Fatal Fungi: Enhancing Mushroom Safety through Machine Learning Identification

Yasmine D. Subbagh

***Abstract*—This project aims to develop a web application that leverages machine learning (ML) to provide a reliable and accessible platform for mushroom foraging. The application integrates two advanced neural network models: a convolutional neural network (CNN) for species identification using images and a backpropagated perceptron neural network for edibility classification based on physical attributes. By combining these models, the platform offers a comprehensive solution to the challenges of accurate mushroom identification and classification, minimizing the risks associated with misidentification, which can have life-threatening consequences.**

The web application is designed to be user-friendly, scientifically accurate, and efficient, featuring a minimalist interface that allows users to upload images and physical characteristics of mushrooms for quick classification. To ensure high levels of accuracy and reliability, the ML models are evaluated using critical performance metrics such as precision, recall, and F1-score. The system is developed using modern web frameworks (Flask and React) and is deployed on AWS to ensure scalability, security, and seamless user experience.

The ultimate goal of this project is to democratize access to safe and effective mushroom foraging tools, empowering both recreational and professional harvesters to make informed, confident decisions. By reducing the need for expert knowledge and minimizing the risks of dangerous misidentification, the platform aims to bridge the gap between traditional foraging practices and modern technology, promoting sustainable and safe interactions with nature. The application also has the potential to be expanded in the future, offering additional features such as user accounts, model improvements, and mobile applications to further enhance its impact on the foraging community.

***Index Terms*—Backpropagated Perceptron Neural Network, Convolutional Neural Network, Foraging Fungi, Mushroom Hunting, Poisonous.**

I. Introduction

MUSHROOM hunting is becoming more popular as many are taking bigger considerations into where and their produce is getting to their dining room table, additionally considerations into if it is ethical or organic. While mushroom foraging can be grounding, allowing an individual to connect with their food pipeline, foraging fungi affords a large risk. Fungi can be deadly if the wrong species is consumed. However, there are thousands of species of fungi, this creates a difficult process for foragers to classify their harvestings. While many avoid this risk by only hunting for specific species, or turning to social media for classification help, many are missing out on the consumable potential of their crop.

There are approximately 14,000 species of mushrooms, but only around 2,200 are edible [2], in Figure 1 a variation of different mushroom species can be seen. That is means only about 16% of all mushroom species are edible, these are bad odds to risking fatal outcomes. A study done by Svanberg and Lindh found that since the 1990s, mushroom hunting and consumption has grown drastically [1]. This is a result of social shifts as well as the desire for forestland leisure activities (ie. hiking). Research done by Anusiya et al. determined that wild mushrooms are rich in antioxidants and antimicrobial [3], slating themselves to be highly desirable by those that value their nutritional and medicinal potential.

Figure 1: Collage of subset (Alaskan Natives) of Mushroom Species

With the rise in foraging popularity, amateurs have begun to go harvesting. However, they lack the expertise to safely identify fungi species. To overcome this hurdle, many choose to perfect their ability to identity a handful of species and only hunt for those; or they turn to social media (like Facebook or reddit) to enlist others to help them identify their crop. There are a few issues with relying on those from the internet to help classify fungi. The main issue again is the heavy decency on

This paper was submitted for review on December 9th, 2024.

Yasmine D. Subbagh is a graduate student in Master of Science in Computer Science and Software Engineering at the University of Washington – Bothell (UWB) is affiliated with UWB's Inclusivity, Diversity, Equity, and Accessibility (IDEA) Lab and worked previously at OneRadio Corporation. (email: ydns@uw.edu)

.

images of the mushroom; lighting, scale, and clarity can all lead to incorrect classifications. Another risk that is incurred is relying on strangers for what could be a life-or-death decision.

In order to ease the time consuming, expertise information driven, and risk inducing efforts of mushroom classifications, a thorough fungi classification application can aid in the heavy lifting. Machine learning (ML) has emerged as a well-rounded tool that can be used in all disciplines. By analyzing complex datasets, ML models can identify patterns and correlations that cannot be done at the same rate by harvesters. This is where the ML models can aid in providing insights on the decision and identification process. The classification app will utilize multiple machines leaning models in order to maximize its abilities in narrowing down the identification field.

