Advances in Artificial Intelligence and Machine Learning; Research (Ahead of Print) Received 03-03-2026; Accepted 22-05-2026; Published online 29-05-2026

Deep Learning Models for Microscopic Fungal Identification in Spacecraft Applications

Wenzhen Fu fuwe0007@e.ntu.edu.sg
School of Mechanical & Aerospace Engineering,
Nanyang Technological University,
Singapore 639798, Singapore.

Yujian Fu yujian.fu@aamu.edu
College of Engineering, Technology Physical Sciences,
Alabama AM University,
Huntsville, Alabama, 35762, USA

Jeanette Jones jeanette.jones@aamu.edu
Department of Biological Sciences
College of Agricultural, Life and Natural Science
Alabama A&M University, Normal, USA 35762

Zhijiang Dong zhijiang.dong@mtsu.edu
Department of Computer Science,
Middle Tennessee State University,
Tennessee, United States.

Corresponding Author: Yujian Fu

Abstract

Fungal contamination on space vessels can impose great risk to the health of astronauts and the durability of the vessels. The manual microscopic classification is both time-consuming and non-feasible in space missions, and thus it is essential to automate the process. The study involves the model development of three deep learning models, namely the baseline CNN, EfficientNetB0, and MobileNetV2, for the classification of five fungal species using a secondary DeFungi microscopic image database. After image augmentation and common preprocessing, each model was tested in terms of precision, recall, F1-score, and ROC-AUC. Findings indicate that the MobileNetV2 model performed better than the other models, which registered 67% test set accuracy and excellent performance across most classes. It is compact and lightweight and can make onboard fungal monitoring easy in resource-limited settings. In this study, the research is used to indicate how to achieve autonomous bio-surveillance advancement through the use of deep learning in long-duration space missions.

Keywords: Fungal Classification, Deep Learning, CNN, MobileNetV2, EfficientNetB0, Microscopic Image Analysis.

Citation: Wenzhen Fu, et al. Deep Learning Models for Microscopic Fungal Identification in Spacecraft Applications. **Advances in Artificial Intelligence and Machine Learning**. 2026. (Ahead of Print) https://dx.doi.org/10.54364/AAIML.2026.63308

1. INTRODUCTION

1.1 Background

Spread of fungi in closed systems such as space vehicles is a major threat to both human beings and the systems. Different harmful fungi like Candida albicans, Aspergillus niger, Trichophyton rubrum, and some others thrive well in closed loops that have regulated humidity and recycled air and have the ability to result in infections or weakening of materials [1]. The current forms of positive diagnosis and classification of fungal infections by microscope and manual identification take up a lot of time and require expert interpretation, and there is an urgency to find an automated and precise solution. Recent developments of deep learning and, specifically, Convolutional Neural Networks (CNNs) have been shown to be largely successful in visual-based medical diagnosis, including fungal diagnosis – when the number of data records is limited, then also CNNs can perform adequately to accurately detect most of the fungal infections from the microscopic image at the cellular level, as observed by Zieliñski and some other researchers, as presented in their paper published in the year 2020. Their research established the fact that feature engineering plays a critical role in the accuracy of models, and it is more effective than traditional image processing methods. Similarly, Rahman and some fellow researchers published their research in 2023 to present on how AI-based frameworks can effectively differentiate among the types of fungal infections by utilising histopathology pictures that establish the high precision of modern deep learning models. Both studies emphasise that advanced AI-based methods are necessary to identify microscopic infections in sensitive and resource-constrained areas like spacecraft missions, where time and accuracy are equally important. Continuing with these works, the current project has examined the classification of fungi with the DeFungi dataset [1]. It integrates a baseline CNN with more advanced models, which are pretrained, such as EfficientNet-B0 and MobileNetV2, to enhance the robustness and accuracy of the model on five fungal species. Incorporation of such AI-based fungus identification systems into space exploration may mark the path to autonomous bio-monitoring that can eliminate much of the dependence on Earth-based diagnostic capabilities and provide greater safety in this long-duration space exploration.

The existence and continuity of fungal contamination in spacecraft-related spaces were also confirmed by recent studies. As an example, Chander et al. (2022) [2], provided the case of how the new strains of fungi with specific morphological and genomic features were discovered in NASA spacecraft assembly facilities during the Mars 2020 mission, proving their survival in harsh cleanroom conditions and possible radiation resistance. Likewise, in the study of Regberg et al. (2020) [3], OSIRIS-REx assembly, test, and launch environment microbial monitoring demonstrated that different fungal communities were present on cleanroom surfaces, which points to the constant risk of contamination by microorganisms despite controlled conditions. The results highlight the need to ensure the spacecrafts have effective and automated systems to detect the fungal presence. Fungal classification via deep learning has been studied in other studies of the earth's biomedical setting, where it has shown excellent results in the identification of microscopic fungal structures based on histopathological images [4, 5]. Nevertheless, such strategies are mostly formulated and tested in terrestrial laboratory environments and fail to consider the environmental limitations that are peculiar to spacecraft systems, including microgravity, limited computing capabilities, and time constraints. This gap is an indicator of the necessity to consider lightweight and deployable deep learning architectures with respect to space-relevant bio-surveillance applications.

1.2 Aim and Research Objectives

Aim:

The goal of this project is to evaluate the performance of state-of-the-art deep learning models in fungal contamination detection from microscopic cellular images as captured in a spacecraft laboratory environment.

Objectives:

According to the aim of the research, the objectives are formulated as:

1. To observe the performance of a baseline Convolutional Neural Network (CNN) to classify fungal species from the microscopic cellular images.
2. To implement fine-tuned pretrained models (EfficientNet-B0 and MobileNetV2) for the same fungal classification task.
3. To compare the performance of CNN and the pre-trained classifiers on a subset of input images based on suitable evaluation metrics, and to conclude about their effectiveness for microscopic fungal classification in spacecraft environments.

1.3 Research Significance

This study aims to make a substantial contribution to the enhancement of health and safety practices during long-duration space missions. Due to the microbial corrosion that can weaken the technologies on board the space ships, as well as endangering the health of the crew members with infections and allergies, the proliferation of fungi in the spacecraft environment is a severe challenge. Conventional methods of fungal screening (like manual microscopy) cannot be used in space because of resource shortages, time, and the need to arrive at fast answers through decisions of personnel involved. This specific project will enable automatic monitoring of fungi, especially at the microscopic level, through the implementation of deep learning models to identify microscopic fungal species. This strategy, based on both baseline CNNs and potent pretrained models such as EfficientNetB0 and MobileNetV2, ensures high levels of accuracy in determining and identifying various fungal species. This promotes early screening and containment activities that minimise the health risks and breakdown of the system. Furthermore, it is possible to reduce the reliance on Earth-based analysis by deploying AI-based classification tools on board, which could be of high priority in future space exploration missions to the Moon, Mars, or other planets. The research is also relevant to the larger body of work on biomedical imaging, on how AI could be used to aid bio-surveillance in extreme, resource-constrained contexts. In general, this project contributes to making space habitats wiser and self-governing in terms of health monitoring solutions.

1.4 Structure of Study

The study is coherently structured into different sections and subsections that start with the background of fungal classification at the microscopic level in the outer-earth environment, with a brief description about the aim, objectives, and significance of the study as presented earlier. Then, a detailed review of literature on the seriousness of fungal contamination, fungal detection methods, state-of-the-art algorithms, and their evaluation metrics is presented with reference to previous reputed research. In the methodology section, the applied preprocessing, modelling, and evaluation process are described in detail, while the results section presents observed performances of the applied models in fungal classifications and interprets their performances. Finally, the conclusion section summarises the entire study with probable limitations and highlights the possible directions for improvement of results in the future.

