JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

A Comparative Study of Machine Learning Methods for Optimizing Mushroom Classification

Muhammad Arslan[1], Muhammad Azam[2], Muazzam Ali[1*], M U Hashmi[1], Amna Kousar[2], and Zarqa Zafar[1]

[1]Department of Basic Sciences, Superior University, Lahore, 54000, Pakistan.
[2]Department of Computer Sceicne, Superior University, Lahore, 54000, Pakistan.
[*]Corresponding Author: Muazzam Ali. Email: muazzamali@superior.edu.pk

Received: July 19, 2024 **Accepted:** September 24, 2024

Abstract: The exact characterization of mushrooms as consumable or noxious is a basic undertaking that has suggestions for general wellbeing and security. Utilizing machine learning to further develop the forecast precision in such double grouping undertakings addresses a critical progression in both food handling and botany studies. This paper investigate the use of ensemble learning strategies to improve prescient execution utilizing a cleaned and pre-handled form of the notable UCI Mushroom Dataset. The dataset incorporates highlights, for example, cap diameter, gill color, and stem shape, each adding to the characterization precision of mushrooms. In this review, we applied a few noticeable gathering techniques: Random Forest, Gradient Boosting, AdaBoost, Extra Trees, and Bagging. Every procedure uses an alternate methodology to total expectations from different models to work on the dependability and exactness of forecasts over utilizing a solitary model methodology. The assessment of these procedures was directed utilizing a complete arrangement of measurements including accuracy, precision, recall, F1 score, ROC AUC, Matthews Correlation Coefficient (MCC), and Cohen's Kappa. The Additional Trees group strategy showed prevalent execution, accomplishing the most noteworthy accuracy of 99.17%, precision of 99.20%, and recall of 99.28%. These outcomes were joined by a F1 score of 99.24% and a ROC AUC of 99.94%, demonstrating exceptionally solid prescient capacities. . Such discoveries feature the capability of gathering techniques in basic applications where the expense of misclassification can be serious. The review not just reaffirms the worth of gathering learning in complex order errands yet in addition gives a guide to additional examination into its applications in different spaces requiring high-stakes navigation.

Keywords: Mushroom Classification; Ensemble Learning; UCI Mushroom Dataset; Predictive Performance; Food Safety; Machine Learning.

1. Introduction

Mushroom grouping into eatable or toxic classifications conveys significant ramifications for general wellbeing, as mistakes can bring about difficult ailment or casualty [1, 2]. Machine learning offers groundbreaking arrangements in this high-stakes parallel characterization task, where accuracy is central [3]. Among different strategies, gathering learning is especially striking for its ability to amalgamate expectations from numerous models, along these lines accomplishing better accuracy and vigor looked at than single-model methodologies [4]. This philosophy is essential in fields where prescient accuracy straightforwardly influences human wellbeing, making it a significant device in dietary security evaluations and general wellbeing choices [5]. In this unique situation, the capacity to precisely foresee mushroom harmfulness forestalls foodborne diseases as well as guides dependable scavenging works on, supporting biodiversity and environmental wellbeing. Broad exploration highlights the viability of gathering techniques in different logical spaces like money, medical services, and natural science [6]. These strategies, which incorporate Random Forest, Gradient Boosting, AdaBoost, Extra Trees, and Stowing, have been commended for their capacity to beat individual models by really conglomerating different choice ways to frame an agreement that diminishes individual model predispositions and fluctuations. Research by Breiman (2001) and Schapire (1999) features how these strategies can alleviate overfitting while at the same time improving the prescient ability of Machine learning frameworks. The guideline behind these techniques is that the aggregate insight of a gathering of prescient models can outperform the accuracy of a solitary model by making up for its constraints and utilizing the strength of the group. Notwithstanding hearty documentation, there exists a shortage in writing tending to thorough assessments across a wide exhibit of measurements. Many examinations have engrossed in dominatingly on accuracy and area under the ROC bend (ROC AUC), frequently dismissing other basic measurements like Matthews Relationship Coefficient (MCC) and Cohen's Kappa, which offer further bits of knowledge into model execution, particularly in unequal datasets where class circulations are slanted. Also, while ensembles methods are broadly concentrated on in numerous spaces, their particular applications to basic wellbeing related fields like toxicology remain underexplored, giving prolific ground to significant examination. This examination looks to expand the utilization of gathering techniques past their conventional areas, underlining their basic job in general wellbeing situations where the differentiation among eatable and harmful mushrooms is essential. By utilizing an extensive arrangement of assessment measurements, this study plans to give a nuanced comprehension of each model's exhibition, subsequently contributing important bits of knowledge to the field of food handling and general wellbeing. Through this methodology, we not just reaffirm the flexibility of ensemble learning strategies yet additionally prepare for their more extensive application in regions where their effect can life-save. The momentum scene of ensemble learning research uncovers a critical hole in precise, relative examinations that consolidate a wide range of assessment measurements [7]. This hole is especially clear in the mushroom characterization setting, where the precision of grouping can be a crucial matter [8]. While existing writing regularly accentuates enhancements to individual group procedures, thorough similar examinations across these strategies utilizing different measurements stay interesting. These measurements are critical for a comprehensive evaluation of model execution, including parts of accuracy, mistake balance, and the prescient certainty of the model across class names.

