



Supervised segmentation on fusarium macroconidia spore in microscopic images via analytical approaches

K. A. Azuddin¹ · A. K. Junoh² · A. Zakaria¹ · M. T. A. Rahman¹ · N. M. I. M. Nor³ · H. Nishizaki⁴ · Z. Latiffah³ · N. F. Azuddin³ · M. Z. Abdullah⁵ · T. P. Terna³

Received: 1 June 2022 / Revised: 23 May 2023 / Accepted: 8 September 2023 /
Published online: 9 October 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Fungi are one of the major causes that contributed to plant diseases. There are lots of fungi species but it is estimated that only 10% have been described. There are two major approaches to identifying fungi species, morphological identification, and molecular test which need cautious clarification to make good interpretations and are time-consuming. In this paper, we propose a Machine Learning approach that involves the use of the K-Means clustering technique, and Decision Tree to highlight the observed fungi spore images taken under the microscopic view and discard background pixels to produce digital images database which later can be used for Deep Learning.

Keywords Artificial Intelligence · Machine Learning · Pattern Recognition · Supervised and Unsupervised Learning Methods

1 Introduction

The fungi domain is one of the most various eukaryotic families on Earth with estimates of numerous million existent species [1–3]. In the field of plant pathology, the level of health of an observed plant can be affected by several factors such as disease infection [4], surrounding environmental conditions, pests, and others. Disease infections against plants can be caused by some biotic or pathogenic living agents such as fungi, bacteria, or viruses, and most plant diseases are usually caused by fungi [5].

Fungi at an early stage are usually difficult to detect because of their small-sized structure and usually may be noticeable when they fruition to mushroom or mold. Some fungi

✉ A. K. Junoh
kadri@unimap.edu.my

¹ Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Arau, Malaysia

² Institute of Engineering Mathematics, Universiti Malaysia Perlis, Arau, Malaysia

³ School of Biological Sciences, Universiti Sains Malaysia, Gelugor, Malaysia

⁴ Department of Mechatronics, University of Yamanashi, Kofu, Japan

⁵ Faculty of Mechanical Engineering & Technology, Universiti Malaysia Perlis, Arau, Malaysia

are pathogenic to plants, when grown on a plant, they can cause diseases such as mildew, rust, and others to the plant [6]. In plantations, if left untreated, the fungal disease can cause farmers to suffer a significant loss in their yields [7]. Not all fungi are pathogenic to plants, some generate a symbiotic relationship with the host plant [8].

Fungi are usually described through morphological identification [9], however, morphological identification of the species is generally difficult due to the significant resemblance among them especially when it involves the same genera due to their morphological complex [10], and molecular test [11, 12] which need to undergo several cautious, hassle procedure and time-consuming process like shown in Fig. 1.

Some past studies have been carried out with images of fungal spores that have been added with color staining in microscope slides to help highlight the fungal spores when viewed under the microscope [10], and this color stain has already distinguished the fungal spores from the background from the beginning making any image segmentation process like Thresholding [13] easier and, and in this research is to try to distinguish the fungal spore from the background without the help of color staining under the microscopic views in order to produce a database that keeps the true nature of the observed fungi spore.

This research is mainly to recognize several fungi species under the *Fusarium* genera with Deep Learning, however, due to the limitation of space to explain all the conducted processes, only the Machine Learning part will be explained in this paper. Currently, by training with only raw images of the observed *Fusarium* spore, in some of the produced high accuracy percentages of CNN's models, only a few of the models act accordingly in the validation process, raising a question where, can a reliable and accurate automated image analysis system for identifying and classifying fungal spores in microscope images if more species were added to in the future. Machine Learning was implemented in order to produce segmented images of the observed *Fusarium* spores under the microscopic. The purpose of these segmented images is to be used as training images later in Deep Learning part, it is expected with the help of these segmented images a higher validation accuracy of CNN's model can be produced. The structure

