Engineered Science
DOI: https://dx.doi.org/10.30919/es1441

Rapid Screening Accuracy: Machine Learning-Based Detection System for Species Identification of Aspergillus from Human Clinical Samples Compared with MALDI-TOF MS

Worada Samosornsuk,[1] Pradya Prempraneerach,[2,*] Chollanant Khattiyawech[1] and Panarat Hematulin[3]

Abstract

Rapid and precise *Aspergillus* species-level identification was needed to guide clinicians in effective treatment. Matrix-assisted laser desorption/Ionization time-of-flight mass spectrometer (MALDI-TOF MS) offers high identification accuracy with more complex processes and high-cost instruments, compared to conventional *Aspergillus* culture and microscopy. This research proposes an *Aspergillus* identification technique at the species level from colony images using machine learning (ML) based image processing methods for rapid diagnosis. Twenty-two strains of *Aspergillus* were confirmed as belonging to eight species using the MALDI-TOF MS and then cultured for 3 days on two types of medium. After fungal-colony image collection and enhancement, three main features, including Laws' texture energy measures (LTE), dominant a^*- and b^*- colors, and statistical feature matrix (SFM), were extracted from colony images for training four ML algorithms: K-nearest neighbors (KNN), Gaussian naive bayes (GNB), linear support vector machine (SVM), and random forest (RF). Training accuracy of all MLs using each feature reveals that dominant colors are the most significant feature; it can be further improved by combining colors with SFM, LTE, or both. The highest average accuracy is around 96% after the KNN and SVM are trained and predicted by dominant colors together with SFM and/or LTE for *Aspergillus* cultured on two different media.

Keywords: *Aspergillus* species; Identification; Color features; Texture Energy; Machine learning; MALDI-TOF MS.
Received: 06 December 2024; Revised: 16 January 2025; Accepted: 20 January 2025.
Article type: Research article.

1. Introduction

Invasive fungal infections in humans are caused by opportunistic pathogens often mutualistic with hosts. An immunocompromised body, associated with risk factors such as medical intervention, treatment-induced immunosuppression, and co-infection, can be susceptible to fungi.[1] WHO is concerned about the increasing number of fungal infections, affecting more than 300 million people and causing about 1.5 million deaths.[2,3] In recent years, many studies have investigated some fungi co-infected with COVID-19 patients, in which *Aspergillus* is one of the most relevant fungi. *Aspergillus* spp. is an omnipresent fungus that is commonly found in contaminated food and the environment; it can produce various mycotoxins such as aflatoxins and ochratoxin A, which affect human and animal health at nano-concentrations as well as cause fungal infections and symptoms such as irritation, allergy, cancer, and immunodepression. Thus, *Aspergillus* species are considered a dangerous fungus that needs to be addressed for better treatments and for co-infection prevention.[4-7] Various studies have revealed the fatal incidences and risk factors of *Aspergillus* spp. The fungal diagnosis is still challenging; poor identification can affect treatment failure and delay, and cause the dissemination of fungal infections and patient deaths.[8-10] In the clinic analysis, many pathogenic identification techniques ranging from conventional to advanced methods can be summarized in Table 1.[11] Each technique needs an expert with specialized knowledge and training to interpret results for the accurate diagnosis of fungal infection.

Fungal identification traditionally relies on visual characteristics of fungus culture, which can take at least four days and requires skilled experts for fungal extraction and

[1] *Department of Medical Technology, Faculty of Allied Health Sciences, Thammasat University, Pathum Thani, 12120, Thailand*
[2] *Department of Mechanical Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathum Thani, 12120, Thailand*
[3] *Microbiology Laboratory Unit, Thammasat University Hospital, Pathum Thani, 12120, Thailand*
**Email:* ppradya@engr.tu.ac.th (P. Prempraneerach)

Table 1: The overview of several clinical fungal identification techniques.

Techniques	Level	Specialist	Time	Sensitivity	limitation	Price
Fungal culture	species	✓	≥ 4 days	Low	Medium selection	Low
Microscopy	-	✓	≥ 4 days	Low	Similar microscopic	Low
Histopathology	-	✓	2-3 weeks	Low	Tissue analysis	Low
MALDI-TOF MS	species	-	~1 hr.	High	Database	Low (sample) High (machine)
PCR	Gene	✓	1 days	High	Selection of primer	High

cultivation.[11,12] Traditional methods may lead to misidentification due to fungal-colony morphology and colour similarities. Modern molecular biological techniques like real-time polymerase chain reaction (PCR) and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) have been utilized for fungi identification with higher reliability.[11,13] Real-time PCR requires complex processes and can be prone to contamination due to its sensitive nature. On the other hand, MALDI-TOF MS offers a rapid, accurate identification procedure, but it requires substantial investment and regular database updates. Therefore, MALDI-TOF MS is currently regarded as the gold standard for fungal identification with high accuracy.

Various artificial intelligence (AI) and machine learning (ML),[14-16] including supervised, unsupervised, semi-supervised, and reinforcement algorithms, have been applied to perform the fungi classification.[17,18] The XMVision Fungus AI,[19] developed based on a ResNet-50 transfer-learning model, was trained with nearly 100,000 morphological fungi images (398 strains, 39 molds). As a result, this system could achieve an identification accuracy of up to 93%, comparable to that of microbiology experts. Using ML method for *Aspergillus* identification,[18] 100 morphology images of four species were first collected, and then an active region-based segmentation was employed to remove backgrounds; the support vector machine (SVM) classifier achieving up to 99% training accuracy using 48 eigenvalues, obtained from principal component analysis (PCA) after feature extraction using histogram of oriented gradients (HOG). However, due to morphological ambiguity, identifying fungal species in clinical specimens remains challenging. Many researchers rely on microscopic images for fungal identification. Tochigi *et al*.[20] used AI to distinguish between *Aspergillus* and mucorales, confirmed by the PCR method, through microscopy and image processing techniques. Features like hyphal edges and angles were plotted in 2D, achieving distinct segregation, while traditional methods required the cultivation on agar and expert evaluation of the dye decolourization. Arredondo-Santoyo *et al*.[21] studied dye colour changes in basidiomycetes and *Trichoderma* spp., creating a dataset of 235 images. Feature extraction using Haralick and HOG was trained by six ML classifiers, and the Resnet-C-SVM model achieved the highest accuracy of 96.5%. Other studies on fungal classification for skin diseases found that ML models using k-nearest neighbors (KNN),[22] SVM, and others with PCA achieved accuracy rates of 94%-98%, outperforming those without PCA. Similarly, Aditya *et al*. employed ML algorithms,[23] including a convolutional neural network (CNN), to predict hair diseases. CNN and random forest (RF) can achieve accuracies of 92% and 88%, respectively. Many studies now focus on isolating fungi for rapid disease identification and cost-effective treatment.[24]

This paper proposes an *Aspergillus* identification technique at the species level from colony images using ML-based image processing methods for rapid diagnosis.

2. Materials and methods

2.1 Collection of clinical fungi

The 34 strains of *Aspergillus,* collected from different patients by Thammasat Hospital Laboratories, were subcultured on potato dextrose agar (PDA) and Sabouraud dextrose agar (SDA) for 4 days and incubated at room temperature. Moreover, all samples consisted of leftover isolates for which the patients' clinical histories were unavailable. *Aspergillus* colonies must be grown in a closed system with no air circulation to reduce the spread of spores and prevent other microbial contamination on agar plates. All isolates were confirmed for genus and species using the MALDI-TOF MS technique as an expert database for ML.

2.2 MALDI-TOF MS

MALDI-TOF MS is a technique for proteomic molecular analysis and the mass spectrum acquisition of an unknown organism, measured to the ratio of mass-to-charge (m/z) and compared with reference databases for identification.[25-27] For appropriate fungal growth, the overall preparation steps for MALDI-TOF MS sample extraction were demonstrated.[28] For MALDI-TOF MS measurement, the spectrum pattern was established by the FlexControl™ software 3.4 on the autoflex maX™ TOF/TOF mass spectrometer (Bruker Daltonik Inc., Bremen, Germany) in linear positive mode (a range of 2k-20k m/z). A bacterial test standard (BTS) such as Escherichia coli ATCC 25922 is a calibrator. All acquired fungal mass spectrums were confirmed as genus and species, as well as reported score values on MALDI Biotyper (MBT) Compass Explorer, library version 4.1 (9,999 entries; Bruker Daltonics),

together with the MT filamentous fungi library 3.0 and MBT Compass library revision L. In addition, the confidentiality in interpreting depends on the score obtained. A score greater than or equal to 2 was highly probable genus and species identification, whereas a score within a range of 1.7-1.9 was probable genus identification except for species, and a score less than 1.7 was not reliable identification.[29,30]

2.3 Dataset and instruments for fungal colony image acquisition

To generate a dataset of fungal colonies, digital images with a resolution of 6048×4024 pixels were taken by a Nikon Z6 camera equipped with a NIKKOR Z MC 105 mm Macro Lens (VR, ISO1250). An isolation box with vertical sliding doors was used to prevent fungal spore dispersion during imaging, as shown in Fig. 1 of the supplementary section. After each imaging session, this box was autoclaved to prevent cross-contamination. Thus, it needed to be built with high-temperature resistant materials, like tempered glass and polyoxymethylene, to withstand up to 120 °C. A total of 800 images (100 images per species) with variations in petri-dish placement locations were collected. As illustrated in Fig. 1 of the supplementary section, white-light illumination was provided through the front tempered-glass plate for a lighting-controlled condition.