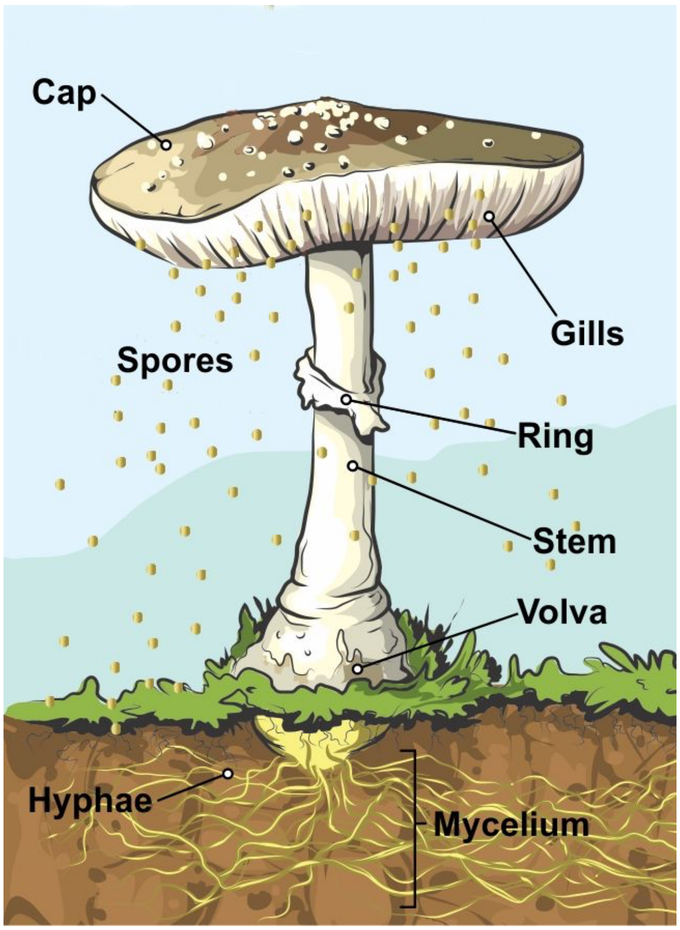

Figure 2: Labeled Anatomy of a Mushroom

The classification web application will be compromised of two main models. One, will identify the specific species of the fungi by input of images using a convolutional neural network. The second model will use the measured physical attributes of the mushroom and classify if it is edible or not by using a backpropagated perceptron neural network. These physical attributes are seen in Figure 2 above, different measurements of select organs of the mushroom anatomy will be used as the input for the model. This combination was used as a double line of defense to help quickly identify poisonous fungi and help harvesters narrow down the pool of species their fungi specifically is. Convolutional neural networks are high performers in classification using images, this is because they are able to self-identify trends and features within dataset images. Backpropagated neural networks are good for classifying quantitative data as they are able to again identify features and trends unsupervised and relearn from their own mistakes.

The projected final product of this project is to have a web application that is able to both identify fungi images as well as classify if it predicted to be edible or poisonous based off the mushroom physical measures attributes (ie. color, height, width, proportions). The current contributions include the mushroom attributes ML model.

II. Related Work

There are a few commercial applications that aim to classify and identify fungi as well as research that has been done on machine learning mushroom classification. The main commercial application that is used by recreational foragers is Seek by iNatrualist [4]. This application serves as an identification app for wildlife, plants, and fungi. While it wildly used, it relies purely on photos and the ML model to make classifications. The issue with just using photos is that lighting and scale can vary greatly from photo to photo and therefore can lead to inaccurate results.

Research conducted by Tongcham et al. focuses on the use of ML to automate the spawn quality classification done on commercial farms. Spawn classification is manually done and it's time consuming, classifiers are looking for defects caused by pathogens and mold. While this project uses ML for classification with fungi, it is classifying an entirely different output.