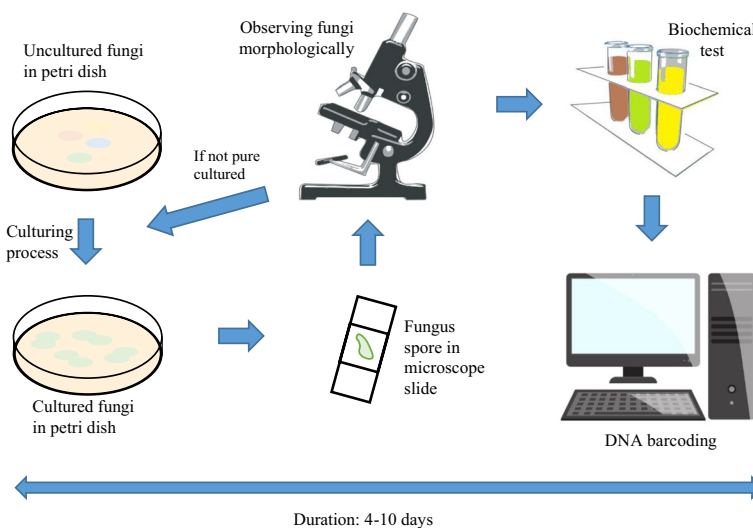


Fig. 1 Process flow in fungi identification with molecular test

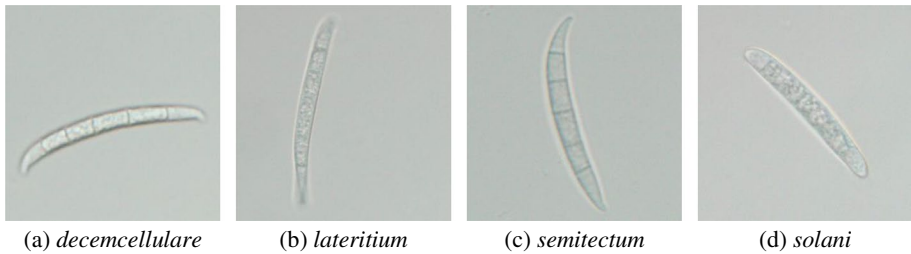


Fig. 2 Random images of four *Fusarium* species

of this paper begins with an introduction, then materials and methods were presented, results and conclusion.

2 Materials and methods

One of the most common fungi is *Fusarium*, and they can be found almost anywhere, be it in soil, on plants [14], and even on humans [15]. They can be pathogenic [16] or endophytic [17] towards plants and in this paper, the observed *Fusarium* fungi are *decemcellulare* (MT296783), *lateritium* (MT296785), *solani* (MT296787) where the fungi were obtained from rattan spine, and *semitectum* (also known as *F. incartantum*) (MT708867) were obtained from corn in peninsular Malaysia (Fig. 2). Further information about the observed fungi can be referred to the NCBI database. All of the fungi had been preliminary pure cultured and validated through DNA molecular test. The fungi spore were morphologically observed with Olympus BX53 microscope with 40 \times magnification, and the microscopic images of the fungi were taken with Olympus DP72 microscope camera. The raw images of all the spores are in resolution of 4140 \times 3096 \times 3. The fungi spore images were taken with three backlighting: daylight, cool white, and warm white. The images then were cropped to contain only a single spore in each image to the size of 100 \times 100. Here randomly selected images for each of the observed fungi are presented (Fig. 3).

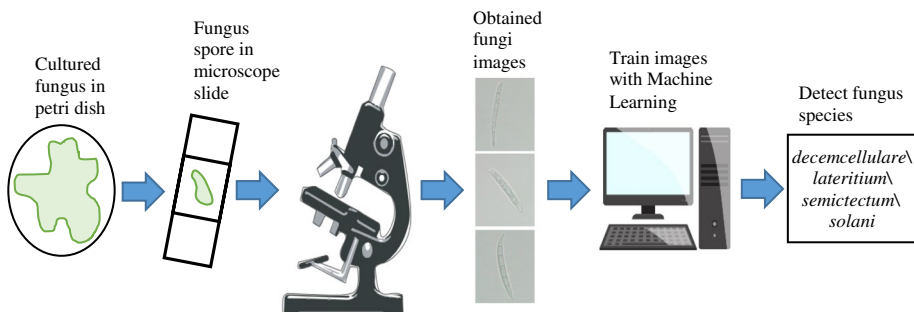


Fig. 3 Process flow of the proposed method in developing *Fusarium* digital images database

3 Machine learning

Machine Learning (ML) is a field of study that allows machines to automatically learn and improve their performance by analyzing and making predictions based on data. The concept of ML was first introduced by Arthur Samuel in 1959. According to Mohri [18], Machine Learning (ML) is defined as a computational method based on information that has been previously trained to improve the accuracy of machine predictions. Various models are used in machine learning, including linear regression, decision trees, support vector machines (SVM), neural networks, and more [19]. This technology has gained significant attention due to its ability to solve complex problems that are difficult to solve using traditional programming methods. From a philosophical and scientific perspective, Machine Learning (ML) is interesting because it can develop an understanding of the underlying principles of intelligence [20]. In this study, we utilized the K-Means clustering and Decision Tree classification methods to solve the problem at hand. Figure 4 illustrates the overall process flow of the Machine Learning (ML) approach employed in this study.

4 K-Means clustering

One of the methods for segmenting an image is with clustering technique [21]. The Clustering tool that was used in this paper is the K-Means clustering. Since the data that was obtained in this study are in form of still images of the *Fusarium* fungi spore, the first thing to do before beginning the clustering process, is to extract all of the pixel RGB values from the fungi spore image. All of the extracted pixel RGB values were stored in a matrix format. The fungi spore image size was reduced to 100×100 pixels. There are 10000 pixels with RGB values were extracted for a single fungus spore image. Now that the pixels' values are extracted, the pixel can be clustered. Generally, the pixels were clustered into two main categories. The first category is the background pixels, followed by the fungal spore pixels, which include the fungus edges and the fungal spores' internal filling.

The pseudocode for the clustering algorithm begin by extract image pixel RGB value, then select number of clusters, N . Next, start the centroid by first rearranging the data set and then selecting N data points at random for the centroid. Repeat the process until there is no change in the centroid. Calculate the squared distance between all the centroid and pixels. Then each pixel will be assign to the closest centroid and became a cluster. The Centroid were obtain by calculate the average of all pixel value in the cluster.

