

# Comparative Evaluation of Ensemble Learning Algorithms for High-Dimensional Data Classification

**Farhan Azmi bin Abdullah**

Associate Professor, University Putra Malaysia, Malaysia

## Abstract

High-dimensional data classification represents a significant area of investigation within machine learning because modern computational applications frequently involve datasets containing large numbers of explanatory variables. Such datasets may exhibit complex interactions, redundant attributes, correlated features, and limited observations relative to the dimensionality of the feature space. These characteristics can influence classification accuracy, computational efficiency, and the generalization capabilities of predictive algorithms. Ensemble learning provides a methodological approach to addressing selected classification difficulties by combining multiple predictive models through structured aggregation procedures. The present paper develops a hypothetical comparative framework for examining the classification characteristics of five ensemble learning algorithms: Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting. The proposed methodology considers a fictional binary classification dataset containing 3,000 observations and 2,400 numerical attributes. Feature selection, model training, hyperparameter optimization, and independent evaluation are incorporated into the proposed research design. Particular attention is given to predictive performance under increasing feature dimensionality and the methodological implications of irrelevant or redundant variables.

Two illustrative analytical tables demonstrate how classification outcomes and feature-dimensionality effects could be reported. The first presents hypothetical classification results using accuracy and F1-score, while the second describes illustrative performance differences across selected feature subsets. These numerical values are manually constructed and do not represent executed simulations or experimentally verified findings. The methodological discussion examines the relationship among ensemble diversity, feature selection, model complexity, computational efficiency, and predictive reliability. The study identifies evaluation bias, excessive dimensionality, and class imbalance as important considerations when comparing learning algorithms. Overall, the proposed framework provides a structured foundation for future empirical investigations involving ensemble-based classification of high-dimensional datasets. Its conclusions concern methodological design rather than demonstrated algorithmic superiority.

**Keywords:** Ensemble Learning, High-Dimensional Data, Data Classification, Random Forest, Gradient Boosting, Bagging, AdaBoost, Feature Selection, Machine Learning, Predictive Accuracy.

## 1. Introduction

The increasing availability of digital information has encouraged the development of computational methods capable of analyzing datasets containing numerous explanatory variables. High-dimensional data arise in scientific research, financial analytics, biomedical computing, image recognition, and other information-intensive applications. Such datasets may contain hundreds or thousands of features describing individual observations. Although these features can provide detailed representations of the underlying phenomena, their presence also introduces difficulties for predictive modeling. Redundant attributes, limited sample sizes, complex variable relationships, and irrelevant information can reduce classifier reliability and increase computational requirements.

High-dimensional classification involves assigning observations to predefined categories using information distributed across extensive feature spaces. The complexity of this task increases when the number of variables becomes large relative to the available training observations. Under such conditions, conventional predictive algorithms may identify relationships that are specific to the training sample rather than representative of new data. Highly correlated variables and noisy measurements can further complicate classification boundaries. Consequently, successful model development requires careful consideration of feature selection, generalization error, training stability, and evaluation design.

Ensemble learning has emerged as an important methodological approach for improving predictive classification through the combination of multiple models. Instead of relying exclusively on one classifier, ensemble procedures construct collections of predictive components and combine their outputs using voting, averaging, or sequential optimization. Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting represent different strategies for developing such systems. Their learning mechanisms vary according to sampling procedures, feature randomization, classifier dependency, and optimization objectives. These differences make comparative evaluation relevant when considering high-dimensional datasets.

Feature dimensionality also influences the relationship between computational complexity and predictive performance. Increasing the number of explanatory attributes may introduce useful information, but it can also expand the search space and increase sensitivity to irrelevant variables. Feature selection procedures attempt to identify informative subsets while preserving meaningful relationships between observations and classification outcomes. However, the effectiveness of dimensionality reduction may differ among ensemble algorithms because their learning procedures respond differently to correlated, noisy, or weakly informative features. Understanding these interactions is necessary for selecting suitable classification strategies.

The present paper proposes a hypothetical comparative investigation of ensemble learning algorithms for high-dimensional binary classification. Five ensemble approaches are examined within a common methodological framework involving feature preparation, model development, and independent evaluation. A fictional dataset with 3,000 observations and 2,400 explanatory attributes provides the basis for illustrating the intended analytical procedure. Two constructed numerical tables demonstrate possible methods of comparing classification performance and feature-dimensionality effects.

