iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51772

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Machine Learning-based Classification of Edible and Poisonous Mushrooms: A Performance Comparison

Irfan Ahamed Essa[1], R. Dhanalakshmi[2]
[1, 2]*Department of Computer Science and Information Technology Kalasalingam Academy of Research and Education, 626126,Tamil Nadu, India*

Abstract: ***In the disciplines of computer vision and machine learning, classifying mushrooms is a key problem. The objective of this study is to create a machine learning model that can correctly categorise different species of mushrooms as either edible or harmful based on their visual characteristics. The Field Guide to North American Mushrooms of the Audubon Society served as the study's data source, and it includes descriptions of hypothetical samples of 23 species of gilled mushrooms from the Agaricus and Lepiota families. Each sample is categorised as either definitely edible, definitely poisonous. To create our machine learning model for this study, we combined five different algorithms: SVM, logistic regression, decision tree, random forest, and KNN. Metrics such as accuracy, precision, recall, and F1 score were used to assess each algorithm's performance. Our findings show that the random forest algorithm outperformed the other algorithms with a 95% accuracy rate. However, by employing feature selection strategies and hyperparameter tuning, we further improved our model. We were able to raise the accuracy of our model to 97% by choosing the most pertinent features and fine-tuning its hyperparameters. The outcomes of this study should offer a practical method for precisely categorising mushrooms and lowering the danger of mushroom poisoning. This paper proposes an ensemble approach utilizing five different machine learning classifiers to detect mushroom classification. The study analyzes various features, such as cap-shape, gill-colour, stem texture and ring-type & stalk features, and the findings reveal that the proposed random forest approach outperforms the base classifiers, achieving the highest accuracy of 99.63%.***

Keywords: ***CART Algorithm, SVM, Machine learning, Classification, Feature extraction, Mushroom, k nearest neighbor.***

I. INTRODUCTION

Many factors, such as climatic factors, soil factors, and biotic factors, influence the diversity of organisms, which is exciting because it can see differences from one organism to another, which is still one level. One of the exciting living organisms to observe is the fungal species. Each country must have distinctive and abundant natural resources and have various levels of living organisms, including at the level of genes, species, and eco-systems.

Mushroom classification is a significant issue in the fields of biology, mycology, and food science. Mushrooms are widely consumed as food, and the demand for edible mushrooms is increasing worldwide. However, the identification of edible and poisonous mushrooms is a complex task, even for experts. Many poisonous mushrooms closely resemble edible ones, making it challenging to distinguish between them based on visual inspection alone. Ingesting poisonous mushrooms can lead to severe health problems and even death, highlighting the importance of accurate mushroom classification. Identifying edible and poisonous mushrooms is a difficult task, even for experts. Many poisonous mushrooms closely resemble edible ones, making it difficult to distinguish between them based on visual inspection alone. Consuming poisonous mushrooms can lead to serious health problems.

However, mushrooms also contain a lot of mycotoxin, which determines whether they are dangerous or edible.Most individuals are unaware of the significant distinction between deadly and edible mushrooms. Every mushroom has unique qualities of its own. These traits are recognised as qualities that can be used to divide mushrooms into two groups: edible and poisonous. Several studies have looked into medical data mining, including the classification of mushrooms. The head diameter and neck length of mushrooms are thought to be the most crucial characteristics for categorising the stubby spines of the mushrooms. According to current estimates, there are 1.5 million different species of mushrooms in the globe. Less than 69,000 species have been officially recognised as existing in the globe. Less than 200,000 different species can be found in Indonesia. According to, just 2000 species of the millions of fungal species that occur worldwide are edible or can be utilised as food ingredients.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

A. "Exploring Techniques for Mushroom Classification: The Promise of Machine Learning and Computer Vision":

Mushroom classification is a challenging task due to the high variability in the appearance and characteristics of different species. To overcome this challenge, various techniques have been explored in the literature. In recent years, machine learning and computer vision techniques have shown great promise in this area. For example, in a study by Huang et al. (2021), a deep learning model based on a CNN architecture was developed for mushroom classification. The model achieved high accuracy in identifying different species of mushrooms based on their images, and outperformed several other existing methods for mushroom classification. Similarly, in another study by Patil et al. (2019), a convolutional neural network-based model was developed for mushroom classification using both visual features and other non-visual features such as geographic location and time of year.