In a research project completed by Tutuncu et al. four machine learning algorithms were used to classify mushrooms based on physical features of the mushrooms [5]. While one these models was able achieve 100% accurate classification of edible vs poisonous fungi, this was done with an AdaBoost algorithm. The paper never goes on to create an application for user use but suggests that an automatic mushroom mobile application can be developed.

In a project completed by Rahman et al. proposes using machine learning and IoT to automate the classification of toxic mushrooms on a farm at the commercial level [6]. This shows the application that the models presented in this project can be applied to. Additionally, it tests many different type of ML models at classification accuracy, however it does not test a backpropagated perceptron neural network. And, while it does cover the projected use of the application to rely on images via a web cam, the data used in the models were physical features that must be measured manually.

This project is novel in that is provides a new modern web platform for fungi identification and edibility classification. There is no other project that combines these two models into a web application to create a simple and in-depth predicter. Datasets from Kaggle, WildFoodUK, and MushroomWorld will be used to have access to pre-processed and identified data, this data however will still need to be cleaned. The impact of this project is to create a product that is able to help mushroom harvesters be able to quickly rule out poisonous crop and help narrow down the species identification. Classification is done two machine learning models within the same application, a backpropagated perceptron and convolutional neural network. Using images and physical features to classify edibility and species type. Allowing users to spend less energy identifying species or weighing risks, and more time harvesting fungi.

III. Project Goals & Vision

This project takes into account the requirements as defined by stakeholders in order to develop a full featured product that meets their needs.

A. *Goal(s) of Completed Work*

The product that is to be completed by the proposed work done in this project is a web application and/or mobile application that allows users to upload pictures and physical attributes of a mushroom in order to classify its edibility and predicted species identity. The outline in this report will focus on the web application over mobile development.

The modern web platform will use a Flask model-view-controller architecture. This architecture was chosen as it's easy to develop and maintain [8]. Different portions of the system can be modified (even overhauled) with little modification needed to the other components in the system to accommodate the updates. Flask is a lightweight and flexible web framework for Python [9]. Using this architecture/framework will allow users to upload mushroom data and quickly receive predictions on edibility and species identification. Seen below in Figure 3 is a diagram of the system architecture.

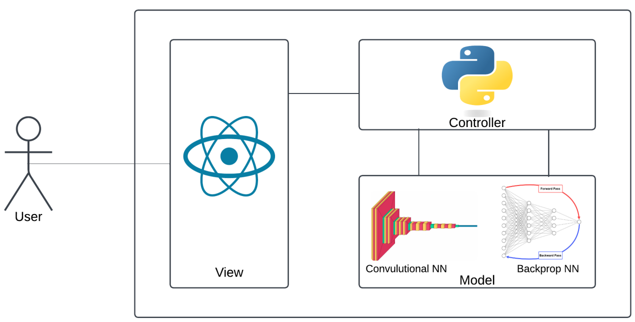

Figure 3: Web Application System Architecture

Additionally, Flask was chosen as the framework for the model as it is built upon Python. Python will be the workhorse coding language for this project as it supports module and many packages that will be required for building the ML models and application. While the deep learning (DL) neural network algorithms will be implemented within the code base of this project in order to minimize computation power and times, many other libraries will be used for data processing, extrapolation, and visualization. The front end (view) of the application will be in React as it is a component-based architecture. It has become an industry standard because of its reusability, maintainability, and cross-platform ability. As development increases to mobile applications, translation will be simple with React Native.

While the initial training process of the DL models will be done locally to reduce computational costs on the cloud, production of the application will be hosted on Amazon Web Services (AWS) public cloud infrastructure. Using Elastic Beanstalk (EBS) will allow for automatic scaling of the application as more traffic influxes in.

