IMPROVISATION OF MUSHROOM TOXICITY BASED ON FEATURES EXTRACTED FROM IMAGES BY USING K-NEAREST NEIGHBOUR ALGORITHM COMPARING GCFOREST ALGORITHM

Sk Yaseen Ahmed[1],T. P. Anithaashri[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The aim is to improvise the detection of Mushroom Toxicity using novel K-Nearest Neighbor algorithm compared to the GCforest algorithm.
Materials and methods: By using novel K-Nearest Neighbor algorithm and GCforest both were identified and performed with the sample size of 45 each and the software tools that were used in this project are jupyter notebook. Accuracy values for identification of toxicity in mushrooms are calculated to quantify the performance of the GCforest algorithm against novel K-Nearest Neighbors algorithm.
Results : The analysis on train dataset and test dataset were successfully performed using SPSS and acquired accuracy for the GCforest algorithm and novel k-nearest neighbor algorithm method which gave the accuracy with the level of significance ($p>0.05$) the resultant data depicts the reliability in independent sample tests.
Conclusion: On the whole process of prediction of accuracy the K-Nearest Neighbor method gives significantly better performance compared with GCforest algorithm. By extracting images from image processing.

Keywords: Novel K-Nearest Neighbor, GCforest algorithm, Machine learning, Mushroom toxicity, Image processing.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. Saveetha University, Chennai, Tamilnadu. India, Pincode:602105
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105.

1. Introduction

In this project, will examine the data and build different machine learning models that will detect if the mushroom is edible or poisonous by its specifications like cap shape, cap color, gill color, etc using different classifiers by image processing and machine learning (Al-Mejibli and Abd 2017). This above process can be solved by using machine learning techniques(Singh and Chauhan 2016). can find out what the features can describe. The class that is poisonous or edible of the mushroom we took a dataset from kaggle. There are a lot of myths around mushrooms and their edibility. Had done exploratory data analysis on the data set in python to bust those myths. Existing approach using image processing only deals with accuracy (Dong and Zheng 2019).

The existing system implemented through Machine learning approaches and image processing Neural Network and Adaptive Neuro Fuzzy inference systems are used for implementation of the classification techniques (Wagner, Heider, and Hattab 2021). Different techniques used for classification are used to categorize different mushrooms as mushroom toxicity edible or non-edible using anaconda software. In addition, disease monitoring by the Bureau of Epidemiology, health care system reported that the number of patients and dead persons from eating poisonous are 1,175 and 6 persons, respectively (Ismail, Zainal, and Mustapha 2018). The number of cases of consuming poisonous mushrooms is increasing, because people living in the northern and northeast of Thailand usually prefer to collect wild mushrooms for cooking by themselves continuously (Al-Mejibli and Abd 2017). The wild mushroom can grow well particularly in the provinces in the northern and northeast of Thailand; therefore, the morbidity rate of northern and northeast regions is higher than the other parts of Thailand (Bennett, Philippides, and Nowotny, n.d.). From the analysis of statistical data, it can be found that the rate of illness and death is sharply high in the period of May to November because this period is suitable for the growth of mushrooms (Seymour 2017). Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap in the existing system is implemented through machine learning approaches. This article deals with a comprehensive overview of recent research in classification of edible and non-edible mushrooms. The objective of this is to describe the pathophysiology of mushroom toxicity, to review the presentation of a patient with mushroom toxicity and summarizing the treatment options for mushroom toxicity. Modalities are explained to improve care coordination among interprofessional team members in order to improve outcomes for patients affected by mushroom toxicity.

2. Materials and Methods

The Study setting of the proposed work is done in the Compiler Design Lab, Saveetha School of Engineering, Saveetha University. The number of groups identified are two. Group 1 is the toxic mushroom and group 2 is the non-toxic mushroom. GCforest was iterated various number of times with a sample size of 45 calculated from clinical website and SPSS analysis is carried out with level of significance $p<0.05$.

