



Radiomics Model for Mycetoma Grains Classification from Histopathological Microscopic Images Using Partial Least Squares Discriminant Analysis (PLS-DA)

Hyam Omar Ali, Romain Abraham, Guillaume Desoubes, Ahmed H Fahal,
Clovis Tauber

► To cite this version:

Hyam Omar Ali, Romain Abraham, Guillaume Desoubes, Ahmed H Fahal, Clovis Tauber. Radiomics Model for Mycetoma Grains Classification from Histopathological Microscopic Images Using Partial Least Squares Discriminant Analysis (PLS-DA). 2021. hal-03247688

HAL Id: hal-03247688

<https://hal.science/hal-03247688>

Preprint submitted on 3 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Radiomics Model for Mycetoma Grains Classification from Histopathological Microscopic Images Using Partial Least Squares Discriminant Analysis (PLS-DA)

Hyam Omar Ali^{1,2,3,4}, Romain Abraham⁴, Guillaume Desoubieux^{5,6}, Ahmed H.Fahal², Clovis Tauber³,

1 Faculty of Mathematical Sciences, University of Khartoum, Sudan

2 The Mycetoma Research Center, University of Khartoum, Khartoum, Sudan

3 INSERM U1253-iBrain, Tours University, Tours, France

4 Institute Denis Poisson (IDP), University of Orleans, Orleans, France

5 Parasitology and mycology Department, Bretonneau Hospital, Tours, France

6 INSERM U1100-Team 3, Tours University, Tours, France

Abstract

Mycetoma is a chronic granulomatous inflammatory disease that causes severe deformities, disabilities, with many impact on patients and family, particularly in advanced disease stages or when treatment fails. The therapeutic disease strategy heavily relies on the identification of the causative organism and the corresponding classification of the disease as eumycetoma or actinomycetoma. Various diagnostic tools are used for mycetoma differential diagnosis. Histopathology is considered to be an efficient, cost and time-effective tool for mycetoma diagnosis in endemic areas. While histology is currently, the most used diagnostic tool, it requires well-trained pathologists, and that lacks in most rural areas where mycetoma is endemic. In this communication, we present a computational method to effectively differentiate between eumycetoma and actinomycetoma from the grains features in histopathological microscopic images that is based on Radiomics and Partial Least Squares Discrimination Analysis (PLS-DA). In this work, the data were collected from the Mycetoma Research Center of Khartoum, and the proposed approach achieved mycetoma types identification with an accuracy of 91.8% and 0.836 Matthew's Correlation Coefficient (MCC). This computational tool could be of great benefit in rural areas with limited access to specialised clinical centres.

Key words:

Mycetoma, Grains, Radiomics, PLS-DA, Histopathology, Image analysis.

Author summary

Mycetoma is a badly neglected tropical disease that commonly affects poor communities in rural areas. It is classified into actinomycetoma and eumycetoma depending on the causative organisms. Several diagnostic tools are used for the diagnosis of mycetoma causative agents, that include cytology, histopathology, culture, and molecular technique. However, the latter tool requires well-equipped centres that are not available in rural endemic regions. Therefore, cytology, culture, and histopathology are more commonly in use. The histopathological technique is more accurate than cytological one and faster than grains culture. Since the histological technique is operator-dependent, we

introduced a novel computational method that has significant discrimination power for mycetoma types using grains features in histopathological microscopic tissue images. We believe that our method will have a valuable impact on mycetoma patients' diagnosis and management.

