	0.4	3	**0.699 ± 0.052**
Titanic	5	0.859 ± 0.035	0.3	1	**0.860 ± 0.035**
Heart Attack	10	0.904 ± 0.034	0.8	6	**0.911 ± 0.030**
Diabetes	5	0.834 ± 0.034	0.1	3	**0.835 ± 0.034**
Hepatitis C virus	7	0.983 ± 0.009	0.5	3	**0.985 ± 0.007**
Ecoli	3	0.897 ± 0.017	0.5	2	**0.900 ± 0.013**
Dermatology	5	**0.998 ± 0.001**	0.1	1	0.998 ± 0.002
Heart Failure	4	**0.926 ± 0.014**	0.1	1	0.925 ± 0.014
Yeast	7	0.879 ± 0.016	0.1	1	**0.881 ± 0.016**
Cirrhosis	5	0.715 ± 0.049	0.4	4	**0.723 ± 0.041**

The results are provided in Table 2. Across the eleven datasets, the modified algorithm achieved higher ROC-AUC scores in 8 cases out of 11, with the strongest improvements observed on datasets known to contain overlapping or noisy class boundaries, such as Haberman's Survival, Heart Attack, and Cirrhosis. Statistical significance was confirmed using both the Wilcoxon signed-rank test, which produced a p-value of 0.005, and the paired t-test, with a p-value of 0.017. These results indicate that the improvements are consistent and unlikely to result from random variation.

Overall, the experimental results confirm that introducing sample duplication near decision boundaries improves the stability and predictive performance of decision-tree ensembles. The method consistently enhances ROC-AUC on most real datasets. These findings show that the proposed approach effectively strengthens robustness in uncertain regions while keeping computational cost and interpretability close to those of the standard random forest.

Conclusion

This paper introduced a simple yet effective modification of the decision tree learning process aimed at improving robustness in overlapping class regions. The method duplicates samples that lie near splitting thresholds so that both child nodes can learn from uncertain regions, with each duplicated sample weighted according to its distance from the threshold. This design preserves the structure, interpretability, and efficiency of the standard CART algorithm while reducing sensitivity to small feature perturbations and unstable decision boundaries.

A comprehensive experimental study on real and synthetic datasets demonstrated that the proposed approach consistently improves classification performance and stability, especially in tasks characterized by high class overlap. Statistical tests confirmed that these improvements are significant. Importantly, the modification requires minimal changes to existing tree-based implementations and can be seamlessly integrated into standard random forest frameworks.

Future work will focus on theoretical bias-variance analysis, extension to regression and survival analysis tasks, evaluation on larger and higher-dimensional datasets, and exploration of adaptive weighting schemes that further optimize learning in uncertain regions.

REFERENCES

1. **Chen J., Wang Z., Wei Z., Huang C., Yang Y., Wei P., Li H., You Y., Zhang S., Dong Z. et al.** An Interpretable Attention Decision Forest Model for Surface Soil Moisture Retrieval. *Remote Sensing*, 2025, Vol. 17, No. 20, Art. no. 3468. DOI: 10.3390/rs17203468

2. **Yan J., Yamada S.** Learning Dual-Path Soft Decision Trees for Vision Prototype XAI. *IEEE Access*, 2025, in print. DOI: 10.1109/ACCESS.2025.3639412

3. **Rodríguez D.M., Cuéllar M.P., Morales D.P.** On the fusion of soft-decision-trees and concept-based models. *Applied Soft Computing*, 2024, Vol. 160, Art. no. 111632. DOI: 10.1016/j.asoc.2024.111632

4. **Rabcan J., Zaitseva E., Levashenko V., Kvassay M.** Incremental fuzzy decision tree based on cumulative mutual information. *Journal of Computational Design and Engineering*, 2025, Vol. 12, Pp. 116–130. DOI: 10.1093/jcde/qwaf088

5. **Habeeb A.S., Hasan H.A., Al-Sinjary A.M.** Fuzzy decision-tree regression model and its application to measure some climate change factors. *Frontiers in Applied Mathematics and Statistics*, 2026, Vol. 12, Art. no. 1732313. DOI: 10.3389/fams.2026.1732313

6. **Liu L., Yang S., Jiang K.** Fuzzy Decision Tree Optimization for Personalized Hybrid English Teaching in Smart Classrooms. *Journal of Circuits, Systems and Computers*, 2025, Vol. 34, No. 11, Art. no. 2550251. DOI: 10.1142/S0218126625502512

7. **Liu C.** Fuzzy decision tree based online precision marketing method for brand products on the internet. *International Journal of Product Development*, 2024, Vol. 28, No. 4, Pp. 241–256. DOI: 10.1504/IJPD.2024.143258

8. **Han H., Yu Q., Zhu Y., Cheng S., Zhang Y.** An Empirical Study of the Impact of Class Overlap on the Performance and Interpretability of Cross-Version Defect Prediction. International *Journal of Software Engineering and Knowledge Engineering*, 2024, Vol. 34, No. 12, Pp. 1895–1918. DOI: 10.1142/S0218194024500414

9. **Lin B., Liu C., Miao D.** An improved decision tree algorithm based on boundary mixed attribute dependency. *Applied Intelligence*, 2024, Vol. 54, Pp. 2136–2153. DOI: 10.1007/s10489-023-05238-4

10. **Alharbi A.A.** Classification Performance Analysis of Decision Tree-Based Algorithms with Noisy Class Variable. *Discrete Dynamics in Nature and Society*, 2024, Vol. 2024, Art. no. 6671395. DOI: 10.1155/2024/6671395

11. **Kusztal K., Przybyła-Kasperek M.** Distributed Data Classification with Coalition-Based Decision Trees and Decision Template Fusion. *Entropy*, 2025, Vol. 27, No. 12, Art. no. 1205. DOI: 10.3390/e27121205

12. **Vallée R., Vallée J.-N., Guillevin C., Lallouette A., Thomas C., Rittano G., Wager M., Guillevin R., Vallée A.** Machine learning decision tree models for multiclass classification of common malignant brain tumors using perfusion and spectroscopy MRI data. *Frontiers in Oncology*, 2023, Vol. 13, Art. no. 1089998. DOI: 10.3389/fonc.2023.1089998

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Andrei V. Konstantinov
Константинов Андрей Владимирович
E-mail: andrue.konst@gmail.com
ORCID: https://orcid.org/0000-0002-1542-6480

Anastasiya P. Elizarova
Елизарова Анастасия Павловна
E-mail: anastasiya.eliz@yandex.ru
ORCID: https://orcid.org/0000-0002-3796-4757

Lev V. Utkin
Уткин Лев Владимирович
E-mail: lev.utkin@gmail.com
ORCID: https://orcid.org/0000-0002-5637-1420

Submitted: 13.01.2026; Approved: 15.03.2026; Accepted: 27.03.2026.
Поступила: 13.01.2026; Одобрена: 15.03.2026; Принята: 27.03.2026.