This paper looks to fill the distinguished exploration holes by directing a comprehensive relative investigation of different ensemble techniques applied to the UCI Mushroom Dataset, which has been fastidiously cleaned and pre-handled for this review. Our exploration not just tests these strategies against ordinary measurements like accuracy, precision, and recall yet in addition stretches out the assessment to incorporate MCC and Cohen's Kappa, giving an adjusted viewpoint on each model's exhibition. Thusly, we mean to portray the qualities and limits of every strategy in exact terms, consequently directing their application practically speaking. The outcomes from this study are planned to act as a benchmark for future exploration in ensemble learning and to help experts in choosing the most fitting models for comparative undertakings in general wellbeing and then some.

2. Materials and Methods

2.1. Dataset Collection

The dataset utilized in this study was gathered from Kaggle, explicitly a cleaned and handled rendition of the notable UCI Mushroom Dataset. This dataset is used for a twofold characterization undertaking to decide if mushrooms are eatable or harmful in light of different actual qualities. It includes 9 highlights, for example, cap diameter, gill color, stem shape, among others, each adding to the characterization accuracy. The dataset incorporates the objective class marked as 0 (edible) or 1 (poisonous), making it an essential asset for preparing Machine learning models to possibly save lives by forestalling the utilization of poisonous mushrooms.

2.2. Data Preprocessing

Prior to applying Machine learning models, the dataset went through a few preprocessing moves toward guarantee ideal model execution. The preprocessing included model imputation to deal with missing qualities and one-hot encoding to change over downright factors into a structure that could be given to ML calculations to more readily foresee the result. Also, z-score standardization was applied to normalize the scope of ceaseless beginning factors so that each component contributes similarly to the distance calculation. This normalization is critical while looking at estimations that have various units.

2.2.1. Hyperparameter Tuning

To streamline the exhibition of the ensemble models, two principal hyper parameter tuning procedures were utilized: Grid Search and Random Search [9]. Grid Search was utilized to deliberately manage numerous blends of boundary tunes, cross-approving as it goes to figure out which tune gives the best presentation. Random Search, then again, tests a given number of up-and-comers from a boundary space with a predetermined dispersion [10]. While Grid Search is careful, Random Search can arrive at better boundaries all the more rapidly. These procedures are fundamental for adjusting the models to accomplish the most noteworthy conceivable accuracy and execution.

2.3. Ensemble Techniques Used

The review utilized a few ensemble learning strategies, each intended to upgrade expectation accuracy by joining the forecasts of different feeble students to frame areas of strength for a [11]. The following is a more profound specialized clarification of every ensemble procedure utilized:

2.3.1. Random Forest

Random Forest is generally utilized ensemble procedure that forms different choice trees during preparing and yields the method of the classes (grouping) or mean forecast (relapse) of the singular trees [12]. It utilizes the stowing (Bootstrap Total) strategy, where each tree is prepared on an alternate random subset of the information with substitution (bootstrapping). This guarantees variety among trees, decreasing the difference of the model.

The random choice of elements at each split point in the trees further decor relates the singular models. By involving various subsets of elements for each tree split, Random Forest limits overfitting, expands power, and makes the model less delicate to clamor in the information. This randomness in highlight determination guarantees that trees don't over focus on specific examples in the preparation information, making Random Forest exceptionally powerful in situations with high-layered highlight spaces.

Key Characteristics:

- **Number of Trees**: Expanding the quantity of trees in the forest assists with working on the speculation of the model yet with consistent losses.
- **Out-of-Bag (OOB) Error**: Random Forest can appraise its own exhibition utilizing OOB tests, which are information brings up left of the bootstrapped preparing test.

2.3.2. Gradient Boosting

Gradient Boosting Machines (GBMS) are a consecutive output strategy that forms models in a bit by bit way, where each ensuring endeavors to address the errors made by the past models [13]. The method works via training weak learners (typically shallow decision trees) in a gradient descent framework to limit a loss function.

At every emphasis, Gradient Boosting fits another model to the residuals (the distinction between the noticed and anticipated values) of the past model, in this way steadily working on its presentation. In contrast to Random Forest, where trees are developed autonomously, Gradient Boosting successively adds trees, and each tree is impacted by the exhibition of the recently developed trees.

Key Characteristics:

- **Learning Rate:** Control the amount of the adjustment the new tree adds to the general model. Lower Learning rates lead to better speculation however require more trees.
- **Loss Function:** Can be altered to limit various kinds of errors, like classification errors, or squared error in regression.

2.3.3. Ada Boost (Adaptive Boosting)

AdaBoost is a boosting strategy that intends to make major areas of strength for a by consolidating different feeble classifiers, regularly choice stumps (trees with a solitary split) [14]. In AdaBoost, each occurrence of the dataset is relegated a weight. During every emphasis, the calculation zeros in additional on cases that were erroneously grouped by expanding their weight, while diminishing the heaviness of accurately characterized occasions.

The following frail learner in the succession is prepared on this re-weighted information, making it center around hard to-arrange occasions. This approach permits AdaBoost to "adjust" to errors by changing weights progressively [15]. The last classifier is a weighted mix of the frail classifiers, where more weight is given to models that performed better on the training data.

Key Characteristics:

- **Weak Learners**: AdaBoost works best with weak learners that marginally beat arbitrary speculating. It will in general diminish both predisposition and change.
- **Instance Weighting**: Instances that are more earnestly to order are given higher weights so the resulting models center more on these troublesome cases.

