**KF5012 Software Engineering Practice**

**Project Ideation**

Eve O’Brien

Word Count: 2,634

**Problem Statement**

We want to reduce mushroom poisoning by creating a binary classifier system that recognises images of mushrooms and categorises then as either edible or poisonous using images taken from the internet via a web scraper created to scrape images of mushrooms from different websites. This will be carried out by identifying which features are most indictive of a poisonous or edible mushroom via image recognition methods and using Deep Learning methods, such as a Convolutional Neural Network (CNN) to understand if a system like this can be created, and to help those who may not be able to identify safe and unsafe mushrooms when partaking in a hobby like Foresting.

CNNs have been widely regarded as having high accuracies when used in image classifications, thus, why one will be used to create our system.

**Problem Motivation**

There has been an increase in outdoor hobbies and activities in recent times, largely due to the ongoing COVID-19 pandemic, and the shutting down of indoor activities like retail shops, restaurants and gyms. This has been shown through a 12% increase of people spending their time outdoors in Scotland [1], and the University of Cumbria has found in a study that lockdown saw all age groups spending more time in nature daily, and that many people in their study took photos outdoors and tried to identify things like plants and mushrooms [2].

A common outdoor activity is Foresting, whereby groups of people go into countryside areas for walks, with ‘Shrooming’ being a part of this activity. ‘Shrooming’ can be described as the hunting of mushrooms, which can be dangerous if a poisonous genus of mushroom is touched or ingested. Due to the probable chance of beginners starting this activity, the proposed program will use machine learning techniques to classify if a mushroom is poisonous or not based on the mushroom features. The potential risk of beginner ‘Shroomers’ eating a miss-classified mushroom could be regarded as being likely. Ingestion of a poisonous mushroom can result in hallucinations, gastrointestinal illness, liver failure and death [3].

Because of this, we want to create an image classification system that will be able to classify mushrooms into either a ‘poisonous’ group, or an ‘edible’ group to help those who may be partaking in ‘Shrooming’ to keep themselves safe and be able to spot a mushroom they should avoid, as manually recognizing mushrooms relies on a high amount of background knowledge that the majority of newcomers of the hobby may not have, therefore creating an automatic system for this will meet a demand that is not currently available and has the aim to protect people from deadly fungus. There is also a lack of existing system to classify mushrooms on if they are poisonous or edible, so we hope to create a novel solution as there is not a large existing amount of them.

**A Review of Machine Learning Methods in the Classification of Mushroom Edibility**

Abstract

The purpose of this review is to explore different studies that focus on the classification and recognition of mushroom species and the edibility of mushrooms. This review seeks to analyse different research papers and compare them on the effectiveness of their proposed systems and machine learning models used to achieve their systems. The main body of this paper will be split up into two halves, with one focusing on systems with Tabular data, and the other focusing on systems that use Image data to achieve their results. Results of this review found that overall, across studies, using Deep Learning methods with Image data provide the best results for classifying mushroom edibility compared to using Supervised models on either tabular or image data, and would have the most effective application into the real world.

Introduction

3% of all known species of mushrooms are poisonous, with symptoms ranging from stomach discomfort to death, depending on the type and how much of the fungus has been consumed [4]. Moreover, in recent years, more than 10,000 people were poisoned by mushrooms in France with 22 deaths occurring between 2010 and 2017, highlighting the need for the public to be informed on how to identify inedible mushrooms [5]. Preventative measures have been put in place, such as alerting the public not to go mushroom picking and avoid consumption, yet the epidemic continues. Because of this, proposals involving machine learning methods, such as Supervised Learning and Deep Learning, have been utilised by researchers. These effective classification and recognition systems have correctly identified edible and inedible mushrooms.

This paper evaluates different studies that have been conducted in the past, comparing different models and datasets used, and their effectiveness to classify types of mushrooms, from edible, and inedible species using different data types.

Main Body

Method A – Tabular Data

Studies focusing on the classification of mushrooms based on their features have shown promising results in separating poisonous mushrooms from the edible. This is highlighted in a research paper conducted by Wibowo et al (2018), where techniques such as Naïve Bayes and Support Vector Machines (SVM) were used to train mushroom data that included over 8000 data points with 22 different features such as ‘cap-size’, ‘stalk-root’ and ‘gill-spacing’ [6]. The dataset used was the Audubon Society Field Guide to North American Mushrooms (1981) [7]. This dataset includes descriptions of hypothetical samples corresponding to species of mushrooms and is regarded as the most comprehensive and credible guide to mushrooms, making it suitable to be used to predict species and their edibility. Using the Naïve Bayes algorithm, a Supervised Learning technique, researches attained results of 95.88% with a RMSE (Root Mean Squared Error) of 0.1718. Another Supervised Learning technique was used in the form of a Support Vector Machine, where the researchers achieved a 100% accuracy of correctly classified instances. This shows that Traditional Supervised Learning techniques can conclusively detect poisonous mushrooms, and edible mushrooms in the given dataset with an extremely high level of accuracy.