2. LITERATURE REVIEW

2.1 Fungal Contamination in Spacecraft Environments

It is common knowledge that fungi are a major threat to the safety and soundness of the environmental conditions of spacecraft because of their plasticity and their capabilities to form spores and undergo resistance to radiation and extreme environments. Fungal contamination of equipment and personnel health can occur in the space mission environment, which is resource-limited and highly confined in nature. They are able to deteriorate the spaceship materials, such as metal parts, insulation, and polymers, through biodeterioration reactions [6]. Additionally, some fungal species have mycotoxin generation potential, which is an immunological and respiratory threat to astronauts in long-term missions. Simões and Antunes (2021) [6], also stress that microbial growth in space habitats is significantly different compared with the growth on Earth because of modified microgravity, radiation, and closed-loop environments. Such conditions have been known to increase microbial virulence and biofilm formation, especially in opportunistic pathogens such as fungi. It has been revealed that space-associated stressors like microgravity can activate fungal biofilm growth, which is known to increase tolerance to cleaners and antibiotics. In turn, fungal surveillance becomes an important attribute of microbial surveillance within spacecraft. To emphasise this point further, Simões et al. (2023) [7], propose a new research area called ‘astromycology’, which refers to the study of fungi in space. According to their findings, fungi can not only survive in a spacecraft but can also adapt both metabolically and structurally in space. The Aspergillus and Penicillium extremotolerant species have already been retrieved from the International Space Station (ISS). Such tough microorganisms are able to colonise surfaces, such as in water systems or air filters, which makes in-time classification and early detection of fungi items important. With the possibilities of health risks and damage to the buildings, the issue of automated fungal monitoring systems needs to be addressed urgently. It can be described that deep learning-based image classification solutions proven in high-resolution fungal datasets may provide fast, non-invasive diagnostics, allowing improved fungal control initiatives during long-term space exploration.

2.2 Microscopic Fungal Classification

Microscopic fungal classification is known as the identification of fungal species on the basis of magnified images of bio-cells. This is an essential activity in the diagnosis of the infection and/or directing antifungal therapy; however, it has classically been dependent on human-based experience, which is subjective and variable in practice. As digital pathology has become mature, there has been a growing use of image processing and deep learning to automate and make more reliable microscopic classification. Kristensen and some fellow researchers achieved this with their study published in 2022 [8], which created image processing methods and machine learning algorithms to perform a classification of microscopic images of Gram stain. As much as their experiment targeted bacteria, the process is very relevant when it comes to the identification of fungi. The analysis indicated the significance of preprocessing procedures like stain enhancement and morphological feature extraction prior to the classification, which is also very essential in the analysis of fungus images. More recently, in the mycological application, Koo et al. (2021) [9], suggested a regional convolutional neural network (RCNN) to identify the superficial fungal infection on microscopic images. Their model was able to detect fungal hyphae and fungal spores precisely and in larger amounts than conventional techniques, and worked faster as well. The researchers pointed out that region-based CNNs will enhance the detection of fine, minuscule structures that characterise fungal infections. As a combination, these studies explain how deep learning, particularly CNN-based models, has revolutionised microscopic classification with the improvement of objectivity and scalability. With respect to fungal diagnostics, this change is especially applicable because of the morphological resemblance between various species and the need to have fast and accurate identifications, either in the clinical or environmental settings. The effectiveness of these techniques will indicate that the future use of other CNN versions (like EfficientNet and MobileNet) could remarkably improve the classification accuracy of fungal datasets under unfavourable conditions, such as space vehicles.

2.3 Deep Learning Based Medical Image Classification

The application of deep learning in medical image classification has transformed the industry to allow high-throughput, automated, and definitive diagnosis frameworks with higher precision levels than conventional methods of image analysis. Particularly, Convolutional Neural Networks (CNNs) showed outstanding results when it comes to extracting hierarchical features directly out of medical images, supplementing the necessity of human intervention. These models can identify low-level patterns in the imaging data, allowing for early and accurate diagnosis of complicated disorders. As Liu et al. (2021a) [10], stress, deep-learning-based strategies for segmentation, such as U-Net or its variations, have grown to become essential components of medical image analysis, with their abilities to learn in an end-to-end manner and higher accuracy in demarcating anatomical structures. The models have been broadly used in areas such as radiology, histopathology, and dermatology. Their power lies in their ability to combine context-sensitive feature extraction and maintain the spatial resolutions of medical images to such fine detail. Going into greater detail, Liu et al. (2021b) [11], pointed out the recent development involving hybrid approaches to deep learning where deep learning and attention mechanisms, residual learning, and transfer learning are combined to improve the classification accuracy and the ability to generalise among various structured and unstructured datasets. The gradual rise in access to annotated medical data sets and their scalable training on cloud platforms has also led to the current adherence of these approaches. Integrating area expertise involves domain-specialised knowledge into deep learning structures, which can enable more reliable and explainable models that raise the confidence of AI-assisted diagnostics. Where the identification of fungi is at hand, these developments will serve as a solid background when it comes to implementing them with the models that have the capacity to identify the complex microscopic patterns characteristic of the genus of fungi. With the further progress of the field of deep learning, the prospects of performing precise fungal classification in practice are high, not only on the use cases level, but also on the deployment level, offering true real-time performance and expected usage in a restricted resource landscape such as the elite spacecraft, where a manual classification is not a viable option. The use of such models would be an excellent way of enhancing the biosecurity and early detection of microbial contamination in long-duration space missions.

2.4 Image Classification by CNN

The basis of convolutional neural networks (CNNs) is that, because of the ability to automatically and efficiently learn to extract hierarchical features of images, they have become key contributors to image classification tasks. Contrary to other machine learning algorithms that use conventional handcrafted feature-based processing, CNNs use stacks of convolution, pooling, and non-linear activation layers to discover spatial representations automatically in raw pixels. This has turned out to make CNNs the model of preference in complex image classification, particularly in the fields of medical and scientific images. CNNs are, at their core, specifically developed to behave like the human brain when it comes to the processing of images, being able to pick up edges, textures, shapes, and patterns at various levels of abstraction [12]. The hierarchical architecture removes much of the need for feature selection and ensures robustness against inherent variation, including orientation, scale, and lighting. Another attribute of CNNs is that they reduce parameter count in comparison to an old dense network, so they are effective in utilising high-dimensional data such as images. According to Chen et al. (2021) [13], CNN-based image classifiers are stronger in accuracy and scalability compared with the classical models. They provide a review that describes the design of CNN architectures in a way that they have become deeper and more efficient. All these advancements have extended the application of CNN in fields such as facial recognition, medical diagnostics, and remote sensing, where accuracy is crucial. CNNs have also demonstrated outstanding results when combined with transfer learning to make use of models trained on other datasets, such as ImageNet (which uses large quantities of labelled data), to fit to a particular task given minimal labelled data. Within the scope of the classification of fungi in the conditions of spacecraft facilities, CNNs will be a stable solution to conduct automation of identification based on microscopic images. They are especially appropriate in the field of biomedical image classification, where precision and efficiency are of utmost importance because they can pick up even the slightest differences in detail between fungal species.