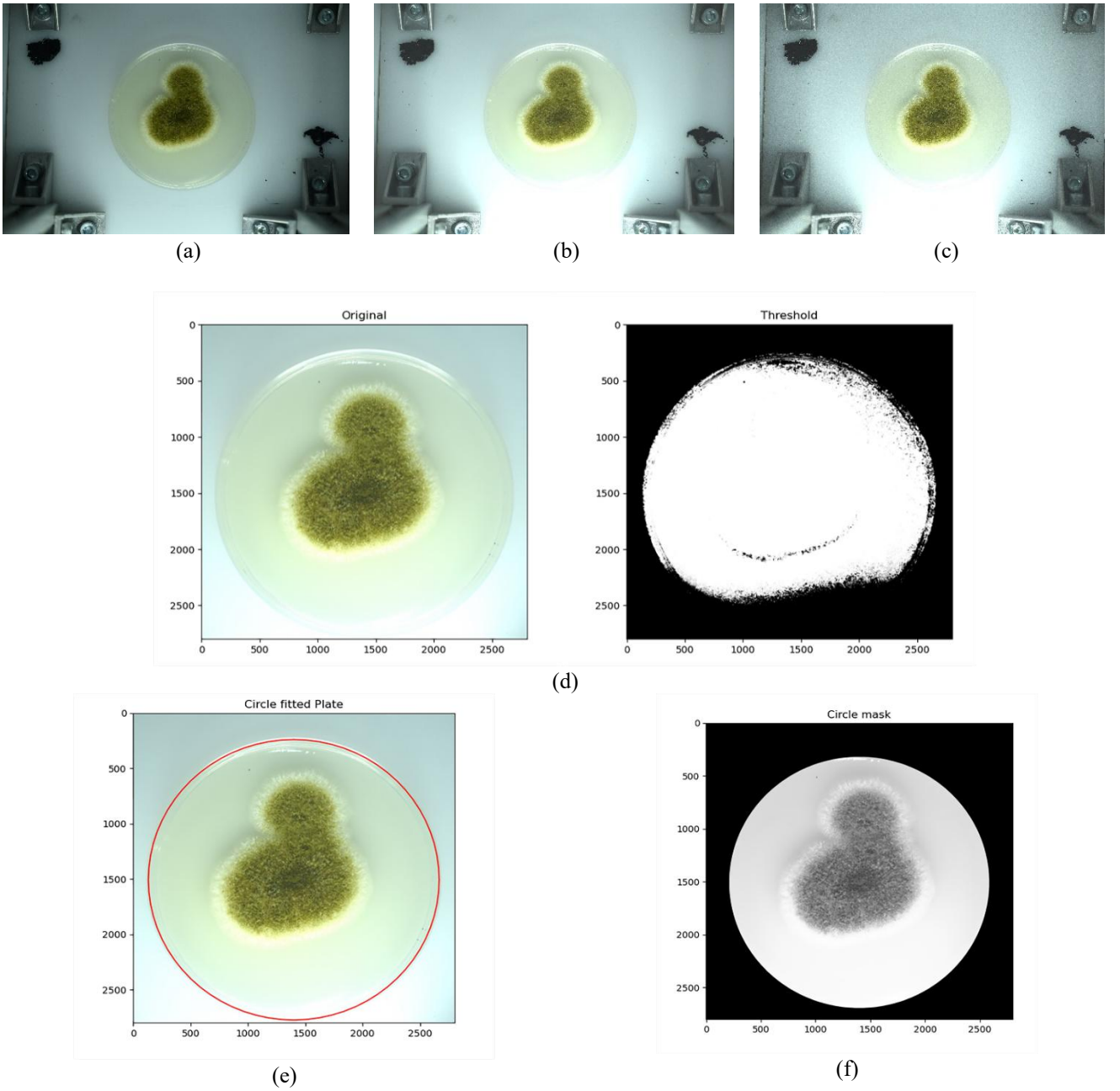


Fig. 1: Original image of 6048×4024 pixel (a), Images after adjusting contrast and brightness (b), Image after enhancing with the 3×3 sharpen kernel (c), Cropped image with resolution of 1800×1800 pixel and thresholding image (d), red circle fitted around the agar plate (e), and segmented agar-plate area for further extracting texture/color features (f).

2.4 Image enhancement and segmentation

Initially, image quality needs to be improved by adjusting brightness/contrast and sharpening all colony images. First, image contrast and brightness can be enhanced by a linear transformation, $g_i(x,y) = \alpha \cdot f_i(x,y) + \beta$, where $f_i(x,y)$ are all original pixels in the standard RGB colour space for $i = 1, 2, 3$ using a contrast scale factor (α) of 1.2 and a brightness scale factor (β) of 40, as demonstrated in Fig. 1(b).[31] Second, all images are then sharpened by convolving with a 3×3 sharpen kernel, shown in Fig. 1(c). Third, after converting an image from the RGB to L*a*b* colour space,[32] a colour-image segmentation using a threshold operation of the *b** channel is applied to the previously enhanced image so that the agar-plate area can be separated from the white background in Fig. 1(d). Fourth, the outer rim of the circular agar plate is automatically detected, and then the colony image is cropped to a square area of 1,800×1,800 pixels with the origin of the circular plate at the image center, as illustrated in Fig. 1(e). Also, the resolution of the cropped circular plate is reduced from 300 dpi to 96 dpi. This color-segmentation technique helps increase processing time in further feature-extraction stages. Additionally, both the perimeter and the area of the culture plate are calculated and used as validation in this colour-segmentation process.

2.5 Feature extraction

Three feature techniques: 1) texture using law's texture energy (LTE), 2) colour from *a** and *b** channels, and 3) statistical information using statistical feature matrix (SFM), are the main attributes for identifying and classifying eight different species of *Aspergillus*. The reasons for selecting these three features are that the roughness and fluffiness of the fungal colony, as well as the colony center, can be characterized by its surface texture and chromatic color, respectively. Furthermore, statistical information, extracted from local regions of the colony surface, can capture repetitive characteristics in either intensity or colour.[33]

2.5.1 LTE measures

LTE measures,[34,35-41] one of the transform-based approaches, which describes an energy variation in terms of statistical information, especially a coarseness, within a fixed-size window. Theoretically, 2D filter kernels larger than five dimensions can be created from three 1D arrays: L3 = [1,2,1], E3 = [-1,0,1], and S3 = [-1,2, -1] for correspondingly detecting Level, Edge, and Spot in an image microstructure. In this study, the 2D kernel dimension for LTE measures is specified beforehand to be 7×7, and the following combinations of 2D kernels, LL, EE, SS, LE, EL, ES, SE, LS and SL, are generated to achieve rotational invariant after convoluting with enhanced images and selected as six textural-energy features. Each letter represents a unique characteristic: L indicates Level or Low-pass (uniform intensity regions), E stands for Edge (edge detection), and S denotes Spot (small, localized features). The pairwise combinations describe interactions between these kernels to extract various texture patterns: LL identifies smooth and uniform areas, EE emphasizes prominent edges, and SS highlights localized spots, while combinations such as LE, EL, ES, SE, LS, and SL, used in this research, capture transitions and interactions between these kernels to extract different texture patterns.

2.5.2 L*a*b* color histogram

The statistical colour distribution in an image can be used to distinguish different fungal colonies among 8 *Aspergillus* species. Generally, the object colour is described in different color spaces, such as RGB, HSV, and L*a*b*,[42-44] each colour space has its advantages and unique characteristics, which might be suitable for diverse applications. In this research, colour distribution from *a** and *b** channels is analyzed using a 1D histogram because the distinct chroma of the fungal colony can be easily identified. After the culture-plate segmentation process, enhanced images are converted from RGB to L*a*b* space. Then, 1D histograms of *a**- and *b**-channels are derived from a center area of images, which are cropped to a resolution of 500×500 pixels, to focus only on the colony chroma. Lastly, peaks of the *a**- and *b**- histograms, called max_*a* and max_*b*, are detected and extracted as the color-feature attributes defined below.

1) max_*a* - a dominant color in the *a**-histogram of the cropped center-area of image.

2) max_*b* - a dominant color in the *b**-histogram of the cropped center-area of image.

2.5.3 SFM

SFM is a spatial interrelation feature for analyzing texture properties,[33,34,39,45-47] derived from inter-sample pixel pairs at various distances. In this research, 4 SFM attributes, including 1) periodicity, 2) contrast, 3) coarseness, and 4) roughness, are employed as primary texture features. Contrast, coarseness, periodicity are correspondingly computed based on 1) a contrast matrix, derived from a mean square difference of intensity at various fixed inter-sample spacing distance (δ), 2) a dissimilarity matrix, calculated from a mean difference of intensity at various δ, 3) a difference between mean of the dissimilarity matrix and its minimum location, thus these three features of SFM could characterize both local and global texture variations of the fungal colony. On the other hand, the roughness feature is derived from a fractal dimension, describing the inter-sample variation at randomized spacing distances.