Another study to highlight the effectiveness of Supervised Learning algorithms regarding mushroom classification was conducted by Chelliah et al (2018), who created all major traditional machine learning algorithms to compare them for inconsistencies, speed and accuracies to understand which learning models suit the problem to identify the mushrooms best [8]. After modelling the data on the data set, the same mushroom data set that Wibowo et al used in their study, they discovered that using a Decision Tree had the best performance, although both Decision Tree and SVM had 100% accuracy, the Tree beat out the SVM on speed. This comparison provides a greater insight into the functionality of Supervised Learning models and the level of classification required. This knowledge may be a starting point for approaching other real-world problems.

Time was taken to thoroughly research and identify more mushroom classification studies, though no papers found discussed the application of Deep Learning techniques.

Method B – Image Data

With the success of Traditional Machine Learning methods being evident for the stated problem, it brings into question how the more modern Deep Learning methods fare. Could an image recognition system trained on a dataset of mushrooms images successfully split the poisonous from the edible?

Techniques such as Neural Networks (NNs), and its subtypes such as Artificial NNs and Convolutional NNs have been proved to outperform these traditional methods in different studies. Such as the record-breaking study conducted by Krizhevsky et al. (2012), where researchers trained a Convolutional Neural Network to classify images into respective categories regarding cancer detection compared to traditional methods and achieved the best results with error rates of 37.5% and 17% respectively [9]. This was a shining moment which highlighted the efficacy of NNs when used in classification tasks. This again, begs the question on if these deep learning methods shine when being tasked to recognize and classify mushroom edibility from poisonous to edible.

When it comes to image recognition, there have also been successes to place mushrooms into categories of poisonous and edible. A paper written by Kang et al (2018) proposed a mushroom recognition system using a Convolutional Neural Network (CNN), with a dataset of 1478 mushroom images, with 38 mushroom species obtained by using web scraping to train the network [10]. Using the network, a high accuracy was achieved of 82.63% on their dataset, showing a high level of confidence that the model can be used to identify species of harmful mushrooms. This shows promise in the use of deep learning techniques, especially CNN’s, which are known for their prowess for image recognition, to be used when creating a system to recognize the toxicity of mushroom species.

Zahan et al (2021) conducted a study using image data containing 8190 images with a training and testing split of 80:20. Pre-processing techniques such as Histogram Equalization were used to ensure a high quality data set, which modify the contrast of an image so that its intensity histogram has a desired shape, which allows the model to recognize the contents of an image with more success [11]. To model the data, Deep Learning techniques such as a Residual Neural network (RNN) and Inceptionv3 (a CNN) were used to obtain high accuracies using the contrast-enhanced images created by the pre-processing techniques. The Adam optimizer was used for this study, with a SoftMax activation function.

The RNN only brought back an accuracy of 58% on the preprocessed mushroom images, a poor performance when compared to the CNN, which showed high performance, with 88% accuracy. This supports the work of Kang et al work with mushroom images, that using CNN’s shows great promise in the results it brings for mushroom edibility classification.

Another study into mushroom classification, conducted by Wang et al (2020) shows great promise with using combined models, instead of using one model at a time. They conducted a study focusing on recognizing the toxicity of different mushroom species using a deep multigrained cascade forest (dgcForest) using web scraped images [12]. This method of using dgcForest has a simple structure and does not need much training data. The model uses a structure where. the information of each layer is processed in an upper layer and the result is delivered to the next layer and so on. This dgcForest method recognized toxic mushrooms with a staggering 98% accuracy, enhancing the real possibility of using a deep learning system to classify mushroom species and help foresters keep themselves safe. The researchers found that by using this deep forest system that feature based learning and iterative classification has the best performance, and any unknown mushroom varieties can be identified quicker compared to the supervised algorithms they tried including Support Vector Machines and Logistic Regression.

As shown by the results of the studies above, using deep learning methods such as Neural Networks and combined models, and the utilization of large datasets can create highly accurate image recognition systems when classifying poisonous and edible mushrooms.

Conclusion

Through thorough research, many sources were analysed that had conducted extensive studies of using classification on tabular data, and recognition using image data. Using traditional supervised learning methods to classify mushrooms, the results received from those models resulted a good amount of accuracy and could potentially be used to classify mushroom species in real world applications. However, these studies were conducted using the same dataset, the Audubon Society Field Guide to North American Mushrooms, meaning their findings cannot be reliably applied and it cannot be stated that these methods can accurately identify and classify mushroom species because there is only one dataset that was tested from, and it could be argued that it is outdated, as it was published 40 years ago. When most studies conducting research into the same topic using the same dataset, wider research needs to be conducted into mushroom classification for the field to gain more reliability in their models.

Moving onto image classification methods and deep learning techniques is where the real promise shows if a system was to be applied to the real world, where those foresting should take a picture of a mushroom and a model would identify it. Although the studies showcasing these findings are and few between, only gathering two studies in total using image recognition instead of tabular data from all available research that could be found.

Using Convolutional Neural Networks and Multi-Grained Cascade Forests, high forms of accuracy were created, which shows real promise, as being a valuable system to create an image recognition system using deep learning methods like a convolutional neural network, or a multi-grained system would be most effective.