2.3.4. Extra Trees (Extremely Randomized Trees)

Extra Trees (or Extremely Randomized Trees) is another ensemble technique like Random Forest, however with a key distinction: rather than involving the most ideal split for every hub in light of a model like Gini or data gain, Extra Trees picks split focuses haphazardly from inside the scope of the element values. This additional arbitrariness makes Extra Trees profoundly hearty and less inclined to overfitting [16]. Also, Extra Trees varies from Random Forest in that it utilizes the whole training dataset for each tree as opposed to bootstrapping [17]. This outcomes in quicker training times since it kills the requirement for developing numerous bootstrapped tests.

Key Characteristics:

- **Increased Randomness**: The random selection of the two highlights and split focuses inside the elements makes the model more strong to commotion and overfitting.
- **No Bootstrapping**: The technique constructs trees on the full dataset, which assists accelerate the model training with handling.

2.3.5. Bagging (Bootstrap Aggregating)

Bagging is one of the easiest and most instinctive ensemble methods [18]. The key thought is to decrease the variance of a model by creating numerous different preparation datasets through bootstrapping (testing with substitution) [19]. For each bootstrapped test, a different model (normally a decision tree) is prepared, and expectations from all models are found the middle value of for relapse undertakings or greater part decided in favor of classification task.

Via training each tree on various subsets of the information, Bagging settles models that have high change, (for example, decision trees) and lessens the probability of overfitting, particularly when the base learners are inclined to it.

- **Bootstrap Sampling:** Random testing with substitution guarantees that a few information focuses are utilized on various occasions, while others are forgotten about in each example [20]. This haphazardness assists with building different models.
- **Aggregation:** The last expectation depends on the total of various forecasts (average or voting), which lessens variance and further develop speculation.
- Random Forest and Extra Trees center on diminishing variance and overfitting by making various decision trees with arbitrary parts or haphazardly chose highlights.
- Gradient Boosting and Ada Boost center on working on weak learners by successively amending errors made by the past models, successfully bringing down both bias and variance.
- Bagging accumulates expectations from different models trained on bootstrapped tests, settling high-variance models.

Every procedure carries an alternate solidarity to the errand, giving adaptability in dealing with different sorts of datasets and working on prescient execution [21]. These strategies are especially valuable when the dataset is perplexing, with various related factors, like the mushroom classification problem.

2.4. Evaluation Metrics

A few metrics were utilized to assess the presentation of the ensemble models, giving an all-encompassing perspective on each model's viability:

- **Accuracy**: Measures the extent of accurately anticipated perceptions to add up to perceptions, giving a fundamental sign of the model's exhibition.
- **Precision**: Demonstrate the proportion of true positives to both true and false positives, featuring the model’s accuracy in anticipating positive occasions.
- **Recall**: The proportion of true positives separated by the amount of true positives and false negatives, showing the model’s capacity to track down every single pertinent occurrence.

- **F1 Score**: A weighted average of Precision and Recall. This scores considers both false positives and false negatives into consideration.
- **ROC AUC**: Mirrors the probability that the model positioning a random positive case higher than a random negative case.
- **Mathews Correlation Coefficient**: Produces a coefficient that shows the nature of binary classifications, perfect for unbalanced datasets.
- **Cohen's Kappa**: Measures the arrangement between two raters who each order things into fundamentally unrelated categories.

This thorough strategic methodology guarantees that the review tests the adequacy of ensemble methods as well as gives significant bits of knowledge into their functional benefits and constraints.

3. Results and Discussions

3.1. Discussions on Correlation Metrics

The connection coefficient matrix gives basic bits of knowledge into the connections between various elements inside the mushroom dataset, which is fundamental for understanding how these highlights interface and impact the objective class — whether a mushroom is edible or poisonous. The following is a conversation of key perceptions from the matrix, which could direct element determination and model improvement.

Cap Diameter and Stem Width: There's a high positive correlation (0.83) between cap diameter and stem width, proposing that mushrooms with bigger covers will more often than not have wider stems. This could suggest that these two highlights convey comparative data, possibly prompting multicollinearity whenever utilized together in certain models.

Gill Attachment and Stem Width: A moderate correlation (0.25) shows some degree of reliance. This could be because of the natural design of mushrooms where the gill attachment could impact or be affected by the width of the stem.

Class and Various Features: Negative correlations with class are seen in cap diameter (-0.17) and cap shape (-0.13), recommending that specific sizes and shapes are more probable related with one or the other edible or poisonous mushrooms.

Positive correlations with class are noted in stem level (0.18) and particularly amazing in negative correlation with stem width (-0.18), featuring these aspects as possibly critical indicators of edibility. The distinguished solid correlations recommend that aspect decrease procedures like Principal Component Analysis (PCA) could be utilized to lessen highlight overt repetitiveness, which could improve model execution by disposing of commotion.

Given the fluctuating levels of relationship with the class, highlight designing could be used to make new elements that better catch the hidden examples connected with mushroom edibility.

This part gives a relative examination of the different ensemble strategies utilized in this review, zeroing in on key assessment metrics: Accuracy, Precision, Recall, F1 Score, ROC AUC, Log Loss, MCC, and Cohen's Kappa. Every metrics gives bits of knowledge into various parts of model execution, guaranteeing a comprehensive perspective on how well each model predicts whether mushrooms are edible or poisonous.