2.5 Transfer Learning for Image Classification

Transfer learning has changed the face of deep learning in image classification, particularly in areas where the available labelled data is very sparse, like medical imaging. This method consists of adopting models that are already trained (firstly on large databases such as ImageNet) and adapting them to specific tasks, either by fine-tuning the layers or using the model as an extractor of features. It saves a lot of training time, is computationally cheap, and is less reliant on copious labelled data. Regarding medical image classification, the transfer learning option has proved to have powerful results in a broad range of tasks, such as diagnostics and segmentation. According to Kim et al. (2022) [14], models such as VGG, ResNet, and Inception, trained on general image data, were successfully used in the medical application area and gave the correct classification under domain-specific limitations as well. Their review reveals that transfer learning not only increases the performance metrics but also helps to avoid overfitting, which is common in small medical datasets. Probably the most promising architecture in this field is the EfficientNetB0, a lightweight CNN model with high accuracy and fewer parameters. It uses a compound scaling algorithm whereby there is a trade-off between network depth, width, and the resolution of the input, leading to optimal efficiency. Mahmood et al. (2025) [15], proved that EfficientNetB0 was useful in their ECG image classification hybrid deep learning model. They used EfficientNetB0 in a hybrid approach together with Fuzzy C-Means clustering to have an enhanced accuracy and noise resistance. The model exploited the capability of EfficientNetB0 to extract hierarchical features, of which it demonstrated its value in complex classification tasks that have ambiguous or overlapping classes. EfficientNetB0 is especially good as a fungal classification inputted by microscopic images in the space conditions, as it achieves excellent results in low-resource, high-variability pictures. Its remarkable generalisation capacity on limited data makes it a very good choice for modelling the biological risk of contamination in spacecraft, where there is little labelled data available but where the opportunity to generalise well is vital.

2.6 Advantages of Lightweight Models for Image Classification

Lightweight models are very useful in the setting of image classification, and specifically in cases where the classification model must run in real-time or on highly constrained systems, such as mobile or embedded systems. These are high-performance and low-memory models that also run on low-computation-cost environments. Consequently, there is a growing demand for resource-light deep learning models that can deliver accurate predictions without straining the hardware. MobileNetV2 is one of the state-of-the-art lightweight models that can be distinguished by its efficient architecture and lower computation complexity. It adds depth-wise separable convolutions, which factorise ordinary convolutions into two simpler tasks. The first is the depth-wise convolutions (using one filter per input channel), and the second task is pointwise convolutions (1×1 convolutions to combine the results). This architectural change cuts the number of parameters and floating-point operations by many magnitudes without harming classification performance. MobileNetV2 also uses an inverted residual structure and linear bottlenecks, which provide another optimisation of memory and inference speed. As illustrated through Gulzar (2023) [16], fruit image classification was done on a MobileNetV2-based TL-MobileNetV2 architecture, delivering a better result against the backdrop of other heavier models, namely, AlexNet, VGG16, and ResNet. The paper pointed out that TL-MobileNetV2 reached an accuracy of 99%, but it is still appropriate to deploy it even on mobile devices, which demonstrates the power of lightweight architecture in precision agriculture. On the same note, Cui et al. (2021) [17], proposed LiteDepthWiseNet, a lightweight CNN framework dedicated to hyperspectral image classification with the focus on demonstrating the efficiency of depth-wise convolutions as a method of minimising computation load without performance loss. Their work indirectly supports the usefulness of lightweight networks in situations where high-resolution inputs and few computing resources become a challenge. In short, when it comes to deploying deep learning in low-resource situations, it is important to have lightweight models such as MobileNetV2. They are effective and, hence, are suited to image classification in real-time due to architectural improvements like depth-wise separable convolutions.

2.7 Performance Evaluation in Multi-class Classification

Validation of model performance forms the epicentre of deep learning models, particularly with problems in multi-class classification in the biomedical domain. The main measures are accuracy, precision, recall, F1-score, and ROC-AUC that provide complementary information on the performance of models. Accuracy is the ratio of correct predictions over total predictions, and it can fail in terms of being misleading in datasets with much larger examples of one class than the other, as observed in biomedical types of data. Precision is the number of positive occurrences that are accurate among those that were predicted to be positive, whereas the recall (sensitivity) determines the adherence between all relevant instances as told by the model. Harmonic mean of precision and recall, known as the F1-score, is a balanced measure, especially where classes are unbalanced. ROC-AUC (area under the Receiver Operating Characteristic curve) evaluates how well the model separates classes since the goal is to show how well a model balances between the perceptions of true positive rate and false positive rate. These metrics are important in the biomedical context where misclassification may have grave consequences. A false negative case may result in the delay of appropriate treatment, but a false positive may result in unnecessary treatment. Consequently, it is not enough to depend on accuracy only. The metrics, such as precision, recall, and AUC under ROC, offer a more detailed picture, as pointed out by Ferrer (2022) [18], particularly in the case where the classes have different clinical significance. Confusion matrices play an important role when explaining multi-class performance. They can give the results of predictions by the classes, which would make it easier to visualise the strengths and weaknesses of the model. According to Yang et al. (2022) [19], AUC under the ROC is a metric that can be considered a better measure than accuracy to evaluate a classifier due to its insensitivity towards the distribution of classes. The common assumption under multi-class AUC is that every pair of classes is well-separated from the distribution of the others to get a reasonably valid performance measurement. Therefore, to ensure a fair and precise estimation of multi-class classifiers in biomedical practice, a set of performance metrics is used in this study by considering the fact that performance under each class needs to be assessed instead of the overall accuracy.

2.8 Research Gap

Although deep learning (DL) is being incorporated extensively in biomedical image analysis, there is still a sharp shortage of studies devoted particularly to the classification of fungal species in space environments with the help of microscopic imagery. Despite the fact that the overall use of AI in medical diagnostics has been covered by multiple studies, there are virtually no works looking into applying DL models to identify fungal contamination in closed, resource-limited facilities like spacecraft, as this study attempts to investigate. The main peculiarity of fungal diseases is that in space, they are a special risk due to microgravity, the limited availability of medical facilities, and the necessity of making autonomous decisions. Nevertheless, the prevailing DL frameworks were developed to operate under terrestrial settings and never tested or adjusted to the peculiarities of the operations of health monitoring systems in space. Moreover, it is obvious that the comparison of various deep learning models on a standard fungal images dataset is lacking. Despite extensive research conducted on different CNN architectures (including more sophisticated pre-trained models) and tested in diverse medical fields, relatively few comparative studies have compared them to each other regarding the fungal classification. Identification of the most appropriate model architecture that leads to accurate, lightweight, and efficient classification during the critical space missions becomes tough. Further, limited deployable AI tool development in real-time and validation have been noted to support the operation of these tools without reliance on computing resources from the ground station on Earth. A wide range of DL implementations, even in health care, remain highly dependent on cloud or manual control, which would be impractical under the long-term missions in space. The need is urgent for models that are both accurate and sufficiently computationally tractable and interpretable, and in addition to these, the models need to be easily adapted to onboard deployment with limited power and available hardware. Such gaps must be bridged to achieve an entirely self-sustaining bio-surveillance system in space with an ability to detect fungus early and prevent its harm to the crew and mission by mitigation.