The software tool used for detecting the accuracy score is using the python sklearn library. Hardware configuration was AMD RYZEN 5 (2.10 GHZ) processor with 8GB ram and 64 bit OS, x64 based processor system. The software configuration was Windows 11 professional. The data was preprocessed after performing many steps such as removing noise data, feature engineering, and feature selection.

GCforest

There are some methods that are involved in GCforest using this algorithm we can calculate the accuracy of the mushroom toxicity detection and obtain the results by extraction of the images using image processing. We propose the GCforest approach, which generates \textit{deep forest} holding these characteristics. This is a decision tree ensemble approach, with much fewer hyper-parameters than deep neural networks, and its model complexity can be automatically determined in a data-dependent way (Auerbach, Donner, and Weiss 2008).

Following are the steps for implementing the GCforest method

There are 8 steps included in the proposed method. Firstly extracting the data, data cleaning, removing nan values, applying SMOTE and finally training the data with the Gcforest method.

Step 1 Start importing the data from a CSV file.

Step 2 Load datasets path. For training, some processing steps are required such as performing EDA, etc.

Step 3 Extracted features values. And removing missing values

Step 4 Apply the support vector machine techniques.

Step 5 Compare with k-nearest neighbor algorithm

Step 6 Identify the toxicity poisonous(1) edible(0)

Step 7 Calculate the accuracy values using SPSS Tool

Step 8 Selecting the top 6 columns for better performance of the model which is highly correlated with the independent variable.

K-Nearest Neighbor Algorithm
According to this method system the mushroom toxicity is selected based on the authorized data that is divided into two groups. One is toxic and the other one is edible based on the sample size and data. Using this novel k-nearest neighbor algorithm we can train the data and compile the output. Novel K-nearest neighbor is more accurate than the support vector machine algorithm because of high efficiency.

Following are the steps for implementing the k nearest neighbor method
There are 8 steps included in the proposed method. Firstly extracting the data, data cleaning, removing nan values, applying SMOTE and finally training the data with the k nearest neighbor method.

Step 1 Start importing the data from a CSV file.

Step 2 extracts the features from the images using machine learning.

Step 3 For training, some preprocessing steps are required such as performing EDA, etc.

Step 4 Training Model. Now in this step, we're going to see model training. Here we are using K-nearest neighbors

Step 5 The dataset being imported.

Step 6 Split Dataset. Next step is to split the dataset into a test and training.

Step 7 Predict the accuracy

Step 8 Then finally applying k nearest neighbor on the cleaned data and finally getting the accuracy of the model is 99.05%.Stop

Anaconda navigator is used for execution of the project code. It helps to manage and access notebook files and any kind of python files. By giving the python environment a command prompt can provide easy access to the code and execution. Main tools that need to be installed in the python environment are keras and tensorflow. Minimum of 4GB RAM is required to compile and execute the project code. Preferred operating systems are windows and ubuntu. The above mentioned method is for users using windows OS. Using anaconda navigator software and anaconda prompt that install the necessary modules.

Statistical analysis
To check the data accuracy and reliability in SPSS statistical software is used with a default alpha value of 0.05 for the sample size of 135. The independent variables for the dataset were blur, varying lighting condition, shadowing effects, image size of the images. Many potential variables are dependent in image classification like spectral signatures, vegetation indices, transformed images, textural or contextual information, multitemporal images, multisensor images, and ancillary data. The image is segmented and binarized to build the function that contains the interest area for detection. The bar graph and the error graph were generated for comparison of differences between the novel K-nearest neighbor and GCforest algorithms.

3. Results

Mushroom toxicity classification is used to reduce the death caused by the poisonous mushrooms and to save human life. For this purpose the support vector machine is compared with the k-nearest neighbor algorithm. By applying these methods the novel K-Nearest Neighbor is giving significantly better accuracy of 99.05% than the support vector machine. The results are collected by performing multiple iterations of the experiment for identifying different scales of accuracy rate. Further performed the statistical calculations using the SPSS tool and obtained the accuracy from the experimented data and independent sample test was performed.
Table1 represents the comparison of accuracy GCforest and K-Nearest Neighbor, by iterating in intrusion detection systems for various numbers of times.
Table 2 represents the sample size(N=45), Mean, Standard deviation and Standard error mean are classified based on the accuracy and loss of the data. The accuracy 100% of the k-nearest neighbor algorithm is significantly higher compared to GCforest algorithm.
Table 3 represents the significance of the data and standard error difference, where significance of GCforest and k nearest neighbor with the confidence interval as 95% and level of significance of 0.05.