Table 1. Computed Results of Evaluation Metrics

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Log Loss	MCC	Cohen's Kappa
Random Forest	0.9907	0.9924	0.9905	0.9915	0.9992	0.0500	0.9812	0.9812
Random Forest Grid Search	0.9906	0.9921	0.9907	0.9914	0.9994	0.0512	0.9811	0.9811
Random Forest Random Search	0.9911	0.9919	0.9919	0.9919	0.9993	0.0493	0.9821	0.9821
Gradient Boosting	0.8740	0.8832	0.8868	0.8850	0.9433	0.3770	0.7458	0.7458

Gradient Boosting Grid Search	0.9896	0.9921	0.9889	0.9905	0.9994	0.0290	0.9791	0.9791
AdaBoost	0.7534	0.7689	0.7849	0.7768	0.8250	0.6842	0.5016	0.5015
Extra Trees	0.9917	0.9920	0.9928	0.9924	0.9994	0.0412	0.9832	0.9832
Extra Trees Grid Search	0.9912	0.9931	0.9907	0.9919	0.9995	0.0684	0.9822	0.9822
Bagging	0.9827	0.9871	0.9813	0.9842	0.9964	0.1273	0.9652	0.9652
Bagging Grid Search	0.9843	0.9860	0.9853	0.9857	0.9985	0.0569	0.9684	0.9684

4. Discussion

4.1. Accuracy

Accuracy is a major metric that actions the extent of accurately anticipated perceptions to the complete perceptions. In this review, most ensembles methods performed especially well, with the extra trees model accomplishing the most noteworthy accuracy at 99.17%, firmly followed by Random Forest at 99.11%. This shows that these ensemble methods are exceptionally successful at accurately arranging mushrooms as edible or poisonous. Conversely, AdaBoost showed altogether lower accuracy at 75.34%, demonstrating it battled with the assignment contrasted with different techniques.

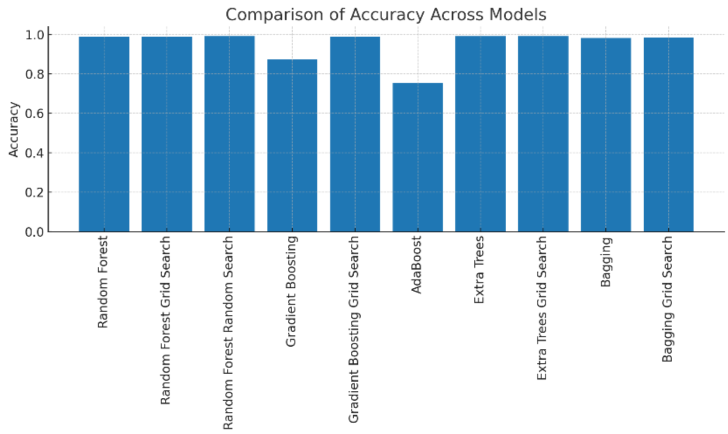

Figure 1. Accuracy

4.2. Precision

Precision returns how well a model predicts true positives without erroneously naming negative cases as positive. The Extra Trees model again leads with a precision of 99.31%, meaning it had exceptionally low false-positive rate. Random Forest and its variations additionally showed solid precision scores above 99%. AdaBoost, with a precision of 76.89%, played out the most terrible, demonstrating more occurrences where non-poisonous mushrooms were misclassified as poisonous.

4.3. Recall

Recall estimates how well the model catches true positive instances, or for this situation, how well it recognize poisonous mushrooms. Extra Trees accomplished the most noteworthy recall of 99.28%, guaranteeing practically all poisonous mushrooms were accurately distinguished. Random Forest firmly followed, and bagging methods performed well as well. AdaBoost battled with recall, accomplishing just 78.49%, demonstrating that it missed a larger number of poisonous mushrooms contrasted with different models.

4.4. F1 Score

The F1 Score gives a harmony among Precision and Recall. `The Extra Trees model succeeded with a close wonderful F1 score of 99.24%, making it the most adjusted model regarding precision and recall. Random Forest and Gradient Boosting additionally performed well. AdaBoost had a lower F1 score of 77.68%, showing that its lower review harmed its general viability.