B. "Machine Learning and Image Processing for Accurate Mushroom Classification: A Study by Bhandari et al. (2017)"

For instance, Bhandari et al. (2017) developed a system for mushroom classification based on image processing and decision tree algorithm. They used a dataset of mushroom images and extracted features such as shape, texture, and color to train the decision tree algorithm. The system achieved high accuracy in identifying different species of mushrooms based on their images, with an accuracy rate of 98.33%. The authors suggested that the system could be used as a tool for rapid and accurate identification of mushrooms in the field, which is particularly useful for amateur mushroom hunters or researchers who need to quickly identify a large number of specimens.

C. "Machine Learning for Mushroom Classification Based on Shape, Color, and Texture Features".

Karthick et al. present a technique for classifying mushrooms that makes use of shape, colour, and texture features derived from mushroom photos in their work titled "Mushroom Classification Based on Shape, Colour and Texture Features Using Machine Learning Techniques" (2020). Although these traits have been utilised in other studies to classify mushrooms, the authors expand on this research by investigating the efficacy of several machine learning methods for this purpose. They take a dataset of 562 photos of mushrooms from four different species, and for each image, they extract 26 shape, colour, and texture attributes.

D. "Classification of Mushroom Species using the ITS1 Sequence Database and BLASTN Model:A Study by Delgado-Serrano"

Various techniques have been employed in a number of prior research for the classification of different species of fungi. Delgado-Serrano [10] carried out a study to categorise different varieties of mushrooms using the ITS1 mushroom sequence database, and the correctness of the data was demonstrated. In that study, the dataset was scaled back because the volume of data per class could impair classification performance; this was done to prevent overfitting and achieve the best possible accuracy [11]. Delgado-Serrano utilises the BLASTN model, which has a 94% accuracy, in contrast.

E. "Decision Tree Algorithms for Analysis: A Comparative Study and Evaluation of their Use in Mushroom Classification"

Algorithm learning and analysis using decision trees have been done before finding data in replacement statistical procedures, extract text, certified medical fields, search engines, and research results show CART is the best algorithm compared to other decision tree algorithms [9]. A decision tree is a decision-making process similar to humans, so that it is easy to understand. If the data does not benefit when it is separated, execution will stop immediately [13], [14]. Types of algorithms that implement a decision tree are C4.5, ID3, and CART. ID3 algorithm can only be used on discrete values, so continuous datasets must be classified in discrete data pools [15]. The C4.5 and CART algorithms can make decisions on continuous data as input for simulation purposes. Using numerical separators, we can construct a tree using CART [16], [17]. Comparing the three algorithms that apply the decision tree is good enough, but further research is needed whether the decision tree method is the best method for the classification of mushroom species. Another study on mushrooms classification has been conducted using the Conventional Neural Network (CNN) Algorithm [18]. The proposed algorithm is robust enough to overcome the challenges presented by a color aberration in achieving accurate estimates. But in doing the classification, it is necessary to apply the Cross-Validation method for dividing and testing the dataset.

III. METHODOLOGY

The project aims to develop a new approach for detecting mushroom classification using machine learning. The methodology involves collecting and preprocessing a dataset, training several base machine learning models using various algorithms, identifying the most effective algorithms