3. METHODOLOGY

3.1 Data Description and Preprocessing

This research is based on the DeFungi Dataset, which consists of a set of 5000 high-resolution microscopic images of five fungal species, i.e., Candida albicans (H1), Aspergillus niger (H2), Trichophyton rubrum (H3), Trichophyton mentagrophytes (H5), and Epidermophyton floccosum (H6) [1]. Data is divided into train, validation, and testing folders, and there are subfolders according to classes in the data. The dataset is perfectly balanced in the training set with 1000 instances for each of the five fungus types; however, in the validation and test sets, the number of instances varies among classes. Typically, in the validation set, the class distribution is the following: H1 (437), H2 (232), H3 (81), H5 (80), and H6 (69). In the test set, there are 437 images for H1, 233 images for H2, 82 images for H3, 80 images for H5, and 70 images for H6. Even though there is class imbalance in both the validation and test sets, there are no explicit rebalancing mechanisms, namely class weighting or synthetic sampling, to maintain the original data balance and test the robustness of the models in a realistic situation. In order to train the data, the data was loaded by using 'ImageDataGenerator' of Keras, in which real-time augmentation was performed on the training data to improve the model generalisation. The augmentation methods involved rotation (range of 20), width and height translation (width shift = 0.1, height shift = 0.1), zoom (range = 0.1), and horizontal flipping, where all the images were resized into 224x224. Test and validation sets were not augmented but normalised only (in [0,1]) by dividing the feature matrices by 255 (as pixel values lie in [0,255]). The batch size was maintained at 32, and the structured data was kept in the form of categorical data through one-hot encoding. For all the fitted classifiers, Adam optimiser with a learning rate of 0.0001. This preprocessing pipeline provides consistency of the input dimensions, class equilibrium, and prevents overfitting of models during training.

3.2 Data Visualisation

After loading the image dataset and resizing each one into 224x224 pixel dimensions, a few equal samples from each class category are visualised in a grid to observe if differences in the type of contamination can be spotted with the naked eye. From the training set folder, three random samples were taken of every one of the five types of fungi (H1, H2, H3, H5, H6), and then the fungal image data was visualised using matplotlib. It is easy to compare the classes visually, as all classes are marked with numbers in a row.

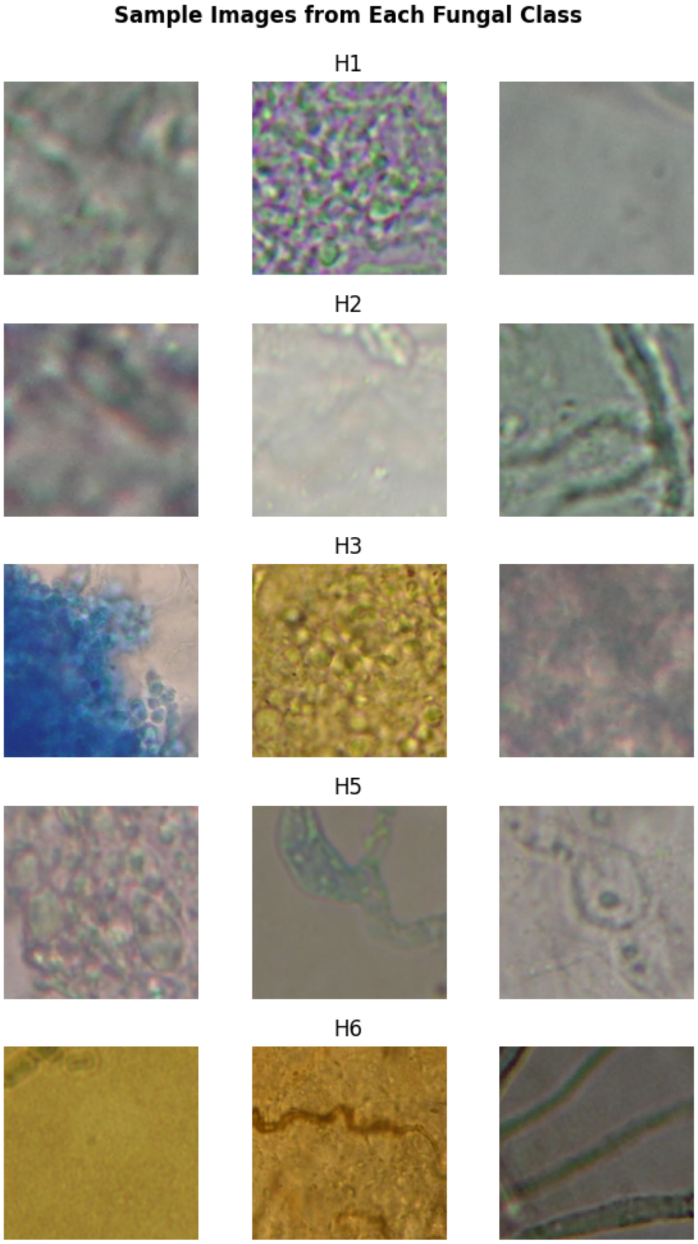


Figure 1: Sample Microscopic Images from Each Class

Based on what was presented in the samples in FIGURE 1, it is possible to observe marked differences in fungal types. The textures composed by Candida albicans (H1) are very faint and smooth, whereas the images of the Aspergillus niger (H2) are more clustered and granular. Trichophyton rubrum (H3) has more dense structures of the cells and variations in colours (blue shade). Trichophyton mentagrophytes (H5) appears to have more spherical elements with a boundary. Epidermophyton floccosum (H6), on the other hand, has a coarser, granular appearance and yellowish colouring. The specific morphological characteristics afford a visual representation on which automated classification can be based, to which models learn to select such characteristics in order to distinguish between fungal classes.

3.3 CNN Modelling

In order to provide a solid basis for the image classification exercises, a basic convolutional neural network (CNN) was built, which was structurally inspired by the study of Shin and two other researchers, as published in 2016 [20]. In their investigation, they examined the effectiveness of several CNN structures and preprocessing strategies to identify facial expression recognition, and finally, the one with the best results was the three-layer CNN with max-pooling and histogram equalisation as the most effective to obtain the maximum accuracy level and generalisation. The base CNN used in this investigation is one with a sequential structure where there are three layers of 3 convolution filter layers whose depths increase sequentially (32-64-128), and they are sequentially followed by max-pooling. The model ends with a flatten layer, a dense layer of 128 units, and a softmax output layer of five classes. The architecture has about 11.2 million trainable parameters and was compiled with Adam optimiser, a learning rate of 0.0001, and categorical cross-entropy set as the loss.

The CNN was trained, and early stopping was turned on to avoid overfitting, which resulted in training for a total of 10 epochs. Multiple metrics were utilised to evaluate the performance of the corresponding models. After printing out a classification report and a confusion matrix, it was possible to identify class-wise precision and recall. Additionally, a multi-class ROC curve was also used to further quantify the discrimination ability of the model with AUC values per class. The diversified evaluation metrics give the advantage of identifying the strengths and weaknesses of each model fitting.

3.4 EfficientNetB0 Modelling

The EfficientNetB0 architecture is concise, as a sequence of CNN layers is optimised to fit the demand of efficient inference capabilities. Unlike traditional ANNs, EfficientNetB0 uses the concept of systematic compound coefficient to scale the neural network parameters, which eventually leads to a small number of parameters that need training and thus greatly improves the performance. Guo et al. (2025) [21], used EfficientNetB0 in thyroid cancer detection and found highly accurate results while testing it on several samples with consistent performance. This essentially proves that the model is able to extract important patterns from medical images that essentially distinguish the cancer subtypes. This makes the model a good choice to explore the detection of fungus types in a spacecraft environment, as performed in this study.

In the project, EfficientNetB0 was utilised with pretrained ImageNet weights. The base layers were kept frozen, and a custom classifier head was appended to the base layers with global average pooling, dropout regularisation, and five-class fungal classification as the softmax output layer. The model has been compiled with Adam optimiser and trained for ten epochs. Early stopping was used to prevent overfitting. The training curves revealed that it was able to steadily curb the accuracy and loss of training data and validation data. Like earlier, the model once fully trained was evaluated on the test set in terms of the same metrics, like accuracy, precision, recall, and F1 score as derived from the confusion matrix, and additionally, the ROC curves for each class.