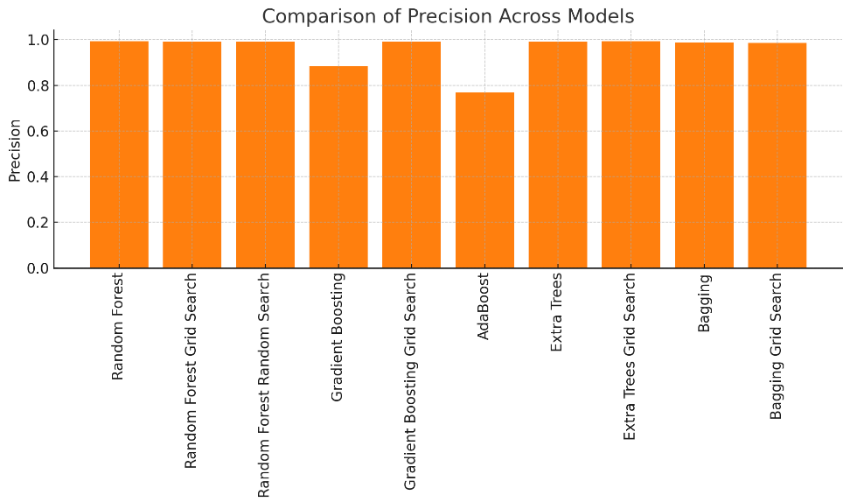

Figure 2. Precision

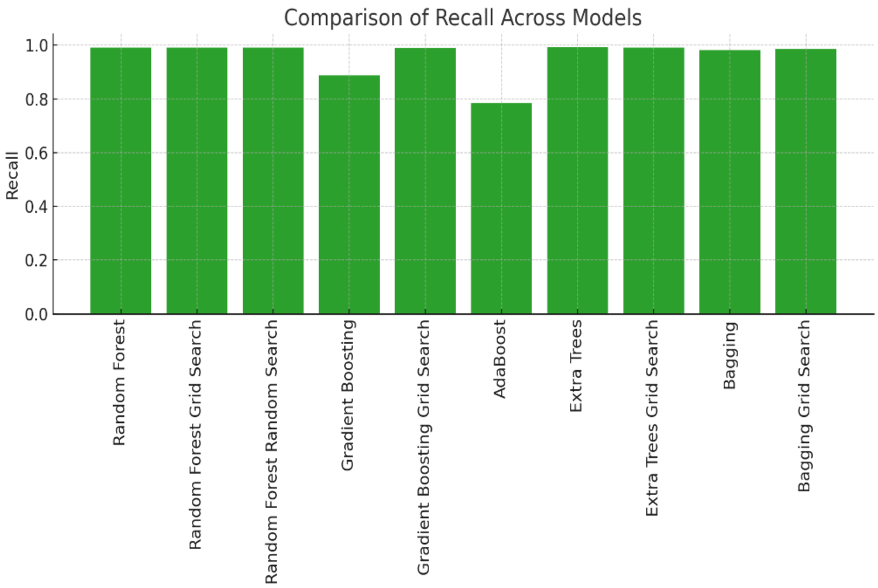

Figure 3. Recall

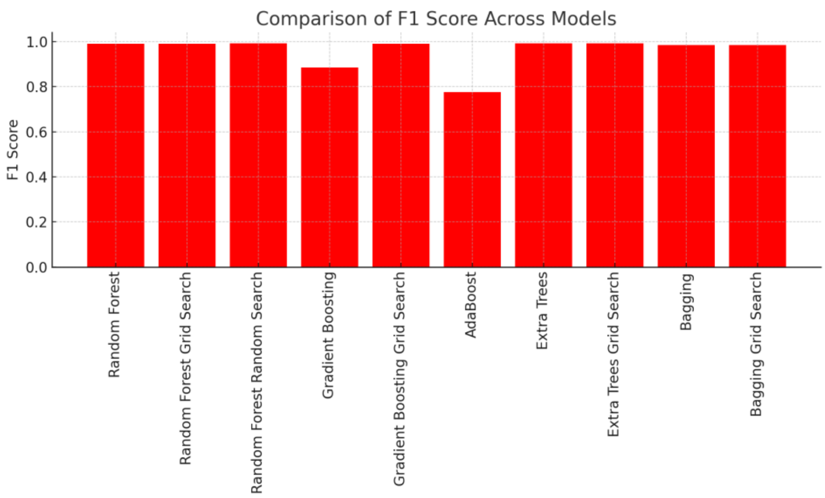

Figure 4. F1 Score

4.5. ROC AUC

ROC AUC estimates the model's capacity to recognize positive and negative classes. Extra Trees and Random Forest accomplished close amazing ROC AUC scores above 0.999, meaning they are fantastic at separating among poisonous and edible mushrooms. Conversely, AdaBoost's ROC AUC score was just 0.8250, showing it has less ability to separate between the two classes.

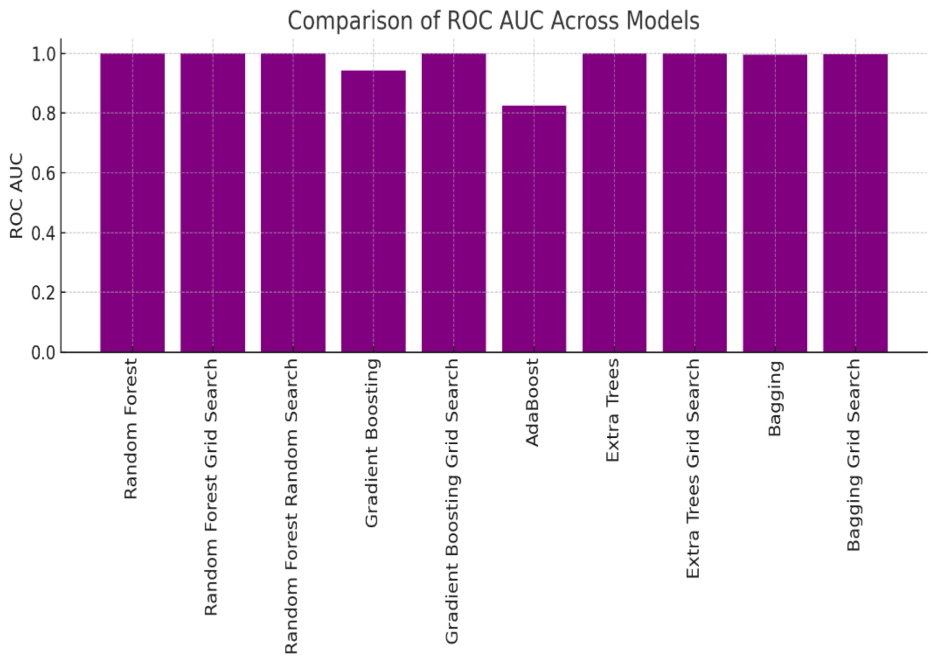

Figure 5. ROC AUC

4.6. Log Loss

Log Loss assesses how close a model's anticipated probabilities are to the genuine qualities, with lower values being better. Extra Trees had the most reduced Log Loss at 0.0412, showing it delivered exceptionally precise likelihood forecasts. Then again, AdaBoost performed ineffectively, with a fundamentally higher Log Loss of 0.6842, showing an absence of trust in its expectations.

