BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 333-340
DOI: 10.2478/bjlp-2022-004035

Detection of Mushroom Insalubrity Based on Features Extracted from Images using K-Nearest Neighbor Algorithm Compared with Support Vector Machine Algorithm

Sk Yaseen Ahmed
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

T. P. Anithaashri
Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The aim is to improve the detection of mushroom insalubrity based on features extracted from images by using novel K-Nearest Neighbor (KNN) algorithm comparing Support Vector Machine (SVM) algorithm. **Materials and Methods:** By using k-nearest neighbor algorithm and support vector machine algorithm, detection of insalubrity is tested over a mushroom datasets with the sample size of 10. Accuracy values for detection of mushroom insalubrity calculated to quantify the performance of KNN compared with SVM. **Results and Discussion:** The analysis on trained dataset and test dataset were successfully performed using SPSS and acquired accuracy for the Support Vector Machine compared to k-nearest neighbor algorithm which gave the accuracy with the level of significance ($p<0.05$) and with G-power about 80%. The resultant data depicts the reliability in independent sample tests. **Conclusion:** On the whole process of prediction of accuracy the novel K-nearest neighbor algorithm gives significantly better accuracy than Support Vector Machines for mushroom toxicity by extracting features in the images.

Keywords
Novel K-Nearest neighbor, Support Vector Machine, Machine learning, Mushroom Toxicity, Image Processing.

INTRODUCTION

In the existing system, it has been examined that the data and build different machine learning models, like k-nearest neighbor, support vector machines that will detect if the mushroom is edible or poisonous by its specifications like cap shape, cap color, gill color, etc. This above process can be solved by using machine learning and techniques (Al-Mejibli and Abd 2017). The process helps to find out what are the features that can be described. The dataset was taken from kaggle.com. The exploratory data analysis on the data set in python paves the way to address these myths using image processing. Existing approach only deals with accuracy (Dong and Zheng 2019) and consumes more time .
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 by consuming the poisonous mushrooms the number of patients and dead persons are 1,175 and 6 persons, respectively (Ismail, Zainal, and Mustapha 2018). The number of cases are 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).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The research gap in the existing system is implemented through machine learning and approaches. The proposed system 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. This system will review the health condition of a patient with mushroom toxicity and summarize the treatment options for the same. The system provides the modalities to improve care coordination among interprofessional team members in order to improve outcomes for patients affected by mushroom toxicity.

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. Support Vector Machine was iterated various number of times with a sample size of 10 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 pre-processed after performing many steps such as removing noise data, feature extraction and feature selection.

Support Vector Machine algorithm

There are some methods that are involved in Support Vector Machine. Novel K-nearest neighbor support vector machine CNN algorithms that can be used for building both regression and classification models (Auerbach, Donner, and Weiss 2008).

Following are the steps for implementing the support vector machine algorithm

Step 1 : Start importing the data from a CSV file.
Step 2 : For training, some processing steps are required such as performing EDA, etc.
Step 3: Extracted features values. Removing missing values
Step 4: Apply the Support Vector Machine techniques. Selecting the top 6 columns for better performance of the model which is highly correlated with the independent variable
Step 5: Compare with K-Nearest Neighbor Algorithm
Step 6: Identify the toxicity poisonous (1) and edible(0)
Step 7: Calculate the accuracy values using SPSS Tool

Step 8: Then finally applying SVM on the cleaned data and finally getting the accuracy of the model is 91.2%

K-Nearest Neighbor Algorithm

According to this method system the mushroom 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 with the trained data and extract the outcome. 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 Algorithm

There are 7 steps included in the proposed method. Initially extracting the data, data cleaning, removing noise ratio, applying Synthetic Minority Over Sampling Techniques(SMOTE). finally training data with the k nearest neighbor method

Step 1: Start importing the data from a CSV file.
Step 2: For training, some processing steps are required such as performing EDA, etc
Step 3: Import Dataset.
Step 4: Split Dataset.
Step 5: The models are trained using K-Nearest Neighbors Algorithm
Step 6: Selecting the top 6 columns for better performance of the model which is highly correlated with the independent variable.
Step 7: Then finally applying k nearest neighbor on the cleaned data and finally getting the accuracy of the model is 99.18%
Step 8: 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 or ubuntu.