**Project Statement**

The plan is to find the most effective model to use to classify the mushroom species, and identify which features are more likely to be found with poisonous mushrooms.

The plan is to create a system, with an effective model to classify whether a mushroom is poisonous or not based on images taken from the internet, using a web scraper to scrape images of mushrooms from websites such as mushroom.world [13], Wikipedia, and Foraging Guide [14], which contain thousands of free to use mushroom images to train our model. Sites like mushroom.world contain extremely detailed data, including images of every species of existing mushroom in the world, and thus why the image data was chosen to be scraped from this website. This method of scraping free to use images was chosen as there is no existing, readily available trustworthy large database of mushroom images that were suitable to help solve our classification problem.

This project will be conducted through a series of tasks. After creating a Literature Review which thoroughly identified and discussed previous solutions of existing work, there was a better understanding of what would be helpful, moving forward to the Solution Design stage of the project. This stage is where diagrams and design of a solution will be created for our model and understanding of what exactly is needed for the dataset and requirements for the Baseline Implementation of our model.

After these technical details have been understood, progress can then be made towards the Baseline Implementation in which a first solution will be created and programmed in the language Python. This will include preparing a first version of the dataset, creating a pipeline, and for the model to be implemented in a way that enables for Iterative Development to be conducted easily, and effectively.

Moving forward to the Iterative Development stage, a larger stage when compared to the others as this where the system will be created into a more finished version of the model, and evaluations will be carried out on the pre-processing pipeline techniques, the model used for the system, and an evaluation of the parameters and tuning methods of the model. As an add-on to improve performance of the model and increase robustness of the proposed solution, extra data well be collected through the Additional Data stage. This stage will introduce additional datasets, with thorough analysis taking place to show how this data has impacted the performance of the chosen model, with the hope of this stage being to increase accuracy and decrease mis-classification rates of the system.

After the creation of the finished system, and the finishing solution being the most effective one possible, the final stage, the Solution Testing stage, will take place. This final task focuses on creating a detailed testing plan to ensure all facets of the system, including the dataset work correctly, and to a high standard, functioning without errors. After this has been performed, recommendations will be written to explain what could be improved or what is needed to be changed in the system so that it runs to the highest degree, and error free.

Finally, a presentation will take place where a run thorough of the system will take place remotely, exploring the final model performance, and if the proposed solution is reliable enough to be used outside of this study.

References

[1] NatureScot (2020). *More people spending time outdoors. New survey reveals importance of nature to Scotland’s recovery from Covid-19*. [online] NatureScot. Available at: https://www.nature.scot/more-people-spending-time-outdoors-new-survey-reveals-importance-nature-scotlands-recovery-covid-19.

[2] Lemmey, T. (2020). *Connection with nature in the UK during the COVID-19 lockdown*. [online] insight.cumbria.ac.uk. Available at: http://insight.cumbria.ac.uk/id/eprint/5639.

[3] Ko¨ppel, C. (1993). Clinical symptomatology and management of mushroom poisoning. *Toxicon*, 31(12), pp.1513–1540.

[4] Health Essentials from Cleveland Clinic. (2018). *Heads Up: Those Wild Mushrooms Growing in Your Backyard Could Be Toxic*. [online] Available at: <https://health.clevelandclinic.org/heads-up-those-wild-mushrooms-growing-in-your-backyard-might-be-toxic/>.

[5] Seattle, F.S.N. 1012 F.A.F.F. and Washington 98104-1008 (2019). *Thousands poisoned by mushrooms in France in recent years*. [online] Food Safety News. Available at: https://www.foodsafetynews.com/2019/12/thousands-poisoned-by-mushrooms-in-france-in-recent-years/.

[6] Wibowo, A., Rahayu, Y., Riyanto, A. and Hidayatulloh, T. (2018). Classification algorithm for edible mushroom identification. *2018 International Conference on Information and Communications Technology (ICOIACT)*.

[7] Lincoff, G. and National Audubon Society (1981). *National Audubon Society field guide to North American mushrooms*. New York: Knopf.

[8] Chelliah, B.J. (2018). Classification of Mushrooms using Supervised Learning Models. *EverScience*, [online] 6(4). Available at: <https://www.ijeter.everscience.org/Manuscripts/Volume-6/Issue-4/Vol-6-issue-4-M-45.pdf>].

[9] Krizhevsky A., Sutskever I., and Hinton G.E., “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jan. 2012, doi: 10.1145/3065386.

[10] Kang, E., Han, Y. and Oh, I.-S. (2018). Mushroom Image Recognition using Convolutional Neural Network and Transfer Learning. *KIISE Transactions on Computing Practices*, 24(1), pp.53–57.

[11] Zahan, N., Hasan, Z., Malek, A. and Sultana Reya, S. (2021). A Deep Learning-Based Approach for Edible, Inedible and Poisonous Mushroom Classification. *2021 International Conference on Information and Communication Technology*.

[12] Wang, Y., Du, J., Zhang, H. and Yang, X. (2020). Mushroom Toxicity Recognition Based on Multigrained Cascade Forest. *Scientific Programming*, 2020, pp.1–13.

[13] <https://www.mushroom.world>

[14] http://www.foragingguide.com