The Python libraries that are to be used within the back end of the application and DL models are: KaggleHub, NumPy, Pandas, Matplotlib, Seaborn, OS, and Scikit-learn. KaggleHub is being used to import the data from Kaggle directly, and OS is being used to effectively retrieve the data into the codebase. The Pandas library is used to handle the data while its being cleaned during the data processing step as well as being passed through the DL models. Both Seaborn and Matplotlib is used by developers and data scientist as visualization tools. Visualizing the data before and after training of the ML models allows for better understanding on the data as well to depict how well the models are performing. NumPy and Scikit-learn libraries are both used for the implementation of the neural networks; NumPy allows for the use and data structures (vectors and matrixes) of linear algebra in Python, and Scikit-learn has tools to aid in the ML process (as in preprocessing).

B. *Identification of Problem and Opportunity*

The problem that this project aims to solve is the inaccessibility of fungi identification, it is a critical challenge for both amateur and professional harvesters. The abundance of fungi species, with only a small fraction being edible, create a significant barrier to safe and efficient harvesting. The process often relies on expertise, careful manual identification, or social media advice; each of these methods has their limitations.

These challenges present an opportunity to leverage modern technology to reduce the reliance on traditional methods and improve the accessibility to accurate identification tools. Machine learning, with its ability to analyze large datasets and identify features in complex patterns, offers a gateway to an automated identification solution. By combing ML with a user-friendly application, mushroom harvesters can quickly and confidently identify species.

C. *Stakeholders and Beneficiaries of Research*

There are several stakeholder groups for this project's product. The key being the core user group: recreational mushroom harvesters. This group lacks the expertise to efficiently or even effectively identify mushroom species and would benefit from a product that is able to accurately identify species and classify if their harvestings are edible or not. Serving this group will be able to minimize their reliance on other stakeholders and lower the accessibility barrier to harvest fungi.

The next key stakeholder is experts within the domain, these groups expel a lot of time, energy, and funding in order to help newcomers stay safe when harvesting. Groups like the Puget Sound Mycological Society (PSMS) have class and workshops in addition to identification clinics [10] in order to educate and protect the community: however, their energy and funding could be spent on other projects if this burden was lifted. Other experts include ecological researchers, if given access to the photo submissions uploaded by users, researchers have a much larger reach. This could give them the possibility to discover previously unidentified mushrooms.

The final stakeholder is the developers and team members of those who intend on implementing the application. The developers will gain (if not previously exposer to) insights and the ability to data science, ML engineering, and full-stack development. Data science methodologies will need to be applied when pre-processing and understanding the dataset. Machine learning engineering techniques and understanding will be gained during the development, training, and accuracy validation of the neural networks. Lastly, full-stack development and cloud engineering skills and knowledge will be a necessity to build the user accessed web application and deployment of the product onto the cloud service.

IV. Completed Project criteria

There are several criteria that the final product that is outlined within this paper must meet to be defined as successful and is solving the problem that was outlined in previous sections.

A. Level of Success

Three levels of success for the project exist: minimum, expected, and aspirational. As resources are allotted to the project, higher levels of success can be obtained. But below, each levels criterion is defined.

The minimum viable product (MVP) will require a basic static web application that is capable of letting users upload physical feature values in order to determine if it inedible or edible with risk (to reduce trust of the system as incorrect outcomes can be fatal). A singular model rather than the two that were outlined earlier, will allow for iterations and initial testing of the system with user use. The backpropagated perceptron neural network will need to be trained and tested with interpretable and high-achieving results including precision, accuracy, recall, and F1-score. These scores will be displayed to the user to allow users make educated and conservative decisions when consuming harvested fungi.

The expected level of success for the product includes everything outlined above in the MVP and more. The user interface (UI) for the web application should be more refined to appease the user's interests while still maintaining minimalism. The platform will implement both neural networks, the backpropagated and convolutional networks. The user will now also be able to upload images on the web app that will automatically identify species. The ML models will again need to be trained and tested with interpretable and high-achieving results including precision, accuracy, recall, and F1-score. These scores will be displayed to the user to allow users make educated and conservative decisions when consuming harvested fungi, comparing results for a better landscape.