Figure 1 represents the analysis of the accuracy of GCforest and k nearest neighbor for better improvisation in the mushroom detection system.
Figure 2 represents the comparison of mean accuracy of GCforest The comparison of accuracy gained. The accuracy of group 1 is 99. 05% and group 2 is 90.75%. The k nearest neighbor has significantly performed better when compared to support vector machines. Group 1 appears to produce the most consistent results with its standard deviation ranging from the lower 93's to higher 95's. Group 2 appears to produce the most variable results with its standard deviation ranging from 85's to 90's. There is a significant difference between GCforest band k nearest algorithm.

4. Discussions

In this study of Mushroom toxicity detection systems we observed that the K-nearest neighbor has higher accuracy of 99.05% in comparison to GCforest 90.75%. K-nearest neighbor has better significance ($p>0.05$) then support vector machine and while using the independent sample t-tests (Goetz 2003; Auerbach, Donner, and Weiss 2008).
Similar work has been carried out by the author (Bennett, Philippides, and Nowotny 2021). (Liu et al. 2021). To predict mushroom toxicity I have used machine learning to gather all the possible information to predict the outcome value, which is the accuracy of the algorithm which we have used in this proposed system (Rahman et al. 2020). Thus, by using the Wrapper method and Filter method, the Key Attributes that contributed to the better classification of mushrooms are identified. The attributes that have been found to be the best ones from both the attribute selection methods are compared. It is found that both the attribute selection methods (Wagner, Dennis, Dominik Heider, and Georges Hattab. 2021) almost gave the same results as the output. Hence by using these attributes as the key attributes, there will be a better accuracy in the classification of mushrooms is edible or poisonous (Adachi et al. 2021). The key attributes were also found to have good Precision, Recall and F-Measure values (Basal, Elfiky, and Eid 2021).

5. References

Adachi, Yoshiyuki, Takashi Kanno, Ken-Ichi Ishibashi, Daisuke Yamanaka, Akitomo Motoi, Masuro Motoi, and Naohito Ohno. 2021. "Binding Specificity of a New Artificial β-Glucan Recognition Protein and Its Application to β-Glucan Detection in Mushroom Extracts." International Journal of Medicinal Mushrooms 23 (4): 1–12.

Al-Mejibli, Intisar Shadeed, and Dhafar Hamed Abd. 2017. "Mushroom Diagnosis Assistance System Based on Machine Learning by Using Mobile Devices." Journal of Al-Qadisiyah for Computer Science and Mathematics. https://doi.org/10.29304/jqcm.2017.9.2.319.

Auerbach, Paul S., Howard J. Donner, and Eric A. Weiss. 2008. "Mushroom Toxicity." Field Guide to Wilderness Medicine. https://doi.org/10.1016/b978-1-4160-4698-1.50046-1.

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." Chemosphere. https://doi.org/10.1016/j.chemosphere.2022.134596.

Basal, Wesam Taha, Abdo Elfiky, and Jehane Eid. 2021. "Chaga Medicinal Mushroom Inonotus Obliquus (Agaricomycetes) Terpenoids May Interfere with SARS-CoV-2 Spike Protein Recognition of the Host Cell: A Molecular Docking Study." International Journal of Medicinal Mushrooms 23 (3): 1–14.

Bennett, James E. M., Andrew Philippides, and Thomas Nowotny. 2021. "Learning with Reinforcement Prediction Errors in a Model of the Drosophila Mushroom Body." Nature Communications 12 (1): 2569.

———. n.d. "Learning with Reward Prediction Errors in a Model of the Drosophila Mushroom Body." https://doi.org/10.1101/776401.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." Bioresource Technology. https://doi.org/10.1016/j.biortech.2022.127234.