2.6 ML algorithms

After deriving 12 attributes: 6 LTE features, two dominant-color features, and 4 SFM features in the feature-extraction process, 4 ML algorithms are employed to identify different *Aspergillus* species according to these features and species labels. In this research, four main supervised ML algorithms

for classifying *8 Aspergillus* species are KNN, Gaussian naive bayes (GNB), SVM, and RF. Due to the high correlation among texture features, these 4 ML algorithms are more appropriate to achieve high prediction accuracy. Initially, 800 datasets are randomly partitioned into two distinct subsets: 536 (or 67%) as a training set and 264 (or 33%) as a validation set.

2.6.1 KNN

The KNN algorithm is one of the fundamental classification techniques for the classification of *n*-dimensional datasets based on a specified distance criterion among all *n*-dimensional training data points.[48-51] This technique operates on the premise that an *I* query point ($q_i \in R^n$), located within a similar proximity of training data points ($p_i \in R^n$) for i =1, ..., m from all m observations, is more likely to belong to the same class. By determining the k nearest neighbors of each q_i data point, the KNN algorithm with k = 5 can facilitate an exploration of the query point according to the Euclidean-distance metric ($d(p,q)$) in Eq. (1) through a search method.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

In a prediction phase, the KNN algorithm can assign a class (y) to a given query point ($q_{i,new}$), based on the majority of the class label ($L(Y)$) among its k nearest neighbors ($N_k\left(q_{i,new}\right)$),[47-50] as described as in Eq. (2).

$$Class\left(q_{i,new}\right) = \arg\max_{c} \sum_{p_i \in N_k(q_{i,new})} L(Y_i = y) \quad (2)$$

2.6.2 GNB

The GNB, one of the probabilistic supervised classifiers, is developed based on the Bayes' theorem that fits Gaussian/normal distributions to all training data in this 12-dimensional feature space under an assumption of independence among all features ($x_i \in R^n$). According to Bayes' theorem, a posterior probability of the y-class hypothesis for a given feature vector $(x_1, x_2, \ldots, x_n)$ can be described as in Eq. (3). Assuming all features (x_i) are uncorrelated for a given y class, Bayes' theorem can be simplified as in Eq. (4).[52-54]

$$P(y|x_1,\ x_2,\ \ldots,\ x_n) = \frac{P(y)P(x_1, x_2, \ldots, x_n|\ y)}{P(x_1, x_2, \ldots, x_n)} \quad (3)$$

$$P\ (y|x_1,\ x_2,\ \ldots,\ x_n) = \frac{P(y)\prod_{i=1}^{n}P(x_i|y)}{P(x_1, x_2, \ldots, x_n)} \quad (4)$$

where $P(x_1,\ x_2,\ \ldots,\ x_n)$ represents an occurrence probability of all features together, derived from the Law of total probability as a fixed variable such that the posterior probabilities are proportional to a product of the prior probability ($P(y)$) and the likelihood of each feature ($P(x_i \mid y)$). The classification rule ($\hat{y}$) employs a maximum a posteriori (MAP), formulated by Eq. (5).

$$\hat{y} = arg\ \max_{y}\ (P(y)\prod_{i=1}^{n}P(x_i|y)) \quad (5)$$

The GNB algorithm calculates both the mean (μ_y) and the standard deviation (σ_y) of the y class and estimates the Gaussian distribution from the training dataset to achieve the highest possible probability.[55,56] Specifically, the likelihood function $\left(P(x_i|y)\right)$ in the MAP estimation can be computed by Eq. (6).

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)}{2\sigma_y^2}\right)^2 \quad (6)$$

2.6.3 SVM

Another prominent supervised ML algorithm is the SVM, which constructs maximum-margin hyperplanes, providing the greatest linear separation among the 12 features of these eight *Aspergillus* species. A given training dataset (x, y) consists of m feature vectors ($\boldsymbol{x}_i \in R^n$) and m target labels of k classes ($y_i \in R^k$) for i = 1, …, m. The separating hyperplane is then defined by the affine function: $\boldsymbol{y} = \boldsymbol{w}^T \cdot \boldsymbol{x} + b$, where $\boldsymbol{w}, b, \boldsymbol{x}$ correspondingly are a normal vector to the linear hyperplane, a bias, and a feature vector. Moreover, the minimum separation distance or margin (γ_i) of this hyperplane from any sample $\mathbf{x}_i$ in the training set (D_N) can be calculated by Eq. (7):

$$\gamma = \min_{x_i \in D_N} \gamma_i = \min_{x_i \in D_N} \frac{y_i(\boldsymbol{w}^T \cdot \boldsymbol{x}_i + b)}{\|\boldsymbol{w}\|} \quad (7)$$

Using the max-min optimization, the optimal $\boldsymbol{w}^*$ and b^* of the maximum-margin hyperplane can be obtained by searching for the maximum separation distance or $\{\ \boldsymbol{w}^*, b^* = \underset{\boldsymbol{w},b}{argmax}\gamma$. From a geometric perspective, maximizing the margin (2γ) between two hyperplanes is equivalent to minimizing $||\boldsymbol{w}||^2 = \boldsymbol{w}^T \cdot \boldsymbol{w}$. Therefore, this optimization problem can be reformulated in Eq. (8) and solved by quadratic programming.[50,57]

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{w}^T x_i + b) \geq 1,\ i\ =\ 1, 2, \ldots, m \quad (8)$$

Even though 12 selected color and texture features of the fungal species are not linearly separable, the linear SVM can still sufficiently and precisely classify them into different classes, similar to the prior studies.[52,53]

2.6.4 RF

Other supervised learning algorithms are based on an ensemble learning method, including the random forest, which aggregates multiple tree-based learner models to improve overall predictive capability and robustness. For the random forest, the base learners are many decision trees that form a hierarchical, tree-like structure where internal nodes correspond to features, branches denote decision rules, and terminal (leaf) nodes represent class assignments.[54-56] For a given training dataset $(\boldsymbol{x}, y)$ with $\boldsymbol{x}_i \in R^n$ and $y_i \in R^k$, the m^{th} node in the decision tree, containing Q_m data, can be split

into left- and right-subsets according to an optimal splitting criterion (θ^*) in Eq. (9).

$$\theta^* = arg\ \min_{\theta} G(Q_m, \theta) \quad (9)$$

where $\theta = (j, t_m)$ is a splitting criterion for a j feature and a threshold of t_m and a split quality at the m node ($G(Q_m, \theta)$) is computed from the "Entropy" or log loss function ($H(Q_m)$), expressed in Eq. (10), to handle uncertainty and impurity in datasets.

$$H(Q_m) = -\sum_k p_{mk} \cdot log(p_{mk}) \quad (10)$$

where p_{mk} denotes a proportion of k-class observations from all training samples in node m, expressed in Eq. (11). An indicator function (I) returns 1, if the y target belongs to class k. Alternatively, I return 0, otherwise.

$$p_{mk} = \frac{1}{n_m} \sum_{x_i \in R_l} I(y_i = k) \quad (11)$$

where n_m represents the total number of training observations belonging to the region R_l. After constructing the decision tree for k-class classification, $\mathbf{x}$ in each region R_l can be classified to the majority class according to the highest number of voting, as described in Eq. (12).

$$k_l^* = arg\ \max_{k} (p_{mk}) \quad (12)$$

A bootstrap resampling can generate numerous modified training subsets from the initial training dataset to train multiple decision trees. Then, the RF creates ensembles of randomized decision trees and iteratively selects random subsets of features to identify the best criteria for splitting and to derive decision nodes. Subsequently, the RF algorithm, comprising 100 trees and a maximum tree depth of 3, is deployed on the fungal training dataset after several parameter tunings.

Fig. 2 summarizes the entire process as a flowchart, from image acquisition to the fungal-species prediction by trained machine-learning models.

3. Results and discussion

3.1 Identification of *Aspergillus* using MALDI-TOF MS

After fungus strains from different patients were identified by MALTOF-MS as *Aspergillus* spp., only eight different species: 1) *Aspergillus niger*, 2) *Aspergillus fumigatus*, 3) *Aspergillus flavus*, 4) *Aspergillus japonicus*, 5) *Aspergillus tamarii*, 6) *Aspergillus terreus*, 7) *Aspergillus nidulans*, and 8) *Aspergillus sclerotiorum* were selected with the following proportion of 18.18, 18.18, 18.18, 18.18, 9.09, 9.09, 4.55, and 4.55%, respectively, according to a cumulative encounter rate in the Thammasat University hospital. The mass spectrum of each strain, which has a high confidence level with a log score above 2.0 from the MALDI-TOF MS identification, was merely chosen in this research to guarantee the certainty of the *Aspergillus* species. The main spectrum profiles (MSPs) dendrogram illustrates the relationships among different fungal strains. Strains belonging to the same species will have a high degree of similarity and will be grouped together. In this study, 34 strains can be categorized into eight species, as shown in a diagram in Fig. 3.

3.2 The morphology and characteristics of fungal colonies

After the MALDI-TOF MS identification, colony images of 8 different *Aspergillus* species in Table 2, cultured on PDA and SDA media, were collected, enhanced, and cropped such that only an area of the Petri plate containing the fungal colony was extracted for further feature-extraction process. All *Aspergillus* colonies in this research shared distinct and similar morphologies and characteristics, such as shape, colour, and texture.