5 Decision tree classification

Decision Tree is one of the decision-making techniques [22] not just in image processing and machine learning but in other areas like Data Mining, pattern recognition, or statistic [23], it has a good interpretation ability involving categorical features, thus making it a good predictive model [24]. Some image processing application has integrated this technique along with Deep Learning [25, 26]. In this study, after all the extracted pixels' RGB values have been clustered, it will then be manually observed and selected which of the clustered pixel belongs within the two main categories, the background pixel, and the fungus pixel. Then only pixels that are labeled as the fungus

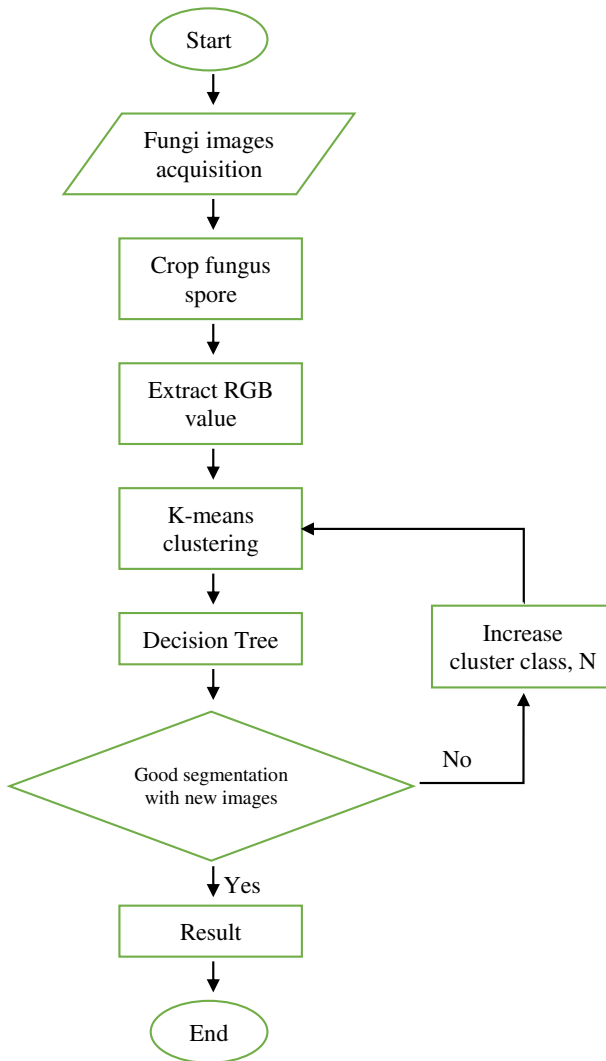


Fig. 4 Machine learning process flowchart

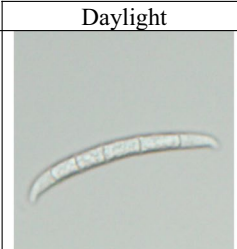
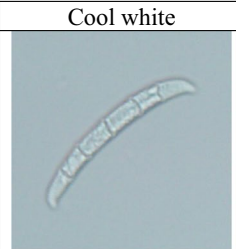
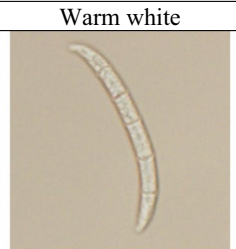









pixels are regarded as the desired information to be used and taught in the Machine Learning (ML) system as a specific *Fusarium* fungus species. Each of the four *Fusarium* fungus species will contain several certain clustered pixel RGB values. Then, Decision Tree were used to determine which combination of clustered pixels represented the specific *Fusarium* fungus species. The Decision Tree consists of several basic elements: Root, Sub nodes, and Leaves. Where the development of the Decision Tree begins with the root and onto the sub-nodes, which represent several features. In the beginning, all RGB value training set, is considered as the root. Features values are then clustered through K-means clustering.

6 Result and discussion

From the presented procedures above, a fungi spore digital images database consisting of the images of four observed fungus species was produced, presented below are some randomly selected of the observed fungi spore with respect to backlight colors and species.

Table 1 above displays fungal spores that were randomly selected and classified based on their respective species, and each spore displayed was emitted with 3 different types of backlight namely daylight, cool white, and warm white. Based on the displayed images, it can be seen that each monitored fungal spore is to some extent semi-transparent, where the colour of the fungus seems it changes according to the colour of the emitted backlight when viewed under a microscope. This may be due to the size of the spores being very small and thin, it could also be due to the intensity of the emitted backlight being stronger so that it allows most of the light to penetrate the structure of the fungus and makes the fungus blend in with the emitted backlight. Based on initial vision observations, it can be seen that there are some areas on the fungal spores that are different in colour from the backlight colour, and that area has been targeted as the desired part to be obtained, while

Table 1 Images of the observed fungi spore with different backlight

	Daylight	Cool white	Warm white
<i>decemcellulare</i>			
<i>lateritium</i>			
<i>semitectum</i>			
<i>solani</i>			