Domondon, Lic Denisa L. 2000. Detection of Fruit Stimulating Biofactors in Waste Used for Mushroom Cultivation.

Dong, Jinhua, and Lixin Zheng. 2019. "Quality Classification of Enoki Mushroom Caps Based on CNN." 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC). https://doi.org/10.1109/icivc47709.2019.8981375.

Goetz, Christopher G. 2003. "Mushroom Toxicity." Encyclopedia of the Neurological Sciences. https://doi.org/10.1016/b0-12-226870-9/01195-3.

Ismail, Shuhaida, Amy Rosshaida Zainal, and Aida

Mustapha. 2018. “Behavioural Features for Mushroom Classification.” 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). https://doi.org/10.1109/iscaie.2018.8405508.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. “A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims.” Soft Computing. https://doi.org/10.1007/s00500-022-06931-1.

Kremmyda, Alexandra, William MacNaughtan, Dimitris Arapoglou, Christos Eliopoulos, Maria Metafa, Stephen E. Harding, and Cleanthes Israilides. 2021. “The Detection, Purity and Structural Properties of Partially Soluble Mushroom and Cereal β-D-Glucans: A Solid-State NMR Study.” Carbohydrate Polymers 266 (August): 118103.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. “Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation.” Journal of Thermal Analysis and Calorimetry. https://doi.org/10.1007/s10973-022-11298-4.

Liu, Yuanchao, Huiping Hu, Manjun Cai, Xiaowei Liang, Xiaoxian Wu, Ao Wang, Xiaoguang Chen, et al. 2021. “Whole Genome Sequencing of an Edible and Medicinal Mushroom, Russula Griseocarnosa, and Its Association with Mycorrhizal Characteristics.” Gene, October, 145996.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. “Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm.” Computational Intelligence and Neuroscience 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. “Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands.” Silicon. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. “A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies.” Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102155.

Rahman, Mohammad Azizur, Shahdat Hossain, Noorlidah Abdullah, and Norhaniza Aminudin. 2020. “Lingzhi or Reishi Medicinal Mushroom, Ganoderma Lucidum (Agaricomycetes), Ameliorates Nonspatial Learning and Memory Deficits in Rats with Hypercholesterolemia and Alzheimer’s Disease.” International Journal of Medicinal Mushrooms 22 (11): 1067–78.

Seymour, Tom. 2017. Foraging Mushrooms Maine: Finding, Identifying, and Preparing Edible Wild Mushrooms. Rowman & Littlefield.

Singh, Rajender, and Mamta Chauhan. 2016. “Effective Management of Agro-Industrial Residues as Composting in Mushroom Industry and Utilization of Spent Mushroom Substrate for Bioremediation.” Advances in Environmental Engineering and Green Technologies. https://doi.org/10.4018/978-1-4666-9734-8.ch008.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. “Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends.” Energy. https://doi.org/10.1016/j.energy.2022.123806.

Wagner, Dennis, Dominik Heider, and Georges Hattab. 2021. “Mushroom Data Creation, Curation, and Simulation to Support Classification Tasks.” Scientific Reports 11 (1): 8134.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. “Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture.” Chemosphere 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. “Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review.” Chemosphere 299 (July): 134390.

6. Conclusion

It is inferred that the K-Nearest Neighbor seems to appear with a better accuracy percentage (99.05%) to detect the toxicity in the mushrooms whether it is edible or poisonous. Then the Gcforest algorithm with the accuracy of (90.62%)

Declarations

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author SYA was involved in data collection, data analysis, and manuscript writing. Author TPA was involved in conceptualization, data validation and critical review of manuscript.

Acknowledgement

The Author would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete this study:

1. RootQuotient Technologies P Ltd, Chennai
2. Saveetha School of Engineering.
3. Saveetha University.
4. Saveetha Institute of Medical and Technical Sciences.

TABLES AND FIGURES

Table 1. Accuracy table for K-Nearest Neighbor and GCforest the accuracy of Method 1 is 99. 05% and Method 2 is 90.75%.