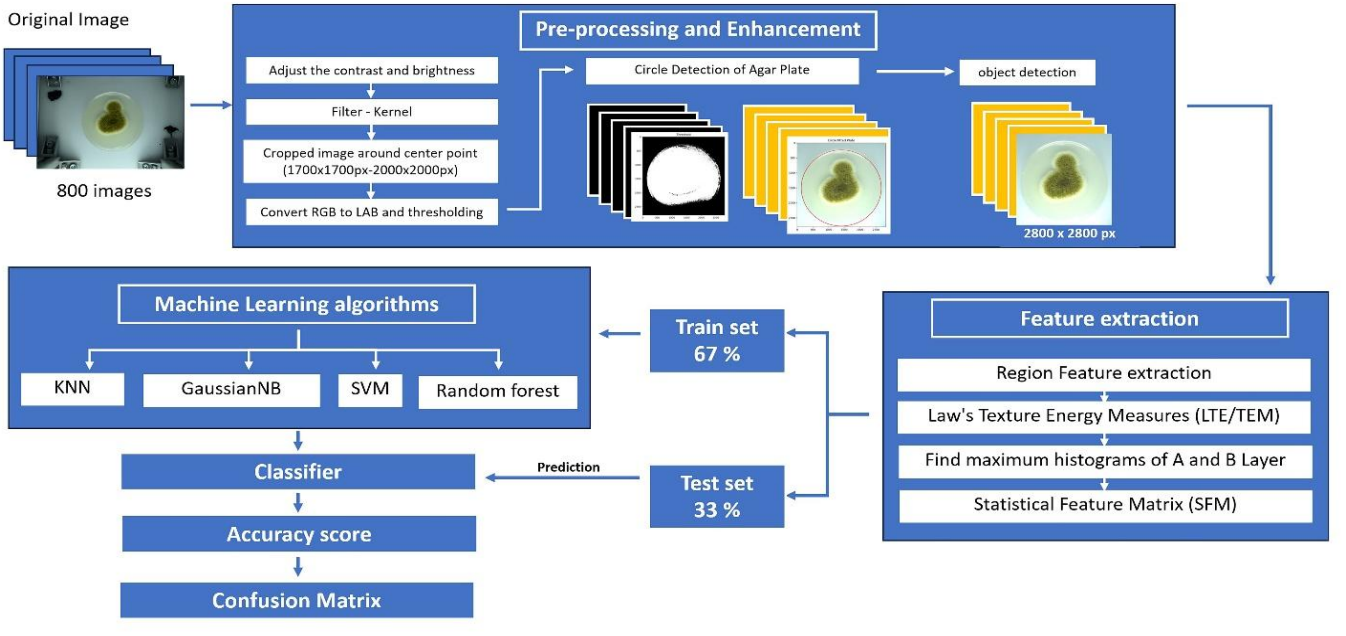


Fig. 2: The overall procedure flowchart of the fungus-image analysis and classification.

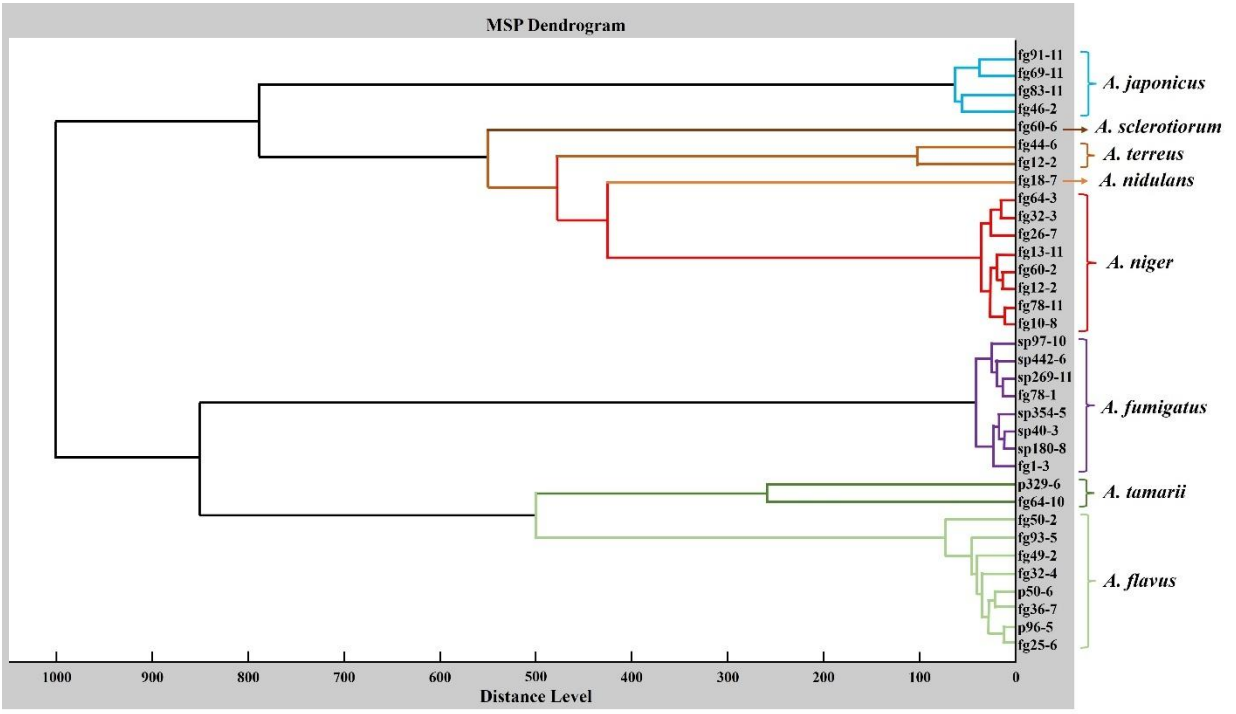
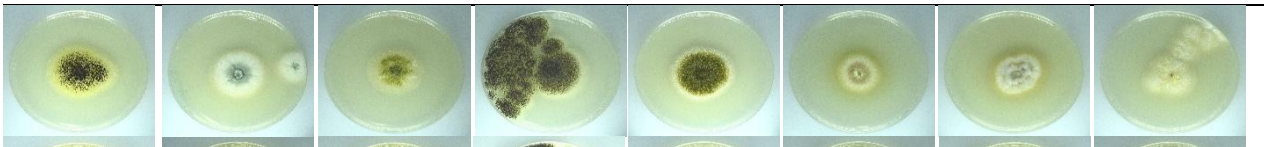
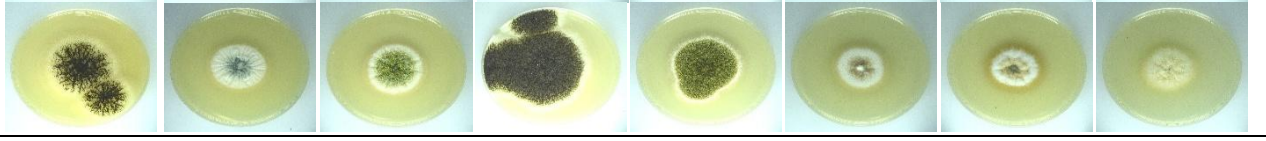


Fig. 3: The MSP Dendrogram, representing the relationships among 8 species of *Aspergillus*, is constructed based on the similarities and differences in the species.

Table 2: Morphologies and characteristics of the *Aspergillus* colonies on PDA and SDA media for specific genus and species, confirmed by the MALDI-TOF MS identification.

Media	*Aspergillus* colonies							
	A. niger	*A. fumigatus*	*A. flavus*	*A. japonicus*	*A. tamarii*	*A. terreus*	*A. nidulans*	*A. sclerotiorum*
PDA								
SDA								

3.3 Featured selection

Numerical results of the texture and colour features were extracted from 22 stains of *Aspergillus* colonies cultured on PDA and SDA media, illustrated in Tables 3 and 4, respectively. Some pairs of *Aspergillus* species, such as (*A. tamarii*, *A. flavus*) as well as (*A. niger*, *A. japonicus*), exhibit similar LTE, SFM texture values, and maximum *a**- and *b**-color values (or max *a***b** colors). In this research, the classification of the *Aspergillus* species can be evaluated based on seven groups of features: 1) LTE, 2) max *a***b** colours, 3) SFM, 4) a combination of LTE and max *a***b** colors, 5) a combination of SFM and max *a***b** colours, 6) a combination of LTE and SFM, and 7) combined LTE, SFM, and max *a***b** colours. Then, four ML algorithms: 1) KNN, 2) GNB, 3) SVM, and 4) RF were trained with these seven groups of features.

3.4 Statistical cross-validation of ML techniques

To statistically assess the four MLs' performance, 1,000 cross-validations with random shuffling were employed to split extracted features into training and validation datasets. According to these four ML algorithms, distributions of these cross-validation accuracies were analyzed based on various feature combinations of fungal colonies cultured on PDA medium (in Table S1) and SDA medium (in Table S2), and their means and standard deviations were derived. Relying only on individual feature-extraction techniques, the KNN, trained with the maximum *a**- and *b**- colors, reveals the highest training accuracy of 94.8% and 84.9% for cultures on PDA and SDA media, respectively. On the other hand, neither individual LTE nor SFM feature is sufficient to differentiate among these eight *Aspergillus* species. After combining either LTE or SFM feature with the maximum-colour feature, average training accuracies could be improved by up to 89% or more using these four ML algorithms. The KNN provides the highest average accuracy of 96.91±0.01% after training with the combination of SFM and maximum-colour features

Table 3: Comparing numerical values of both texture and colour features among 8 different *Aspergillus*-species colony images on the PDA medium.

