Artificial Intelligence in the Diagnosis of Invasive Mold Infection: Development of an Automated Histologic Identification System to Distinguish Between Aspergillus and Mucorales

Naobumi Tochigi1, Sota Sadamoto1,2, Shinji Oura1, Yasuko Kurose1, Yoshitsugu Miyazaki2, and Kazutoshi Shibuya1

1 Department of Surgical Pathology, Faculty of Medicine, Toho University
2 Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases

ABSTRACT

Background: Histopathological identification is usually required since the sensitivity of fungal culture is not sufficient for accurate diagnosis. On the other hand, pathological diagnosis, especially of molds, often is not accurate, even when performed by an experienced pathologist. This is particularly true in the differentiation between mucormycosis and aspergillosis, which have different drugs of choice and medical management. The diseases can easily become severe in a short period of time in accordance with the severity of the underlying disease or predisposing factors. Therefore, correct diagnosis is extremely important and should be entrusted to the pathologist.

Aim: To develop an artificial intelligence (AI)-based automated histological diagnostic system for mold infection to support the diagnosis by general pathologists, especially for distinguishing between Aspergillus and Mucorales.

Method: We used two indicators for the diagnostic system; namely, the angle of independent hyphae and tortuosity of each hypha.

Results and conclusion: We collected 147 and 67 image samples respectively from standard cases of aspergillosis and mucormycosis. All the images were successfully analyzed by automatic recognition of the two indicators. The independent areas divided by the threshold curve generated by two-dimensional plots of the data clearly include the test data obtained from the cases of Aspergillus and Mucorales. The present study demonstrates the usefulness of our newly developed AI-based diagnostic system. Further investigation is required for its practical use.

Key words: AI method, Aspergillus, invasive mold infection, Mucorales, Python

Introduction

In the 20th century, systemic mycosis was considered as a lethal disease by most physicians. Pathologists usually encountered patients suffering from generalized fungal infection at the autopsy room. Therefore, pathological determination was one of the important procedures to identify the causative fungi1,2. However, now pathologists should identify the infectious agents, given that some cases of Mucorales infection can be treated with prompt surgery2.

In general, the diagnosis of infectious disease requires microbiological confirmation, largely by culture. However, the confirmation rate of fungi by culture is low. Therefore, direct microscopic examination is also usually required. On the other hand, whereas the usefulness of gene-assisted diagnostic techniques, such as polymerase chain reaction (PCR) and in situ hybridization, has been emphasized3-4, they have not been established with sufficient accuracy for practical use. Therefore, until now, it is still common for filamentous fungi to be diagnosed or identified by histopathology, which is the classical method of examination. Meanwhile, it has been empirically known by pathologists in the field of diagnostic pathology that Aspergillus tends to elongate its hyphae straight.
through the tissue more than other filamentous fungi. Therefore, we first considered the calculation of the angles by which hyphae intersect as meta-information for quantifying this pathological finding. To prove the pathologists’ empirical observation, we initially analyzed these angles using a semi-automated procedure to confirm the distinction between *Aspergillus* and Mucorales genera. Next, since many pathologists also empirically know that the folding or tortuosity of hypha is more pronounced in Mucorales, we calculated the ratio of the long to short sides of the minimum bounding rectangle and decided to quantify this as the ratio of the area of the bounding rectangle to the area of the actual hyphae. In the present study, using these two indicators, we devised an algorithm to quantitatively distinguish *Aspergillus* and Mucorales. For these two indicators, we applied an automated recognition system for bounding rectangles. This system has received ample recognition in the field of automatic shape recognition, a field that has been advancing remarkably in recent years. Furthermore, in this study, teacher data were inputted to induce deep learning to enhance our artificial intelligence (AI) system and to set thresholds for *Aspergillus* and Mucorales by two-dimensional plotting.

Surgical pathologists are aware of the day-to-day variability of histopathological diagnosis and want to ensure objective diagnosis. Thus, if AI can improve current diagnostic procedures, it will be useful in designing disease management strategies. This paper describes an attempt to use AI technology to analyze images of filamentous fungal infection cases and differentiate the etiological fungus.

**Materials and methods**

We aimed to develop an AI system that can determine the hyphal angle and bounding rectangle of each hypha by Python and OpenCV systems, both of which exhibit these capabilities (Fig. 1a–d).

**Image analysis and data (indicators) extraction**

First, images of samples of sections stained with Modified Gomori Methenamine-Silver Nitrate (GMS) were obtained with a microscopic image capturing system (Olympus DP27, Olympus co. ltd, Tokyo, Japan) at x20 objective and saved as jpg files. After binarization of these captured images, the outer edge of each individual hypha that can be identified (approximately up to 200 per jpg file) was highlighted by hollowing out the interior portion.