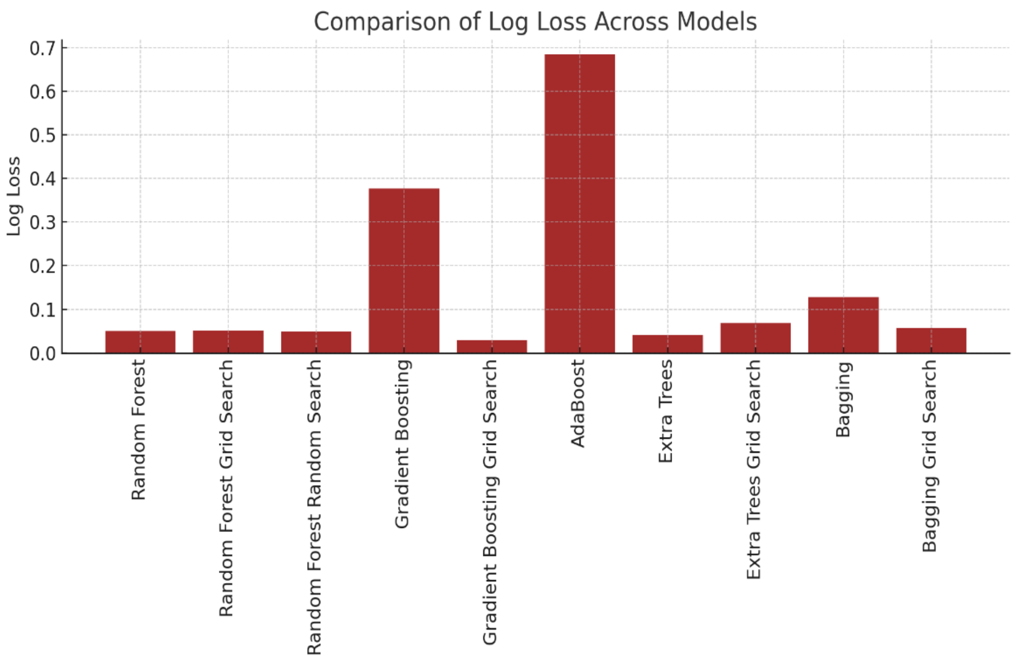

Figure 6. Log Loss

4.7. Mathews Correlation Coefficient (MCC)

MCC gives a decent measures that records for true and false positives and negatives. Once more Extra Trees performed best with a MCC of 0.9832, showing amazing execution across the two classes. AdaBoost, with a MCC of 0.5016, showed fundamentally less fortunate execution in adjusting expectations across both edible and poisonous mushrooms.

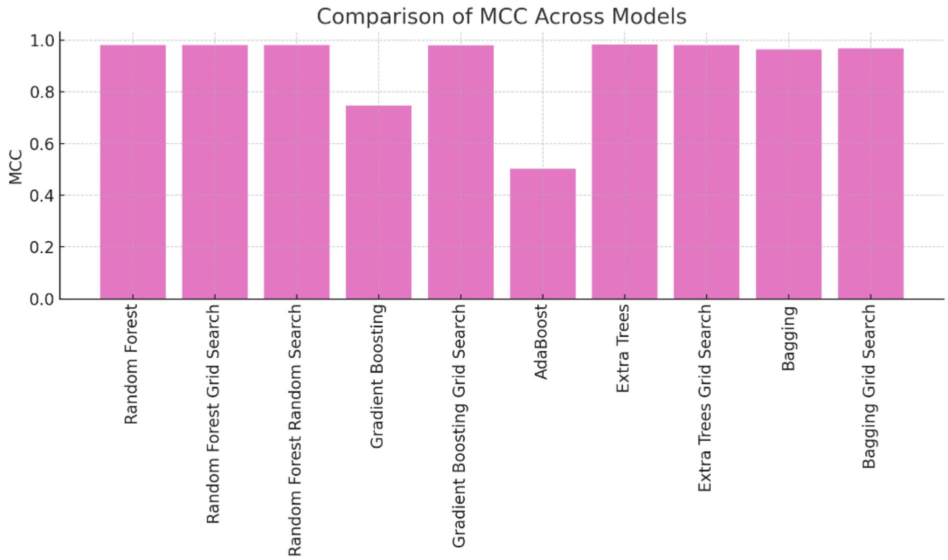

Figure 7. MCC

4.8. Cohen's Kappa

Cohen's Kappa measures between rater unwavering quality, and how well the model's forecasts match the true labels. Extra Trees accomplished the most noteworthy Kappa score of 0.9832, demonstrating solid arrangement among expectations and actual classifications. Random Forest variations likewise scored well, while AdaBoost fallen behind with a Kappa score of 0.5015, building up the way that its forecasts were not as dependable.