A. Proposed Method

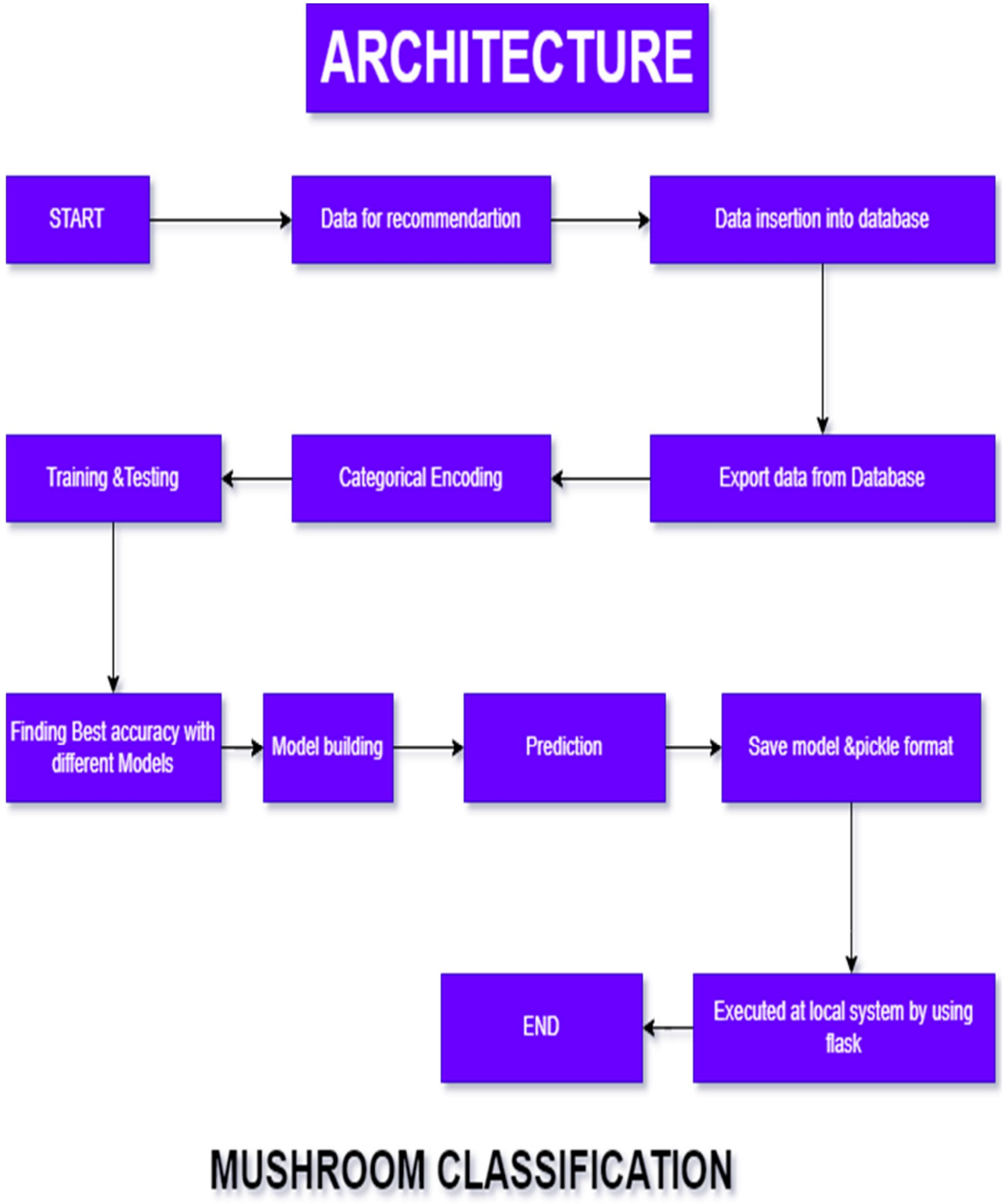
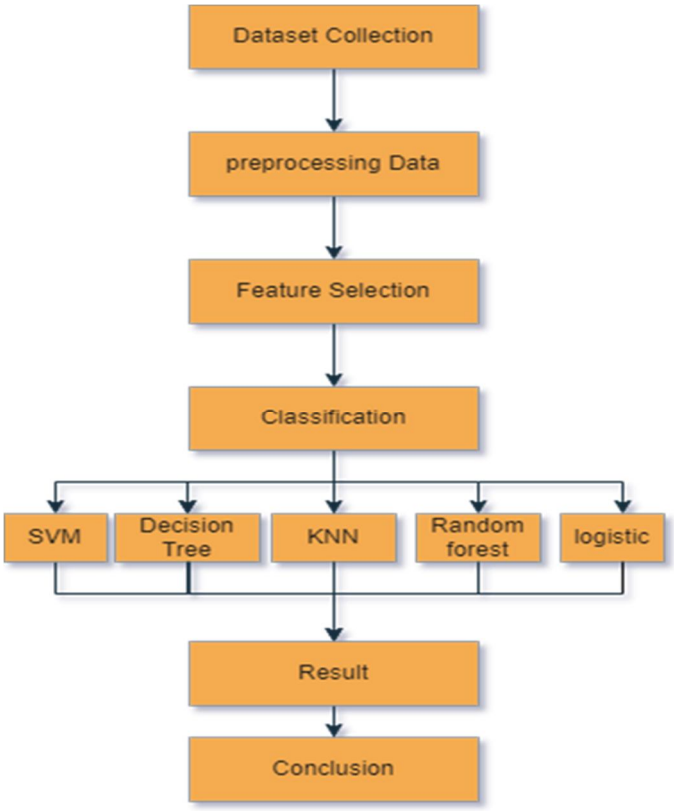