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 10. 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 was generated for comparison of differences between the K-nearest neighbor algorithm and Support vector machine algorithm.

RESULTS

By applying these methods K-nearest neighbor algorithm is giving significantly better accuracy of 95% than Support Vector Machine. The results are collected by performing multiple iterations of the experiment for identifying different scales of accuracy rate. Further performing the statistical calculations using the SPSS tool and obtaining the accuracy from the experimented data and independent sample test is performed.

Table 1 represents the comparison of accuracy Support Vector Machine and K-nearest neighbor, by iterating in intrusion detection systems for various numbers of times.

Table 2 represents the sample size (N=10), Mean, Standard deviation and Standard error mean are classified based on the accuracy and loss of the data. The accuracy 99.18% of the K-Nearest Neighbor algorithm is significantly higher compared to the Support Vector Machine.

Table 3 represents the significance of the data and standard error difference, where significance of Support Vector Machine and k nearest neighbor with the confidence interval as 95% and level of significance of (< 0.05).
Fig. 1 represents the comparison of mean accuracy of the Support Vector Machine. The comparison of accuracy gained. The accuracy of group 1 is 99.18% and group 2 is 91.2%. The K-Nearest Neighbor has significantly performed better when compared to Support Vector Machines. There is a significant difference between Support Vector Machine and K-nearest algorithm.

DISCUSSIONS

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.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Bennett, James E. M., Andrew Philippides, and Thomas Nowotny. n.d. "Learning with Reward Prediction Errors in a Model of the Drosophila Mushroom Body." https://doi.org/10.1101/776401.

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.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

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.

Menser, Gary P. 2016. *Hallucinogenic and Poisonous Mushroom Field Guide*. Ronin Publishing.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Seymour, Tom. 2017. *Foraging Mushrooms Maine: Finding, Identifying, and Preparing*

Edible Wild Mushrooms. Rowman & Littlefield.
Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
Wagner, Dennis, Dominik Heider, and Georges Hattab. 2021. "Mushroom Data Creation, Curation, and Simulation to Support Classification Tasks." *Scientific Reports* 11 (1): 8134.

CONCLUSION

It is inferred that the K-Nearest Neighbor algorithm seems to appear with a better accuracy percentage (99.18%) detecting the toxicity in the mushrooms whether it is edible or poisonous than the Support Vector Machine with the accuracy of (91.2%).

DECLARATIONS

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author SYA was involved in data collection, data analysis, 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. AAIC Technologies Private Limited, Hyderabad
2. Saveetha School of Engineering.
3. Saveetha University.
4. Saveetha Institute of Medical and Technical Sciences.

REFERENCES

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.
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.

ISSN 2029-0454
2022

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

Nadya Chitayae, Andi Sunyoto, "Performance Comparison of Mushroom Types Classification Using K-Nearest Neighbor Method and Decision Tree Method", Information and Communications Technology (ICONTACT) 2020 3rd International Conference on, pp. 308-313, 2020.

Narumol Chumuang, Kittisak Sukkanchana, Mahasak Ketcham, Worawut Yimyam, Jiragorn Chalermdit, Nawarat Wittayakhom, Patiyuth Pramkeaw, "Mushroom Classification by Physical Characteristics by Technique of k-Nearest Neighbor", Artificial Intelligence and Natural Language Processing (iSAI-NLP) 2020 15th International Joint Symposium on, pp. 1-6, 2020.

Hui Zhao, Fuhua Ge, Pengfei Yu, Haiyan Li, "Identification of Wild Mushroom Based on Ensemble Learning", Big Data and Artificial Intelligence (BDAI) 2021 IEEE 4th International Conference on, pp. 43-47, 2021.