No. of Iterations	K-Nearest Neighbor Algorithm	GCForest Algorithm
1	99.05	90.75
2	98.01	91.01
3	98.03	91.02
4	98.05	91.03
5	98.07	91.04
6	98.09	91.05
7	98.11	91.06
8	98.13	91.07
9	98.17	91.08
10	98.19	91.09

Table 2. Statistical Analysis of Mean, Standard Deviation, and Standard Error of accuracy of GCforest and K-Nearest Neighbour. There is a statistically significant difference in accuracy between the methods. K-Nearest Neighbour has the highest accuracy (99.05%) and GCForest (90. 75%).

Group	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Algorithms				
K-Nearest Neighbor	45	99.1850	.29217	.09239
GCForest	45	91.0200	.09832	.03109

Table 3. Comparison of Significance Level with value p>0.05. Both GCforest and K Nearest Neighbour have a confidence interval of 95% with the significance level of accuracy is >0.05.

		F	**sig.**	**t**	**df**	**sig. (2-tailed)**	**Mean difference**	**Std. Error Difference**	**95% Confidence interval of the differenceLower**	**95% Confidence interval of the difference Upper**
Accuracy	Equal variance assumed	1.892	.186	83.759	18	.001	8.165	.09748	7.96020	8.36980
Accuracy	Equal variances not assumed			83.759	11.013	.001	8.165	.09748	7.95047	8.37953

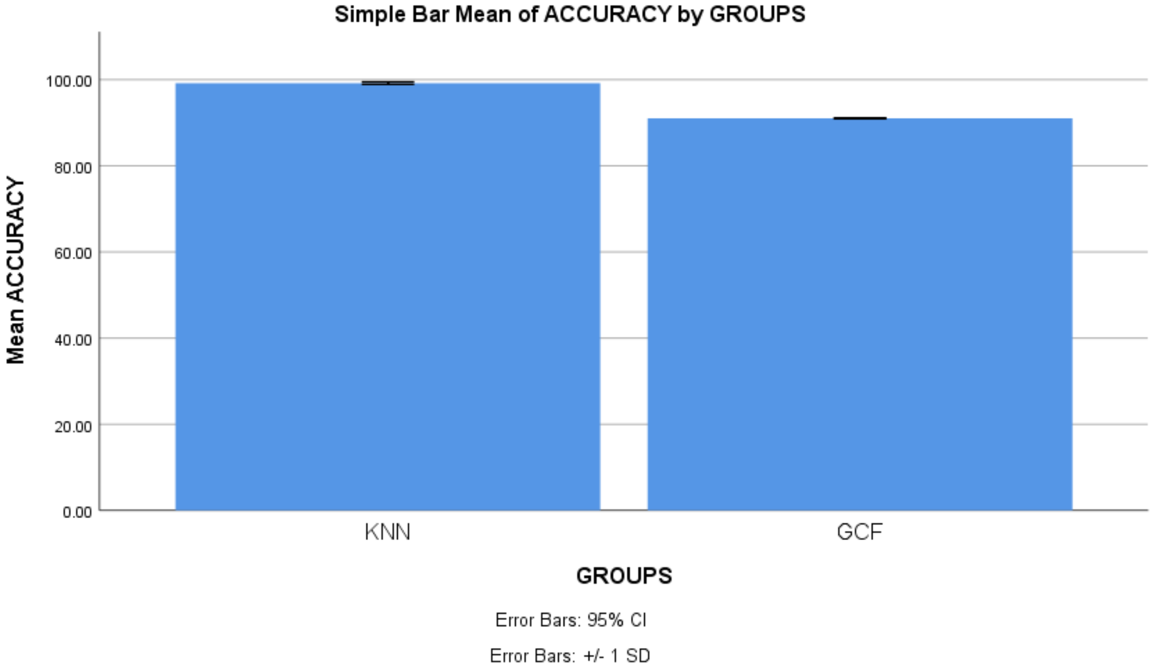

Fig. 1. Comparison of mean accuracy between K-nearest neighbor algorithm over GCforest algorithm, where the former is better than the latter with 90.75% increase. X axis gives the algorithms and Y Axis: Mean accuracy of detection ± 1 SD.