Fig 1.Architecture of the research Framework

There are several stages and methods for completing this research. Details of the research framework can be seen in Fig. 1

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com



B. Data Collection

A popular dataset for classifying mushrooms as edible or harmful is the Mushroom Data Set. Jeff Schlimmer created the dataset in 1981, and it was updated in 2016. It has 22 separate attributes that characterise various aspects of the mushrooms, including cap form, cap color, odor, and gill size, and 8124 instances, each of which represents a unique fungus. The UCI Machine Learning collection, a public collection of datasets for machine learning research, is where the dataset was obtained. The dataset, which is accessible for download on a number of platforms, including Kaggle.com, has been used extensively in studies on machine learning methods for mushroom classification.

C. Data Preprocessing

Data preprocessing phase of the research involved cleaning, normalizing, selecting, and engineering features of the collected datasets. Various Python libraries such as Pandas, NumPy, Scikit-learn, Seaborn, and Matplotlib were used for data preprocessing. Pandas was used to handle and transform data, NumPy for array manipulation and mathematical operations, Scikit-learn for data preprocessing, model selection, and evaluation, Seaborn for data visualization, and Matplotlib for creating different types of visualizations. These libraries provide a range of functionalities for handling, transforming, analyzing, the mushroom dataset.

D. Feature Engineering

The first step in creating a machine learning model for classifying mushrooms is feature engineering, which involves selecting and extracting important characteristics from the dataset that can most effectively distinguish between edible and deadly mushrooms. 22 different features from the mushroom data set describe various aspects of mushrooms in this context. The dataset 22 features that differentiate edible and poisonous mushroom based on some important features. The study concluded that "gill-color," "spore print color," and "population and gill-size" features are the most significant in mushroom classification. In order to prepare the dataset for machine learning, it is important to preprocess and clean the data. This involves handling missing values, converting categorical features into numerical ones, and scaling the features to ensure that they have a similar range. Once the data has been preprocessed, feature selection techniques such as correlation analysis, principal component analysis, and recursive feature elimination can be used to identify the most important features for the classification task.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Some of the most important features for differentiating between edible and poisonous mushrooms include:

1) Cap shape and color
2) Gill size, color, and spacing
3) Odor
4) Bruising
5) Stalk shape and color
6) Habitat
7) Spore print color
8) Veil type

E. Architecture of Proposed Approach

The proposed model's architecture for detecting edible or non edible is illustrated in Fig. 2.

The proposed machine learning methodology includes three stages: training , model building, and detection. The training stage involves feature selection , preprocessing , transformation, and training of 5 classifiers. The ranking stage selects the top four performing algorithms for the ensemble learning model. The detection stage uses the trained five machine learning model to classify into edible or poisonous ,mushrooms, and based on identifying features of mushroom dataset in real-time. Web application.