3.5 MobileNetV2 Modelling

MobileNetV2 is a lightweight deep learning framework aimed at efficient calculations in resource-constrained conditions, while keeping the classification performance at an equal level. It employs depthwise separable convolutions that come at a much lower parameter count and computational cost than typical convolutional layers. It also uses inverted residual blocks together with the use of a linear bottleneck to guarantee faster inference and training. MobileNetV2 has demonstrated a competitive performance in classifying complex visual features using minimal resources in the issue of biomedical imaging. As an example, Indraswari et al. (2022) [22], showed its applicability in the area of melanoma image classification, thus evidencing its potential in medical diagnostics and the ability to perform precise and real-time predictions. In this case, MobileNetV2 has been applied with ImageNet pretrained weights, but with the exclusion of the top layer. A custom head for classifying the five categories of fungi using a global average pooling layer and a dense output layer was added. The convolutional base was kept frozen to preserve learned features, and the classification head model was trained on the DeFungi dataset. The early stopping was used to avoid overfitting, and the model was trained with ten epochs. The assessment was done through an accuracy and loss graph, a report of classification, a confusion matrix, and ROC-AUC curves. The findings have shown a good generalisation and separation ability of classes, which qualifies MobileNetV2 as a potential model in fungal classification under resource limitation.

4. RESULTS AND FINDINGS

4.1 CNN Classifier Evaluation

The CNN classifier model training progress is displayed to observe the rate of improvement of accuracy and loss at the same time in both training and validation sets. Then, after completion of training, its performance is presented through a classification report, a multi-class confusion matrix, and ROC-AUC curves corresponding to each class. The CNN model is observed to have a total of 11,169,605 network parameters constituting a size of 42.61 MB, and its inference time on the test set is observed to be around 143 ms per step, which takes 5 seconds in total, making it suitable for deployment in spacecraft missions.

As observed in FIGURE 2, the CNN classifier was trained for 10 epochs with early stopping criteria, and after 10 epochs, it was observed that the loss could not be decreased further significantly. A gradual increase in train set accuracy is observed, whereas in the validation set, the improvement is oscillatory in nature, and as a result, a similar pattern is observed for losses in both sets.

FIGURE 3, shows that the CNN model had a decent level of test performance, producing about 65% overall accuracy. It was good at classifying H5 and H6 types (F1: 0.68 and 0.75, AUC: 0.95 and 0.98, respectively), but not so well with H2 and H3 due to the misclassification of a large part of it

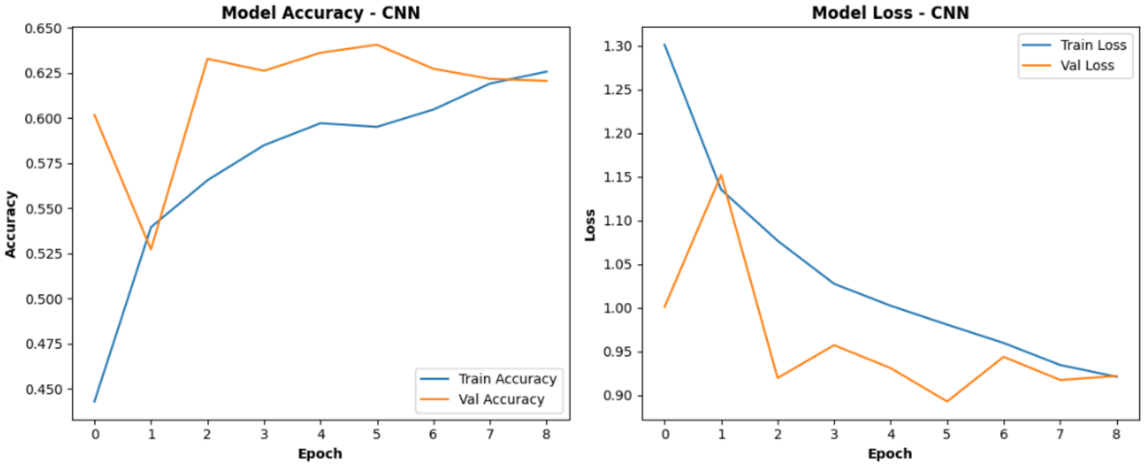
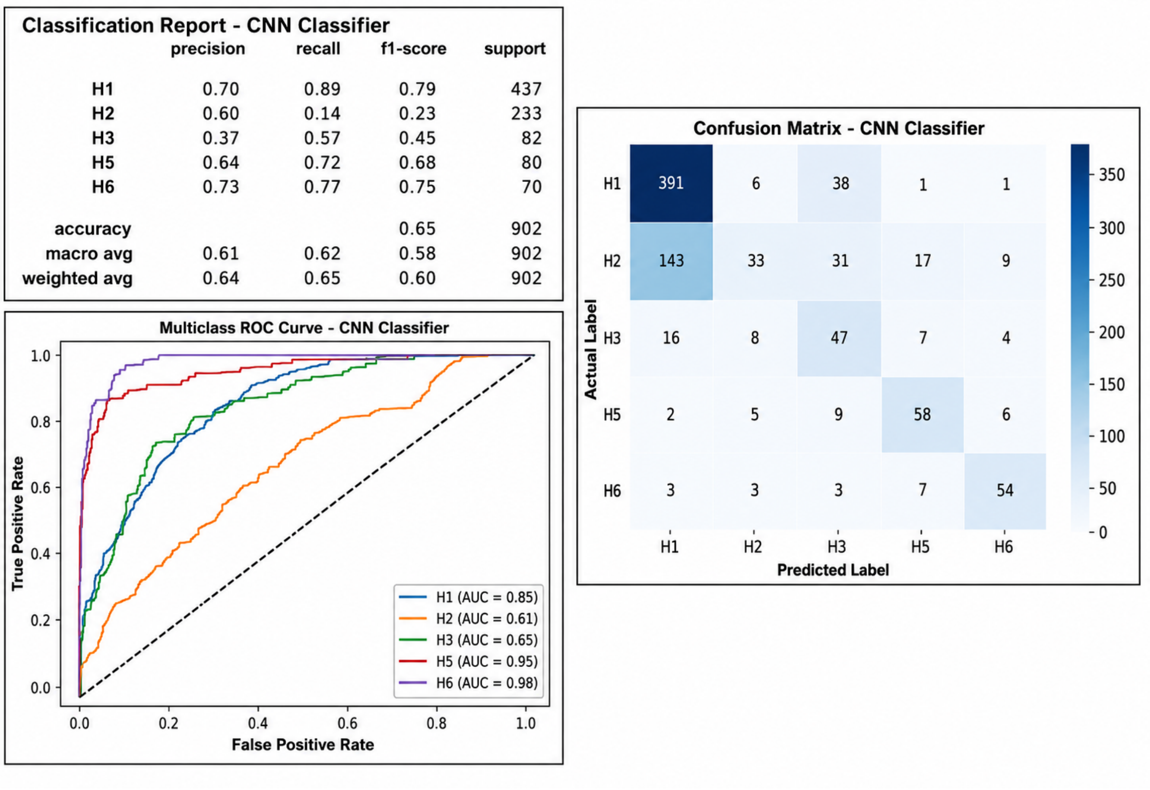


Figure 2: CNN Classifier Training Progress



Classification Report - CNN Classifier

	precision	recall	f1-score	support
H1	0.70	0.89	0.79	437
H2	0.60	0.14	0.23	233
H3	0.37	0.57	0.45	82
H5	0.64	0.72	0.68	80
H6	0.73	0.77	0.75	70
accuracy			0.65	902
macro avg	0.61	0.62	0.58	902
weighted avg	0.64	0.65	0.60	902

Confusion Matrix - CNN Classifier

Actual Label \ Predicted Label	H1	H2	H3	H5	H6
H1	391	6	38	1	1
H2	143	33	31	17	9
H3	16	8	47	7	4
H5	2	5	9	58	6
H6	3	3	3	7	54

Figure 3: CNN Performance on Test Set

as H1. In general, the CNN model is observed to provide significantly better performance towards

H5 and H6 types, but for other fungal types, the performance is low, indicating the generalisability of the model is not quite high.

