VOLUME - 10, ISSUE - 12, DECEMBER - 2021 • PRINT ISSN No. 2277 - 8160 • DOI : 10.36106/gjra

Original Research Paper

Microbiology

ARTIFICIAL NEURAL NETWORKS (ANNs) APPLIED TO ATR-FTIR SPECTRA TO CLASSIFY MEDICALLY IMPORTANT *Trichosporon* SPECIES

Abhila Parashar*	Research scholar Department of Microbiology, Jawaharlal Nehru Medical College JLNMC & Hospital, Ajmer Rajasthan, India. *Corresponding Author
Dr. Vijaylatha Rastogi	Sr. Professor, Department of Microbiology, Jawaharlal Nehru Medical College JLNMC & Hospital, Ajmer Rajasthan, India.
Mitanshu Sharma	Research scholar, Department of Smart Healthcare Devices, Indian Institute of Technology IIT, Jodhpur Rajasthan, India.
Dr. Monica Bhatnagar	Professor, Department of Microbiology, Maharshi Dayanand Saraswati University MDSU, Ajmer Rajasthan, India.

ABSTRACT To distinguish clinically significant fungus, Fourier transform infrared spectroscopy (FTIR) was used. In this work, 75 *Trichosporon* strains from five different species were cultivated on SDA media and FTIR attenuated total reflection (ATR) readings was taken. The classification (FTIR spectra) results of cluster analysis were compared to artificial neural network (ANN) analysis (supervised approach). Validation of training set showed that both techniques properly categorized 100% of the spectra, at least for *T. asahii* (n = 62) and *T. inkin* (n = 8). With the addition of *T. loubieri* (n=1) and *T. asteroids* (n=1), the ANN's accuracy became reliant on the training database, resulting in 90% to 100% classification.

KEYWORDS : Trichosporon, ATR-FTIR, Cluster analysis, Artificial Neural Network.

INTRODUCTION

Trichosporon species are yeast-like basidiomycete fungus that are commonly found in the human body[1]. This species has been identified as an emerging pathogen that causes fatal fungal invasive infections in malignancy, neutropenia, or severely sick patients receiving antibiotics/corticosteroids, or undergoing invasive medical operations in recent years[1–3]. *Trichosporon* is the most important arthroconidial yeast that causes invasive fungal infections in humans, after *Cryptococcus*[1,4]. In reference to recent taxonomic revision of *Trichosporon* species, there are only 20 species in this genus that can be properly recognized by sequencing IGS1 rDNA[5]. Some therapeutically important species, such as *Cutaneotrichosporon dermatis*, *Apiotrichum mycotoxinivorans*, and *Cutaneotrichosporon cutaneum*, have been moved to other genera[6].

The technique of attenuated total reflectance FTIR is used to acquire an IR spectrum in mode of absorbance/ transmittance of a variety of substances. Many microbial species have been detected and differentiated using this technique [7]. The necessity of analyzing vast amounts of data is a major issue when using ATR-FTIR for microbiological diagnosis. Artificial neural networks (ANNs) are used to solve this challenge. ANNs adapt their structure during data analysis, allowing them to reduce classification error. The goal of this study was to employ artificial neural networks to classify *Trichosporon* species using ATR-FTIR spectroscopic data.

2. MATERIALS & METHODS

2.1. Study isolates

A collection of 75 clinical isolates were previously characterized by phenotypic and PCR-based sequencing techniques and related to 5 different *Trichosporon* species [*T. asahii* (62), *T. inkin* (8), *Cutaneotrichosporon dermatis* (formerly *T. dermatis*) (3), *T. asteroids* (1), *Apiotrichum loubieri* (formerly *T. loubieri*) (1)] of *Trichosporon* were exposed to ATR-FTIR analysis. All the isolates were already submitted to GenBank [8]. ATCC strains (*S. aureus* ATCC 6538, *Candida albicans* ATCC 90028, *C. krusei* ATCC 6258) were used as outliers. Culture stocks were thawed and sub-cultured on SDA containing chloramphenicol and gentamicin and incubated at 30°C temperature for 48 hours for ATR-FTIR spectroscopy.