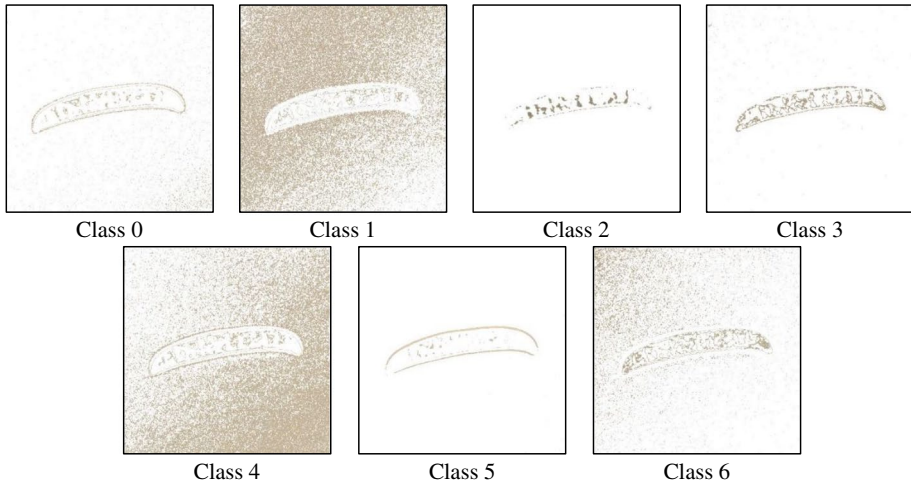


Fig. 5 Image segmentation for *solani* with warm white background in cluster N=7

other areas that are vague and contain colour similarities with backlight are considered as background pixel and need to be discarded.

By undergoing the process of the proposed Machine Learning approach, the fungi spore image pixels RGB values were extracted. These pixel values then proceed to be clustered into several class with K-means clustering. The clustering process was conducted starting from the number of cluster classes, N=3, N=4, N=5, N=6, N=7, N=10, N=15, N=20, N=25, and N=30. In clustering, any cluster classes associated with background pixel values are ignored, and clustered pixels containing pixel values representing fungi are then sorted together using the Decision Tree technique.

Figure 5 above are one of the image segmentation for *Fusarium solani* species with the warm white background done by K-means clustering up to 7 cluster class as labeled in the figure above. The cluster class that is deemed to be labeled as fungus spore pixel by human vision assessment, are classes 2, 3, and 5. While the other 4 cluster classes are labeled as the background pixel, based on qualitative assessment, class 0 shows the similarity that it relates to fungal pixels, however, if observed closely, there is still some background pixel noise in the segmented image and other comparison as shown in Table 2 below.

If Class 0 is chosen to be clustered together with fungal pixel clusters, when tested with a new image with the same species and background, will result in a miss-segmented image where it will contain a lot of background pixel noise making it less successful at segmenting the spore object from the background back.

Table 2 Quantitative comparison between reference image and each segmented image class

	MS Error	Structural Similarity Index (SSIM)	Number of pixels similarity	
			Same	Not same
0	15.86	0.7623	301,714	120,786
2	8.24	0.9540	389,422	33,078
3	8.41	0.9210	379,229	43,271
5	9.76	0.9246	377,722	44,778

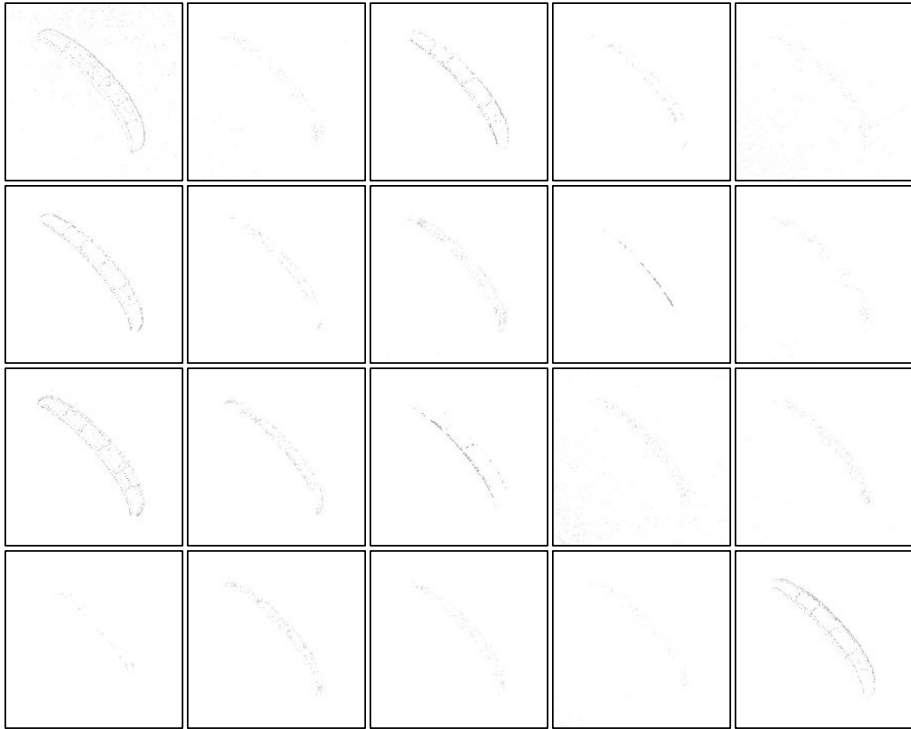


Fig. 6 Image segmentation by K-means clustering for *decemcellulare* spore with cool white background

Figure 6 below shows a sample of the image segmentation result by K-means clustering with $N=30$ for *Fusarium decemcellulare*, where a total of 20 classes are labeled as the fungus pixel while the other 10 classes are labeled as the background pixel group.