For each extracted hypha, a bounding rectangle was set up as the primordial data for our investigation and is highlighted with a yellow line in Fig. 1b. Since it is unsuitable to extract approximated straight lines from hyphae that show extreme folds, for analysis, we selected those hyphae whose bounding rectangles had short sides in the range of 10 to 30 pixels. The line segment was extracted as an approximated straight line of the target hypha that connects the two points of contact between the short sides of the bounding rectangle, and its slope and distance were measured for each of the points of contact (Fig. 1b).

The slopes of the approximated line segments were added up and all crossing angles composed of straight lines were measured (Fig. 1c).

The ratio of the area of the bounding rectangle to the area of the actual hyphae was measured as an index to determine tortuosity reflecting the degree of folding (Fig. 1d).

Using sigmoid function as an activation function, we constructed a logistic regression analysis system based on the perceptron as the fundamental theory and used it to generate an optimized discrimination curve for the two different fungi on a two-dimensional plot with the two indicators.

**PCR method for confirmation of histopathologically proven autopsy cases used as teacher data for deep learning in the AI analysis**

First, we reviewed the autopsy cases of histopathologically proven systemic fungal infection recorded in Toho University Omori hospital to prepare the teacher data for the AI analysis. Cases of both aspergillosis and mucormycosis, of which hyphae observed in the histological sections indicated the most typical pattern, were selected by three pathologists. Second, to confirm the diagnosis of these histopathologically proven cases, we carried out the PCR method with sequencing on their formalin-fixed and paraffin-embedded (FFPE) tissues. We used the PCR method with sequencing to identify fungal DNA as previously described with minor modification. Four panfungal primers (ITS1-ITS2, ITS3-ITS4, ITS1-ITS4) targeting the ITS region and two panfungal primers (NL1-NL4) targeting the D1/D2 region of fungal rDNA were tested. In addition to these panfungal primers, species-specific primers (*Aspergillus*: β tub1-β tub2 and Mucorales: ZM1-ZM2, ZM1-ZM3) were also tested, respectively (Table 1).

We detected one *Aspergillus fumigatus* case and one *Cunnighamella* spp. (one of the Mucorales) case; both were identified by sequencing fungal DNA, and these two cases were selected as teacher data for AI analysis. The case of *Cunnighamella* spp. was used only as teacher data; histologic data in this study have been used for a case presented in another study published by Sadamoto et al.

**Test data preparation for AI analysis**

Test data are needed to verify the validity of the threshold values calculated from the teacher data. We decided to use the histopathologically proven but non-PCR-confirmed cases of aspergillosis and mucormycosis that have typical morphology.
Fig. 1. Image processing procedures for automatic recognition.

- **a:** Capturing, binarization, and hollowing of hyphal silhouette.
  First, images were obtained at x20 objective and saved as jpg files (Step 1). After binarization of the captured images (Step 2), the outer edge of each individual hypha was highlighted (Step 3).

- **b:** Setting up a bounding rectangle and extracting a straight line to approximate the hyphal shape.
  The line segment was extracted as an approximated straight line of the target hypha that connects the two points of contact between the short sides of bounding rectangle, and its slope and distance were measured for each of them.

- **c:** Measurement of intersection (crossing) angles.
  The slopes of the approximated line segments measured in 1b were accumulated and all intersection (crossing) angles composed of straight lines were measured.

- **d:** Measurement of tortuosity (degree of hyphal folding).
  The ratio of the area of the bounding rectangle created in 1B to the area of the actual hyphae was measured as an index to determine tortuosity (degree of folding).