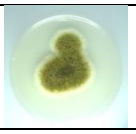
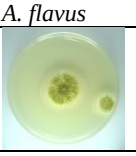
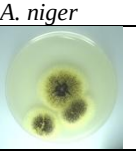
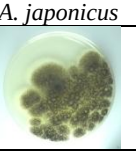
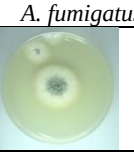
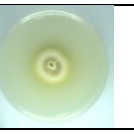
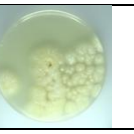
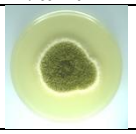
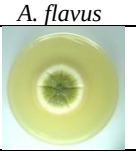
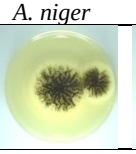
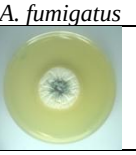
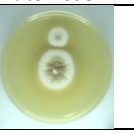
On PDA	*A. tamarii*	*A. flavus*	*A. niger*	*A. japonicus*	*A. fumigatus*	*A. nidulans*	*A. terreus*	*A. sclerotiorum*
Colony image								
LTE								
LL	439966.86	402887.58	408976.17	383659.54	398973.22	376574.62	402143.13	390734.92
EE	1244.89	1091.05	3022.70	2570.25	870.37	908.92	945.84	1016.82
SS	274.28	263.60	556.42	473.73	225.03	241.48	237.62	256.04
LE	6931.79	5901.34	14640.05	13617.17	4916.57	4974.54	5381.38	5478.38
ES	557.60	512.93	1256.43	1051.28	421.21	447.46	451.48	487.53
LS	2426.55	2103.58	5664.85	4989.94	1638.84	1709.30	1790.56	1887.17
Color								
max *a**	118	111	126	123	114	117	116	114
max *b**	160	176	131	136	138	141	157	153
SFM								
Coarseness	17.34	16.95	8.65	7.015	20.39	18.87	19.66	17.97
Contrast	15.35	14.38	28.45	27.17	12.61	13.26	13.38	13.63
Periodicity	0.32	0.21	0.40	0.47	0.16	0.16	0.17	0.19
Roughness	2.73	2.82	2.60	2.55	2.88	2.87	2.87	2.85

Table 4: Comparing numerical values of both texture and color features among 8 different *Aspergillus*-species colony images on the SDA medium.

On SDA	*A. tamarii*	*A. flavus*	*A. niger*	*A. japonicus*	*A. fumigatus*	*A. nidulans*	*A. terreus*	*A. sclerotiorum*
Colony image								
LTE								
LL	408714.2	390616.8	422782.0	413039.5	369115.8	371867.0	366317.7	388132.4
EE	1507.41	1490.69	2182.47	1855.36	918.01	920.61	999.80	996.16
SS	324.92	323.93	424.75	371.13	240.38	237.90	252.29	245.85
LE	7822.47	7404.64	10909.13	9769.50	5126.10	5132.93	5447.06	5466.63
ES	671.86	670.48	926.46	794.44	447.36	444.72	479.78	471.63
LS	2861.79	2803.28	4093.44	3594.50	1723.54	1725.85	1883.11	1854.90
Color								
max *a**	116	110	125	125	116	117	117	116
max *b**	151	164	131	133	137	149	148	148
SFM								
Coarseness	13.39	14.21	11.14	10.54	18.68	18.99	17.91	18.92
Contrast	17.39	16.88	21.98	20.73	13.02	13.16	13.55	13.23
Periodicity	0.32	0.27	0.37	0.43	0.16	0.16	0.18	0.19
Roughness	2.71	2.75	2.63	2.60	2.87	2.87	2.85	2.84

of *Aspergillus* colonies cultured on the PDA medium. The SVM provides the best average accuracy of 96.32±0.01% after training with a combination of all extracted features from *Aspergillus* colonies on the SDA medium. On average, SVM and KNN algorithms outperform RF and GNB methods for these feature datasets.

3.5 Evaluation of *Aspergillus* identification performance

To assess the prediction performance of trained KNN and SVM models for identifying *Aspergillus* species, 50 additional colonies images of the eight *Aspergillus* species, which obtained from 12 new strains (4 new strains of each species: *A. flavus, A. niger*, and *A. fumigatus*) and 10 previously used strains (Resubculture and photograph *A. japonicus*, *A. nidulans*, *A. sclerotiorum*, *A. tamarii*, and *A. terreus*) on both PDA and SDA media were employed. The predictive-classification performance of trained ML models on these unseen colony images can be assessed by the confusion matrix.

Confusion matrices can evaluate the identification accuracy of *Aspergillus* colonies on the PDA medium using the KNN model and on the SDA medium using the SVM model, as shown in Figs. 4 and 5, respectively. Based on the confusion matrix, True positive (TP) along a diagonal, False positive (FP) within a lower triangular, and False negative (FN) within an upper triangular can reveal an amount of correct and incorrect identification of 8 *Aspergillus* species as well as illustrate no overfitting and underfitting problems using the trained ML models. In both Figs. 4 and 5, yellow color and green color correspondingly represent 8-image TP/correct predictions and 7-image TP/correct predictions, while dark purple color indicates low value or zero image of both FP and FN predictions. According to trained KNN and SVM models, 48 out of 50 images are correctly classified, resulting in an overall predictive accuracy of 96%. Thus, both confusion matrices demonstrate the trained models' high predictive performance and low classification error across these 8 *Aspergillus* species.

According to Table 5, performance metrics of the classification report, including F1 score, accuracy, and weighted-average F1 score,[58,59] summarize the overall species-identification performance of the trained models. The F1-score metric is computed from a combination of precision and recall, and accuracy is defined as the ratio of the correct identifications to the total number of identifications. At the same time, a weighted-average F1-score is calculated from the mean of all per-species F1-scores weighted by a normalized support/occurrence of each species, which sums up the predictive identification performance. Employing 50 *Aspergillus*-colony test images on the PDA medium, the KNN model, trained with the combination of SFM and maximum-colour features, achieves both high predictive accuracy and a high weighted-average F1 score of 96.00%. However, the model misidentification occurs between *A. japonicus* (or class 2) and *A. niger* (or class 4) with low prediction rates, as illustrated in Table 5. The predictive performance of the SVM model, trained with all 32 features, was evaluated on another 50 test images of *Aspergillus* colonies on the SDA medium; the same predictive accuracy and weighted-average F1 score of approximately 96.00% are reported in Table 5. Likewise, the misidentification between *A. niger* (or class 4) and *A. japonicus* (or class 2) are still observed as well. The main reason for the misidentification between *A. niger* and *A. japonicus* using either trained KNN or SVM model is that colony images of these two species exhibit very similar colour, texture, and morphology within the first three days of the incubation. The maturity of fungal colonies can be one of the key misidentification sources. If the sample image is taken from an immature fungus, its morphological and texture features may be complicated for correct species identification.

3.6 Discussion on fungal-species identification

Until now, the ML techniques for fast fungi-species identification relying on fungal-colony images are still absent in the literature. Most existing researches on species-identification machine learning utilize microscopic images of fungi spores for analytical and classification purposes. These ML and traditional microscopic approaches were associated with several limitations, including a need for staining before examination under the microscope and an expert

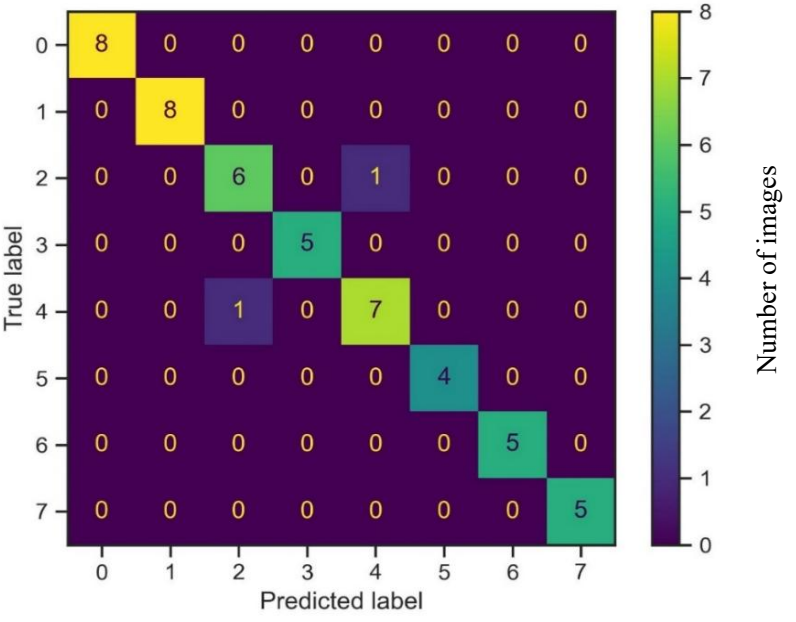


Fig. 4: Confusion matrix illustrates the effectiveness of *Aspergillus*-species identification using the trained KNN model, when 50 test images are acquired from unseen fungi colonies, cultured on the PDA medium. Using this trained KNN model, high TP predictions along the confusion-matrix diagonal, represented by yellow and green colors, and low FP and FN predictions, illustrated by dark purple color, indicate the *Aspergillus*-species identification efficacy. Note: label numbers from 0 to 7 denote these 8 *Aspergillus* species in the following order: *A. flavus, A. fumigatus, A. japonicus, A. nidulans, A. niger, A. sclerotiorum, A. tamarii,* and *A. terreus* species, respectively.