The Tables 3, 4 and 5 above is the result of K-means clustering and shows the number of clusters associated with fungal pixels for three categories, lowest cluster number, $N=3$, highest cluster number, $N=30$, and $N=7$ cluster, number of clusters considered better than the others. Here the total number of clusters, $N=7$ indicates that there is no significant difference between the ratio of the total number of fungal pixel clusters and the background pixel clusters in each species and the backlighting. However, there was a large difference in the number of clusters $N=30$ as shown in the table above.

Presented Fig. 7 above is the result for *Fusarium decemcellulare* spore with respect to N , number of clusters, for daylight background, where Decision Tree has discarded all the pixels that are related to the background pixel and left only pixels that are deemed by Decision Tree, belongs to fungus pixel. From the produced images, it can be seen that K-means

Table 3 Number of cluster related to fungus for cool white background

	$N=3$	$N=7$	$N=30$
<i>decemcellulare</i>	2	4	20
<i>lateritium</i>	1	3	10
<i>semitectum</i>	1	3	11
<i>solani</i>	1	3	14

Table 4 Number of cluster related to fungus for daylight background

	N=3	N=7	N=30
<i>decemcellulare</i>	2	4	17
<i>lateritium</i>	1	3	10
<i>semitectum</i>	2	4	19
<i>solani</i>	2	4	17

clustering produces a non-linearity outcome whereas at first it was expected that the higher the number of clusters, N, the more accurate the images will be produced. However, in this case, it can be seen, that N=3 contain more fungus pixel than N=4 but there are some background pixel noise were able to be seen, N=5 shows less noise than the other two and with more fillings inside the fungus pixels while N=6 contain more pixel noise and less fungus pixel fillings even though it has already been carefully observed and selected during the image segmentation process. For N=15, 20, and 25, they contain less noise with more fillings inside the fungus pixel area than the other clusters but when observed with other species and background, the result is not the same, some contain pixel noise, some with less fungus pixel fillings. Cluster N=7 was chosen as the desired reference data since the images produced contain less pixel noise, and more fungus pixel fillings and are stable with other species and background than the other cluster as shown in Table 6. However, based on observations, for cluster N=7, it shows that the cluster is very sensitive. For example, Fig. 5, shows the image segmentation for *Fusarium solani* fungus spore, if there is even one error in determining the pixel class between the background pixel and the fungus pixel, it will result in a very large effect when the Machine Learning tries to highlight the fungus image where most of the produced images will contain lots of background pixel noise thus making it unable to highlight the fungus in the images.

After deciding on the selection of the number of clusters to be used as the desired reference to be included in the *Fusarium* digital database, the produced database then was used to test against 50 pieces of new images to see how many images that were able to successfully highlighting the fungus in the image.

Table 7 displays the total number of fungus images that were successfully highlighted and discarded the background pixels through the Decision Tree for 3 different numbers of clusters, which are, the smallest number of clusters, N=3, the largest number of clusters, N=30, and N=7, number of clusters which was selected as the ideal number of clusters to be used as reference data in this study.

Then from the produced image segmentation result, the images were examined and determined which images among the produced segmented image belong to the background pixel, and the desired object, the fungus pixel. All the pixel RGB values that are related to the fungus category are compiled together forming reference data in the form of RGB pixel values. Then from the compiled RGB pixel value reference data, Decision Tree were implemented in order to determine and only choose to display any pixel value that is related to

Table 5 Number of clusters related to fungus for warm white background

	N=3	N=7	N=30
<i>decemcellulare</i>	2	3	18
<i>lateritium</i>	1	3	10
<i>semitectum</i>	1	3	15
<i>solani</i>	2	3	14

