Semi-supervised mold differentiation using typical laboratory results as label data

Henrik Pichler[1], Janis Keuper[1], and Matthew Copping[2]

[1] Hochschule Offenburg
`hpichler@stud.hs-offenburg.de`
[2] Biostates GmbH
`m.copping@biostates.de`

Abstract. This study applies semi-supervised learning to automate the differentiation of mold colonies, thereby reducing the time and cost associated with air quality assessments. EfficientNet V2 and Normalization-Free Net (NfNet) were trained on a dataset of mold colony images, created in a semi-supervised way. NfNet demonstrated superior performance, particularly on non-padded images, with explainable AI techniques enhancing interpretability. The models exhibited generalization capabilities to environmental samples, indicating the potential for automating mold identification and streamlining air quality monitoring, thereby reducing manual effort and costs. Future work will focus on refining species handling and integrating the system into laboratory workflows.

Keywords: Artificial Intelligence; Biology; Semi-supervised learning; Object detection; Explainable artificial intelligence

1 Introduction

1.1 Problem Definition

Air quality in workplaces and production facilities affects employee well-being and operational efficiency. In accordance with VDI Guideline 6022, the quality of air entering premises via ventilation must be maintained [1]. Many factors affect indoor air quality, including VOCs, CO2, humidity, and mold spores [2]. Mold contamination is a key factor in maintaining a healthy indoor environment due to its ability to trigger allergic reactions and respiratory issues. This is crucial in vulnerable settings like hospitals [3]. To assess air quality, customers provide air samples using Petri dishes, which are incubated for 5-7 days to allow for mold growth. These colonies are then counted and differentiated to evaluate air quality. Especially the task of differentiating, often including the usage of a microscope, is time-consuming and costly.

1.2 Goal

To reduce the time and cost of evaluating the samples, a project was carried out with the specific goal of developing a more efficient method for differentiating macromorphologically distinguishable mold colonies. It was determined that colonies requiring microscopic analysis would be excluded from the project, as this would involve a more complex setup, that is beyond the scope of this task. A central objective of the project was to demonstrate the feasibility of training machine learning (ML) models on novel forms of data, even when such data is scarce and lacks comprehensive annotation. Furthermore, the project aimed to demonstrate that existing ML models can be effectively extended and adapted to perform more complex tasks, rather than being entirely rebuilt from scratch. The models were trained and evaluated within this semi-supervised context to assess their effectiveness in achieving the goal of streamlining the differentiation process and enhancing overall efficiency.

1.3 Related Work

Semi-supervised Learning The subject of semi-supervised learning examines how to use partially labeled datasets for training machine learning models. [4] presents a method for classifying test data without prior exposure to some of the classes.

Explainable AI The field of explainable AI seeks to understand the behavior of complex machine learning and deep learning models. This is crucial, especially in biological and medical contexts, where such applications demand high levels of accuracy and trust. Consequently, numerous studies have been conducted with the aim of explaining a model's behavior. [5] utilizes attribution maps to generate pseudo ground truths for semi-supervised semantic and instance segmentation.

Transfer Learning Transfer learning reduces training time and cost by using previously trained models. [6] reviews progress in transfer learning for classification, regression, and clustering.

Machine learning models in a biological/medical context There is limited research on the automatic differentiation of molds. However, numerous studies address biological and medical sample analysis, underscoring the significance of this topic. [7] discusses current approaches to AI in the field of nephrology. [8] utilizes a hyperspectral imaging device and a support vector machine (SVM) to identify the growth stage of molds on wheat grains.

2 Approach

2.1 Dataset Creation

To guarantee a minimal labeling effort and to quickly start a training on the novel data type, a dataset was constructed in a semi-supervised manner. The dataset comprises images of Petri dishes, captured from above with a camera with a resolution of 2,144x2,144 pixels and 17.9 pixels/mm. The Petri dishes were prepared using the three-point inoculation method, whereby individual known mold colonies were extracted from one dish and transferred onto another dish [9]. This enables the regulation of the quantity and variety of mold growth on a given sample. A total of 600 Petri dishes were prepared and incubated, with each dish containing three mold colonies, as per the method's specifications. The samples comprised of five distinct mold species, with each distributed across 100 Petri dishes. The remaining 100 samples were colonized with 10 additional mold species, the actual species of which were inconsequential. This approach would enable the models to consider a more extensive set of mold morphologies. Consequently, mold colonies, for which an uncertainty would remain, could be classified as ”other” and examined for revision when required. Following the incubation and image capture, a YoloV7 model was applied to detect the mold colonies without any classification, as this was a problem on which the model had previously been trained on [10]. This resulted in distinct bounding boxes for each mold colony. The predictions were then reviewed, to correct any errors that had occurred, which was a minor issue. As only one type of mold, which was also known, grew on each Petri dish, all annotations for one image could simply be assigned with the same class. The original dimensions of the mold colonies were maintained by utilizing this approach, with each image accurately representing the natural, unaltered dimensions of the colonies. This resulted in a diverse range of sizes for each mold image within the dataset. Henceforth, the dataset will be referred to as the "clean culture" dataset.