4.2 EfficientNetB0 Classifier Evaluation

Like earlier, the variation of accuracy and loss for train and validation sets is plotted for the EfficientNetB0 model throughout its training, and then the performance evaluation is presented for the test set using the same evaluation metrics. The EfficientNetB0 architecture is much lower in size, with only 6,405 trainable parameters, and a total size of 15.47 MB, which takes about 399 msec per step or 17 secs in total for inference in the test set, which is of moderate level for deployment in space missions.

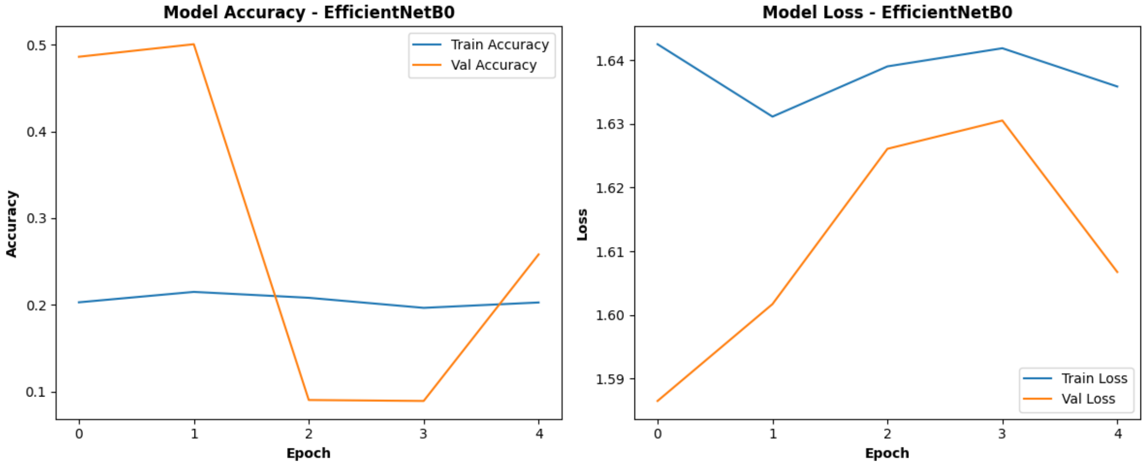
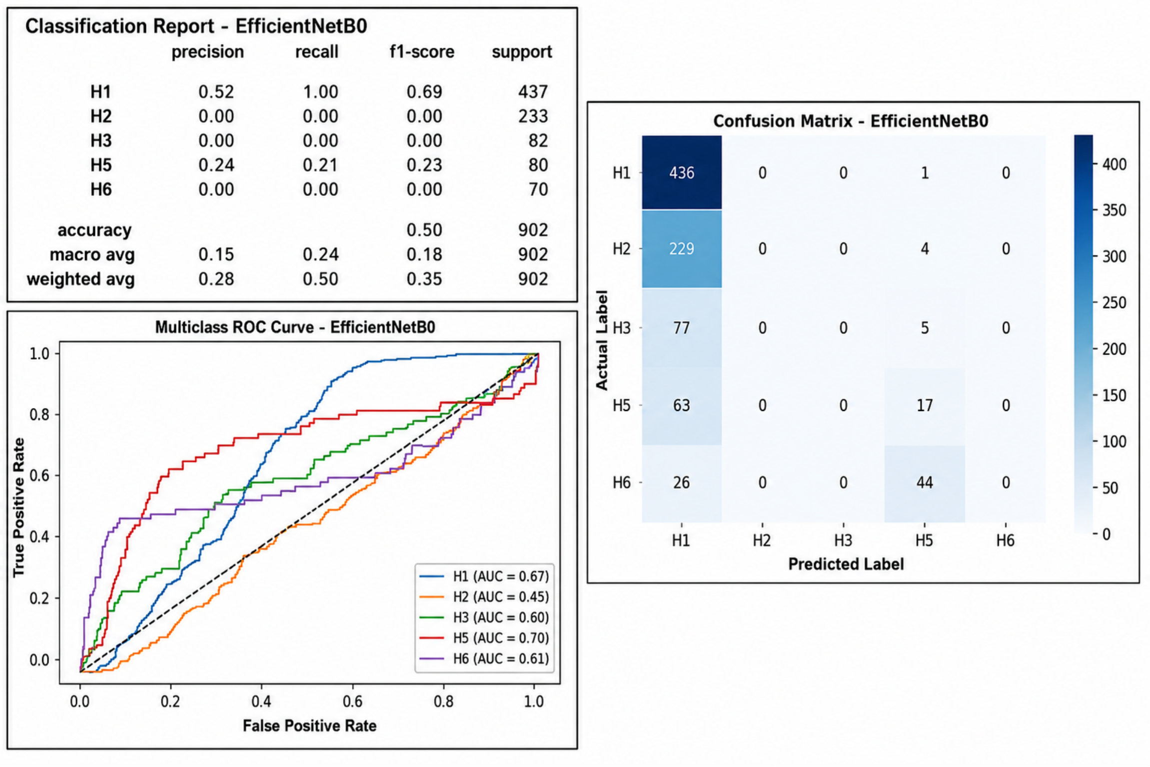


Figure 4: EfficientNetB0 Training Progress

FIGURE 4, shows that during the training of EfficientNetB0, the accuracy remained around 20% in the training set, and the loss was also constant, around 1.63 to 1.64. However, for the validation set, a fluctuation of accuracy in the range 10% to 50% was observed, and as a result, the loss also varied between 1.59 and 1.63.

FIGURE 5, implies, in general, that EfficientNetB0 did not work well, as its total accuracy was 50%, and it had a persistent inclination towards class H1. It also did not manage to detect any instances of H2, H3, and H6 (zero F1-score), indicating a huge disparity with the identification of the H1 class, and thus, it is possible that the model was not able to detect unique features in fungal images that separated other types from the H1 type. It also showed low to moderate ROC-AUC scores on minority classes (H2, H3, and H6), which shows low discriminative power. This behaviour of EfficientNetB0 is possibly due to the fact that freezing of pretrained layers prevents domain adaptation, and the comparatively small size of the dataset restricts the model to learn discriminative features of the minor classes.



Classification Report - EfficientNetB0

```
              precision    recall  f1-score   support

          H1       0.52      1.00      0.69       437
          H2       0.00      0.00      0.00       233
          H3       0.00      0.00      0.00        82
          H5       0.24      0.21      0.23        80
          H6       0.00      0.00      0.00        70

    accuracy                           0.50       902
   macro avg       0.15      0.24      0.18       902
weighted avg       0.28      0.50      0.35       902
```

Confusion Matrix - EfficientNetB0

Actual Label \ Predicted Label	H1	H2	H3	H5	H6
H1	436	0	0	1	0
H2	229	0	0	4	0
H3	77	0	0	5	0
H5	63	0	0	17	0
H6	26	0	0	44	0

Figure 5: EfficientNetB0 Performance on Test Set

4.3 MobileNetV2 Classifier Evaluation

The process is the same for MobileNetV2, which starts with visualisation of training progress and ends with evaluation of the trained model on the test set on the same evaluation metrics. The fitted MobileNetV2 model has only 6,405 trainable network parameters with the smallest total size of about 8.64 MB out of the three fitted models. The model takes about 259 msec per step or 10 secs in total for inference in the test set, which is also of a moderate level for deployment in fungus contamination detection in space missions.