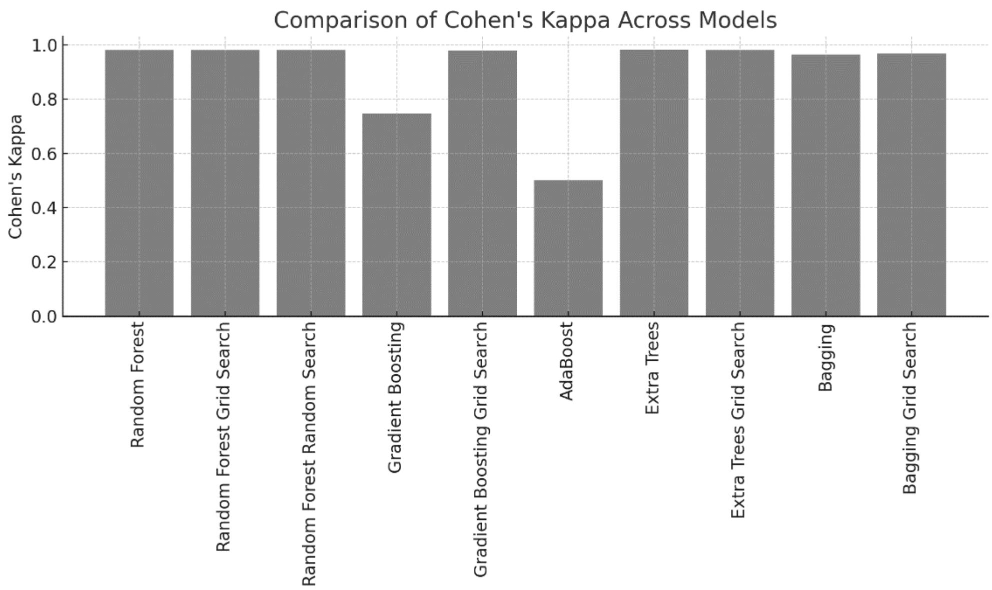

Figure 8. Cohen's Kappa

In general, the Extra Trees model exhibited the best execution across virtually all metrics, making it the most solid and exact gathering procedure for the mushroom classification task. Random Forest and its variations likewise performed emphatically, with just slight contrasts in their metrics. Gradient Boosting, while compelling, showed marginally lower execution in examination. AdaBoost reliably showed less fortunate execution across all metrics, demonstrating that it is less appropriate for this particular classification task. This relative examination features the viability of ensemble techniques, especially Extra Trees, in handling complex order issues where accuracy and reliability quality are basic.