## 2. Objectives of the Study

The principal objective of this study is to examine the methodological suitability of ensemble learning algorithms for classifying high-dimensional datasets. The proposed investigation focuses on classification performance, feature dimensionality, and computational reliability. Three specific

objectives guide the comparison of Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting.

## 2.1 Assessment of Ensemble Classification Performance

The first objective is to examine how different ensemble learning algorithms may perform when applied to high-dimensional classification problems. Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting are selected to represent alternative ensemble construction strategies. The proposed comparison considers their potential differences in classification accuracy, precision, recall, and F1-score. Particular attention is given to understanding how the aggregation of multiple predictive components may influence classification reliability when observations contain numerous explanatory attributes and complex relationships among variables.

The investigation further aims to establish a fair analytical procedure for comparing ensemble classifiers. Each algorithm would receive equivalent training information and an appropriate opportunity for hyperparameter optimization. Predictive performance would be assessed using an independent evaluation partition, thereby limiting the influence of model-selection decisions on reported outcomes. This objective emphasizes the importance of distinguishing genuine differences in classification capability from variations caused by inconsistent preprocessing, unequal computational resources, or inappropriate evaluation procedures during comparative machine learning research.

## 2.2 Examination of Feature Dimensionality Effects

The second objective is to investigate how changes in the number of available features may influence ensemble classification performance. High-dimensional datasets frequently contain attributes that provide limited additional predictive information or repeat relationships represented by other variables. The proposed study considers three feature configurations involving selected subsets and the complete feature space. These configurations provide a conceptual basis for examining whether reducing dimensionality may improve predictive reliability while decreasing unnecessary computational complexity.

The investigation also considers differences in the sensitivity of ensemble algorithms to irrelevant and correlated attributes. Randomized tree-based models may respond differently to extensive feature spaces than sequential boosting approaches. Consequently, identical feature subsets may not provide equivalent benefits across classifiers. This objective emphasizes the relationship between feature representation, model diversity, and predictive generalization. Any conclusions concerning dimensionality effects would require controlled experiments in which feature selection procedures are applied consistently and evaluated without accessing independent test information.

## 2.3 Identification of Computational and Evaluation Constraints

The third objective focuses on identifying methodological difficulties that may influence the comparison of ensemble learning algorithms. These difficulties include excessive feature dimensionality, training instability, computational overhead, class imbalance, and inappropriate parameter selection. The proposed framework considers how such conditions may affect classification outcomes and the interpretation of performance measures. Particular attention is given to distinguishing the consequences of algorithmic design from those associated with dataset characteristics and the availability of representative training observations.

A further objective involves examining procedures that support reproducible and statistically defensible comparisons. Repeated model development, independent testing, and transparent documentation of computational settings are considered necessary components of reliable evaluation. The study also recognizes that small differences in classification metrics may not represent meaningful improvements. Consequently, this objective emphasizes uncertainty estimation, appropriate statistical testing, and careful interpretation of reported performance differences rather than selecting algorithms exclusively according to the highest observed numerical score.

### 3. Review of Literature

#### 3.1 Development of Ensemble Classification Methods

Dietterich (2000) examined ensemble methods and described how combinations of classifiers may improve predictive performance through complementary learning behavior. Breiman (2001) developed Random Forests by combining randomized decision trees and examining their generalization properties. Bühlmann and Yu (2002) investigated Bagging from a statistical perspective and considered its influence on prediction stability. These contributions established important foundations for understanding how the aggregation of multiple models can influence classification reliability, particularly when individual predictors exhibit sensitivity to training data variations.

Friedman (2001) developed Gradient Boosting as an approach involving sequential predictive optimization, while Friedman (2002) examined stochastic modifications to the boosting procedure. Kuncheva (2004) presented a systematic treatment of classifier combination strategies and ensemble design. Polikar (2006) discussed ensemble-based systems within decision-making applications, emphasizing the integration of multiple predictive components. Collectively, these contributions demonstrate that ensemble methods employ different mechanisms for improving predictions and require evaluation according to the characteristics of the classification problem.