2.2 Classifier Training

Two classification models, EfficientNet V2 [11] and Normalization Free Net (NfNet) [12], were selected for training. Both models were initialized with pre-trained weights based on ImageNet [13]. The mold colonies were extracted from the clean culture dataset using the provided bounding boxes. Two training approaches were investigated. For the first approach, the mold colony crops were padded using zero padding or resized to a size of 1000x1000 pixels, depending on whether the image crop was larger or smaller than 1000x1000 pixels. The resulting images will be referred to as "padded" images. The second approach involved leaving the images unaltered, maintaining their original dimensions. However, they were made square. These images will be referred to as "non-padded" images. Following the padding/resizing, the images were fed through their respective models, resulting in a single classification per mold colony. In accordance with the concept of progressive learning as outlined in [3], adaptive image augmentation was employed to augment the training images and increase data variety. This meant that larger images, and consequently mold colonies, were augmented with a higher magnitude than smaller ones. Cross-entropy loss was utilized as the loss function throughout the training process. The models were subjected to continuous validation during training using a combination of the F1-Score and the loss on the validation data.

2.3 Feature Inspection

The field of explainable artificial intelligence (XAI) plays a crucial role in guaranteeing the transparency and reliability of model predictions, thus enabling the deployment of these models in practical applications [14]. In this work, Grad-CAM (Gradient-weighted Class Activation Mapping) [15] was used to provide visual explanations of the models' predictions. The method highlights the areas, that contributed most to the decision, enabling a visual inspection of the model through heatmaps.

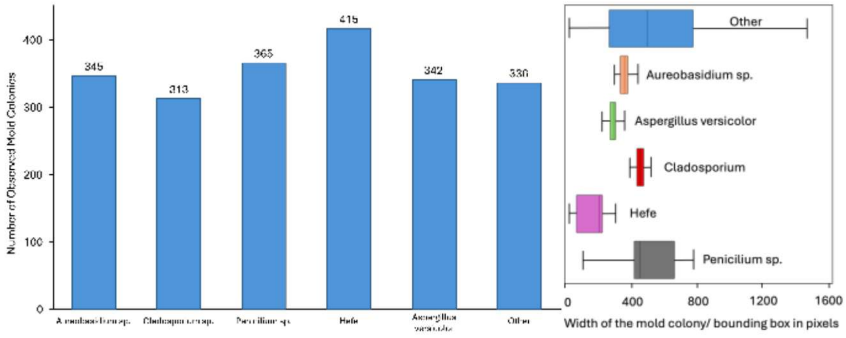
2.4 Classifier Evaluation

The models were evaluated on both the clean culture Dataset and an "environmental" dataset, which represented real-world mold samples and included 640 additional annotated images with a total of 12,472 individual mold colonies. The environmental Dataset was employed exclusively for the purpose of evaluating the models' performance in more complex, real-world conditions. In a practical application, the dataset would not be built from the ground up, but rather developed over time to improve the model further. This served the purpose of a proof of concept, demonstrating the generalization ability of the trained models. The primary metrics, used to evaluate the classification performance on the test datasets, were the accuracy, precision and recall. Particular interest was paid to the handling of unknown mold colonies and uncertain decisions respectively. These cases should be classified as "Other". The correct handling of such cases by a model would facilitate the refinement of predictions by a reviewer. The evaluation also considered the performance differences between models trained on padded and non-padded images, respectively. The evaluation of models trained on padded data was conducted exclusively using padded data, while the models trained on non-padded data were evaluated only with non-padded data.

3 Results

3.1 Dataset Statistics

In total 600 images were taken for the clean culture dataset, which resulted in 2,116 single mold colonies from 589 Petri dishes.



(a) (b)

Fig. 1: Statistics for the clean culture dataset. (a) shows the distribution of classes for the dataset. The Acronym "sp." denotes "and related species". (b) shows the distribution of the different mold species based on their colony sizes, measured by the dimensions of the bounding boxes surrounding each colony.