2.2. Spectral analysis

A "Spectrum two-UATR"(Perkin Elmer, Waltham, USA) spectrometer was used to measure the IR spectra. For all measurements, spectral data was collected under regulated microbiological and physical conditions. Using a sterile disposable loop, a single colony was extracted and put directly to ATR sample window. A spectrum was created by combining 100 scans in the spectral region of 4,000-400 cm-1 with a spectral resolution of 0.5 cm-1 and subtracting a blank background spectrum acquired from the clean sample surface. After spectral acquisition, the ATR sampling surface was disinfected by wiping it with lint-free paper dampened with 70 percent ethanol. Three samples of each culture were used in the process [9]. Unsupervised multivariate analyses (HCA) and(PCA)) were used to study infrared spectra. For supervised learning approach, MATLAB software was used for data processing and the design of artificial neural networks.

2.3. Pre- processing of raw spectra

The Baseline Subtracted system-generated raw spectra (for each isolate in triplicate) were subjected to Pearson coefficient analysis for the removal of outlier spectrum among triplicates. Averaging the remaining spectra for the single isolate was done to obtain the final unsmoothed spectrum.

2.4. Spectral smoothening and normalization

The average spectra were smoothened using Savitsky-Golay algorithm with nine point window along with second-degree polynomial (Jupyter notebook interface library versions-pandas 1.0.3, NumPy 1.19.5 & scipy 1.5.4in Python language) and then vector normalized[9].

2.5. Spectral window selection

The lipids, proteins, and polysaccharides that make up the membrane, cell wall, and capsules have a big influence on the yeasts' systematics and phylogeny. The spectrum obtained in the 4000-400cm^{-1}region was divided into 4 windows viz. a). 900-400cm^{-1}, the fingerprint region. b). 1500-900cm^{-1} with dominant signatures of carbohydrates, c). 1800- 1500cm^{-1} the amide region, d). 4000-2800cm^{-1} representing water & lipid region and further analyzed.

2.6. Spectral data input for Artificial Neural Network (ANN) model development

The classifications of selected 75 spectra from section 2.5 were systematically assigned by integer values to make it logical by ANN software (**Table.1**).

Table. 1 Classifiers assigned for spectral data used for ANN model development

Species (Isolates n=75)	Assigned value or Target
T. asahii (n = 62)	0
T. inkin (n = 8)	1
Cu. dermatis (n=3)	2
T. asteroides (n=1)	3
T. loubieri (n=1)	4

2.7. Artificial neural network (ANN)

In this study, the general technique for ANN analysis involved training and improving network models, then evaluating the classifiers with independent validation data sets. On the basis of IR spectra with known class assignment of study isolates, training and validation were conducted out[10]. Non-linear correlations between signals and classes could be investigated using an artificial neural network (ANN) model. Finally, classification approaches can uncover spectral markers that can reveal information about a disease's origin. Cluster and ANN analysis were used to evaluate each spectrum.

For developing the ANN, the total 75 spectra were randomly divided by MATLAB GUI of neural network (nnstart) software into training (n=53), validation (n=11) and testing (n=11) data set. The training data set was supplied to the network during training, and the network was then changed based on its output. The validation data set was used to assess network generalization and to end training when generalization degraded. Finally, because the testing data set has no bearing on training, it may be used as an independent assessment of network performance both during and after training. The same process was further repeated with only *T. asahii* and *T. inkin* as numbers of samples sufficed to run ANN program.

Second derivative spectral data of carbohydrate signatures (1200–900 cm-1 range) were vector-normalized for species level classification as per the earlier study[8]. The training of the connected feed-forward ANNs was performed with 5 neuron layers with one output layer by the Levenberg-Marquardt back-propagation algorithm as shown in **Figure.1**.

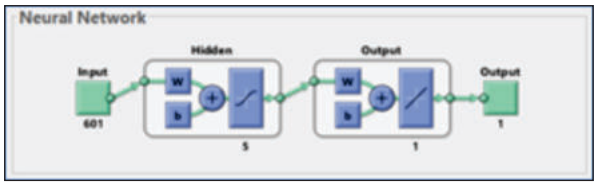

Figure 1. Connected feed-forward Artificial neural network program

The above-mentioned process in Figure.1 was repeated 100 times with random training, validation and test database distribution to check the robustness of the so formed ANN model.

