Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Edible Mushroom Classification Using Advanced Machine Learning Approaches

Md. Ashikuzzaman Ashik
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mdashikuzzaman004@gmail.com

Zakirul Islam
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
zakirul011@gmail.com

MD Sabbir Ahammed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mdsa134867@gmail.com

Sakib Imtiaz
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
sakibimtiaz1998@gmail.com

Md. Musfiqur Rahman Mridha
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mdmridha100730@gmail.com

Md. Fatin Nibbrash Nakib
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
fatinnibbrash@gmail.com

***Abstract*—Mushrooms, as a dietary component, offer immense nutritional and medicinal benefits. However, their classification into edible or poisonous categories is critical due to the severe health risks associated with consuming toxic varieties. Misidentification can result in adverse effects ranging from gastrointestinal distress to fatal poisoning. This study utilizes Machine Learning (ML) algorithms to tackle the problem through analysis of an extensive dataset comprising 22 mushroom attributes. The dataset was analyzed using seven machine learning models: Logistic Regression (LR), Support Vector Machine (SVM), LightGBM, XGBoost, AdaBoost, Random Forest (RF), and k-Nearest Neighbors (KNN). Most models achieved perfect classification with 100% accuracy and an AUC of 1.00, demonstrating their ability to distinguish between edible and poisonous mushrooms effectively. AdaBoost exhibited near-perfect performance with minor misclassifications. These results highlight the robustness of ML-based systems in ensuring food safety and preventing mushroom-related poisoning incidents. Future work will focus on scaling this approach to larger datasets, incorporating explainable AI techniques, and deploying these models in real-world applications for automated mushroom identification.**

***Index Terms*—Identification of Mushrooms, Machine Learning (ML), Food Safety**

I. Introduction

Mushrooms, the fruiting bodies of fungi, have been an integral part of human diets for centuries due to their high nutritional value and medicinal properties. Rich in proteins, vitamins (such as riboflavin and niacin), and antioxidants, mushrooms are widely consumed globally. However, not all mushrooms are safe to eat; some species are highly toxic and can cause severe poisoning or even death if ingested. This dual nature of mushrooms (nutritious yet potentially dangerous) makes accurate identification crucial for safe consumption. The challenge of distinguishing between edible and poisonous mushrooms lies in their morphological similarities. Traditional methods of identification often rely on visual inspection or biochemical analysis. While biochemical methods are reliable, they are time-consuming and impractical for everyday use. Visual identification, on the other hand, is prone to errors when performed by non-experts, leading to accidental poisonings. This underscores the need for automated systems capable of accurately classifying mushrooms. The significance of this research extends beyond academic interest; it has practical implications for public health and safety. Accurate mushroom classification systems can prevent accidental poisonings, promote sustainable foraging practices, and empower individuals with reliable tools for identifying edible species.

In this study, we present a comprehensive exploration of ML techniques for the classification of edible and poisonous mushrooms, addressing a critical public health concern. Our contributions are as follows:

- **Comprehensive Evaluation:** We implemented and evaluated seven advanced ML models Logistic Regression, Support Vector Machine (SVM), LightGBM, XGBoost, AdaBoost, Random Forest, and k-Nearest Neighbors (KNN) on a well-established mushroom dataset.
- **Perfect Classification Performance:** Most models achieved perfect classification with 100% accuracy and an AUC of 1.00, demonstrating their ability to reliably distinguish between edible and poisonous mushrooms.
- **Efficient Data Preprocessing Pipeline:** We employed advanced preprocessing techniques such as label encoding, one-hot encoding for categorical features, and feature scaling to ensure compatibility with ML algorithms and enhance model performance.

II. Related Works

Balika J. Chelliah et al. [1] conducted a comparative study on classifying poisonous or non-poisonous mushrooms using supervised ML models. They applied algorithms such as Decision Trees, Random Forests, Support Vector Machines, Logistic Regression, Gaussian Naive Bayes on a dataset from the UCI repository and concluded that Decision Trees performed best in terms of accuracy and reliability.

Prashant Sharma et al. [2] proposed an integrated ML model using the UCI Mushroom Dataset, combining decisions from the most accurate methodologies, achieving a 95% accuracy rate.

K. Kousalya et al. [3] compared Naive Bayes, Support Vector Machine (SVM), Decision Tree (C4.5) and Logistic Regression for classifying mushrooms as edible or poisonous using the Kaggle dataset and achieved the highest accuracy of 93.34% using the C4.5 algorithm.

Pranjal Maurya et al. [4] developed a mushroom classification method based on texture features using ML, achieving 76.6% accuracy with an SVM classifier, outperforming other classifiers like KNN, Logistic Regression, and Decision Trees.