### Table 1. Nucleotide sequences of primers

| Primer | Sequence | Reference |
|--------|----------|-----------|
| ITS1   | 5'-TCC GTA GGT GAA CCT GCG G-3' | 11        |
| ITS2   | 5'-GCT GCG TTC TTC ATC GAT GC-3' | 11        |
| ITS3   | 5'-GCA TCG ATG AAG AAC GCA GC-3' | 11        |
| ITS4   | 5'-TCC TCC GCT TAT TGA TAT GC-3' | 11        |
| NL1    | 5'-GCA TAT CAA TAA GCG GAG GAA AA-3' | 12       |
| NL4    | 5'-GGT CCG TGT TTC AAG ACG G-3' | 12        |
| β tub1 | 5'-AAT TGG TGC CGC TTT CTG G-3' | 13        |
| β tub2 | 5'-AGT TGT CGG GAC GGA ATA G-3' | 13        |
| ZM1    | 5'-ATT ACC ATG AGC AAA TCA GA-3' | 14        |
| ZM2    | 5'-TCC GTC AAT TCC TTT AAG TTT C-3' | 14        |
| ZM3    | 5'-CAA TCC AAG AAT TTC ACC TCT AG-3' | 14        |
We took images randomly. Test data were automatically recognized and extracted from images taken by the same procedure from one autopsy case. Data for each case of mucormycosis and aspergillosis were plotted on a two-dimensional plot diagram created with the teacher data. This study protocol was approved by the Ethics Committee of the Faculty of Medicine, Toho University (Approval number: A20040_A17172 and M21205), and the details of the study are available on our hospital’s webpage.

Results

In total, 147 image samples were collected from a standard case of mucormycosis and 67 from that of aspergillosis. In the former, the average number of hyphae that could be automatically recognized in a single image sample was 137.1 with a standard deviation of 65.5. In the latter, the average number of those that could be automatically recognized was 215.7 with a standard deviation of 73.6. We then combined and plotted the two-dimensional (2-D) data from the two control cases as teacher data (Fig. 2a). All data, which included the two automatically recognized indicators from each hypha, were plotted after omitting the noise, such as heavy hyphal mass or collagen fibers. Representative images of *Aspergillus* and Mucorales are shown in Fig. 2b and 2c. Then, we investigated the details of the effect on the shape of the threshold curve of the number of automatically recognizable hyphae per image. The number of hyphae suitable for analysis in each image taken was different; we assumed that the quality of the analysis would improve with more hyphae suitable for analysis in a single image. Therefore, in addition to a plot of data from all images that were recognizable (Fig. 2a), we also plotted data extracted from images for which our system was able to recognize more than 100 hyphae per image (Fig. 3a), and data extracted only from images containing more than 150 recognizable hyphae (Fig. 3b). As a result, the threshold curve, which was a quadratic function curve, became closer to a straight line as more specimens with a larger number of hyphae were plotted from a single image.

Finally, to validate the accuracy of our automated diagnostic system, we randomly chose other cases from our...
records. Data extracted from each case of histopathologically proven but with no PCR products were obtained for aspergillosis (Fig. 4a) and mucormycosis (Fig. 4b) and were plotted on the 2-D plot generated from the teacher data (Fig. 2a). Both the Aspergillus and Mucorales regions were sufficiently far from the threshold curve (Fig. 4c).

Discussion

Whereas Aspergillus and Mucorales are both filamentous fungi, it has been known that their responses are different against anti-fungal agents\(^\text{10}\). In addition, invasive fungal infections usually occur as opportunistic infections in patients with serious predisposing diseases, such as leukemia or autoimmune disease, and, recently, co-infection with coronavirus disease 2019 (COVID-19). The latter has been referred to as COVID-19-associated pulmonary aspergillosis (CAPA) and/or mucormycosis (CAM), which have become clinically important fungal infections\(^\text{11, 12}\). Therefore, we wanted to establish rapid and accurate histological diagnosis procedures that are equally accessible to all pathologists, especially those engaged in the diagnosis of infectious diseases.

It is important to note that the shape or outline of fungi that pathologists observe is usually influenced and/or deformed by host defense reaction mechanisms and anti-fungal agents. In a previous study, we analyzed images of Aspergillus and Mucorales on tissue sections stained with GMS and found that the average hyphal crossing angle of Aspergillus was acute, with a standard deviation less than that of Mucorales\(^\text{5}\). We manually recognized the fungi in the previous study, so a more objective viewpoint was needed. In the present study, the presence of fungi was recognized by our AI system, which calculated the crossing angles generated by hyphae themselves and the tortuosity of each hypha.

For each case, we assessed multiple images. If the number of plotted hyphae was not sufficient, the crossing angle analysis was not definitive. Therefore, we separately plotted additional graphs with more than 100 and more than 150 hyphae from each image. With more than 150 hyphae (Fig. 3b), interestingly, the threshold curve, which was a quadratic function curve, became closer to a straight line as more recognized points from a larger number of hyphae were plotted from a single image. This result suggests that the ideal threshold curve with a linear shape can be attained by inputting more information from PCR-proven teacher data.

In the present study, we used data for known cases of aspergillosis and mucormycosis that were confirmed by PCR and illustrated the characteristics of both cases. These findings objectively support the conclusions of our previous study.