An aspirational level of success includes everything from the expected level of success and more. A goal for future development is creating accounts for users to be able to save their fungi classifications in order to compare results to previous predictions. Another reach goal is including more models that can give the user a larger landscape of classifications (or verify one another), models like AdaBoost utilizes a mix of models into one and has previously shown itself as a high performer. With donations or volunteers, having mycologist identify (hand label data) user data that can be used to retrain the model for better accuracy, will allow for less speculation from the community and create a more accurate platform. With additional help from mycologists, bios and guides can be crafted for the application to help harvesters better understand results in addition to their own crop. Lastly, creating accompanying mobile applications (iOS and Android)

B. Quality and Associated Measurement Metrics

There are two main metric groups to consider for this project, accuracy and performance of the ML models and the user experience of the applications.

The ML models need to be able to perform at a classification that can be trusted by users. In order to determine if the models are performing at a rate that is acceptable, accuracy and overfitting will be analyzed. To ensure that overfitting is not occurring, visuals will be created that need to demonstrate that the training and testing error loss continue to decrease. Other ways to ensure there is no overfitting is by checking that the testing loss is never more than the training loss, and lastly that the models are able to predict a decent accuracy. Accuracy of the models should be close to perfect; this can be analyzed using classification accuracy, precision, recall, and F1-score. These metrics for the pre-trained backpropagated perceptron NN model can be seen below in Figure 4.

```
Test Accuracy: 0.9990

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      6820
           1       1.00      1.00      1.00      5365

    accuracy                           1.00     12185
   macro avg       1.00      1.00      1.00     12185
weighted avg       1.00      1.00      1.00     12185
```

Figure 4: Accuracy and Performance Metrics for Trained Backpropagated Perceptron Neural Network Model

The application is useless if no one uses it. To ensure that the application is meeting user needs, user-centered research will be conducted by the team to iterate on their findings to create an app that meets the users' needs. Other ways to validate that the app is succeeding is collecting qualitative metrics on incoming user traffic, traffic origin, screen time, task completion time, user clicks, and coded user comments/reviews.

C. Targets

The targets for this project are defined in terms of measurable outcome, focusing on the applications core functionalities, usability, and broader impact. The ML model accuracies should hit high on key performance metrics. The precision rate of the backprop NN needs to be >= 0.95 (for both positive and negative) while the convolutional NN only needs to perform at >= 0.80, we can aim lower with the imagery classification. The recall scores for the backprop model needs to be >= 0.98, this is because a false negative (classified edible when not) can lead to users making dangerous decisions. As we can see from above however, the pre-trained backprop model is performing well above these baseline metrics. To ensure that these metrics hold truth behind the numbers, the dataset that is supporting the backprop model needs to be above 70,000 records, and the convolutional NN needs at least 42,000 records. This is to ensure that each species is tracked at least once if not multiple times and allow for a more complex feature extraction.

The application has must be capable of certain tasks too. It must be able to allow for seamless upload and processing of the images for species classification, ideally it would be able to do so within five seconds. And the physical attribute-based backprop model should be able to input, process, and output a result within three seconds. The application should be able to handle errors and direct users to solutions.

V. Positioning of Project

This capstone research project focuses solely on creating a new modern platform for running DL models and mushroom identification. It does not focus on other types of plants (like flowers or trees) nor is it able to classify animal species. Existing approaches, as highlighted in the existing work section, focus purely on the physical attributes of the mushroom to determine edibility and lack giving users knowledge by identifying the species. Making them impractical and hard to trust and rely on.

This project introduces machine learning approach to toxicology which aims to offer an accessible and reliable alternative to traditional identification process. The project's outcomes could revolutionize the ecology sector for foragers and researchers by providing an automated tool.

VI. Constraints, risks, and resources

Due to the nature of the application at hand, there are physical risks to the users. In addition to the constraints of the application during research, development, and production.

A. Key Constraints

The number one constraint of the project is the ability to acquire quality data, a ML is only as good as its data. The effectiveness and accuracy to the real world of the ML models heavily depends on the quality of the data, and enough variation in the data to represent the landscape of fungi. While the data from Kaggle, WildFoodUk, and MushroomWorld allows for a comprehensive representation of mushroom species, we are relying on the labeling of the data to be accurate.