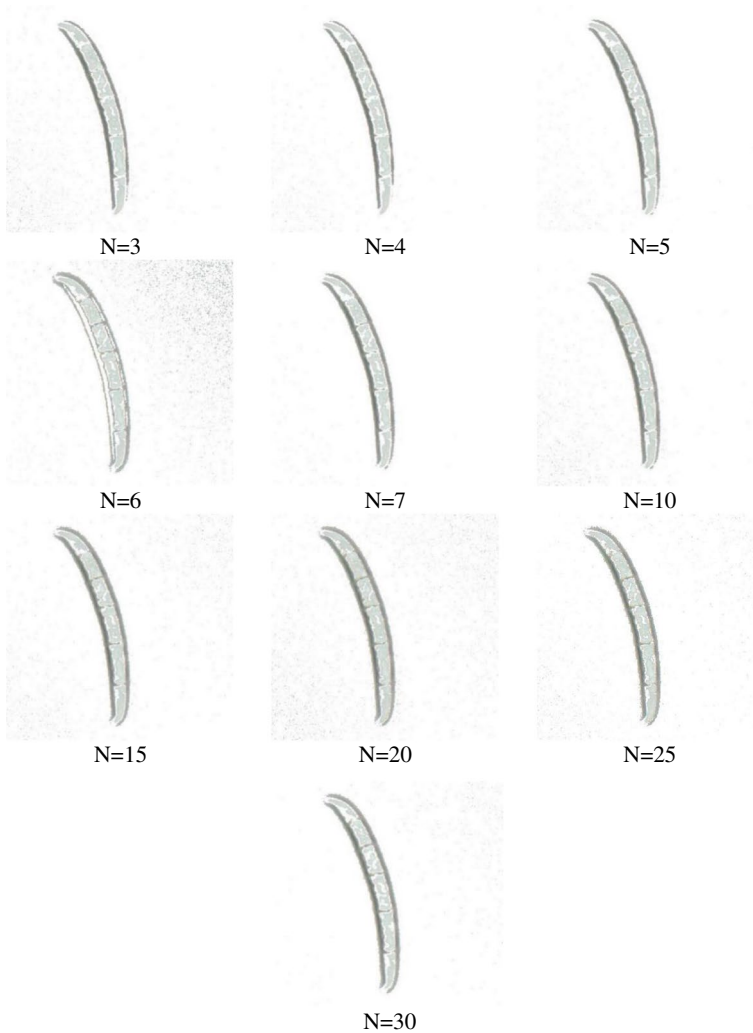


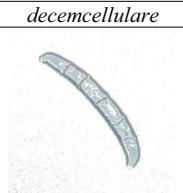
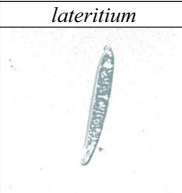
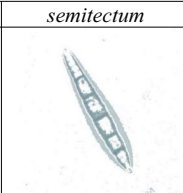



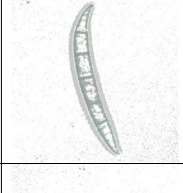





Fig. 7 Produced *Fusarium decemcellulare* with daylight background images from Decision Tree

the fungus category, and discarded any pixel value that is related to the background category. The figure below shows some of the produced fungi spore images that have discarded all pixels that are related to the background pixel by Decision Tree.

Table 8 shows the images of the resulting fungi spore images after going through the Decision Tree with $N=7$ clusters from K-Means. Here, all pixels that have an RGB value associated with the backlight are removed, and only leave pixels with RGB value that is related to the fungi pixel. The remaining pixels is representing the desired object of the monitored fungi in this study then the images were filtered with a median filter to reduce any pixel noise that was still caught in the produced images.

One of the famous image segmentation techniques is Thresholding [13], it basically divides the original image into 2 images based on the mean value of the extracted pixels. Then repeat the process on the 2 separate images produced, until all the generated

Table 6 Segmented images for cluster N=7

	<i>decemcellulare</i>	<i>lateritium</i>	<i>semitectum</i>	<i>solani</i>
Cool white				
Daylight				
Warm white				

separate images produce images that do not contain background pixels, and those images will be considered and clustered as pixels related to fungi. The same Decision Tree classifier is then applied to the produced Threshold database to try and segment between spore-related pixels and background pixels. Here the result of the segmented image between the K-Means clustering and Thresholding are presented as shown below.

Figure 8 above show the produced segmented image between the proposed method and Thresholding for *decemcellulare* with daylight background. The reference image was produced by manually remove the background photo editor. Based on visual evaluation, it can

Table 7 Total number of test images that were highlighted based on the number of clusters

	N=3	N=7	N=30
Cool white background			
<i>decemcellulare</i>	20	34	20
<i>lateritium</i>	11	31	20
<i>semitectum</i>	14	34	33
<i>solani</i>	14	29	10
Daylight background			
	N=3	N=7	N=30
<i>decemcellulare</i>	40	48	24
<i>lateritium</i>	16	16	16
<i>semitectum</i>	14	32	11
<i>solani</i>	7	18	6
Warm white background			
	N=3	N=7	N=30
<i>decemcellulare</i>	24	36	30
<i>lateritium</i>	13	18	17
<i>semitectum</i>	11	17	25
<i>solani</i>	23	38	27

Table 8 Fusarium spore images before and after performing segmentation

		Daylight		Cool white		Warm white	
<i>decemcellulare</i>	Original						
	Result						
<i>lateritium</i>	Original						
	Result						
<i>semitectum</i>	Original						
	Result						
<i>solani</i>	Original						
	Result						

be seen that K-means produces a better segmentation where the image shows that it has no background noise pixels and only a small part of the spore pixels are not displayed while the segmented image by Thresholding shows that there is a lot of background noise pixels and its spore pixels less refined and rough.

The chart in Fig. 9 above shows the Mean Squared Error (MSE) values for both K-means and Thresholding segmentation methods for several randomly selected images from each species and background color. Both segmentations are compared with each reference image respectively, from the graph it can be clearly seen that the K-means segmentation produces a lower MSE value for all randomly selected images compared to the Thresholding segmentation method.