Before concluding, we would like to note some important
limitations of this study. In this method, crowded huge masses of hyphae must be excluded because they cannot be analyzed as bounding rectangles, but conversely, a large number of such masses can lead to an extremely small sample size, making it difficult to analyze them adequately. Also, we had to use autopsy subjects for the present study, of which materials tend to be affected by excessive formalin fixation. This procedure causes the destruction of proteins and nucleic acids (known as nicks and gaps) that usually induces false negative PCR results. In addition, these materials may disintegrate over several years. These limitations illustrate the need to develop a more robust system that can differentiate fungal genera. This study may be considered as the first step toward that goal. Going forward, we will improve this method so that it will contribute to adequate diagnosis and therapy.

Conclusion

In the future, many analyses will use machine learning, and AI systems will become part of diagnostic procedures. Although these methods are ideal for medical use, we should avoid making diagnoses that are not clear-cut. Future studies will aim to express the findings in numerical terms.

Acknowledgments

This work was supported by AMED 21fk0108094h0003 and 22fk0108135h0303, and MHLW Research Program on Emerging and Reemerging Infectious Diseases (Grant Number JPMH21HA2011).

Authors’ contributions

Naobumi Tochigi designed this study. Sota Sadamoto designed the PCR methods. Shinji Oura built the AI system. Yasuko Kurose carried out the analyses. Yoshitsugu Miyazaki and Kazutoshi Shibuya oversaw the study.

Conflicts of interest

None.

References

1) Guarner J, Brandt ME: Histopathologic diagnosis of fungal infections in the 21st century. Clin Microbiol Rev 24: 247-280, 2011.
2) Okubo Y, Ishiwatari T, Izumi H, Sato F, Aki K, Sasai D, Ando T, Shinozaki M, Natori K, Tochigi N, Wakayama M, Hata Y, Nakayama H, Nemoto T, Shibuya K: Pathophysiological implication of reversed CT halo sign in invasive pulmonary mucormycosis: a rare case report. Diagn Pathol 8: 82, 2013.

3) Shinozaki M, Tochigi N, Sadamoto S, Yamagata Murayama S, Wakayama M, Nemoto T: Histopathological diagnosis of invasive fungal infections in formalin-fixed and paraffin-embedded tissues in conjunction with molecular methods: comparison of reproducibility and reliability of histopathological evaluation, polymerase chain reaction, and in situ hybridization. Med Mycol J 59: E7-E18, 2018.

4) Sadamoto S, Shinozaki M, Nagi M, Nihonyanagi Y, Ejima K, Mitsuda A, Wakayama M, Tochigi N, Murakami Y, Hishima T, Nemoto T, Nakamura S, Miyazaki Y, Shibuya K: Histopathological study on the prevalence of trichosporonosis in formalin-fixed and paraffin-embedded tissue autopsy sections by in situ hybridization with peptide nucleic acid probe. Med Mycol 58: 460-468, 2020.

5) Tochigi N, Sadamoto S, Shinozaki M, Wakayama M, Shibuya K: Comparison in quantities from including angles comprising lines of hypha themselves in histological images between Mucorales and Aspergillus. Med Mycol J 60: 85-89, 2019.

6) Eunju K, Ayman H: Automatic representation and reconstruction of DBM from LiDAR data using Recursive Minimum Bounding Rectangle. ISPRS J Photogramm Remote Sens 93: 171-191, 2014.

7) Rosenblatt F: The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65: 386-408, 1958.

8) Austin PC, Merlo J: Intermediate and advanced topics in multilevel logistic regression analysis. Stat Med 36: 3257-3277, 2017.

9) Sadamoto S, Mitsui Y, Nihonyanagi Y, Amemiya K, Shinozaki M, Murayama SY, Abe M, Umeyama T, Tochigi N, Miyazaki Y, Shibuya K: Comparison approach for identifying missed invasive fungal infections in formalin-fixed, paraffin-embedded autopsy specimens. J Fungi (Basel) 8: 337, 2022.

10) Patterson TF, Thompson GR 3rd, Denning DW, et al: Practice guidelines for the diagnosis and management of aspergillosis: 2016 update by the Infectious Diseases Society of America. Clin Infect Dis 63: e1-e60, 2016.

11) Lai CC, Yu WL: COVID-19 associated with pulmonary aspergillosis: a literature review. J Microbiol Immunol Infect 54: 46-53, 2021.

12) Pal R, Singh B, Bhadada SK, Banerjee M, Bhogal RS, Hage N, Kumar A: COVID-19-associated mucormycosis: an updated systematic review of literature. Mycoses 64: 1452-1459, 2021.