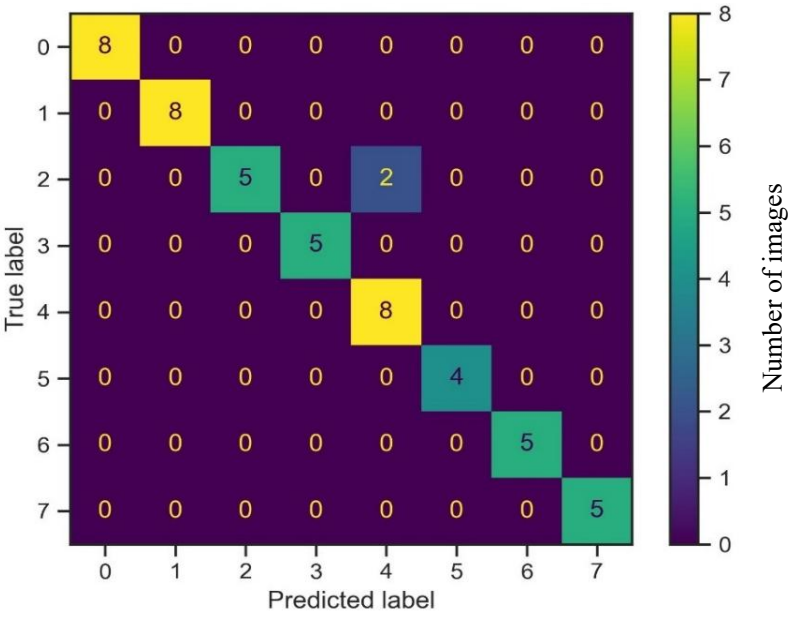


Fig. 5: Confusion matrix illustrates the effectiveness of *Aspergillus* species identification using the trained SVM model, when 50 test images are acquired from unseen fungi colonies, cultured on the SDA medium. Using this trained SVM model, high TP predictions along the confusion-matrix diagonal, represented by yellow and green colors, and low FP and FN predictions, illustrated by dark purple color, indicate the *Aspergillus*-species identification efficacy. Note: label numbers from 0 to 7 denote these 8 *Aspergillus* species in the following order: *A. flavus, A. fumigatus, A. japonicus, A. nidulans, A. niger, A. sclerotiorum, A. tamarii,* and *A. terreus* species, respectively.

Table 5: Performance metrics, based on accuracy, F1 score, and the weighted-average F1 score, are derived from the predictive identification by the trained models using 50 unseen images of 8-species *Aspergillus* colonies, cultured either on PDA or on SDA medium.

Medium type	PDA	SDA
Class	F1-score	F1-score
0	1.00000	1.000000
1	1.00000	1.000000
2	0.85714	0.833333
3	1.00000	1.000000
4	0.87500	0.888889
5	1.00000	1.000000
6	1.00000	1.000000
7	1.00000	1.000000
Accuracy	0.96000	0.960000
Weighted- average F1-score	0.96000	0.958889

interpretation/labelling for the species classification. On the other hand, the accuracy of *Aspergillus* identification at the species level by the MALDI-TOF MS using the Bruker Filamentous Fungi Database v3.0 can achieve up to 85.7%.[60,61] The MSI-2 database can be added to improve the identification accuracy further.[61] Once the proper database has been included, the misidentification of the *Aspergillus* species rarely occurs.

However, this research emphasizes the rapid and accurate fungal-species identification within a minimum duration of three days. Thus, this approach, offering a viable alternative to the other methods in the literature, enables clinicians to identify pathogens quickly and to provide initial treatment before confirmation of the fungi species from the conventional methods. Consequently, it has the strong potential to mitigate treatment delays and reduce the incidence of drug resistance in patients. Furthermore, *Aspergillus* species-level identification directly impacts the urgency level of the treatment because of the toxin production, depending on the *Aspergillus* species. Some species of *Aspergillus*, such as *A. niger, A. fumigatus, A. flavus, A. nidulan*s, and *A. terreus* can produce harmful toxins. Mycotoxins and their direct effects on patient health were discussed in several studies.[62-64] In contrast, some species do not generate any toxic substance, namely some strains of *A. flavus*, which inhibit aflatoxin synthesis.[65,66] Moreover, some species, such as *A. giganteus*, *A. longivesica*, and *A. clavatus* can be utilized in antibacterial, antiviral, and antiprotozoal activities.[62]

The maximum *a**- and *b**- color is the most significant feature for identifying/classifying these eight *Aspergillus* species according to fungal-colony morphology and characteristics. Combining the maximum *a**- and *b**- color feature with either the SFM or LTE feature can further improve the identification accuracy. By integrating all 12 extracted features, the average training accuracy of these four ML algorithms ranges between 83% and 95%, which is comparable to those using a combination of the maximum colour feature and either SFM or LTE alone. Among these four

ML methods, the GNB method consistently provides low average accuracies after training with these extracted features. This result suggests that the GNB method is more suitable to well-separated features or large datasets.[67] In addition, the cultural medium for culturing fungal colonies can affect the learning accuracy of the ML algorithms. The growth of isolated *Aspergillus* colonies cultured on different cultural media could exhibit the distinguishing morphological and cultural characteristics, as described in the research by Nalawade *et al*.[68]

4. Conclusion

Several methods for clinical fungal identification, including fungal culture, microscopic imaging, MALDI-TOF MS, and PCR, have been available in hospitals worldwide. Even though the MALDI-TOF MS and PCR are two of the fastest fungal-identification techniques that can generate high-quality laboratory reports, the accessibility of MALDI-TOF MS tends to be limited to large private hospitals due to its high investment and maintenance costs. Therefore, the rapid identification of *Aspergillus* spp. might not be widely accessible for clinical diagnosis and could delay the proper treatment. Our primary research development focuses on 1) providing a low-cost alternative for *Aspergillus* species identification, 2) improving identification accuracy and time within the first 3 days, and 3) reducing the delayed diagnosis and misdiagnosis for further better patient treatment. After identifying eight common clinically found *Aspergillus* species, including *A. flavus, A. fumigatus, A. japonicus, A. nidulans, A. niger, A. sclerotiorum, A. tamarii,* and *A. terreus* species, using the MALDI-TOF MS, texture and colour features such as LTF, SFM, and maximum *a**-and *b**-colors are extracted from enhanced colour images of the fungal colonies, cultured on the PDA and SDA media for 3 days. Then, four supervised ML techniques, including KNN, GNB, SVM, and RF, with derived targets/labels from the MALDI-TOF MS identification, are trained with the individual features and combinations of these three features.

Even though these four ML algorithms can yield better average training accuracy of about 80% using only colour features than average training accuracy of about 60-70% using either LTE or SFM features alone, which implies that the maximum *a**-and *b**- color is more prominent feature than the texture of these eight *Aspergillus* species. The developed fungus-colony image identification can achieve the best training accuracy of 96% using the KNN and linear SVM algorithms when the fungal colonies were cultured on PDA and SDA media, respectively. Furthermore, these trained KNN and linear SVM models demonstrate high predictive identification with the accuracy and weighted-average F1-score of 96% on the unseen *Aspergillus*-colony images. In the literature, few researchers have applied ML or deep learning algorithms for fungal-species identification only from microscopic images to assist patient diagnosis, despite growing attention from the medical communities. Identifying fungal species using ML algorithms and utilizing fungal-colony images has not been previously explored. Nevertheless, the fungi-colony images from the classical method can offer several advantages, including widespread usage and low equipment cost, compared to other molecular techniques. Thus, this proposed fungal-species identification system using ML could be easily scaled up and made accessible to all hospitals, especially in rural areas.

Further exploration of transfer learning algorithms, such as CNNs or visual geometry group-16 (VGG16), can be applied for feature extraction as well as classification of various *Aspergillus* species, which might achieve similar or potentially better accuracy than this proposed fungal-species identification system. Owing to the diversity and complexity of the fungal taxonomy, other fungal pathogens that cause serious infections might exhibit similar morphology and colony characteristics to *Aspergillus* species. An additional collection of colony images, combined with the MALDI-TOF MS identification, should be further investigated in ongoing research to enable quick and accurate identification of the fungal species.

Acknowledgements

This research is financially supported by the Thailand Toray Science Foundation (TTSF) for 2022 and by the Thammasat University Research Fund, Contact No TUFT 68/2566 and Thailand Science Research and Innovation (TSRI) Fundamental Fund, fiscal year 2025.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Applicable.

References

[1] C. Firacative, Invasive fungal disease in humans: are we aware of the real impact?, *Memorias Do Instituto Oswaldo Cruz*, 2020, **115**, 1-9, doi: 10.1590/0074-02760200430.

[2] M. C. Fisher, D. W. Denning, The WHO fungal priority pathogens list as a game-changer, *Nature Reviews Microbiology*, 2023, **21**, 211-212, doi: 10.1038/s41579-023-00861-x.