1 Introduction

2 Mycetoma is a WHO recognised neglected tropical disease that was included in the WHO/
3 NTD list in 2016. It is a chronic granulomatous inflammatory disease. It is reported
4 worldwide but endemic in tropical and sub-tropical areas. Majority of cases occur in
5 the "Mycetoma belt" stretching between the latitudes of 15° South and 30° North. The
6 belt countries are Sudan, Somalia, Senegal, India, Yemen, Mexico, Venezuela, Colombia,
7 Argentina [1,2]. Sudan, India and Mexico reported the greatest number of cases [3,4].
8 The most susceptible group of mycetoma infection is young adults in remote rural areas.
9 Mycetoma largely affects field labourers, agriculturalists and herdsmen [1,4,5]. The
10 lower extremity and hands are the frequently infected sites comparing to the other body
11 sites [2–4]. The infection might spread to involve the deep structures and bone resulting
12 in destruction, deformity, loss of function, and occasionally mortality [5–7].
13 Mycetoma is an inflammatory, painless, and slowly progressive disease caused by certain
14 types of bacteria (Actinomycetoma) or fungi (Eumycetoma). While the mode of myce-
15 toma transmission is still unknown [8], the literature suggests that causative organisms
16 are present in the soil, thorns, or animal dung and can enter the subcutaneous tissue
17 through minor trauma [2,8]. Mycetoma is characterised by subcutaneous mass with
18 multiple sinuses discharge grains containing colonies of the causative organism, as shown
19 in Fig1. These grains are considered as a unique characteristic of the disease [4,6,7,9].
20 Identification of causative organisms plays a significant role in the treatment of mycetoma,
21 which requires prolonged administration of anti-fungal or antibiotics drugs depending
22 on the mycetoma type [1,4,5,10,11]. Incorrect diagnosis of mycetoma can have serious
23 consequences on the patient and the disease prognosis and outcome. Currently, there are
24 many efforts aiming at establishing an early identification of the causative organism **12**



Fig 1. Massive foot Actinomycetoma showing subcutaneous mass with multiple sinuses and discharge.

25 Several diagnostic tools are used to identify causative organisms and to ascertain
26 the disease extension along the tissue planes. These tools include; imaging, cytology,
27 histological, culture, and molecular techniques [10,12]. The imaging techniques such
28 as X-ray, CT and MRI, define the extent of disease [13–16], while the other tools are
29 employed to recognise the causative organisms. Grain culture, cytological and histological
30 are commonly used in endemic areas [17].

31 Grain culture is a core tool for organisms' identification, but it is time-consuming and
32 requires expert microbiologists to obtain accurate results. Also, this method is vulnerable

33 to false-positive results due to contamination. Although molecular techniques provide
34 authenticated results, it is expensive and cannot be afforded by the majority of patients,
35 and require well-equipped infrastructures which are not available in most of the mycetoma
36 endemic area [6, 9]. On the other hand, cytological and histopathological techniques
37 are simple, rapid, cheap methods and commonly used in rural areas where most of the
38 affected populations located [2, 6, 9]. However, false-negative results are common in
39 cytology because fine-needle aspiration for cytology (FNAC) is blindly performed, and it
40 is possible to miss grain pockets in the tissues. A recent comprehensive study conducted
41 at the Mycetoma Research Center (MRC) showed that histological technique is more
42 accurate than the cytological one in organism identification [18].

43 The histopathological method can only differentiate between mycetoma fungal and
44 bacterial types conditionally by the availability of grains in tissue sections [9]; this because
45 the tissue reactions are similar in both types of mycetoma and to other non-specific
46 chronic granuloma diseases [6]. This discrimination mainly relies on the knowledge and
47 experience of pathologists on the microscopic appearance of the organism [6, 10, 18–20].
48 Due to the neglect of mycetoma and its high prevalence in the tropical regions, especially
49 rural areas, it is rare to find well-trained pathologists with adequate experience in
50 mycetoma diagnosis. To tackle this, the present work was conducted to provide an
51 open-source tool for objective identification of the mycetoma causative organisms. This
52 work introduces a new computational mycetoma differential diagnostic method based on
53 histopathological image analysis.