Nadya Chitayae et al. [5] utilized the UCI Mushroom Dataset to compare K-Nearest Neighbor and Decision Tree methods for classifying edible and poisonous mushrooms and the Decision Tree method outperformed KNN, achieving an accuracy of 91.93%, along with a precision of 0.9227, recall of 0.9193, and an F1 score of 0.9210.

Sedat Metlek et al. [6] developed a classification system to distinguish between poisonous and edible mushrooms using Random Forest, Decision Tree and Logistic Regression algorithms on a dataset of 8124 samples with 22 features. Optimizing parameters with GridSearchCV, the Random Forest algorithm achieved the best performance, with precision, recall, and F1 scores of 0.93, 0.98, and 0.95 respectively.

Md. Samin Morshed et al. [7] explored mushroom edibility classification using efficient feature selection and nine ML methods, achieving the best performance with k-NN, which attained 99% accuracy and an F1-score of 99%.

III. Methodology

This section outlines the methodology employed to classify mushrooms as either edible or poisonous using a variety of ML algorithms. The process consists of several key steps, including data preprocessing, feature scaling, model selection, hyperparameter tuning, and performance evaluation, etc. as shown in Fig. 1. The implementation was carried out in Python using libraries such as Scikit-learn, LightGBM, XGBoost, and others.

A. Dataset and Preprocessing

The dataset used for this study was sourced from an openly available mushroom classification dataset [8]. It contains categorical features describing various morphological characteristics of mushrooms along with a binary target variable indicating edibility (edible [0] vs. poisonous [1]). A detailed overview of each feature, including its description and possible values, is provided in Table I.

- **Label Encoding:** The target variable (class) was encoded into binary numeric values using the LabelEncoder from Scikit-learn.
- **One-Hot Encoding:** All categorical features were transformed into numerical representations using one-hot encoding to ensure compatibility with machine learning models.

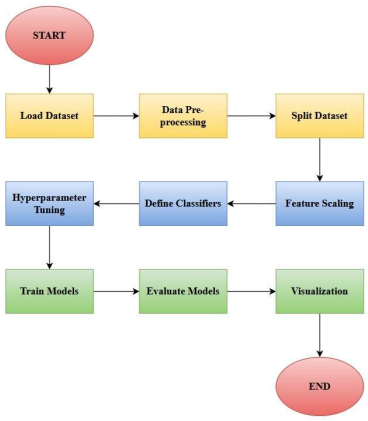

Fig. 1. Flowchart of the Overall Experiment.

- **Feature Scaling:** To standardize the feature space and improve model performance, the features were scaled using the StandardScaler.

TABLE I
FEATURES USED FOR CLASSIFYING EDIBLE AND POISONOUS MUSHROOMS

Features	Considered Variables
cap surface	fibrous, grooves, smooth, scaly
cap shape	bell, convex, conical, flat, knobbed, sunken
cap color	brown, buff, cinnamon, gray, green, purple, pink, red, white, yellow
bruises	bruises, no bruises
gill attachment	attached, descending, free, notched
gill spacing	close, crowded, distant
gill size	broad, narrow
gill color	black, brown, buff, chocolate, gray, green, orange, pink, purple, red, white, yellow
odor	almond, anise, creosote, fishy, foul, musty, none, pungent, spicy
stalk shape	enlarging, tapering
stalk root	bulbous, club, cup, equal, rooted, rhizomorphs, missing
stalk surface above ring	fibrous, scaly, silky, smooth
stalk surface below ring	fibrous, scaly, silky, smooth
stalk color above ring	brown, buff, cinnamon, gray, orange, pink, red, white, yellow
stalk color below ring	brown, buff, cinnamon, gray, orange, pink, red, white, yellow
veil type	partial, universal
veil color	brown, orange, white, yellow
ring number	none, one, two
ring type	cobwebby, evanescent, flaring, large, none, pendant, sheathing, zone
spore print color	black, buff, brown, chocolate, green, orange, purple, white, yellow
habitat	grasses, leaves, meadows, paths, urban, waste, woods
population	abundant, clustered, numerous, scattered, several, solitary

B. Data Splitting

The dataset was split into three subsets in a stratified manner to preserve class distribution across each subset. The training set, comprising 60% of the data, was used primarily for model

training. A separate validation set, comprising 20% of the data, was used for hyperparameter tuning during grid search, ensuring that models were fine-tuned without overfitting to the training data. Finally, the remaining 20% of the data constituted the test set, which was held out for the final evaluation of model performance.

C. Model Selection

A diverse set of machine learning algorithms was selected to evaluate their performance on the mushroom classification task. These include: Logistic Regression, Support Vector Machines (SVM), LightGBM, XGBoost, AdaBoost,Random Forest and k-Nearest Neighbors (KNN); Where applicable, GPU acceleration was utilized to speed up training for LightGBM (device='gpu') and XGBoost (tree_method='gpu_hist').