Figure 10 above shows the SSIM value for both K-means and Thresholding methods, in line with the MSE value, it shows that the K-means segmentation method produces a segmentation image with a high SSIM value for all the images compared to



Fig. 8 Produced segmented image between K-Means and Thresholding

the Thresholding method. From these two charts, it can be said that the K-means segmentation method is able to produce a segmentation image with less background pixels and a high similarity to the original image. However, from the full visual observation results as in Table 7, the K-Means segmentation method is indeed capable of segmenting the fungal images but not all images are successful, this is due to the phenomenon of lens flare from the microscope backlight during the image collection process, and for the situation in this study this can be avoided in the future by reducing the image size during the process image collection because small images are less affected by the lens flare phenomenon and other issue like the overlapping spores are to be considered with other overlapping algorithm [27]

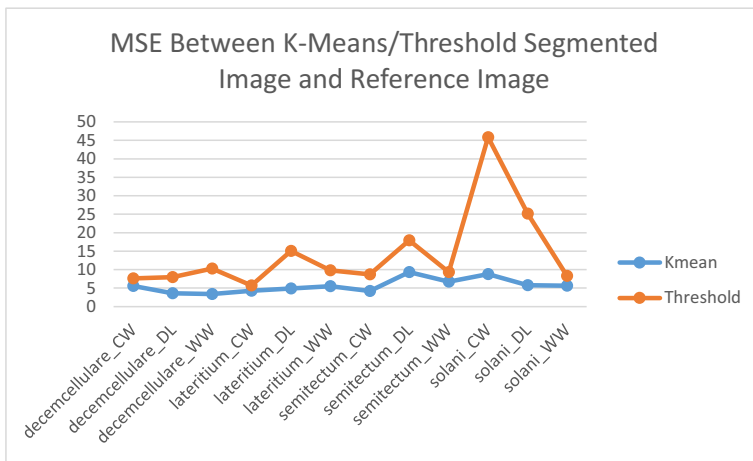


Fig. 9 Mean Squared Error chart between K-Means and Thresholding

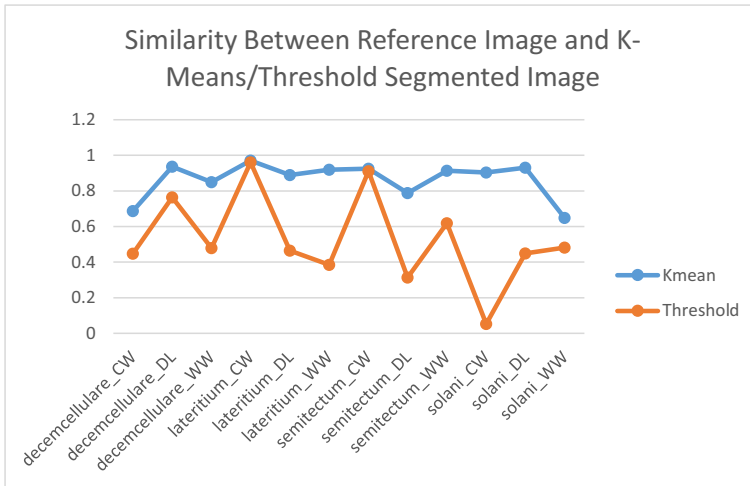


Fig. 10 Structural Similarity Index Measure (SSIM) between K-Means and Thresholding

7 Conclusion and future work

The proposed method in this research study is to build a digital images database for Fusarium group of fungi where this database can later be used as a data library in building an image processing fungi identification. Thus, will help shorten the period of fungi species identification. Reducing the cost of time and equipment used in conventional fungi identification method, since the proposed method will allow researchers to skip some of the procedures involved in the currently used method (molecular test) which used more disposable equipment in order to identify fungus species. K-means Clustering shows that it is able to separate between fungi pixels and background pixels, however, it seems that K-means clustering is not really dependable when segmenting images that have high significant similarities like those tested in this study where the tested fungi are semi-transparent and able to blend with the backlighting. However, the database generated through K-means clustering may be further refined by selecting the different total number of clusters, N on certain species and certain backgrounds and not just maintaining the selection of the total number of clusters for each tested fungus and background. Thresholding methods are proposed for future work to try to produce a better fungal database. The produced segmented fungi images database will then be used to be trained in Deep Learning.

Acknowledgements The writing of this report is fully funded by Malaysia's Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2022/STG01/UNIMAP/02/1) and with the collaboration and assistance of Universiti Sains Malaysia, Universiti of Yamanashi, Japan, and Nandemo Dagang Sdn. Bhd.

Data availability The data images of Fusarium spore generated in this study are not yet publicly available in any repository as this study is currently still ongoing. However, the information regarding the fungi observed in this study is available in National Center for Biotechnology Information database.

Declarations

Conflict of interest The authors declare no conflict of interest. No funding was received for conducting this study.