54 The proposed method seeks at identifying causative organisms of mycetoma with an
55 efficiency that is comparable to expert-based diagnosis in specialised clinical centres.
56 We developed a quantitative method for the discrimination of the mycetoma types from
57 grains properties in histopathological microscopic images. In the proposed approach,
58 grains in microscopic images are first characterised through *Radiomics features* [21–24]
59 and then classified using Partial Least Square Discriminant Analysis (PLS-DA) [25–27].
60 The radiomics are used to extract a large number of quantitative features, which vary
61 from classical features such as first-order statistics to advanced ones that involve texture
62 and spatial characteristic of the grains. Once the features of the grain are calculated, we
63 estimated a PLS-DA model to identify new variables, which are a linear combination
64 of the original features that can be used to discriminate the grains based on their
65 causative organism. The proposed method can be implemented as a robust, reliable and
66 user-friendly software for the analysis of histopathological images, providing a solution
67 for medical needs for mycetoma diagnosis.

68 **Materials and methods**

69 **Samples and Images Acquisition**

70 Two sets of data were included in this study. The main set of images acquired following
71 a unique protocol that was used to train and validate the model. The secondary set,
72 composed of images acquired in various conditions, to evaluate the robustness of the
73 proposed approach. The main data set included 55 patients with confirmed different
74 mycetoma types. Surgical biopsies were obtained from patients with various mycetoma
75 types, duration and clinical presentations seen at the MRC or from the field surveys in
76 Sudan after written informed consent. Nine patients had biopsies devoid of any grains
77 and were excluded from the study.

78 The patients were randomly selected among patients seen at the MRC during the last five
79 years to ensure homogeneity and accuracy of the diagnosis. There were 31 patients with
80 eumycetoma and 24 patients with actinomycetoma. The collected surgical biopsies were
81 fixed in 10% formal solutions, followed by paraffin-embedded tissue blocks preparation.

Rotary microtome was used to acquire 2–3 sections with $(3–5)\mu$ thickness. All sections were stained with Hematoxylin and Eosin stain (H&E) according to standard routine laboratory procedures at Bretonneau Hospital (CHRU) through Tissue-Tek Prisma instrument.

The main data set contained a total of 327 tissue microscopic images from the 55 patients which were used for model training and validation. The number of grains that could be used for each patient’s biopsy varied between one and six. Images were captured in RGB colour space with Nikon Eclipse 80i microscope by the conditions given in Table 1 and labelled with consideration for the patient ID in order to avoid statistical bias.

Table 1. Microscopic Acquisition Conditions

Parameter	Value
Brightness control	Knob 5/10 ND8 On ND32 On
Field diaphragm	Highest level
Magnification	10X
Dimension and Quality	800×600
Colour	Enhance and white auto
Field diaphragm knob	Highest level
Filter	6
NCB11 Filter	Off

The secondary set of data was composed of 14 actinomycetoma and 14 eumycetoma photomicrographs. This data set has diverse acquisition parameters and staining method. The H&E staining process was performed manually. Olympus microscope was used for images capturing with 10X magnification, while lighting and tuning conditions were not unified. This secondary data set covered several sources of technical variability that can limit the performance of predictive models [28].

Preprocessing and Features Extraction

Grains were manually segmented from the tissue photomicrograph using ImageJ software. Images were converted into weighted grey images before features extraction in order to limit the colour influence from the staining process [29]. Fig.2 shows illustrative eumycetoma and actinomycetoma microscopic images in the first and second column, respectively, along with the segmented grains.

Radiomics features were composed of 102 variables divided into nine shape descriptors, 18 first-order statistics, and 75 texture features. Shape features were extracted from the grains masks while all other features were based on grey intensity values at the pixel level in microscopic images classes. Features were standardised by auto-scaling and used as an input to build the predictive model. PyRadiomics package version 2.2.0 was used for feature extractions [30].

Modelling and Analysis

A PLS-DA model was adopted for grains classification. It is a supervised classification method that combines the properties of the Partial Least Square (PLS) regression model with a classification technique. For PLS-DA modelling, the features set X of images is analysed. The classes membership is translated into a dummy column vector Y by values of "1" and "−1" that indicates if a sample is eumycetoma (FM) and actinomycetoma (BM). Each row in $X_{m \times n}$ represents the different extracted features (m) of one individual sample, where n indicates number of samples.