[3] G. Janbon, J. Quintin, F. Lanternier, C. D'Enfert, Studying fungal pathogens of humans and fungal infections: fungal diversity and diversity of approaches, *Genes & Immunity*, 2019, **20**, 403-414, doi: 10.1038/s41435-019-0071-2.

[4] V. Navale, K. R. Vamkudoth, S. Ajmera, V. Dhuri, Aspergillus derived mycotoxins in food and the environment: Prevalence, detection, and toxicity, *Toxicology Reports*, 2021, **8**, 1008-1030, doi: 10.1016/j.toxrep.2021.04.013.

[5] J. Shankar, An overview of toxins in Aspergillus associated with pathogenesis, *International Journal of Life Sciences Biotechnology and Pharma Research*, 2013, **2**, 16-31.

[6] J. Pemán, A. Ruiz-Gaitán, C. García-Vidal, M. Salavert, P.

Ramírez, F. Puchades, M. García-Hita, A. Alastruey-Izquierdo, G. Quindós, Fungal co-infection in COVID-19 patients: should we be concerned?, *Revista Iberoamericana de Micología*, 2020, **37**, 41-46, doi: 10.1016/j.riam.2020.07.001.
[7] J. Salmanton-García, R. Sprute, J. Stemler, M. Bartoletti, D. Dupont, M. Valerio, C. Garcia-Vidal, I. Falces-Romero, M. Machado, S. de la Villa, M. Schroeder, I. Hoyo, F. Hanses, K. Ferreira-Paim, D. R. Giacobbe, J. F. Meis, J. P. Gangneux, A. Rodríguez-Guardado, S. Antinori, E. Sal, X. Malaj, D. Seidel, O. A. Cornely, P. Koehler, COVID-19-associated pulmonary aspergillosis, March-August 2020, *Emerging Infectious Diseases*, 2021, **27**, 1077-1086, doi: 10.3201/eid2704.204895.
[8] L. Martins-Santana, C. P. Rezende, A. Rossi, N. M. Martinez-Rossi, F. Almeida, Addressing microbial resistance worldwide: challenges over controlling life-threatening fungal infections, *Pathogens*, 2023, **12**, 293, doi: 10.3390/pathogens12020293.
[9] M. Egger, L. Bussini, M. Hoenigl, M. Bartoletti, Prevalence of COVID-19-associated pulmonary aspergillosis: critical review and conclusions, *Journal of Fungi*, 2022, **8**, 390, doi: 10.3390/jof8040390.
[10] W. Hurt, J. Youngs, J. Ball, J. Edgeworth, P. Hopkins, D. R. Jenkins, S. Leaver, A. Mazzella, S. F. Molloy, S. Schelenz, M. P. Wise, P. L. White, H. Yusuff, D. Wyncoll, T. Bicanic, COVID-19-associated pulmonary aspergillosis in mechanically ventilated patients: a prospective, multicentre UK study, *Thorax*, 2024, **79**, 75-82, doi: 10.1136/thorax-2023-220002.
[11] A. Mendonça, H. Santos, R. Franco-Duarte, P. Sampaio, Fungal infections diagnosis-Past, present and future, *Research in Microbiology*, 2022, **173**, 103915, doi: 10.1016/j.resmic.2021.103915.
[12] J. Baker, D. W. Denning, The SSS revolution in fungal diagnostics: speed, simplicity and sensitivity, *British Medical Bulletin*, 2023, **147**, 62-78, doi: 10.1093/bmb/ldad011.
[13] A. Haider, M. Ringer, Z. Kotroczó, C. Mohácsi-Farkas, T. Kocsis, The Current level of MALDI-TOF MS applications in the detection of microorganisms: a short review of benefits and limitations, *Microbiology Research*, 2023, **14**, 80-90, doi: 10.3390/microbiolres14010008.
[14] R. D. Jathanna, D. Acharya, L. E. Lewis, K. Makkithaya, Early detection of late onset neonatal sepsis using machine learning algorithms, *Engineered Science*, 2023, **26**, 976, doi: 10.30919/es976.
[15] W. Zhang, S. E. Barykin, T. V. Kirillova, I. V. Kapustina, N. S. Lukashevich, A. Zaytsev, An innovative resource management framework using logit-boosted machine learning algorithms for vehicular ad hoc networks (VANETs), *Engineered Science*, 2023, **26**, 980, doi: 10.30919/es980.
[16] P. R. Rane, S. Vincent, Landslide susceptibility mapping using machine learning algorithms for nainital, India, *Engineered Science*, 2021, **17**, 142-155, doi: 10.30919/es8d600.
[17] S. Wainaina, M. J. Taherzadeh, Automation and artificial intelligence in filamentous fungi-based bioprocesses: A review, *Bioresource Technology*, 2023, **369**, 128421, doi: 10.1016/j.biortech.2022.128421.
[18] N. R. Raudah Mohamed Radzuan, H. Jaafar, F. N. Zabani, An application of principal component analysis in aspergillus species identification, 2022 IEEE 10th Conference on Systems, Process & Control, December 17, Malacca, Malaysia, IEEE, 2022, 296-301, doi: 10.1109/ICSPC55597.2022.10001780.
[19] R. Jing, X. Yin, X. Xie, H. Lian, J. Li, G. Zhang, W. Yang, T. Sun, Y. Xu, Morphologic identification of clinically encountered moulds using a residual neural network, *Frontiers in Microbiology*, 2022, **13**, 1021236, doi: 10.3389/fmicb.2022.1021236.
[20] N. Tochigi, S. Sadamoto, S. Oura, Y. Kurose, Y. Miyazaki, K. Shibuya, Artificial intelligence in the diagnosis of invasive mold infection: development of an automated histologic identification system to distinguish between *Aspergillus* and mucorales, *Medical Mycology Journal*, 2022, **63**, 91-97, doi: 10.3314/mmj.22-00013.
[21] M. Arredondo-Santoyo, C. Domínguez, J. Heras, E. Mata, V. Pascual, M. S. Vázquez-Garcidueñas, G. Vázquez-Marrufo, Automatic characterisation of dye decolourisation in fungal strains using expert, traditional, and deep features, *Soft Computing*, 2019, **23**, 12799-12812, doi: 10.1007/s00500-019-03832-8.
[22] S. R. Nalamwar, S. Neduncheliyan, Development of intelligent skin disease classification system using machine learning, 5th International Conference on Inventive Research in Computing Applications, August 3-5, Coimbatore, India, IEEE, 2023, 884-890, doi: 10.1109/ICIRCA57980.2023.10220856.
[23] S. Aditya, S. Sidhu, M. Kanchana, Prediction of Alopecia Areata using Machine Learning Techniques, 2022 IEEE International Conference on Data Science and Information System, July 29-30, Hassan, India, IEEE, 2022, 1-6, doi: 10.1109/icdsis55133.2022.9915804.
[24] A. F. Lau, Matrix-assisted laser desorption ionization time-of-flight for fungal identification, *Clinics in Laboratory Medicine*, 2021, **41**, 267-283, doi: 10.1016/j.cll.2021.03.006.
[25] Y. Cai, C. Lin, Y. Wang, Theoretical study of general principle of high-resolution MALDI-linear time-of-flight mass spectrometry in middle to high m/z ranges, *International Journal of Mass Spectrometry*, 2023, **489**, 117052, doi: 10.1016/j.ijms.2023.117052.
[26] O. Ogundairo, O. Ayo-Farai, C. P. Maduka, C. C. Okongwu, A. O. Babarinde, O. T. Sodamade, Review on maldi mass spectrometry and its application in clinical research, *International Medical Science Research Journal*, 2023, **3**, 108-126, doi: 10.51594/imsrj.v3i3.642.
[27] R. Patel, A moldy application of MALDI: MALDI-ToF mass spectrometry for fungal identification, *Journal of Fungi*, 2019, **5**, 4, doi: 10.3390/jof5010004.
[28] J. Yansombat, S. Samosornsuk, C. Khattiyawech, P. Hematulin, T. Pharamat, S. L. Kabir, W. Samosornsuk, Colletotrichum truncatum, an endophytic fungus derived from Musa acuminata (AAA group): antifungal activity against Aspergillus isolated from COVID-19 patients and indole-3-acetic acid (IAA) production, *Asian-Australasian Journal of Bioscience and Biotechnology*, 2023, **8**, 23-29, doi: 10.3329/aajbb.v8i2.66927.