#### 3.2 Randomized Ensembles and High-Dimensional Features

Geurts et al. (2006) introduced Extremely Randomized Trees as an approach employing substantial randomization during tree construction. Banfield et al. (2007) compared alternative decision-tree ensemble techniques and examined differences in their classification behavior. Rokach (2010) reviewed ensemble-based classifiers and discussed their methodological characteristics. These contributions establish a foundation for comparing tree-based ensembles that differ in their sampling, feature selection, and decision aggregation procedures. Their relevance is particularly apparent when classification involves extensive feature spaces and potentially complex variable interactions.

Guyon and Elisseeff (2003) examined feature selection methods and their relevance to predictive accuracy, computational efficiency, and interpretability. Saeys et al. (2007) reviewed feature selection techniques in high-dimensional biological datasets, while Fan and Lv (2008) investigated dimensionality reduction for ultrahigh-dimensional feature spaces. Strobl et al. (2007) examined potential bias in Random Forest variable importance measures.

#### 3.3 Comparative Evaluation and Generalization Reliability

Bengio and Grandvalet (2004) examined uncertainty associated with cross-validation estimates and demonstrated important limitations in variance estimation. Demšar (2006) investigated statistical procedures for comparing classifiers across multiple datasets. Varma and Simon (2006) examined bias

arising when cross-validation is used for model selection and performance estimation. These contributions emphasize the importance of distinguishing model development from independent evaluation. Their methodological implications are particularly relevant when numerous algorithms and hyperparameter configurations are compared using a limited number of observations.

Hastie et al. (2009) provided a comprehensive statistical treatment of predictive learning and generalization. Cawley and Talbot (2010) investigated overfitting associated with model selection procedures, while Bergstra and Bengio (2012) examined randomized hyperparameter optimization. Zhou (2012) presented theoretical and practical foundations of ensemble learning, and Chen and Guestrin (2016) investigated scalable tree boosting. These contributions support rigorous comparative design, transparent parameter optimization, and careful assessment of whether observed performance differences remain reliable beyond the original training data.

## 4. Research Methodology

### 4.1 Research Design and Ensemble Selection

The present paper proposes a hypothetical comparative quantitative research design for evaluating ensemble learning algorithms under high-dimensional classification conditions. Five methods are selected: Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting. These approaches represent different combinations of resampling, randomization, aggregation, and sequential optimization. The proposed classification problem contains two outcome categories and numerous explanatory attributes. This methodological framework has not been implemented through computational experiments and is intended exclusively to demonstrate a structured research procedure.

The intended experimental design would involve controlled model development using equivalent training observations and predefined evaluation criteria. Each classifier would receive appropriate hyperparameter optimization within the same computational budget. Tree-based base learners would be configured according to the methodological requirements of the corresponding ensemble. The comparison would examine predictive performance under the complete feature space and selected lower-dimensional representations. Independent testing would be conducted after model configurations had been finalized, thereby reducing unnecessary bias in comparative evaluation.

### 4.2 Hypothetical Dataset and Dimensionality Reduction

For illustrative purposes, the proposed study assumes a fictional numerical dataset containing 3,000 observations and 2,400 explanatory variables. The target outcome consists of two classes, with 80% of observations assigned to the majority category and 20% to the minority category. The features are assumed to include informative, correlated, and irrelevant attributes. These characteristics are constructed solely to illustrate a high-dimensional classification problem. No actual observations were collected, computationally simulated, or analyzed during manuscript preparation.

The proposed preprocessing procedure would involve missing-value assessment, identification of duplicated attributes, appropriate feature scaling where necessary, and training-only variable ranking. Three feature configurations would be considered: 200 selected attributes, 800 selected attributes, and the complete set of 2,400 attributes. Feature selection would occur within the model development process, without using independent test outcomes. This structure would allow investigators to examine how reducing dimensionality influences different ensemble algorithms while limiting information leakage and preserving comparability.

## 4.3 Model Training and Performance Evaluation

The hypothetical dataset would be divided into training, validation, and independent testing partitions containing 70%, 15%, and 15% of observations, respectively. Stratified partitioning would preserve the intended class proportions. Model selection would rely on training and validation information, with appropriate internal cross-validation where required. Hyperparameters would include ensemble size, tree depth, learning rate, and feature sampling settings. These choices would be adjusted according to the selected algorithm while maintaining consistent evaluation conditions.