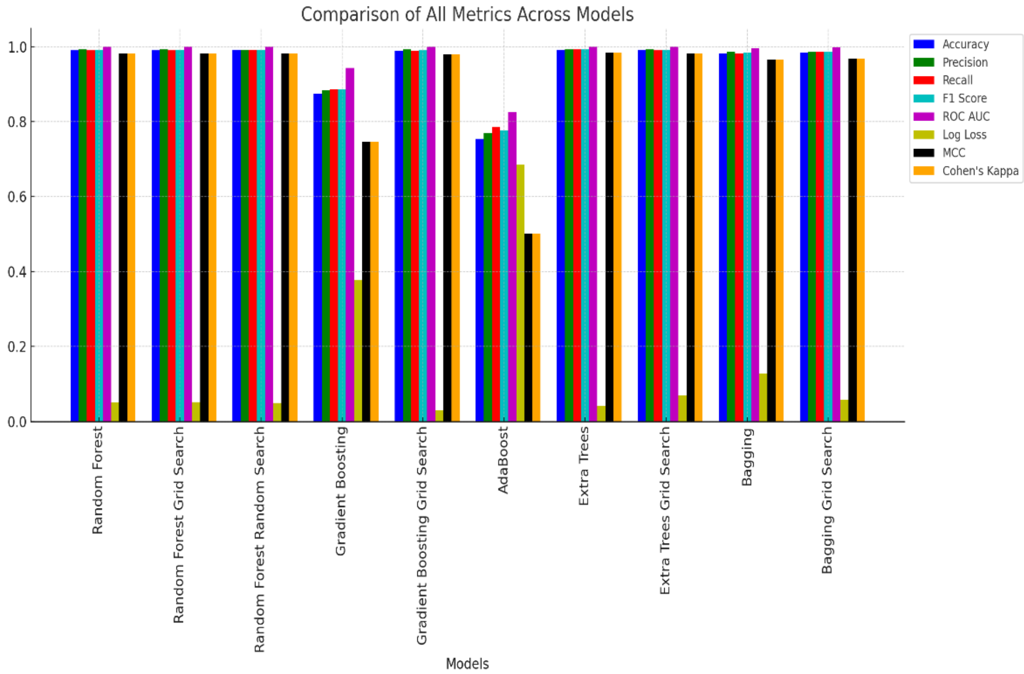

Figure 9. All Metrics Comparison

5. Conclusions

This research planned to evaluate the exhibition of few ensemble learning methods — Random Forest, Gradient Boosting, AdaBoost, Extra Trees, and Bagging — on the binary classification task of recognizing edible and poisonous mushrooms. By utilizing complete arrangement of assessment metrics, including accuracy, precision, recall, F1 score, ROC AUC, MCC, and Cohen's Kappa, the research gave a nitty gritty correlation of how these models performed on the UCI Mushroom Dataset. The discoveries showed that the Extra Trees model reliably beated other ensemble strategies, accomplishing the most noteworthy scores across most metrics, including accuracy (99.17%), precision (99.20%), recall (99.28%), and a great ROC AUC of 0.9994. This model's capacity to keep up with high precision and recall features its solidarity in limiting false positives and catching true positives effectively, which is essential in general wellbeing related classification tasks like mushroom poisonousness. The Random Forest technique likewise performed well, however marginally fallen behind Extra Trees with regards to precision and recall. Gradient Boosting, while viable, showed diminished execution contrasted with the top models, and AdaBoost fundamentally failed to meet expectations across most metrics, making it less appropriate for this particular assignment. These outcomes highlight the worth of ensemble strategies, especially Extra Trees, in basic applications where the expense of misclassification is high. The research builds up the significance of using an extensive variety of assessment metrics to completely survey model execution, especially in cases including lopsided or complex datasets. This exploration exhibits the reasonableness of utilizing ensemble learning procedures to take care of high-stakes classification issues. The prevalent presentation of Extra Trees and Random Forest proposes that these models can be dependably utilized in applications where misclassification could prompt serious outcomes, like in food handling or clinical conclusion. Experts can profit from the hearty order abilities of these models, particularly while working with complex datasets with related factors. While this research gave important bits of knowledge, there were restrictions connected with the dataset and the extent of models tried. Future examination could investigate crossover ensemble strategies or profound learning methods to additionally advance classification execution. Furthermore, assessing these models on other general wellbeing datasets or extending the scope of assessment metrics could give a more far reaching comprehension of their materialness in different situations.

References

1. Marley, Greg A. Chanterelle dreams, Amanita nightmares: The love, lore, and mystique of mushrooms. Chelsea Green Publishing, 2010.
2. Money, Nicholas P. Molds, Mushrooms, and Medicines: Our Lifelong Relationship with Fungi. Princeton University Press, 2024.
3. Richardson, Mary, et al. "Computer-based tests and machine marking: candidates' perceptions and beliefs about the test taking experience." (2023).
4. Valladares, Fernando, and Ülo Niinemets. "The architecture of plant crowns: from design rules to light capture and performance." Functional plant ecology. CRC Press, 2007. 101-150.
5. Boardman, Jonathan. "A New Kind of Data Science: The Need for Ethical Analytics." (2022).
6. Strang, Veronica. "Integrating the social and natural sciences in environmental research: a discussion paper." Environment, Development and Sustainability 11 (2009): 1-18.
7. Lanusse, François. "The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys." Publications of the Astronomical Society of Australia 40 (2023): e001.
8. Cotter, Tradd. Organic mushroom farming and mycoremediation: Simple to advanced and experimental techniques for indoor and outdoor cultivation. Chelsea Green Publishing, 2015.
9. Belete, Daniel Mesafint, and Manjaiah D. Huchaiah. "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results." International Journal of Computers and Applications 44.9 (2022): 875-886.
10. Arfeen, Zeeshan Ahmad, et al. "A comprehensive review of modern trends in optimization techniques applied to hybrid microgrid systems." Concurrency and Computation: Practice and Experience 33.10 (2021): e6165.
11. Mienye, Ibomoiye Domor, and Yanxia Sun. "A survey of ensemble learning: Concepts, algorithms, applications, and prospects." IEEE Access 10 (2022): 99129-99149.
12. Frey, Jennifer-Carmen. Using data mining to repurpose German language corpora. An evaluation of data-driven analysis methods for corpus linguistics. Diss. University of Bologna, 2020.
13. Akib, A., S. M. Zaman, and Fabiha Farzana. A Comprehensive Investigation into Detecting Schizophrenia from EEG Signals Using a Machine Learning Approach. Diss. Department of Electrical and Elecrtonics Engineering (EEE), Islamic University of Technology (IUT), Board Bazar, Gazipur-1704, Bangladesh, 2023.
14. Ferreira, Artur J., and Mário AT Figueiredo. "Boosting algorithms: A review of methods, theory, and applications." Ensemble machine learning: Methods and applications (2012): 35-85.
15. Santos, Silas Garrido Teixeira de Carvalho, and Roberto Souto Maior de Barros. "Online AdaBoost-based methods for multiclass problems." Artificial Intelligence Review 53.2 (2020): 1293-1322.
16. Murphy, Aidan. Evolving provably explainable fuzzy pattern tree classifiers using grammatical evolution. Diss. University of Limerick, 2021.
17. Ahmad, Muhammad Waseem, Jonathan Reynolds, and Yacine Rezgui. "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees." Journal of cleaner production 203 (2018): 810-821.
18. Yadav, Sunil, and Munindra Kumar Singh. "Hybrid machine learning classifier and ensemble techniques to detect Parkinson's disease patients." SN Computer Science 2 (2021): 1-10.
19. Michelucci, Umberto, and Francesca Venturini. "Estimating neural network's performance with bootstrap: A tutorial." Machine Learning and Knowledge Extraction 3.2 (2021): 357-373.
20. Molenberghs, Geert, et al. "On random sample size, ignorability, ancillarity, completeness, separability, and degeneracy: Sequential trials, random sample sizes, and missing data." Statistical Methods in Medical Research 23.1 (2014): 11-41.
21. Ishwarya, M. V., et al. "Innovations in Artificial Intelligence and Human Computer Interaction in the Digital Era." Computational Imaging and Analytics in Biomedical Engineering. Apple Academic Press, 2024. 105-145.