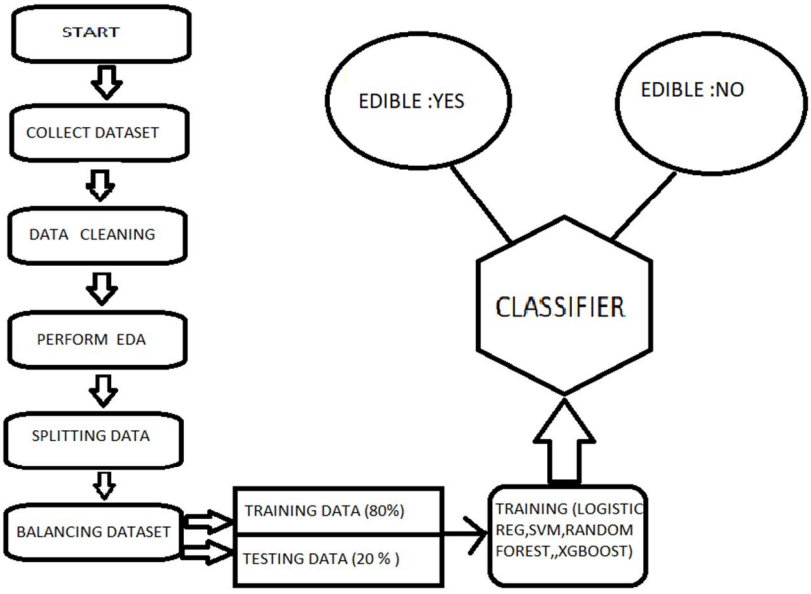


Fig. 2 Architecture of the proposed model

IV. RESULTS AND DISCUSSION

The study aimed to assess the effectiveness of a stacking ensemble machine learning technique in detecting cloud-based phishing attacks. The dataset was divided into training and testing sets, and various machine learning models, including Random Forests, Support Vector Machines, K-Nearest Neighbors, Decision Tree, Gradient Boosting Classifier, CatBoost Classifier, XGBoost Classifier, and Multi-layer Perceptron, were trained and tested. The best four models were combined using a stacking ensemble approach. The results showed that the proposed method achieved high accuracy, precision, and recall in detecting phishing websites, outperforming other methods and individual models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Evaluation Metrics

Evaluation metrics are measures used to evaluate machine learning model performance. The metrics assess the accuracy of the model in making accurate predictions on a given dataset. Metrics like accuracy, precision, recall and F1-score are used to evaluate phishing detection system effectiveness. Confusion matrix is used to calculate these metrics and provide detailed breakdown of true positives, false positives, true negatives, and false negatives for the classification model predicting phishing websites.

B. Experiment Results

This study compared the accuracy of various traditional machine learning classifiers for mushrooms. Results showed that the proposed stacking ensemble approach outperformed other classifiers in all datasets, achieving the highest accuracy scores of 97.6%, 86.2%, 88.6%, and 98.8% for Datasets 1 to 4, respectively. The K-Nearest Neighbors classifier had the lowest accuracy score in all datasets. The proposed stacking ensemble approach was found to be more effective than traditional machine learning classifiers for detecting phishing attacks. Table 1 shows evaluation metrics on dataset 1.

TABLE 1
EVALUATION METRICS OF MUSHROOM DATASET 1

ML Model	ACCURACY	PRECISION	RECALL	F1 SCORE
LOGISTIC REGRESSION	0.8344	0.853	0.791	0.821
K NEAREST NEIGHBOR	0.9833	0.974	0.991	0.982
SVC	0.0952	0.966	0.926	0.949
DECISION TREE	0.9760	0.973	0.976	0.975
RANDOM FOREST	0.9963	0.996	0.996	0.996
Gradient Boosting Classifier	0.9384	0.954	0.915	0.934

Fig. 2 shows evaluation of the proposed approach and other classification algorithms based on their accuracy, precision, recall, and F-measure on mushroom Dataset.