The proposed evaluation would report classification accuracy, precision, recall, and F1-score, supplemented by computational time and model stability in an actual implementation. Particular attention would be given to the minority category because aggregate accuracy can conceal unequal classification performance. Repeated experiments and confidence intervals would be necessary before interpreting small differences as meaningful. The present manuscript uses manually constructed numerical examples to demonstrate reporting procedures. No training, testing, or statistical significance analysis was performed.

## 5. Analysis

### 5.1 Illustrative Comparison of Ensemble Classification Outcomes

The comparative evaluation of ensemble learning methods requires consistent classification measures and clearly defined reference outcomes. Table 1 presents hypothetical classification results for five ensemble algorithms evaluated against the same fictional test partition. The test set contains 450 observations, including 90 minority-class cases and 360 majority-class cases. True-positive and false-positive counts are provided to demonstrate how classification accuracy and F1-score could be calculated. All values are constructed examples rather than observations generated through actual model execution.

The illustrative results demonstrate how different ensemble approaches might produce variations in minority-class recognition and overall classification accuracy. Gradient Boosting is assigned the highest F1-score within this fictional comparison, while Random Forest and Extremely Randomized Trees receive equivalent illustrative accuracy. These differences should not be interpreted as evidence of genuine predictive superiority. An implemented study would require repeated training, independent evaluation, confidence intervals, and appropriate statistical comparisons before determining whether any performance differences are reproducible or practically important.

**Table 1: Hypothetical Comparative Classification Performance of Ensemble Algorithms**

| Ensemble Algorithm         | True Positives | False Positives | Accuracy (%) | F1-Score |
|----------------------------|----------------|-----------------|--------------|----------|
| Bagging                    | 66             | 24              | 89.3         | 0.733    |
| Random Forest              | 72             | 18              | 92.0         | 0.800    |
| Extremely Randomized Trees | 70             | 16              | 92.0         | 0.795    |
| AdaBoost                   | 65             | 27              | 88.4         | 0.714    |
| Gradient Boosting          | 74             | 20              | 92.0         | 0.804    |

**Source:** Hypothetical results manually constructed for academic illustration. The assumed test set contains 90 positive and 360 negative cases. Accuracy and F1-score are derived from the hypothetical classification counts and rounded. No classifiers were actually trained or tested.

The table illustrates why models with identical accuracy may exhibit different F1-scores. In this example, Random Forest, Extremely Randomized Trees, and Gradient Boosting all have approximately 92% accuracy, but their positive predictions and false-positive counts differ.

## 5.2 Illustrative Influence of Feature Dimensionality

Feature dimensionality may influence classification performance because increasing the number of attributes can introduce redundant information, correlated variables, and unnecessary computational complexity. Table 2 presents hypothetical F1-scores across three feature configurations. The constructed values illustrate a possible situation in which selected feature subsets provide more favorable performance than the complete 2,400-variable representation. These differences are not experimental findings. They merely demonstrate how feature-dimensionality effects could be organized within a comparative ensemble learning investigation.

The proposed analysis emphasizes that feature reduction should be evaluated in relation to individual learning algorithms rather than assumed universally beneficial. Randomization mechanisms, classifier diversity, and optimization procedures may influence how models respond to high-dimensional information. An actual study would require independently fitted feature selectors and repeated model evaluation for every dimensionality condition. Computational time, memory consumption, and predictive stability would also need assessment. The hypothetical values therefore provide a reporting example without establishing the effectiveness of any specific feature selection strategy.

**Table 2: Hypothetical F1-Score Comparison Across Feature Dimensions**

| Ensemble Algorithm         | 200 Features | 800 Features | 2,400 Features |
|----------------------------|--------------|--------------|----------------|
| Bagging                    | 0.820        | 0.770        | 0.733          |
| Random Forest              | 0.834        | 0.820        | 0.800          |
| Extremely Randomized Trees | 0.839        | 0.822        | 0.795          |
| AdaBoost                   | 0.801        | 0.758        | 0.714          |
| Gradient Boosting          | 0.843        | 0.821        | 0.804          |

**Source:** Fictional comparative values developed for methodological illustration. Feature subsets of 200 and 800 variables are assumed to be selected using training-only procedures. The 2,400-feature column corresponds to the hypothetical full-feature comparison in Table 1. No dimensionality reduction experiments were conducted.