References

- O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol* 71(9):5544
- Blackwell M (2011) The Fungi: 1, 2, 3... 5.1 million species? *Am J Bot* 98(3):426–438
- Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, Ruess RW (2014) A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecol Monogr* 84(1):3–20
- Wei Z, Gu Y, Friman VP, Kowalchuk GA, Xu Y, Shen Q, Jousset A (2019) Initial soil microbiome composition and functioning predetermine future plant health. *Sci Adv* 5(9):eaaw0759
- Nazarov PA, Baleev DN, Ivanova MI, Sokolova LM, Karakozova MV (2020) Infectious plant diseases: etiology, current status, problems and prospects in plant protection. *Acta Naturae* 12(3):46
- Jain A, Sarsaiya S, Wu Q, Lu Y, Shi J (2019) A review of plant leaf fungal diseases and its environment speciation. *Bioengineered* 10(1):409–424
- Sukmawati D, Miarsyah M (2017) Pathogenic activity of *Fusarium equiseti* from plantation of citrus plants (*Citrus nobilis*) in the village Tegal Wangi, Jember Umbulsari, East Java, Indonesia. *Asian J Agric Biol* 5(4):202–213
- Truong C, Mujic AB, Healy R, Kuhar F, Furci G, Torres D, Niskanen T, Sandoval-Leiva PA, Fernández N, Escobar JM, Moretto A, Palfner G, Pfister D, Nouhra E, Swenie R, Sánchez-García M, Matheny PB, Smith ME (2017) How to know the fungi: combining field inventories and DNA-barcoding to document fungal diversity. *New Phytol* 214(3):913–919. <https://doi.org/10.1111/nph.14509>
- Seifert KA, Rossman AY (2010) How to describe a new fungal species. *IMA Fungus* 1(2):109–111
- Zieliński B, Sroka-Oleksiak A, Rymarczyk D, Piekarczyk A, Brzychczy-Włoch M (2020) Deep learning approach to describe and classify fungi microscopic images. *PLoS one* 15(6):e0234806
- Gherbawy Y, Voigt K (2010) *Molecular identification of fungi*. Springer, Heidelberg Dordrecht, London, New York, pp 1–512
- Manamgoda DS, Cai L, Hyde KD (2012) A phylogenetic and taxonomic reevaluation of the *Bipolaris* - *Cochliobolus* - *Curvularia* Complex. *Fungal Diversity* 56:131–144
- Tahir MW, Zaidi NA, Blank R, Vinayaka PP, Lang W (2016) Fungus detection system. 2016 IEEE International Conference on autonomic computing (ICAC). IEEE, pp 227–228
- Havrlentová M, Šliková S, Gregusová V, Kováčsová B, Lančaričová A, Nemeček P, Hendrichová J, Hozlár P (2021) The influence of artificial fusarium infection on oat grain quality. *Microorganisms* 9(10). <https://doi.org/10.3390/microorganisms9102108>
- Zhang N, O'Donnell K, Sutton DA, Nalim FA, Summerbell RC, Padhye AA, Geiser DM (2006) Members of the *Fusarium solani* species complex that cause infections in both humans and plants are common in the environment. *J Clin Microbiol* 44(6):2186–2190
- Maryani N, Sandoval-Denis M, Lombard L, Crous PW, Kema GHJ (2019) New endemic *Fusarium* species hitch-hiking with pathogenic *Fusarium* strains causing Panama disease in small-holder banana plots in Indonesia. *Persoonia: Mol Phylogeny Evol Fungi* 43:48–69
- Khan N, Afroz F, Begum MN, Roy Rony S, Sharmin S, Moni F, Mahmood Hasan C, Shaha K, Sohrab MH (2018) Endophytic *Fusarium solani*: A rich source of cytotoxic and antimicrobial naphthaquinone and aza-antraquinone derivatives. *Toxicol Rep* 5:970–976. <https://doi.org/10.1016/j.toxrep.2018.08.016>
- Mohri M, Rostamizadeh A, Talwalkar A (2018) *Foundations of machine learning*. MIT press
- Mahesh B (2020) Machine learning algorithms-a review. *Int J Sci Res (IJSR)* 9:381–386
- Kim KG (2016) Book review: deep learning. *Healthc Inform Res* 22(4):351. <https://doi.org/10.4258/hir.2016.22.4.351>
- Dhanachandra N, Chanu YJ (2017) A new approach of image segmentation method using K-means and kernel-based subtractive clustering methods. *Int J Appl Eng Res* 12(20):10458–10464
- Quinlan JR (1990) Decision trees and decision-making. *IEEE Trans Syst Man Cybern* 20(2):339–346. <https://doi.org/10.1109/21.52545>
- Rokach L, Maimon O (2005) Decision trees. *Data mining and knowledge discovery handbook*. Springer, Boston, MA, pp 165–192

24. Günlük O, Kalagnanam J, Li M, Menickelly M, Scheinberg K (2021) Optimal decision trees for categorical data via integer programming. *J Global Optim* 81:233–260
25. Yoo SH et al (2020) Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front Med* 7:427
26. Jiao L, Zhao J (2019) A survey on the new generation of deep learning in image processing. *IEEE Access* 7:172231–172263
27. Liu X, Wang S, Lin JCW, Liu S (2022) An algorithm for overlapping chromosome segmentation based on region selection. *Neural Comput & Applic*. <https://doi.org/10.1007/s00521-022-07317-y>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.