D. Hyperparameter Tuning

Hyperparameter tuning was conducted using grid search with 5-fold cross-validation to optimize model performance. The hyperparameter grids for each model were carefully designed based on commonly used configurations. The best-performing hyperparameters for each model were identified based on validation accuracy.

IV. Results and Analysis

This section presents the results of the mushroom classification task using various ML models. The performance of each model is evaluated using metrics such as accuracy, confusion matrices, and AUC scores. The results demonstrate that most models achieved near-perfect classification performance.

TABLE II
CLASS-WISE PERFORMANCE METRICS FOR XGBOOST ALGORITHM

Algorithm	Class	Accuracy	Precision	Recall	F1 Score
XGBoost	0 (edible)	100%	1.00	1.00	1.00
	1 (poisonous)	100%	1.00	1.00	1.00

The class-wise performance of XGBoost algorithm is shown in Table II.

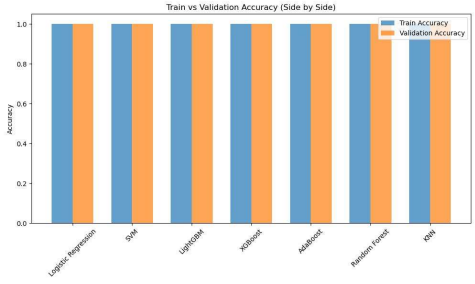

Fig. 2. Training vs Validation Accuracy

A. Training vs Validation Accuracy Analysis

Fig. 2 illustrates the training and validation accuracies for all models. It demonstrates that all models achieved near-perfect accuracy on both the training and validation sets, indicating strong generalization and no signs of overfitting.

B. Confusion Matrix Analysis

The confusion matrices for selected models are presented in the following figures. These matrices provide insight into the classification performance of each model.

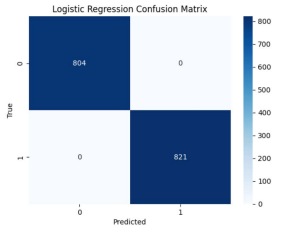

Fig. 3. Logistic Regression Confusion Matrix

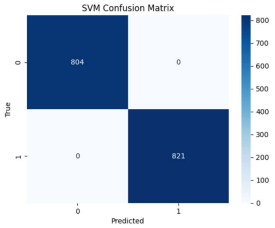

Fig. 4. SVM Confusion Matrix

Fig. 3 illustrates the performance of the Logistic Regression model on the test set. It shows perfect classification, with 804 true negatives (correctly classified edible mushrooms) and 821 true positives (correctly classified poisonous mushrooms), and no misclassifications.

The confusion matrix (Fig. 4) for the Support Vector Machine (SVM) model similarly reflects perfect classification, with 804 true negatives and 821 true positives, highlighting its effectiveness in this classification task.

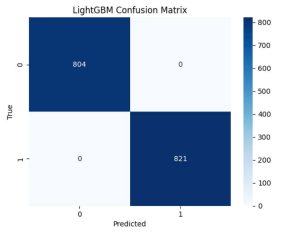

Fig. 5. LightGBM Confusion Matrix

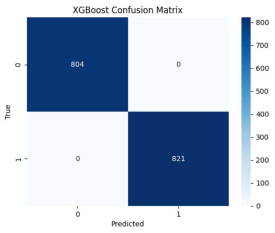

Fig. 6. XGBoost Confusion Matrix

The confusion matrix for the LightGBM model (Fig. 5) also shows perfect classification, with 804 true negatives and 821 true positives, confirming its high accuracy in distinguishing between edible and poisonous mushrooms.

In Fig. 6, the XGBoost model also achieved perfect classification, with 804 true negatives and 821 true positives, confirming its robustness in distinguishing between edible and poisonous mushrooms.

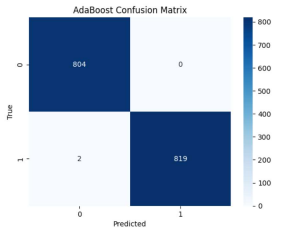

Fig. 7. AdaBoost Confusion Matrix

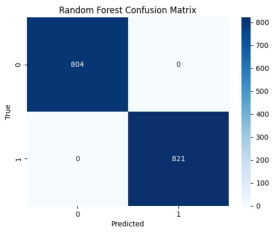

Fig. 8. Random Forest Confusion Matrix

The AdaBoost model's confusion matrix (Fig. 7) shows near-perfect classification, with 804 true negatives and 819 true positives. However, it misclassified 2 poisonous mushrooms as edible, indicating a slight drop in performance compared to other models.

Fig. 8 demonstrates the Random Forest model's performance on the test set, achieving perfect classification with 804 true negatives (edible mushrooms correctly classified) and 821 true positives (poisonous mushrooms correctly classified), and no misclassifications.