The illustrative pattern suggests a possible research hypothesis concerning the relationship between irrelevant attributes and classification performance. However, confirming this relationship would require actual experiments involving documented feature selection methods and representative high-dimensional datasets.

## 6. Discussion

### 6.1 Ensemble Diversity and Predictive Classification

Ensemble learning combines multiple predictive components to produce a collective classification outcome. Its effectiveness depends partly on the relationships among individual learners and the errors they produce. When component models exhibit complementary predictive behavior, their combination may reduce dependence on a single unstable classifier. The proposed framework examines this principle through Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting. Each algorithm introduces diversity through different construction or optimization mechanisms that may influence classification reliability.

The methodological comparison also recognizes that additional ensemble components do not automatically improve predictive performance. Highly correlated learners may provide limited complementary information, while excessive complexity can increase computational requirements. Consequently, ensemble diversity should be evaluated alongside the quality of individual predictions and the characteristics of the dataset. The hypothetical results illustrate possible performance differences but cannot establish their causes. A genuine experimental study would need to examine how learner diversity contributes to classification outcomes under controlled conditions.

### 6.2 Feature Selection and High-Dimensional Learning

Feature selection represents an important consideration when the number of explanatory variables becomes large relative to available observations. Irrelevant or highly correlated attributes may increase computational demands and complicate the identification of reliable predictive relationships. The proposed framework evaluates this issue through three hypothetical feature configurations. By considering the same ensemble algorithms across different dimensionalities, the intended comparison provides a methodological basis for examining whether selected variable subsets produce more stable classification performance.

However, reducing dimensionality can also remove useful information when feature selection procedures fail to identify complex interactions. A variable that appears weak when examined independently may contribute predictive information through its relationship with other features. Consequently, simple ranking methods may not always identify the most useful feature combinations. The proposed investigation therefore emphasizes appropriate feature selection design and independent testing. Any observed improvement must be interpreted cautiously because selection procedures can introduce additional model complexity and evaluation bias.

### 6.3 Comparative Performance and Statistical Interpretation

Comparing ensemble algorithms requires careful interpretation of performance differences because classification metrics are influenced by dataset composition, model configuration, and random variation. The hypothetical results assign slightly different F1-scores to Random Forest, Extremely Randomized Trees, and Gradient Boosting. Such differences are insufficient to establish methodological superiority without repeated experiments.

A reliable evaluation would require uncertainty estimates and statistical procedures appropriate to the number of models, datasets, and experimental repetitions involved in the comparison. Demšar (2006) discussed methods for comparing classifiers across multiple datasets, while Cawley and Talbot (2010) examined performance bias associated with model selection. These contributions emphasize that

statistical evaluation must account for the design of the experiment. The proposed framework therefore recommends repeated assessment and appropriate control of hyperparameter optimization. The interpretation of predictive performance should consider the magnitude and uncertainty of differences rather than focusing exclusively on rankings derived from isolated numerical estimates.

## 7. Challenges

### 7.1 Excessive Feature Dimensionality and Redundancy

High-dimensional datasets frequently contain explanatory variables that are redundant, weakly informative, or strongly correlated with other attributes. These characteristics can complicate classifier development by increasing the number of possible relationships that a learning algorithm must examine. Tree-based ensembles may respond to such complexity through feature randomization, but their effectiveness remains dependent on the distribution of relevant information. Excessive dimensionality may also increase computational requirements and make the interpretation of predictive relationships more difficult.

A further difficulty arises when the number of available features approaches or exceeds the number of training observations. Under these conditions, classifiers may identify relationships that do not generalize beyond the development dataset. Feature selection provides one potential response, but inappropriate selection can remove informative variables or introduce bias. Consequently, dimensionality reduction procedures require independent evaluation. The proposed framework emphasizes preserving useful predictive information while reducing unnecessary complexity through carefully controlled variable selection and model validation.

### 7.2 Computational Cost and Hyperparameter Optimization

Ensemble algorithms may require substantial computational resources because they construct and combine multiple predictive components. Training time depends on factors such as ensemble size, tree depth, learning rate, feature count, and the number of available observations. High-dimensional data can increase the cost of evaluating potential splitting variables and configuring model parameters. Consequently, comparisons based exclusively on classification accuracy may overlook important differences in computational efficiency and practical implementation requirements.