As observed from FIGURE 6, the MobileNetV2 shows excellent training progress over 10 epochs, in FIGURE 6, as accuracy in both sets gradually increased and finally settled around 65% in the train set and around 60% for the validation set. In alignment the loss in both sets also gradually decreased and finally settled around 0.9 at the final epoch in both sets.

FIGURE 7, indicates that MobileNetV2 performed best on average, as well as having balanced precision and recall values in all the fungal classes at an accuracy of 67%. It performed well on the minority classes (H3, H5, H6) and recorded high ranks of ROC-AUC in most directions above 0.9, demonstrating high discriminative power and stability in the classification of fungi.

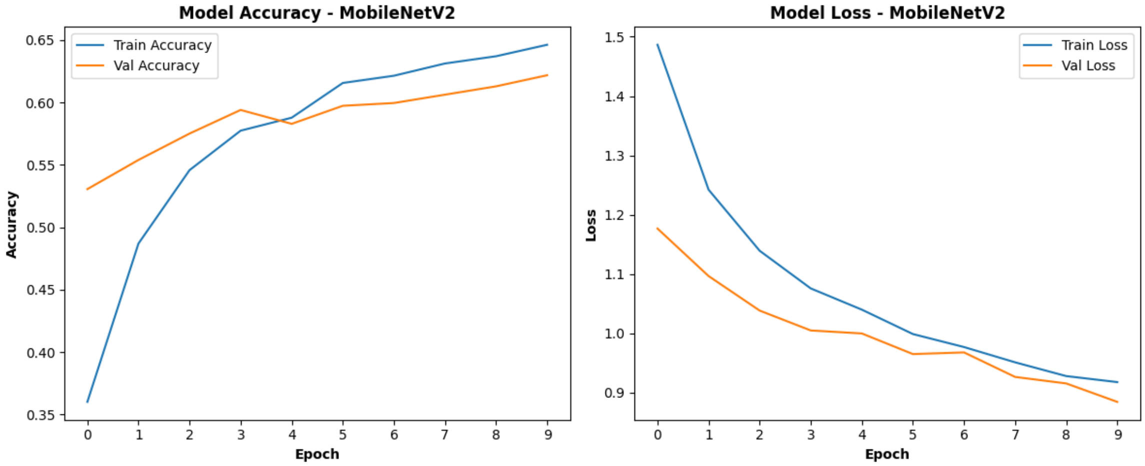
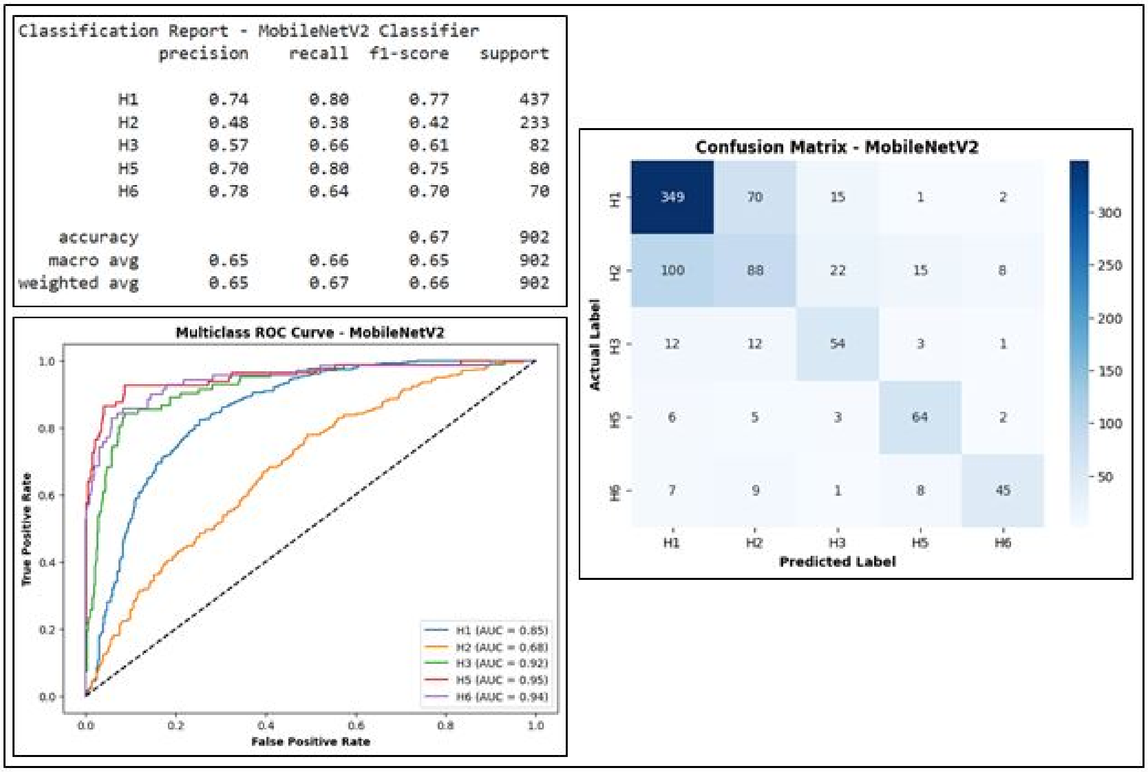


Figure 6: MobileNetV2 Training Progress Plot



```
Classification Report - MobileNetV2 Classifier
              precision    recall  f1-score   support

          H1       0.74      0.80      0.77       437
          H2       0.48      0.38      0.42       233
          H3       0.57      0.66      0.61        82
          H5       0.70      0.80      0.75        80
          H6       0.78      0.64      0.70        70

    accuracy                           0.67       902
   macro avg       0.65      0.66      0.65       902
weighted avg       0.65      0.67      0.66       902
```

	H1	H2	H3	H5	H6
H1	349	70	15	1	2
H2	100	88	22	15	8
H3	12	12	54	3	1
H5	6	5	3	64	2
H6	7	9	1	8	45

Figure 7: MobileNetV2 Performance on Test Set

4.4 Performance Comparison of Models

The important metrics, as found through evaluation of fitted models on the test set, are summarised in TABLE 1, for easier comparison of fungal classification performances. This essentially helps to select the best deep learning algorithm for detecting the type of infection on organic cells based on versatile criteria.

Table 1: Comparing Test Set Performances of Three Models

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Avg. ROC-AUC**
CNN	0.65	0.61	0.62	0.58	0.86
EfficientNetB0	0.5	0.15	0.24	0.18	0.61
MobileNetV2	0.67	0.65	0.66	0.65	0.87

TABLE 1, shows clearly that, of all the models, MobileNetV2 brought out the best performance criteria gaps with respect to the other models. It had an optimal performance in terms of accuracy (0.67), precision (0.65), recall (0.66), and F1-score (0.65), and a mean ROC-AUC of 0.87. On the contrary, EfficientNetB0 had a low generalisation, especially in recall and F1-score, indicating that it is hard to recognise a variety of classes. Baseline CNN had a decent performance, yet it was dominated by MobileNetV2. MobileNetV2 has a great balance between high precision and computational efficiency, making it optimal to use in fungal type classification to solve the constrained nature as encountered in spacecrafts, among other models, where the performance and resource consumption are primary concerns.