J. H. Tegzes and B. Puschner, ""Toxic mushrooms," the veterinary clinics of north America," Veterinary Clinics of North America: Small Animal Practice, vol. 32, no. 2, pp. 397–407, 2002.

C. Lei, W. Tangkanakul, and L. Lu, "Mushroom poisoning surveillance analysis," OSIR Journal, vol. 1, no. 1, pp. 8–11, 2006.

J. White, S. A. Weinstein, L. De Haro et al. , "Mushroom poisoning: a proposed new clinical classification," Toxicon, vol. 157, pp. 53–65, 2019.

J. H. Diaz, "Evolving global epidemiology, syndromic classification, general management, and prevention of unknown mushroom poisonings," Critical Care Medicine, vol. 33, no. 2, pp. 419–426, 2005.

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 (Shokuhin Eiseigaku Zasshi), vol. 42, no. 3, pp. 185–189, 2001.

M. Lu, "Present status and future prospects of the mushroom industry in China," Acta Edulis Fungi, vol. 13, no. 1, pp. 1–5, 2006.

J. Zhao, M. Cao, J. Zhang, Q. Sun, Q. Chen, and Z. -R. Yang, "Pathological effects of the mushroom toxin α-amanitin on BALB/c mice," Peptides, vol. 27, no. 12, pp. 3047–3052, 2006.

TABLES AND FIGURES

Table 1. Accuracy table for K-Nearest neighbor and Support Vector Machine, the accuracy of Method 1 is 99. 05 % and Method 2 is 91. 2 %

No. of iterations	K-Nearest neighbor Algorithm	Support Vector Machine Algorithm
1	99.05	91.21
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	99.13	91.07

ISSN 2029-0454

9	99.17	91.08
10	99.18	91.09

Table 2. Statistical Analysis of Mean, Standard Deviation, and Standard Error of accuracy of Support Vector Machine and K-Nearest neighbor . There is a statistically significant difference in accuracy between the methods. K-Nearest neighbor has the highest accuracy (99.18%) and Support Vector Machine (91.2%).

Group	N	Mean	Std. Deviation	Std. Error Mean
Algorithms				
K-Nearest neighbor	10	99.1850	.29217	.09239
Support Vector Machine	10	91.2070	.49715	.15721

Table 3. Comparison of Significance Level with value p<0.05. Both Support Vector Machine algorithm and K-Nearest neighbor 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 difference Lower	95% Confidence interval of the difference Upper
Accuracy	Equal variance assumed	.658	.042	43.751	18	.001	7.9780	.18235	7.59490	8.36110

Accuracy	Equal variances not assumed			43.751	14.554	.001	7.9780	.18235	7.58829	8.36771

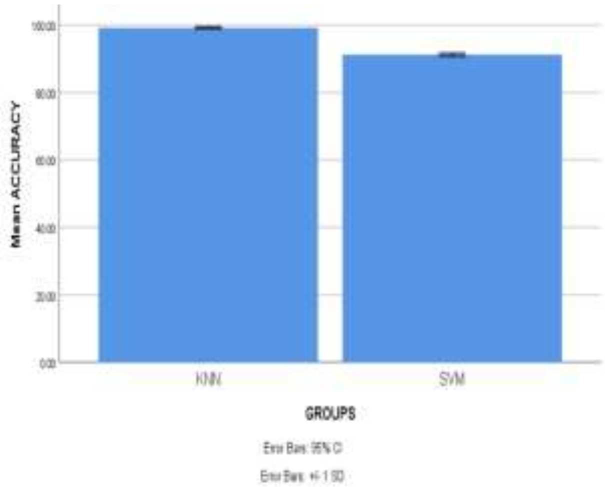

Fig. 1. Comparison of mean accuracy between K-Nearest neighbor algorithm over Support Vector Machine algorithm, where the former is better than the later with an increase of 7.96%. X-Axis gives the algorithms and Y-Axis: Mean accuracy of detection ± 1 SD.

`