3. RESULTS

The Band assignment of main ATR-FTIR peaks of *Trichosporon* given in **Table. 2** for the reference of spectral window selection.

Table.2. Band assignment of main ATR-FTIR peaks of *Trichosporon* [8]

Characteristic stretch/bend[a]	*Trichosporon*
C-O str. (Carbohydrate)	1035
PO_2 sym. str. (Nucleic acid)	1081
C-O-C str. (Glycosidic linkages)	1141
C-N str. (Protein)	1404
CH deformation of >CH2 in lipids protein	-
C=O str. Amide II (Protein)	1541
C=O str. Amide I	1630
Of β-pleated sheet structures of proteins	
C=O str. (Lipid)	1740
CH_2, CH_3 str. sym (Lipids-fatty acids))	2849
CH2, CH3 str. asym (Lipids)	2917
C-H tr. (Lipids)	3000
O-H str. Band of water	3286

asym. = asymmetric; sym. = symmetric; str. = stretching; def. = deformation.

Artificial neural networks were created to analyze spectral data. The neural networks had an input layer, a hidden layer with five neurons, and an output layer with one neuron. The ANN model randomly partitioned the input data into three subsets: a training set (70 percent of spectra), a test set (15 percent of spectra), and a validation set (15 percent of spectra).

The designed network differentiates *Trichosporon* species with variable accuracy as shown in **Figure. 2 & 3**.

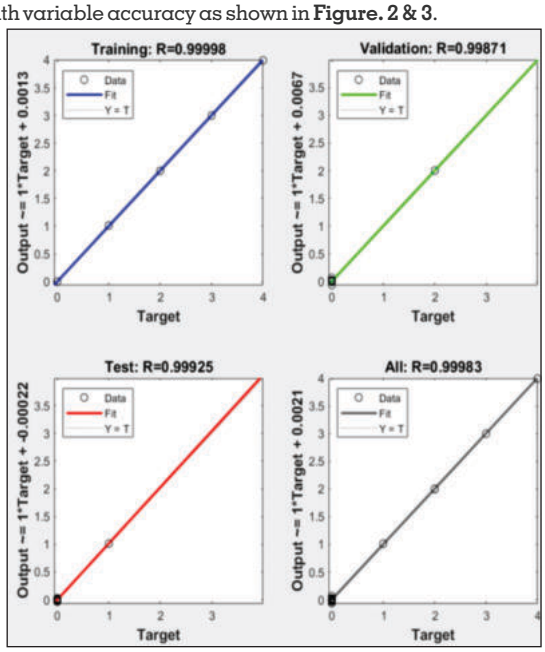

Figure 2. Complete correlation of Training, validation & testing database with ANN regression model

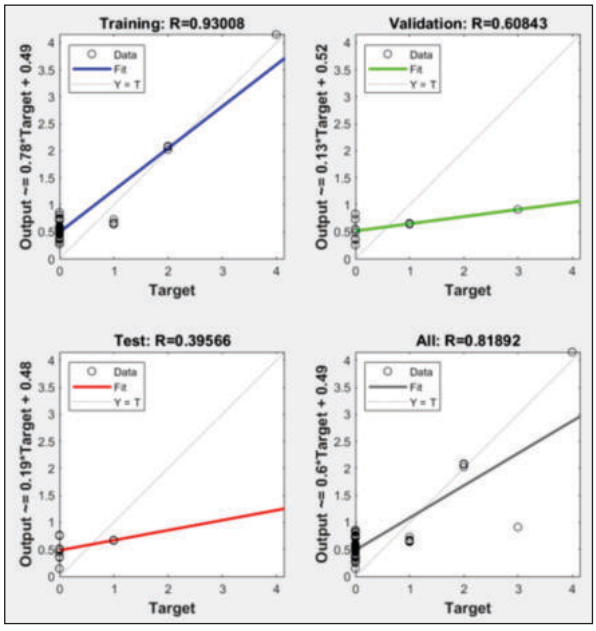

Figure.3 low level correlation of Training, validation & Testing ANN regression model

In **Figure. 2 & 3** the Regression (R) graphs represent a correlation between output of ANN (Y-axis) and classification provided on the basis of sequencing results (X-axis).