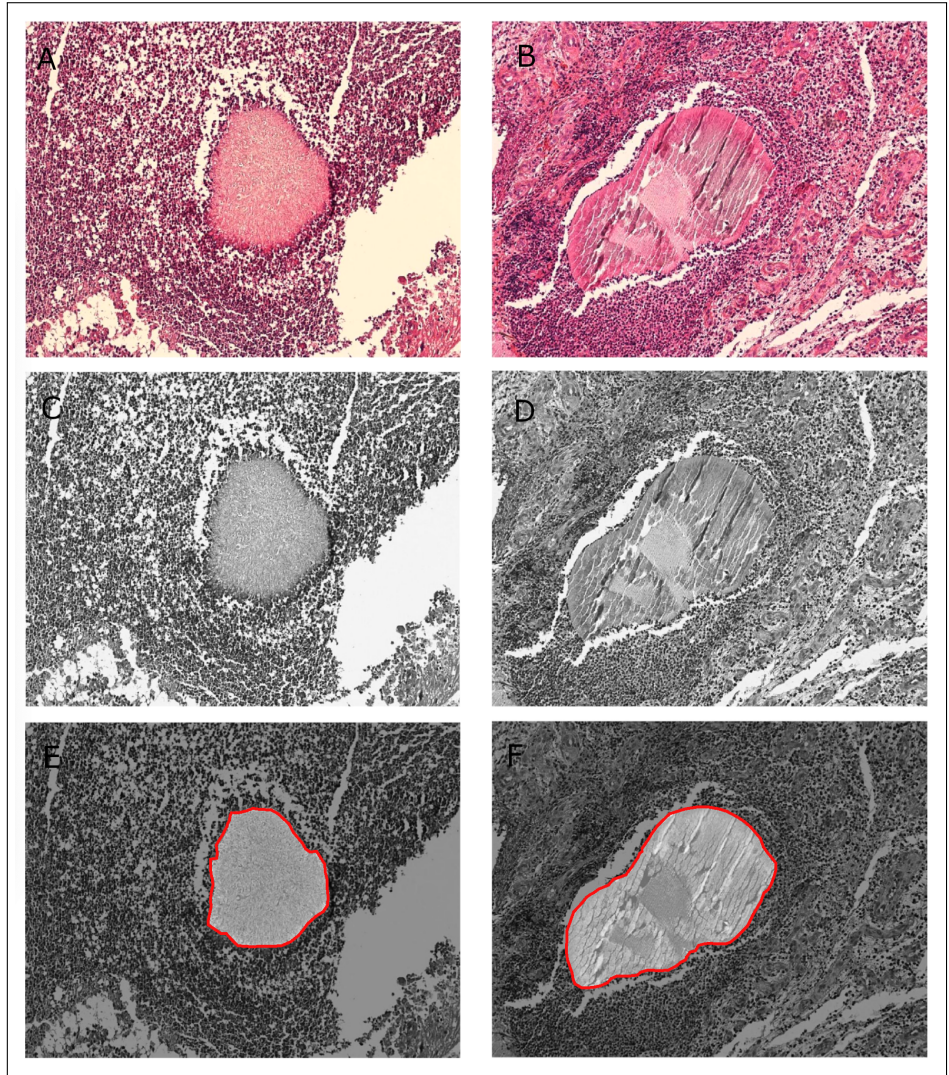


Fig 2. Image Preprocessing. In the first row, sections stained with H&E shows mycetoma grains, the second row exhibits the conversion into grey images, and the last one demonstrates the segmentation of grains. (A, C, E): eumycetoma, and (B, D, F): actinomycetoma