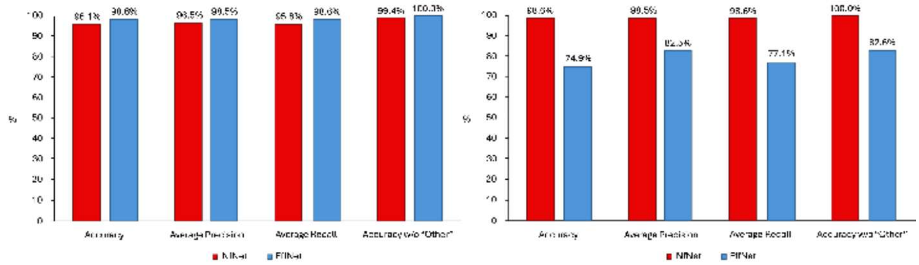
After the creation and review of the clean culture dataset, the following distribution of mold species was found (Fig. 1a): 345 Aureobasidium sp., 313 Cladosporium sp., 365 Penicilium sp., 415 Hefe, 342 Aspergillus versicolor* and 336 other mold colonies with 10 different species.

A size distribution of 1,500 pixels was found across the dataset (Fig. 1b). The largest observed mold colony was 1520 pixels wide ("Other"), the smallest 20 pixels ("Hefe"). The mean size of the colonies was 347 pixels. The size refers to the width of the bounding box around the mold colony.

A large size distribution was found across the dataset (Fig. 1b). This refers to the size of the bounding box around the colony. The largest observed mold colony was 1520 pixels large ("Other"), the smallest 20 pixels ("Hefe"). The mean size of the colonies was 347 pixels.

3.2 Training results

Both the EfficientNet V2 as well as the Normalization-Free Net (NfNet) were trained on the clean culture dataset (chapter 2.2).



(a) (b)

Fig. 2: Comparison of training both the EfficientNetV2 (EffNet) as well as the Normalization-Free Net (NfNet) on the clean culture dataset using both padded images (a), as well as non-padded images (b).

The training of the models on the clean culture dataset resulted in high accuracy across all models (Fig. 2). In the case of the padded images, the EfficientNet V2 demonstrated an accuracy of 98.6%, which represents a 2.5% improvement over the performance of the NfNet on the same dataset. In contrast, the NfNet demonstrated a higher accuracy of 98.6% for the non-padded images, representing a 23.7% improvement over the EfficientNet V2 on the same dataset. The NfNet demonstrated the same accuracy on the non-padded images, as the EfficientNet V2 on the padded images (98.6%).

3.3 Feature Inspection

A further inspection of the obtained results was performed, utilising Grad-CAM to identify the most important features for the models' predictions.
Fig.3 shows the visualizations generated using Grad-CAM for both the EfficientNetV2 and NfNet models, with comparisons between predictions on a single mold colony sample from the class "Cladosporium sp.". The Grad-CAM visualizations illustrate that EfficientNet V2 produces coarse highlighted regions. In the padded input, it highlights the corners and edges of the image. The NfNet highlights specific regions of interest in both the non-padded and padded images, without highlighting the corners or edges of the padded input.

3.4 Evaluation on environmental Data

Fig.4 shows a comparison between a sample from the clean culture dataset and a sample from the environmental dataset. It shows, that the mold colonies growing on the environmental sample are much smaller and occur in larger quantities.
Fig.5 shows the evaluation of both models on the environmental dataset (chapter 2.4). EfficientNet V2 demonstrated superior performance on the padded images, achieving higher overall accuracy (43.3%) and accuracy without the class "Other" (60.0%) compared to NfNet (27.1% and 31.3%). In the case of the non-padded images, NfNet demonstrated

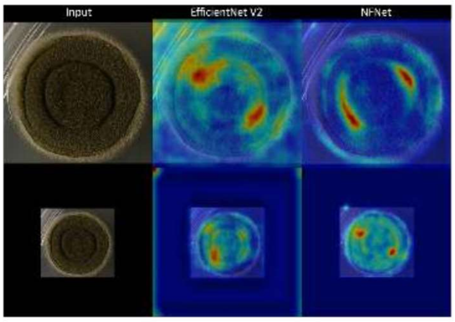
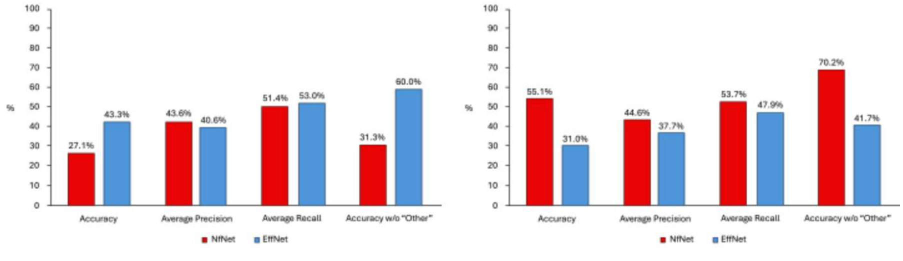