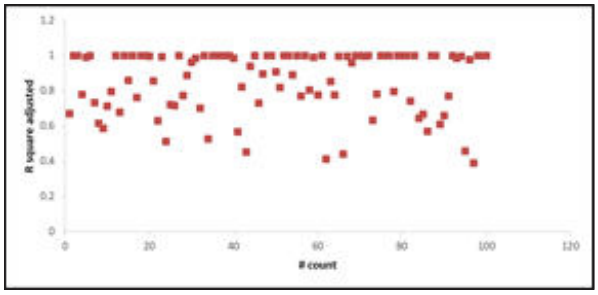

Figure. 4 Results with 100 ANN modeling with spectral data of all five species of *Trichosporon* showing partial correlation (R^2) with training data

The accuracy varied (47 out of 100 cases provided 100% accuracy in classification) due to the random selection of limited samples of *T. loubieri* and *T. asteroides*. When the entire process (100 ANN modeling) was repeated with only considering *T. asahii* (n = 62) and *T. inkin* (n=8). So in this case, for 99 out of 100 ANN modeling, the classification was 100% as shown in **Figure 5.**

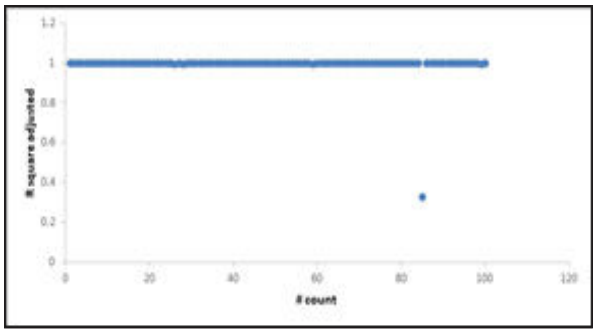

Figure. 5 100 ANN modeling with spectral data of two predominant species of Trichosporon showing complete correlation (R2) with training data

4. DISCUSSION AND CONCLUSION

In this present study, ATR-FTIR spectra of *Trichosporon* isolates were analyzed by neural networks. Unlike unsupervised classification, supervised classification (ANNs) necessitates a prior knowledge of the spectral data's class structure[11]. In this study, with the help of prior knowledge (training data-set), the ANNs could identify five distinct *Trichosporon* species with a variable accuracy of 47-100%. Direct interpretation of Trichosporon IR spectra does not allow for species level identification[12]. It is very difficult to link a specific carbohydrate/protein in the fungal cell to a specific peak in the IR spectra. The ATR-FTIR method can determine the type and quantity of functional groups in a sample[8,10,12–14].

From the above results, the accuracy of ANN depends highly on numbers of samples provided to the neural network. It could be a risk for classification using limited samples rather, one should go with multivariate (HCA, PCA & R&T calculation) analysis as suggested by other studies[8,10,15–19]. But if the database is sufficient, then ANN can classify with 100% accuracy[20]. In conclusion, we demonstrated that with a bigger dataset, the ATR FT-IR with ANNs may differentiate between various *Trichosporon* species.

Acknowledgement:
This work was supported by a post fellowship start up grant (SUG) to Dr Vijaylatha Rastogi from Department of Health Research (DHR), MOHFW, India [V.25011/ 285-HRD/ 2016-HR, dated 24-5-2018]. We are thankful to MRC (material research Centre) Lab, MNIT (Malviya National Institute of Technology), Jaipur to provide testing facility.