The computational power to train, test, and continuously deploy the ML models can be very costly. Deep learning is very intensive and can put a large strain, in addition to time, on the hardware. GPUs may not always be available to the developers and cloud computing is far too expensive. Additionally, the model training needs to be validated for accuracy as well as overfitting, if the models are overfitting, they will not be able to perform once feed user data.

The nature of the application of the product is sensitive. User's must use the information presented to them by the platform with a grain of salt and make smart decisions. If users become too reliant on the classification defined by the ML models, they could potentially make poor decision than can become dangerous.

B. Resources Needed for Success

There are several aspects that need to be met to achieve a successful platform. Obtaining quality data is the highest priority to the platform's success. Gaining access, collecting, and pre-processing the data is imperative to performance of the ML models. Having an efficient development process and pipeline will maximize the energy and time spent on the project. Using the correct development methodology paired with tools that can simplify the implementation costs. Monetary support is also key in the longevity of the application, the lower the cloud services costs, the longer the application can be hosted. Lastly, Feedback and guidance from users as well as mycologist is key to a better understanding of industry norms and users' needs and feedback. By taking in feedback from the mushroom harvesting community, the product can be iterated upon to better achieve the goals of professionals and amateurs alike.

C. Anticipated Risks

The highest priority risk is the user becoming too reliant on the service and making unsafe decisions that can harm themselves. Experts in the area are wary if the ability of ML models that automate the identification process [11]. Not all species can be identified just by the look of the fungi, some require other senses by experts to accurately identity the fungi, something that currently cannot be replicated using the ML models. This could lack of information to the model could cause it to incorrectly identify a mushroom that can lead to a fatal result; one eight of a death cap kill you.

Cost is a large risk for this project, as cloud services are not cheap and can quickly rack up a large bill. Since the production level application would like to be deployed on AWS elastic beanstalk in order to allow automatic scaling as the application grows, remaining on AWS Free Tier is impossible. The project management may not be able to sustain paying the cloud services fees if they grow too large, this puts the application use at risk.

Other risks include the development of the platform, including the ability to obtain the data as well as the implementation of the applications. If not enough data is collected, or worse if the data is biased or skewed, the algorithms will not be able to build models that are accurate reflections of reality or that are able to correctly classify new user data. Additionally, if the developers are unable to build the models accurately or design, implement, integrate, and deploy the application, use will cease to exist. Having competent and experienced developers is key.

VII. Conclusion

Mushroom foraging is an activity that connects individuals to their food sources and provides opportunities to explore nature while enjoying its bounty. However, the inherent risks of misidentifying fungi, many of which can be poisonous, present significant barriers for amateurs and experts alike. Traditional methods of identification—through expertise, reliance on social media, or by focusing on a few well-known species—are often unreliable, time-consuming, and restrictive. This project proposes a novel approach by leveraging machine learning (ML) to address these challenges and create a robust fungi identification and edibility classification platform.

The web application outlined combines two powerful neural network models: a convolutional neural network (CNN) for species identification using images and a backpropagated perceptron neural network for edibility classification based on physical attributes. By integrating these technologies, the platform ensures a comprehensive, user-friendly, and scientifically accurate tool for mushroom foragers. This double-layered approach enhances user confidence and mitigates the life-threatening risks associated with fungi misclassification.

Through careful development and rigorous testing, the application aims to meet high standards of accuracy and reliability, as measured by performance metrics like precision, recall, and F1-score. Furthermore, the integration of modern web frameworks like Flask and React, along with deployment on AWS, ensures scalability, accessibility, and ease of use.

The ultimate goal of this project is to democratize access to safe and efficient mushroom foraging tools, empowering both recreational and professional harvesters. By reducing reliance on expertise and minimizing risks, the platform bridges the gap between technology and traditional foraging practices, encouraging sustainable and enjoyable interactions with nature.

VIII. Future Work

While this project provides a promising foundation for mushroom identification and edibility classification using machine learning, there are several areas for potential expansion and enhancement. In future iterations, we aim to refine and extend the platform to offer even more accurate, reliable, and user-friendly solutions for mushroom foragers. The project road map, including future work, can be seen in Figure 5.