Fig. 3: Grad-CAM visualizations comparing the EfficientNet V2 and NfNet model predictions for a single mold colony image. The top row shows the results for the non-padded input, whereas the bottom row displays the results obtained with a padded input image. The heatmaps visualizations highlight the regions, on which the models focused. Dark red indicates a stronger focus of the model on that region, and conversely, dark blue indicates an uninteresting region. The depicted mold colony belongs to the class "Cladosporium sp."

(a) (b)

Fig. 4: Comparison between a clean culture sample dish (a) with three colonies of Penicilium sp. and a environmental sample dish (b) with a combination of different mold species



(a) (b)

Fig. 5: Comparison of evaluating both the EfficientNetV2 (EffNet) as well as the Normalization-Free Net (NfNet) on the environmental dataset using both padded images (a), as well as non-padded images (b).

superior performance in comparison to EfficientNet V2. The former achieved a higher overall accuracy (55.1%) and accuracy without the class ”Other” (70.2%) in comparison to the latter (31.0% and 41.7%). Overall, NfNet demonstrated an overall performance increase in comparison to EfficientNet V2 of 24.1% in accuracy, 6.9% in average precision and 5.8% in average recall.

4 Discussion

During the dataset creation (chapter 2.1) 11 samples had to be discarded due to contamination by additional mold growth. Furthermore, on several plates, the mold grew more extensively than anticipated, resulting in increased number of samples per dish. This resulted in a slight imbalance in the class distribution, which was effectively mitigated by applying class weights to the loss function during training.

The introduction of the ”Other” class, with the creation of 100 additional samples, notably enhanced the diversity of the clean culture dataset, both in terms of mold size and morphology (Fig.1b). The introduction of a broader range of mold has led to an improvement in the classification task, as the models were able to gain a more complete understanding of the diversity of the dataset. Training on this expanded dataset yielded accurate and consistent results, particularly with NfNet, which demonstrated superior performance in handling data variability compared to EfficientNet V2. While EfficientNet V2 encountered challenges with non-padded images due to the size variance of the mold colonies, NfNet’s resilience to such variability enabled it to outperform EfficientNet V2, particularly on the non-padded images.

Upon examination of the feature maps, it became evident, that there were notable discrepancies between the two models. EfficientNet V2, in particular, appeared to rely excessively on the padding of images, with a corresponding reduction in focus on the mold colonies themselves. This made it less suitable for the task at hand, as it did not consistently extract meaningful features from the mold colony images. In contrast, NfNet performed significantly better by concentrating on the actual features of the mold colonies, such as the characteristic rim of the colony, as seen in Fig. 3. This enabled NfNet to classify molds with greater reliability, especially on non-padded images.

To further evaluate the models’ performances, especially in a real-world context, a supplementary annotation of 640 images was performed. These additional annotations were crucial for validating the models’ real-world applicability. The results confirmed several key points. NfNet exhibited a clear preference for non-padded images (55.1% accuracy), whereas EfficientNet V2 demonstrated superior performance with padded images (43.3% accuracy). Nevertheless, Nfnet demonstrated superior overall performance, attaining higher accuracy on non-padded images than EfficientNet V2 on padded images. However, NfNet demonstrated difficulty in generalizing its knowledge to padded images, likely due to the introduction of artificial boundaries by padding, which interfered with its ability to recognize features across different scales. Conversely, EfficientNet V2 exhibited superior performance in handling padded images due to its reliance on padding, though this came at the cost of its feature extraction capabilities on non-padded images.