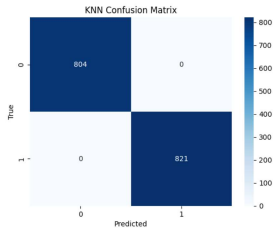

Fig. 9. K-Nearest Neighbors (KNN) Confusion Matrix

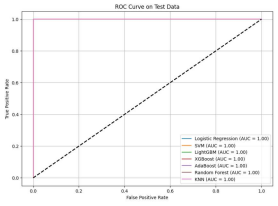

Fig. 10. ROC Curve on Test Data

The KNN model achieved perfect classification, as reflected in Fig. 9, with 804 true negatives and 821 true positives, and no misclassifications.

C. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve for all models on the test set is shown in Fig. 10. All models achieved an Area Under the Curve (AUC) of 1.00, indicating excellent discriminatory power with no trade-offs between sensitivity and specificity.

D. Overall Performance

All models demonstrated exceptional accuracy on both training and validation datasets, as shown in Table III. The Random Forest, Logistic Regression, SVM, LightGBM, XGBoost, and k-Nearest Neighbors models achieved perfect classification on the test set, while AdaBoost exhibited a slight drop in performance due to two misclassifications which can be observed from the confusion matrix of Fig. 7.

TABLE III
PERFORMANCE METRICS OF ALGORITHMS

Algorithms	Accuracy	Precision	Recall	F1 Score
Logistic Regression	100%	1.00	1.00	1.00
Support Vector Machine	100%	1.00	1.00	1.00
LightGBM	100%	1.00	1.00	1.00
XGBoost	100%	1.00	1.00	1.00
AdaBoost	100%	1.00	1.00	1.00
Random Forest	100%	1.00	1.00	1.00
KNN	100%	1.00	1.00	1.00

Table IV shows the comparison of our results (XGBoost is shown for its minimal misclassification) compared to related studies of recent experiments.

TABLE IV
PERFORMANCE COMPARISON WITH RELATED WORKS.

Models	Accuracy (%)
DT, RF, SVM LR, GNB [1]	-
Integrated Machine Learning Algo [2]	95
NB, DT(C4.5), SVM, LR [3]	93.45
SVM, KNN, LR, DT [4]	76.6
KNN, DT [5]	91.93
RF, DT, LR [6]	95
Nine ML Models [7]	99
XGBoost (Base Model)	**100**

V. CONCLUSION AND FUTURE WORKS

This study successfully demonstrated the use of machine learning models for mushroom classification, achieving exceptional results. Most models, including Random Forest, Logistic Regression, SVM, LightGBM, XGBoost, and KNN, achieved perfect classification with an accuracy of 100% and an AUC of 1.00. AdaBoost exhibited near-perfect performance with only two misclassifications. The ROC curve and confusion matrices validated the robustness of the models for distinguishing edible and poisonous mushrooms.

Future work could explore applying this methodology to larger datasets with more complex features to evaluate scalability. Additionally, integrating explainable AI techniques can enhance interpretability by identifying key features influencing predictions. Extending this work to real-world applications such as automated mushroom identification systems could significantly improve public safety and food security.

REFERENCES

[1] B. J. Chelliah, S. Kalaiarasi, A. Anand, G. Janakiram, B. Rathi, and N. K. Warrier, "Classification of mushrooms using supervised learning models," *International Journal of Emerging Technologies in Engineering Research (IJETER)*, vol. 6, no. 4, 2018.

[2] P. Sharma and C. Gupta, "Computer vision and machine learning based mushroom types classification," *Journal of Data Acquisition and Processing*, vol. 39, no. 1, pp. 876–893, 2024.

[3] K. Kousalya, B. Krishnakumar, S. Boomika, N. Dharati, and N. Hemavathy, "Edible mushroom identification using machine learning," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 2022, pp. 1–7.

[4] P. Maurya and N. P. Singh, "Mushroom classification using feature-based machine learning approach," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, B. B. Chaudhuri, M. Nakagawa, P. Khanna, and S. Kumar, Eds. Singapore: Springer Singapore, 2020, pp. 197–206.

[5] N. Chitayae and A. Sunyoto, "Performance comparison of mushroom types classification using k-nearest neighbor method and decision tree method," in *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 2020, pp. 308–313.

[6] S. Metlek and H. Çetiner, "Classification of poisonous and edible mushrooms with optimized classification algorithms," in *International Conference on Applied Engineering and Natural Sciences*, vol. 1, no. 1, 2023, pp. 408–415.

[7] M. S. Morshed, F. Bin Ashraf, M. U. Islam, and M. S. R. Shafi, "Predicting mushroom edibility with effective classification and efficient feature selection techniques," in *2023 3rd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2023, pp. 1–5.

[8] "Mushroom," UCI Machine Learning Repository, 1981, DOI: https://doi.org/10.24432/C5959T.