117 The procedure of the proposed PLS-DA model used is given in pseudocode(1) [25–27].
 118 The source of variability was modelled by Latent Variables (LVs) which are linear
 119 combinations of the extracted features in X . The maximum variation which secures
 120 from X is determined by the weight vector w . The whole set of features X was utilised
 121 for grains classification due to the ability of the PLS-DA model in reducing the impact
 122 of the irrelevant features. Hence, the loading vectors (p) and (q) are the coefficients
 123 assigned to features in their linear combination with various magnitude based on the
 124 importance of features, so loading vectors indicate the influence of each feature on each
 125 LV. Similarly, X-Score (t) represents the coordinates of samples in the LVs projection.
 126 Each LV generates a variation which sums up to the total of variation secured by the
 127 other LVs. The residual variation which has not been estimated by the current LV
 128 is updated as a new features set. Eventually, the matrix B provides the regression
 129 coefficients which describe the relationship between mycetoma grains features X and

130 mycetoma classes Y .

Algorithm 1 PLS-DA Model Construction.

Input: Features Set (X), dummy variable (Y), and number of LVs (A)

Output: Regression Matrix (B)

for i **in** A **do**

1. Weight Vector: $w = X'Y$.
2. X-Score: $t = \frac{Xw}{\sqrt{\sum w^2}}$.
3. X-Loading: $p = \frac{t'X}{\sqrt{\sum t^2}}$.
4. Y-Loading: $q = \frac{Y't}{\sqrt{\sum t^2}}$.
5. Regression Coefficient: $b_i = w(pw)^{-1}q$.
6. Residual of X : $res_x = X - tp$.
7. Residual of Y : $res_y = Y - tq$.
8. $X = res_x$.
9. $Y = res_y$.

end for

Regression Matrix B : $B = \{b_1, b_2, \dots, b_A\}$.

For the prediction purpose, the model utilises the regression coefficients B and features set X of unknown samples to predict whether they are FM or BM as follows:

$$Y_{pred} = XB$$

131 Since the PLS-DA model is inherited from the PLS model, the estimated Y_{pred} is never
 132 an integer with an exact membership (i.e 1 or -1). Several decision rules can be used to
 133 convert the predicted values into their essential classes [25]. In this study, the threshold
 134 for classes separation is calculated based on Bayes's theory and used to define perfect
 135 class membership. The Bayesian threshold calculation assumes the predicted values of
 136 both classes fit into a Gaussian distribution. This gives the probability of any sample
 137 belonging to class FM/BM from its predicted value Y_{pred} . The estimated threshold
 138 value is selected at the point where the number of false positives and false negatives is
 139 minimised [31–34].

140 The images of the different patients were randomly split into training/validation with
 141 70/30 proportions. A summary of the data set is presented in Table 2.

Table 2. Data Split.

	FM	BM	Total
Training	131	98	229
Validation	63	35	98
Total	194	133	327

142 All the analysis was performed using MATLAB software version R2017b and PLS
 143 Toolbox software version 88 from Eigenvector Technologies [35].

144 Quantitative evaluation

145 The proposed model was assessed in two different ways according to recommended
 146 practices [25, 36].

147 First, we performed Cross-Validation (CV) as an internal validation method to assess
 148 the complexity of the model by determining the optimum number of LVs. This allows
 149 to evaluate the complexity of the model considering the predictive ability of the model
 150 itself [25, 32]. The proposed model was trained with venetian blinds 10 folds-Cross-
 151 Validation (10 CV), and the minimum CV classification error was considered to select
 152 the LVs number [32, 37–39]. In our experiments, the smallest error was associated with
 153 6 LVs, Fig.3. However, using the high number of latent variables can often be associated
 154 with overfitting. Meaning that the model would have a good performance on the learning
 155 data, but would fail on other data sets. To prevent overfitting and to have a stronger
 156 generalisation capacity, we opted for a lesser number of variables and considered three
 157 latent variables. The difference of CV error between using 6 LVs (0.0556) and 3 LVs
 158 (0.07281) was small in the training data, and the model behaved better on testing data.