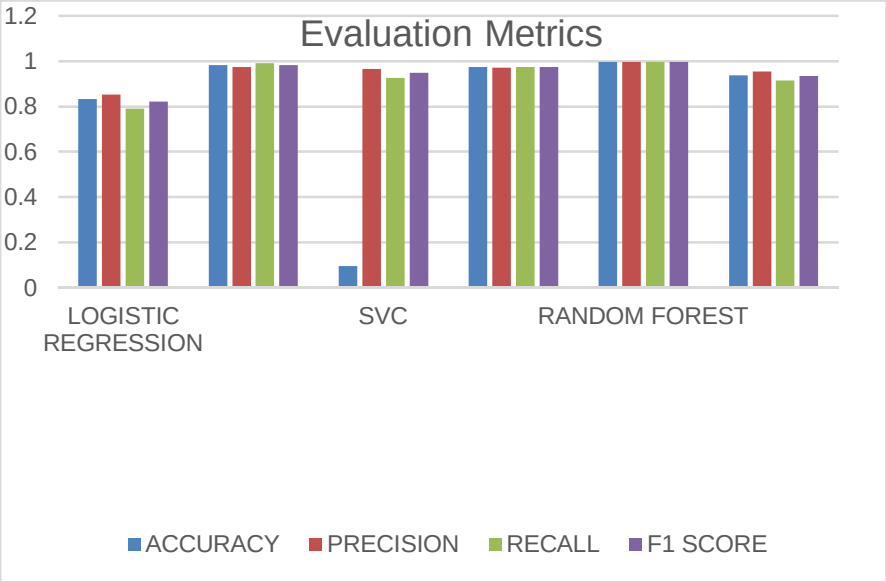


Fig. 2 Evaluation metrics

This study's results are the performance of each method, such as accuracy, precision, recall, and F1-score.

1) Accuracy is the level of closeness between the predicted value and the actual value.
2) Precision is the accuracy between the information requested by the user and the system's answers.
3) Recall (sensitivity) is the model's ability to find all relevant cases or data points in the dataset.
4) F-Measure or F1 Score is an evaluation calculation in information retrieval that combines recall and precision.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. CONCLUSIONS

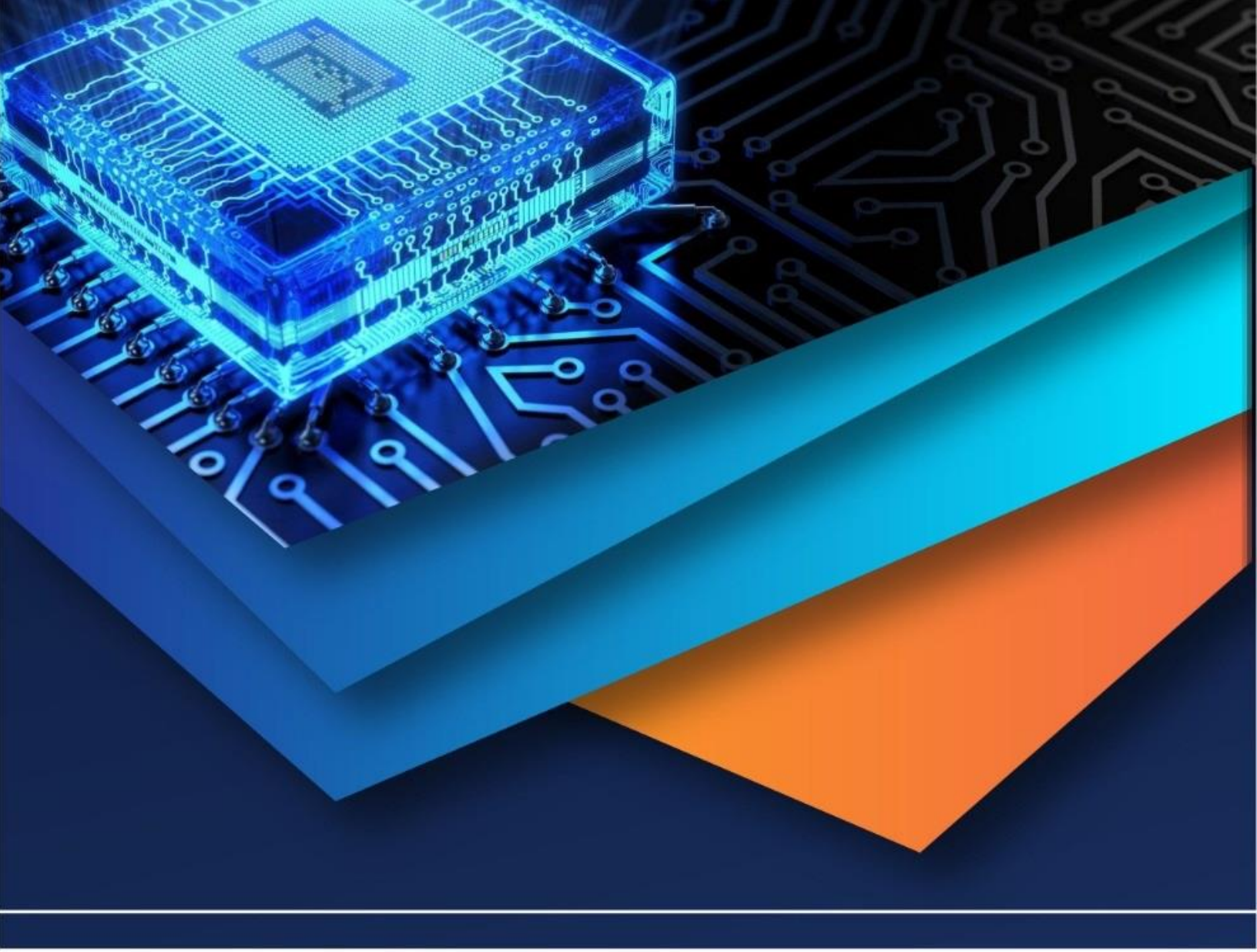
It may be inferred from the research that machine learning approaches can be successfully applied for mushroom classification. In order to identify distinct species of mushrooms, a variety of classification methods like Random Forest, Support Vector Machine, and K-Nearest Neighbours have shown to have good accuracy rates.and also accurately identifying with high levels of precision, recall, accuracy, and F1 score. In conclusion, our study emphasises the significance of both feature selection and dataset size in raising the precision of machine learning-based mushroom classification. To further improve the effectiveness of mushroom classification systems, future research should continue to investigate the use of sophisticated feature selection and extraction methods as well as larger and more diverse datasets.