Hyperparameter optimization introduces additional computational demands because several configurations may need evaluation before a suitable model is selected. Randomized search and other optimization procedures can help organize this process, but they still require appropriate computational budgets and validation strategies. Excessive parameter exploration may also increase the risk of overfitting to development data. The proposed methodology therefore recommends predefined search spaces, consistent resource allocation, and independent evaluation before concluding that a particular ensemble configuration provides superior performance.

### 7.3 Evaluation Bias and Classification Stability

Evaluation bias can arise when feature selection, preprocessing, or model optimization procedures use information that should remain reserved for independent testing. Such information leakage may produce overly favorable classification results and undermine the credibility of comparative conclusions. High-dimensional datasets are particularly vulnerable because extensive variable spaces create numerous opportunities for identifying accidental relationships. Reliable research therefore requires clearly

separated data partitions and careful control of all preprocessing decisions throughout model development.

Classification stability presents another challenge because ensemble algorithms may produce different predictions when training samples, random seeds, or parameter configurations change. Small differences in observed performance may reflect experimental variation rather than genuine algorithmic advantages. Consequently, single-run comparisons provide limited evidence concerning generalization. The proposed framework recommends repeated testing, transparent reporting, and appropriate uncertainty estimation. These procedures would help determine whether performance differences remain consistent across independent datasets and alternative experimental conditions.

## 8. Conclusion

Ensemble learning provides a relevant methodological approach to classification problems involving extensive feature spaces and complex relationships among explanatory variables. The present paper developed a hypothetical comparative framework for examining Bagging, Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting. Its principal emphasis was the organization of consistent evaluation procedures and the consideration of feature dimensionality. The proposed research design demonstrates how ensemble algorithms could be compared without assuming that computational complexity necessarily produces better classification performance.

The illustrative results demonstrate possible methods for reporting classification accuracy, F1-score, and differences across feature configurations. However, the numerical values were constructed exclusively for academic demonstration and do not represent executed experiments or statistically supported findings. Meaningful comparisons would require authentic or computationally generated datasets, reproducible model training, independent testing, and suitable uncertainty analysis. The proposed framework therefore serves as a methodological example rather than evidence establishing the superiority of a particular ensemble learning algorithm.

Overall, the study identifies feature redundancy, learner diversity, computational cost, parameter optimization, and evaluation bias as important considerations in high-dimensional classification. Future empirical research should investigate these factors through controlled comparisons across datasets with different dimensionality and sample characteristics. Ensemble methods may provide useful predictive capabilities when their configurations and evaluation procedures are appropriately designed. Nevertheless, reliable conclusions concerning

## References

1. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140. <https://doi.org/10.1007/BF00058655>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
4. Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer. [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)

5. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
6. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
7. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182. <https://www.jmlr.org/papers/v3/guyon03a.html>
8. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
9. Huang, J. Z., Li, M., & Salloum, S. (2023). Observation points classifier ensemble for high-dimensional imbalanced classification. *CAAI Transactions on Intelligence Technology*. <https://doi.org/10.1049/cit2.12100>
10. Kim, J., Kang, J., & Sohn, M. (2021). Ensemble learning-based filter-centric hybrid feature selection framework for high-dimensional imbalanced data. *Knowledge-Based Systems*, 220, Article 106901. <https://doi.org/10.1016/j.knosys.2021.106901>
11. Bolón-Canedo, V., & Alonso-Betanzos, A. (2019). Ensembles for feature selection: A review and future trends. *Information Fusion*, 52, 1–12. <https://doi.org/10.1016/j.inffus.2018.11.008>
12. Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. *Artificial Intelligence Review*, 57, Article 137. <https://doi.org/10.1007/s10462-024-10759-6>
13. Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15, 3133–3181. <https://www.jmlr.org/papers/v15/delgado14a.html>
14. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
15. Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms*. Wiley. <https://doi.org/10.1002/0471660264>
16. Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33, 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
17. Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, 5, 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
18. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
19. Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. <https://doi.org/10.1109/34.709601>
20. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46, 389–422. <https://doi.org/10.1023/A:1012487302797>