While machine learning models can significantly aid in mushroom identification, human expertise is still essential for ensuring the accuracy of species classification and edibility assessment. To further enhance the platform's reliability, we plan to integrate expert feedback through partnerships with mycologists. By leveraging expert guidance, we can continually refine the models, correct misclassifications, and ensure the platform provides more accurate and trusted information. Experts could also help in annotating and validating user-submitted data, which would allow the platform to evolve in real time.

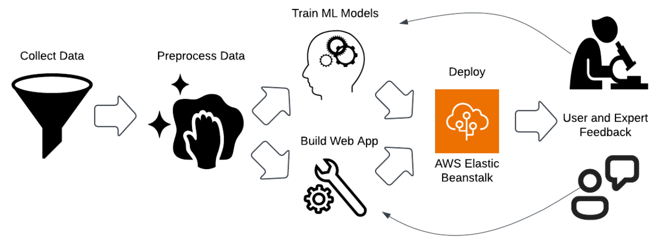

Figure 5: Project Road Map

To enhance user experience, future iterations of the platform could introduce user accounts, allowing users to save their identification results, track foraging history, and receive personalized recommendations. For example, users could receive notifications about the edibility of specific mushrooms they've encountered before, or the platform could provide suggestions for safe foraging practices based on user data. This personalization will also enable a community-driven approach, where users can share their experiences and feedback, further enriching the platform.

References

[1] K. C. Semwal, S. L. Stephenson, and A. Husen, *Wild Mushrooms and Health: Diversity, Phytochemistry, Medicinal Benefits, and Cultivation*. CRC Press, 2023.
[2] I. Svanberg and H. Lindh, "Mushroom hunting and consumption in twenty-first century post-industrial Sweden," *J Ethnobiology Ethnomedicine*, vol. 15, no. 1, p. 42, Aug. 2019, doi: 10.1186/s13002-019-0318-z.
[3] G. Anusiya *et al.*, "A review of the therapeutic and biological effects of edible and wild mushrooms," *Bioengineered*, vol. 12, no. 2, pp. 11239–11268, doi: 10.1080/21655979.2021.2001183.
[4] "Seek by iNaturalist · iNaturalist," iNaturalist. Accessed: Dec. 06, 2024. [Online]. Available: https://www.inaturalist.org/pages/seek_app
[5] "Edible and Poisonous Mushrooms Classification by Machine Learning Algorithms | IEEE Conference Publication | IEEE Xplore." Accessed: Dec. 06, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9797212
[6] H. Rahman *et al.*, "IoT enabled mushroom farm automation with Machine Learning to classify toxic mushrooms in Bangladesh," *Journal of Agriculture and Food Research*, vol. 7, p. 100267, Mar. 2022, doi: 10.1016/j.jafr.2021.100267.
[7] P. Tongcham, P. Supa, P. Pornwongthong, and P. Prasitmeeboon, "Mushroom spawn quality classification with machine learning," *Computers and Electronics in Agriculture*, vol. 179, p. 105865, Dec. 2020, doi: 10.1016/j.compag.2020.105865.
[8] A. Leff and J. T. Rayfield, "Web-application development using the Model/View/Controller design pattern," in *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference*, Sep. 2001, pp. 118–127. doi: 10.1109/EDOC.2001.950428.
[9] "Introduction to Web development using Flask - GeeksforGeeks." Accessed: Dec. 07, 2024. [Online]. Available: https://www.geeksforgeeks.org/python-introduction-to-web-development-using-flask/
[10] "Puget Sound Mycological Society | About PSMS | History & Mission." Accessed: Dec. 08, 2024. [Online]. Available: https://www.psms.org/id-clinics.php

[11] J. Vincent, "A 'potentially deadly' mushroom-identifying app highlights the dangers of bad AI," The Verge. Accessed: Dec. 08, 2024. [Online]. Available: https://www.theverge.com/2017/7/28/16054834/mushroom-identifying-app-machine-vision-ai-dangerous