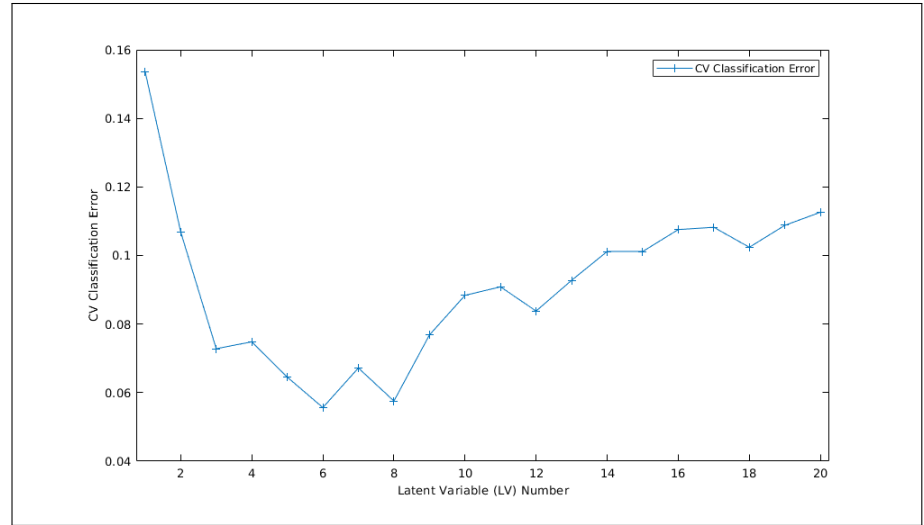


Fig 3. Classification Error: The plot shows the error made by the model according to the involved number of LV.

159 For further assessment of the model robustness, we employed a permutation test to
 160 determine the signification of the relation between grains features set and predicted
 161 classes. We conducted a permutation test with 100 cycles whereby the mycetoma classes
 162 of the images are randomly shuffled while maintaining the features set unchanged in each
 163 cycle and building the model with the same parameters as the original one [25, 40, 41].
 164 Furthermore, it is important to evaluate the effect of samples associated with grains
 165 features on the model prediction ability. The residual Q and the Hotelling T^2 values
 166 are calculated for this purpose. T^2 value measures the variation in each sample within
 167 the PLS-DA model. A larger value indicates a greater influence of the sample on the
 168 model. Q demonstrates the goodness of samples fit the model. This allows the detection
 169 of outliers, that can be removed from the learning sample set [32, 33].
 170 Variable Importance in Projection (VIP) score was analysed to understand the importance
 171 of each feature in the PLS-DA model and how strongly they contribute to the classification.
 172 Usually, it is considered that VIP scores have a threshold value of 1, meaning that the
 173 features which score greater than 1 in the model are significant for the prediction
 174 ability [?].
 175 Finally, the performance of the model was evaluated through sensitivity, specificity, area
 176 under the curve (AUC), accuracy, and Matthew's Correlation Coefficient (MCC).

Results

The data shown in Fig.4 depict the scores plot and projection of samples with regards to the three LVs used for predication model. LVs represent the combinations of the grains features that best discriminate between eumycetoma and actinomycetoma. The plot demonstrates a separation for the discriminated classes in the estimated feature space, where two loose clusters can be identified for the mycetoma classes.