VI. FUTURE WORK

Future work should concentrate on improving the interpretability of the stacking ensemble technique and identifying the crucial elements that are essential for mushroom classification using machine learning, according to the study's conclusions. To increase accuracy, especially given the changing properties of mushrooms, it is also advised to integrate this method with cutting-edge strategies like deep learning and natural language processing. Overall, the proposed method shows encouraging results, but more research is required to improve the method make it feasible.

REFERENCES

[1] "How many edible/poisonous mushrooms are there?", Mushroomthejournal.com, 2020. [Online]. Available:http://www.mushroomthejournal.com/greatlakesdata/TopTen/Quest19.html. [Accessed: 10- Dec- 2020].
[2] "Supervised Learning - an overview | ScienceDirect Topics", Sciencedirect.com, 2020. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/supervisedlearning. [Accessed:10- Dec- 2020].
[3] T. Fukuwatari, E. Sugimoto, K. Yokoyama, and K. Shibata,"Establishment of animal model for elucidating the mechanism of intoxication by the poisonous mushroom Clitocybe acromelalga,"Journal of the Food Hygienic Society of Japan, vol.42, no.3, pp.185–189, 2001.
[4] . Eyad Sameh , M. Khaled A. , M. Mohamad, G. Mohannad, S. Bassem,A. Samy S.," Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network," InternationalJournal of Academic and Applied Research (IJAAR), Vol. 3 (2), pp. 1-8, 2019.
[5] Al-Mejibli and D. Hamed Abd, "Mushroom Diagnosis Assistance System Based on Machine Learning by Using Mobile Devices" Intisar Shaded AlMejibli University of Information Technologyand Communications Dhafar Hamed Abd Al-Maaref University College,vol. 9, no. 2, pp. 103– 113,2017.
[6] Chowdhury and S. Ojha, "An Empirical Study on Mushroom Disease Diagnosis : A Data Mining Approach," International Research Journal of Engineering and Technology(IRJET), vol. 4, no.1, pp. 529–534, 2017.
[7] . M. Ottom, "Classification of Mushroom Fungi Using Machine Learning Techniques", International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 5, pp. 2378-2385, 2019.
[8] "Google Drive: Sign-in", Drive.google.com, 2021. [Online]. Available: https://drive.google.com/drive/folders/1X29L94ZluLNjwhFO_dFQiKoDvpr8sBI. [Accessed: 12- Mar 2021].
[9] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," pp. 334–337, 2013
[10] P. Maurya and N. P. Singh, "Mushroom classification using featurebased machine learning approach," in Proceedings of 3rd International Conference on Computer Vision and Image Processing,Singapore:Springer Singapore, 2020, pp. 197–206.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)