Conflict of interest: None
Permission for IEC: Yes

REFERENCES

1. Colombo AL, Padovan ACB, Chaves GM. Current Knowledge of Trichosporon spp. and Trichosporonosis. Clinical Microbiology Reviews. 2011 Oct 1;24(4):682–700.
2. Girmenia C, Pagano L, Martino B, D'Antonio D, Fanci R, Specchia G, et al. Invasive infections caused by Trichosporon species and Geotrichum capitatum in patients with hematological malignancies: a retrospective multicenter study from Italy and review of the literature. J Clin Microbiol. 2005 Apr;43(4):1818–28.
3. de Almeida Júnior JN, Hennequin C. Invasive Trichosporon Infection: a Systematic Review on a Re-emerging Fungal Pathogen. Front Microbiol [Internet]. 2016 [cited 2020 Jan 23];7. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2016.01629/full
4. Pfaller MA. Molecular approaches to diagnosing and managing infectious diseases: practicality and costs. Emerg Infect Dis. 2001;7(2):312–8.
5. Liu X-Z, Wang Q-M, Göker M, Groenewald M, Kachalkin AV, Lumbsch HT, et al. Towards an integrated phylogenetic classification of the Tremellomycetes. Studies in Mycology. 2015;81:85–147.
6. Liu X-Z, Wang Q-M, Göker M, Groenewald M, Kachalkin AV, Lumbsch HT, et al. Towards an integrated phylogenetic classification of the Tremellomycetes. Studies in Mycology. 2015 Jun 1;81:85–147.
7. Yu C, Irudayaraj J. Spectroscopic characterization of microorganisms by Fourier transform infrared microspectroscopy. Biopolymers. 2005 Apr 15; 77(6): 368–77.
8. Rastogi V, Parashar A, Bhatnagar M, Sharma M, Rudramurthy SM V. Evaluation of Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) Spectroscopy for Characterization of Trichosporon Species Isolated from Clinical Samples. Int J Pharm Clin Res. (in press).
9. Lasch P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemometrics and Intelligent Laboratory Systems. 2012 Aug 1;117:100–14.
10. Lasch P, Diem M, Hänsch W, Naumann D. Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. J Chemom. 2007 Mar 28;20(5):209–20.
11. Dayhoff JE. Neural Network Architectures: An Introduction. Van Nostrand Reinhold; 1989. 259 p.
12. Lechowicz Ł, Adamus-Białek W, Kaca W. Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy and Artificial Neural Networks Applied to Differentiate Escherichia coli Strains. Severcan F, editor. Journal of Spectroscopy. 2012 Dec 3;2013:538686.
13. Costa FSL, Silva PP, Morais CLM, Arantes TD, Milan EP, Theodoro RC, et al. Attenuated total reflection Fourier transform-infrared (ATR-FTIR) spectroscopy as a new technology for discrimination between Cryptococcus neoformans and Cryptococcus gattii. Anal Methods. 2016 Oct 6;8 (39): 7107–15.
14. Ferreira ICC, Aguiar EMG, Silva ATF, Santos LLD, Cardoso-Sousa L, Araújo TG, et al. Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) Spectroscopy Analysis of Saliva for Breast Cancer Diagnosis. Journal of Oncology. 2020 Feb 10;2020:e4343590.
15. Liu D, Caliskan S, Rashidfarokhi B, Oldenhof H, Jung K, Sieme H, et al. Fourier transform infrared spectroscopy coupled with machine learning classification for identification of oxidative damage in freeze-dried heart valves. Sci Rep. 2021 Jun 10;11(1):12299.
16. Naumann A. A novel procedure for strain classification of fungal mycelium by

cluster and artificial neural network analysis of Fourier transform infrared (FTIR) spectra. Analyst. 2009 Jun 1;134(6):1215–23.
17. Bhat R. Potential Use of Fourier Transform Infrared Spectroscopy for Identification of Molds Capable of Producing Mycotoxins. International Journal of Food Properties. 2013 Nov 17;16(8):1819–29.
18. Biancolillo A, Marini F. Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. Front Chem [Internet]. 2018 [cited 2021 Sep 8];0. Available from: https://www.frontiersin.org/articles/10.3389/ fchem. 2018. 00576/full
19. Helm D, Labischinski H, Schallehn G, Naumann D. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. Microbiology. 1991 Jan 1;137(1):69–79.
20. Lasch P, Naumann D. Infrared Spectroscopy in Microbiology. In: Encyclopedia of Analytical Chemistry [Internet]. American Cancer Society; 2015 [cited 2021 Sep 18]. p. 1–32. Available from: https:// online library. wiley. com/ doi/abs/10.1002/9780470027318.a0117.pub2