A principal component of the evaluation process was the comparison of the accuracy of the models in question, with and without the class named ”Other”. This comparison proved highly beneficial in understanding the models’ability to handle cases of uncertainty or unfamiliar mold colony samples. Without the ”Other” class, the models would have been forced to classify these ambiguous cases into one of the existing classes, which would have increased the likelihood of misclassification. The incorporation of the ”Other” class enabled the models to express uncertainty, thereby enhancing both generalization and precision. This proved to be a particularly beneficial approach in the environmental data, as it provided a safety net for uncertain cases and reduced the overall error rate. It is noteworthy that the highest accuracy achieved on the environmental data was 70% for the NfNet when the "Other" class was excluded. This illustrates that for predictions where the model is certain, the accuracy is already considerable, leaving only a few mold colonies that require manual review. This provides a robust basis for deploying the models on environmental data, specifically NfNet, as it suggests that the majority of classification work could be automated, with only a small portion requiring further inspection. It also offers a promising starting point for further improvements to the models, particularly in refining their handling of uncertain cases.

5 Future Work

One avenue for future research would be to only utilize the environmental data, which includes not only images and bounding boxes but also general class labels for mold colonies, to start training a model. It is important to note that the bounding boxes currently present in the environmental data set were manually annotated. In theory, the dataset would consist solely of the raw images and general class labels, without any bounding box information. The primary objective is to investigate the potential of an iterative approach to construct a comprehensive dataset from the available information, thereby eliminating the necessity for extensive manual annotation.

The classification model can be extended to more complex samples through training on single-class samples. In the initial stage, all colonies within a sample are identified and those belonging to known classes are excluded. If the remaining colonies belong to only a single unknown class, the sample can be utilized for further training. The model will classify the known colonies, leaving the unknown ones labeled as "other," which can then be assigned to the remaining class.

This method could result in a comprehensive dataset comprising all mold species, reducing the necessity for manual annotation and enabling the dataset to evolve. This approach requires a sufficient number of initial samples, particularly "clean culture" samples, that comprise only of mold colonies the same class, to learn effectively. Some human supervision is needed, but it offers a scalable solution that reduces manual effort.

6 Conclusion

This study shows that semi-supervised learning can be used to differentiate mold colonies. A clean culture dataset was developed and used to train classifiers that can identify mold species based on their macromorphological characteristics. The environmental dataset showed that the models can generalize to real-world conditions, although performance declined in more complex scenarios.

Automating this process could reduce costs. The ability to quickly identify mold colonies could cut air quality evaluation costs, making routine inspections more affordable and accessible. This could lead to more frequent assessments, contributing to cleaner air in workplaces and public buildings.

Semi-supervised learning can enhance the efficiency of the mold differentiation process, reducing time and cost. Future enhancements could prioritize the management of unknown species and integrate the system into standard lab workflows. This automation could facilitate broader adoption of air quality monitoring, making cleaner air a more attainable goal.

7 Acknowledgements

Thanks to Melanie Priemer for creating the clean culture samples. Special thanks to Nadja-Stella Voce for the extensive annotation of the environmental data, which made the evaluation of the models possible.

References

1. VDI: Vdi 6022 blatt 1 - raumlufttechnik, raumluftqualität - hygieneanforderungen an raumlufttechnische anlagen und geräte (vdi-lüftungsregeln) (2018)
2. Agency, U.S.E.P.: Care for your air: A guide to indoor air quality. US EPA (aug 2024)
3. Held, J.: Measuring mold and the impact of mold on human health (August 2024)
4. Cao, K., Brbic, M., Leskovec, J.: Open-world semi-supervised learning
5. Lee, J., Yi, J., Shin, C., Yoon, S.: Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation (2021)
6. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10) (2010) 1345–1359
7. Niel, O., Bastard, P.: Artificial intelligence in nephrology: core concepts, clinical applications, and perspectives. American Journal of Kidney Diseases **74**(6) (August 2019) 803–810
8. Sun, Y., Ye, Z., Zhong, M., Wei, K., Shen, F., Li, G., Yuan, J., Xing, C.: Rapid and nondestructive method for identification of molds growth time in wheat grains based on hyperspectral imaging technology and chemometrics. Infrared Physics & Technology **128** (December 2022) 104532
9. DTU Denmark: Team: Dtu-denmark/measurement - 2020 igem (2020)
10. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2023) 7464–7475
11. Tan, M., Le, Q.V.: Efficientnetv2: Smaller models and faster training. arXiv (April 2021)
12. Brock, A., De, S., Smith, S.L., Simonyan, K.: High-performance large-scale image recognition without normalization. In: Proceedings of the 38th International Conference on Machine Learning, PMLR (July 2021)
13. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2009)
14. Minh, D., Wang, H.X., Li, Y.F., Nguyen, T.: Explainable artificial intelligence: A comprehensive review. Artificial Intelligence Review **55**(5) (2021) 3503–3568
15. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE (2017)