5. CONCLUSION

5.1 Summary

This paper presented an investigation into the automation of the fungal classification process on microscopic pictures in space-like conditions using deep learning models. The performances of three models, CNN, EfficientNetB0, and MobileNetV2, were tested through the DeFungi dataset. The CNN was fairly accurate, whereas EfficientNetB0 was not good because of its bias towards H1 fungal type, resulting from its ability to detect unique properties of fungal images that separate other fungal types from H1. MobileNetV2 was more accurate than the others, with high precision and recall, balanced accuracy, and the ROC-AUC score was high on all classes. It has a lightweight desktop and discriminating ability, which makes it the best candidate for resource-constrained and real-time onboard bio-surveillance. The results justify the incorporation of AI-based diagnostics to improve astronaut safety, as well as spacecraft integrity.

5.2 Limitations

There are various limitations to this research. The first issue entails the small number of images and the detailing of picture quality of fungal types in the DeFungi dataset, which possibly led to a deficiency in generalisation and the performance of some models, like CNN and EfficientNetB0, for rare fungus types. Second, the models have not been tested with real-world conditions in spacecraft, as they were trained and tested in a simulated environment, and will be potentially very different in reality, such as by considering lighting, variability in imaging performance, and limited degrees of freedom in operation. Third, the use of transfer learning constrains flexibility to new fungal species that are not learnt during pretraining. Moreover, the image data are the only data considered in the study, and there is a potential for including other biological signals that would make image data more accurate in the real application of detecting and classifying fungal conditions. Lastly, the DeFungi dataset is publicly available and widely used to benchmark, but the size and the imbalance of classes of the validation and test set could limit the generalisability of the models. This research alleviates this limitation using data augmentation and transfer learning, but larger and more varied datasets, especially those measured in space-relevant conditions, are required for future validation. Computational limitations and the size of the dataset prevented the use of k-fold cross-validation and statistical significance testing, which can be one of the future directions.

5.3 Future Work

Future research can be devoted to further fine-tuning of the models, such as MobileNetV2, and similar pretrained classifiers to improve their accuracy through mobile-specific attention mechanisms and domain-specific augmentation. Model generalisation would be enhanced by a larger sample size, which can include additional species of the fungus in varied imaging settings. The level of trust and openness on the critical mission decisions might also be boosted by the implementation of explainable AI methods. The highly precise and efficient models can be deployed in space missions where real-time laboratory experiments can be performed remotely from Earth. Moreover, explainability methods like Grad-CAM and SHAP will be included in future work to increase the transparency of the model and aid in safety-related decision-making. Finally, cooperative AI systems that combine several biosensors can present a more comprehensive and encompassing detection of microbial threats to long-term space exploration.

5.4 ACKNOWLEDGEMENT

This research study is supported by the National Science Foundation under Grant No. 2426293. [NSF EiR 2402017, IUSE 2236002, ExpandAI 2426293]

5.5 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

5.6 DATA AVAILABLE ON REQUEST FROM THE AUTHORS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] https://www.kaggle.com/datasets/anshtanwar/microscopic-fungi-images?utm_source=chatgpt.com

[2] Chander AM, Teixeira MD, Singh NK, Williams MP, Simpson AC ,et al. Description and Genome Characterization of Three Novel Fungal Strains Isolated From Mars 2020 Mission-Associated Spacecraft Assembly Facility Surfaces – Recommendations for Two New Genera and One Species. J Fungi. 2023;9:31.

[3] Regberg AB, Castro CL, Connolly HC, Davis RE, Dworkin JP, et al. Prokaryotic and Fungal Characterization of the Facilities Used to Assemble Test and Launch the OSIRIS-REX Spacecraft. Front Microbiol. 2020;11:530661.

[4] Rahman MA, Clinch M, Reynolds J, Dangott B, Meza Villegas DM, et al. Classification of Fungal Genera From Microscopic Images Using Artificial Intelligence. J Pathol Inform. 2023;14:100314.

[5] Zieliñski B, Sroka-Oleksiak A, Rymarczyk D, Piekarczyk A, Brzychczy-Włoch M. Deep Learning Approach to Describe and Classify Fungi Microscopic Images. PLOS One. 2020;15:e0234806.

[6] Simões MF, Antunes A. Microbial Pathogenicity in Space. Pathogens. 2021;10:450.

[7] Simões M, Cortesão M, Azua-Bustos A, Bai FY, Canini F, et al. The Relevance of Fungi in Astrobiology Research – Astromycology. Mycosphere. 2023;14:1190-1253.

[8] Kristensen K, Ward LM, Mogensen ML, Cichosz SL. Using Image Processing and Automated Classification Models to Classify Microscopic Gram Stain Images. Comput Methods Programs Biomed Update. 2023;3:100091.

[9] Koo T, Kim MH, Jue MS. Automated Detection of Superficial Fungal Infections From Microscopic Images Through a Regional Convolutional Neural Network. PLoS ONE. 2021;16:e0256290.

[10] Liu X, Gao K, Liu B, Pan C, Liang K, et al. Advances in Deep Learning-Based Medical Image Analysis. Health Data Sci. 2021a;2021:8786793.

[11] Liu X, Song L, Liu S, Zhang Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability. 2021b;13:1224.

[12] Elngar AA, Arafa M, Fathy A, Moustafa B, Mahmoud O, et al. Image Classification Based on CNN: A Survey. JCIM 2021;6:18-50.

[13] Chen L, Li S, Bai Q, Yang J, Jiang S, et al. Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sens. 2021;13:4712.

[14] Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, et al. Transfer Learning for Medical Image Classification: A Literature Review. BMC Med Imaging. 2022;22:69.

[15]] Mahmood MA, Alsalem K, Elbashir MK, Abd El-Ghany SA, El-Aziz AA. Leveraging Fuzzy C-Means and EFFICIENTNETB0 for Improved Ecg Image Classification. Int J Comput Commun Control. 2025;20.

[16] Gulzar Y. Fruit Image Classification Model Based on MobileNetV2 With Deep Transfer Learning Technique. Sustainability. 2023;15:1906.

[17] Cui B, Dong XM, Zhan Q, Peng J, Sun W. Litedepthwisenet: A Lightweight Network for Hyperspectral Image Classification. IEEE Trans Geosci Remote Sens. 2022;60:1-15.

[18] Ferrer L. Analysis and Comparison of Classification Metrics. 2022. arXiv preprint: https://arxiv.org/pdf/2209.05355v1

[19] Yang Z, Xu Q, Bao S, Cao X, Huang Q. Learning With Multiclass AUC: Theory and Algorithms. IEEE Trans. Pattern Anal Mach Intell. 2022;44:7747-7763.

[20] Shin M, Kim M, Kwon DS. Baseline CNN Structure Analysis for Facial Expression Recognition. In2016 25th IEEE international symposium on robot and human interactive communication. RO-MAN. IEEE. 2016:724-729.

[21] Guo H, Zhang J, Li Y, Pan X, Sun C. Advanced Pathological Subtype Classification of Thyroid Cancer Using EfficientNetB0. Diagn Pathol. 2025;20:28.

[22] Indraswari R, Rokhana R, Herulambang W. Melanoma Image Classification Based on MobileNetV2 Network. Procedia Comput Sci. 2022;197:198-207.