[29] P. M. Luethy, A. M. Zelazny, Rapid one-step extraction method for the identification of molds using MALDI-TOF MS, *Diagnostic Microbiology and Infectious Disease*, 2018, **91**, 130-135, doi: 10.1016/j.diagmicrobio.2018.01.015.
[30] M. R. Vidal-Acuña, M. Ruiz-Pérez de Pipaón, M. J. Torres-Sánchez, J. Aznar, Identification of clinical isolates of Aspergillus, including cryptic species, by matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), *Medical Mycology*, 2018, **56**, 838-846, doi: 10.1093/mmy/myx115.
[31] J. C. Russ, F. B. Neal, Image enhancement in the spatial domain, The Image Processing Handbook, CRC Press, Boca Raton, FL, 2016, 244-248, ISBN: 978-1498740265.
[32] P. Lin, X. Li, D. Li, S. Jiang, Z. Zou, Q. Lu, Y. Chen, Rapidly and exactly determining postharvest dry soybean seed quality based on machine vision technology, *Scientific Reports*, 2019, **9**, 17143, doi: 10.1038/s41598-019-53796-w.
[33] W. K. Mutlag, S. K. Ali, Z. M. Aydam, B. H. Taher, Feature extraction methods: a review, *Journal of Physics*: *Conference Series*, 2020, **1591**, 012028, doi: 10.1088/1742-6596/1591/1/012028.
[34] A. Humeau-Heurtier, Texture feature extraction methods: a survey, *IEEE Access*, 2019, **7**, 8975-9000, doi: 10.1109/ACCESS.2018.2890743.
[35] A. S. Setiawan, Elysia, J. Wesley, Y. Purnama, Mammogram classification using law's texture energy measure and neural networks, *Procedia Computer Science*, 2015, **59**, 92-97, doi: 10.1016/j.procs.2015.07.341.
[36] R. Kvyetnyy, O. Sofina, A. Olesenko, P. Komada, J. Sikora, A. Kalizhanova, S. Smailova, Method of image texture segmentation using Laws' energy measures, Proceedings of Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments, August 7, Wilga, Poland, 2017, 1784-1792, doi: 10.1117/12.2280891.
[37] S. Dash, U. R. Jena, Multi-resolution laws' masks based texture classification, *Journal of Applied Research and Technology*, 2017, **15**, 571-582, doi: 10.1016/j.jart.2017.07.005.
[38] M. Manav, M. Goyal, A. Kumar, Role of optimal features selection with machine learning algorithms for chest X-ray image analysis, *Journal of Medical Physics*, 2023, **48**, 195-203, doi: 10.4103/jmp.jmp_104_22.
[39] N. Giakoumoglou, PyFeats: Open-source software for image feature extraction, (1.0.11), Zenodo, 2021.
[40] S. Dash, U. R. Jena, Multi-resolution laws' masks-based texture classification, *Journal of Applied Research and Technology*, 2017, **15**, 571-582, doi: 10.1016/j.jart.2017.07.005.
[41] H. Singh, V. Sharma, D. Singh, Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor, *Visual Computing for Industry, Biomedicine, and Art*, 2022, **5**, 3, doi: 10.1186/s42492-021-00100-1.
[42] A. Humeau-Heurtier, Color texture analysis: a survey, *IEEE Access*, 2022, **10**, 107993-108003. doi: 10.1109/ACCESS.2022.3213439.
[43] E. M. Jawad, H. J. M. Hazim, G. Daway, Retinal image enhancement by using adapted histogram equalization based on segmentation and lab color space, *Intelligent Engineering and Systems*, 2022, **15**, 614-622, doi: 10.22266/ijies2022.0630.52.
[44] D. J. Bora, A. K. Gupta, F. A. Khan, Comparing the Performance of L*A*B* and HSV color spaces with respect to color image segmentation, *Emerging Technology and Advanced Engineering*, 2015, **5**, 192-203, doi: 10.48550/arXiv.1506.01472.
[45] K. Mishra, A. Balodi, Comparative analysis of neural network algorithms for severity analysis of valvular regurgitation, 2022 IEEE Delhi Section Conference, February 11-13, New Delhi, India, IEEE, 2022, 1-5, doi: 10.1109/DELCON54057.2022.9752904.
[46] M. Singh, D. Singh, V. Sharma, Breast tumor detection and classification in Mammograms: Gabor wavelet vs. statistical features, *Preprints,* 2018, 1-13, doi: 10.20944/preprints201806.0343.v1.
[47] C. Wu, Y. Chen, Statistical feature matrix for texture analysis, *CVGIP*: *Graphical Models and Image Processing*, 1992, **54**, 407-419, doi: 10.1016/1049-9652(92)90025-S.
[48] S. Zhang, J. Li, KNN classification with one-step computation, *IEEE Transactions on Knowledge and Data Engineering*, 2023, **35**, 2711-2723. doi: 10.1109/TKDE.2021.3119140.
[49] S. B. Imandoust, M. Bolandraftar, Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background, *International Journal of Engineering Research and Applications*, 2013, **3**, 605-610.
[50] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, S. Apoorva, Feature extraction using convolution neural networks (CNN) and deep learning, 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, May 18-19, Bangalore, India, IEEE, 2018, 2319-2323, doi: 10.1109/RTEICT42901.2018.9012507.
[51] N. Sawant, D. R. Khadapkar, Comparison of the performance of GaussianNB algorithm, the K neighbors classifier algorithm, the logistic regression algorithm, the linear discriminant analysis algorithm, and the decision tree classifier algorithm on same dataset, *International Journal for Research in Applied Science and Engineering Technology*, 2022, **10**, 1654-1665, doi: 10.22214/ijraset.2022.48311.
[52] Y. Gao, S. Sun, An empirical evaluation of linear and nonlinear kernels for text classification using Support Vector Machines, 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, August 10-12, Yantai, China, IEEE, 2010, 1502-1505, doi: 10.1109/FSKD.2010.5569327.
[53] R. Nariswari, H. Pudjihastuti, Support vector machine method for predicting non-linear data, *Procedia Computer Science*, 2023, **227**, 884-891, doi: 10.1016/j.procs.2023.10.595.
[54] M. Schonlau, R. Y. Zou, The random forest algorithm for statistical learning, *The Stata Journal*: *Promoting Communications on Statistics and Stata*, 2020, **20**, 3-29, doi: 10.1177/1536867x20909688.
[55] A. B. Shaik, S. Srinivasan, A brief survey on random forest ensembles in classification model, International Conference on

Innovative Computing and Communications: Proceedings of ICICC 2018, Springer Singapore, 2018, **2**, 253-260, doi: 10.1007/978-981-13-2354-6__27.
[56] Z. Hu, Suspect prediction based on naive Bayesian method, *IOP Conference Series: Materials Science and Engineering*, 2019, **631**, 032054, doi: 10.1088/1757-899x/631/3/032054.
[57] A. Roy, S. Chakraborty, Support vector machine in structural reliability analysis: A review, *Reliability Engineering & System Safety*, 2023, **233**, 109126, doi: 10.1016/j.ress.2023.109126.
[58] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, S. Parasa, On evaluation metrics for medical applications of artificial intelligence, *Scientific Reports*, 2022, **12**, 5979, doi: 10.1038/s41598-022-09954-8.
[59] M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview, arxiv preprint arxiv: 2008.05756, 2020, 1-17, doi: 10.48550/arXiv.2008.05756.
[60] C. Honsig, B. Selitsch, M. Hollenstein, M. G. Vossen, K. Spettel, B. Willinger, Identification of filamentous fungi by MALDI-TOF mass spectrometry: evaluation of three different sample preparation methods and validation of an in-house species cutoff, *Journal of Fungi*, 2022, **8**, 383, doi: 10.3390/jof8040383.
[61] K. R. Barker, J. V. Kus, A. C. Normand, F. Gharabaghi, L. McTaggart, C. Rotstein, S. E. Richardson, A. Campigotto, M. Tadros, A practical workflow for the identification of aspergillus, fusarium, mucorales by MALDI-TOF MS: database, medium, and incubation optimization, *Journal of Clinical Microbiology*, 2022, **60**, 1-15, doi: 10.1128/jcm.01032-22.
[62] Z. Ráduly, L. Szabó, A. Madar, I. Pócsi, L. Csernoch, Toxicological and medical aspects of *Aspergillus*-derived mycotoxins entering the feed and food chain, *Frontiers in Microbiology*, 2020, **10**, 2908, doi: 10.3389/fmicb.2019.02908.
[63] B. Mousavi, M. T. Hedayati, N. Hedayati, M. Ilkit, S. Syedmousavi, Aspergillus species in indoor environments and their possible occupational and public health hazards, *Current Medical Mycology*, 2016, **2**, 36-42, doi: 10.18869/acadpub.cmm.2.1.36.
[64] W. P. Pfliegler, I. Pócsi, Z. Győri, T. Pusztahelyi, The aspergilli and their mycotoxins: metabolic interactions with plants and the soil biota, *Frontiers in Microbiology*, 2020, **10**, 2921, doi: 10.3389/fmicb.2019.02921.
[65] K. Raksha Rao, A. V. Vipin, G. Venkateswaran, Mechanism of inhibition of aflatoxin synthesis by non-aflatoxigenic strains of Aspergillus flavus, *Microbial Pathogenesis*, 2020, **147**, 104280, doi: 10.1016/j.micpath.2020.104280.
[66] H. M. Martins, M. L. Martins, F. Almeida Bernardo, Interaction of strains of non-toxigenic Aspergillus flavus with Aspergillus parasiticus on aflatoxin production, *Brazilian Journal of Veterinary Research and Animal Science*, 2000, **37**, 439-443, doi: 10.1590/s1413-95962000000600003
[67] N. Çolakoğlu, B. Akkaya, Comparison of multi-class classification algorithms on early diagnosis of heart diseases, y-BIS Conference 2019: Recent Advances in Data Science and Business Analytics, September 25-28, İstanbul, Turkey, 2019, 162-172.
[68] R. R. Nalawade, M. S. Joshi, R. R. Rathod, A. P. Suryawanshi, Morphological and cultural characters of fungi isolated from copra and copra oil, *International Journal of Current Microbiology and Applied Sciences*, 2019, **8**, 2256-2261, doi: 10.20546/ijcmas.2019.809.260.