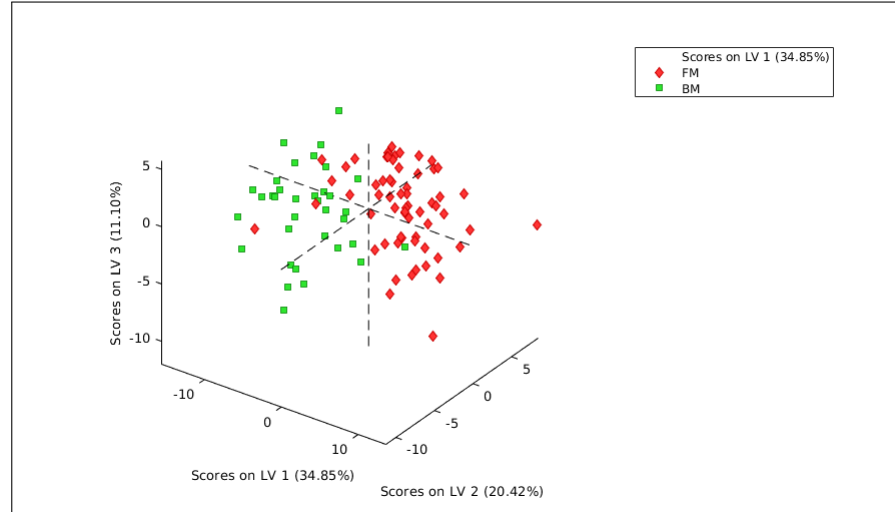


Fig 4. Samples projections: 3D projection of mycetoma images for the three latent variables (LV1, LV2, and LV3)

To detect the outliers, the plot of hotelling T^2 versus Q at 95% confidence levels was used, Fig.5. While a few samples were found slightly higher than the confidence limits for the mycetoma classes, a single FM sample was indicated as a clear outlier with high value for both Q and T^2 . A visual analysis led to consider that the sample exhibited similar properties compared to the other images of the data set, and quantitative analysis showed that the features of the sample were in line with the other samples' features which led us to maintain it within the data set.

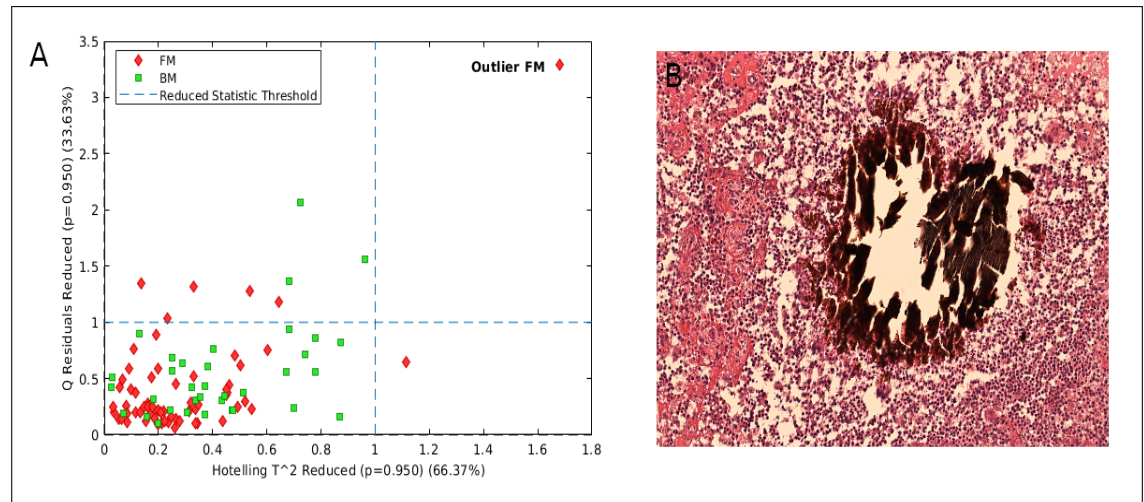


Fig 5. Evaluation of Outliers. (A):Residuals vs Hotelling, (B):Outlier FM sample.

Analysing VIPs score of grain features with a threshold value equals to 1 qualified more than 40% of features to be significant for differential diagnosis, Fig.6. Therefore, we consider 1.2 thresholds to highlight the most important features.

In our experiments, the role of shape features was trivial, while texture features were dominant. We observed that most of the significant features were related to variance, entropy or complexity of grains. The peak of VIPs measures the joint distribution of low grey value and spatial connectivity between a pixel and its neighbours within the small surround. This feature is an indicator for the homogeneity of textures and the tendency for closer blocks to have similar spatial variation. The data shown in Fig.7 illustrates the great differences in shapes that can be found for similar organisms. This shape features, which obtained a very low VIP, were not important to achieve a good classification. On the contrary, Fig.8 represents samples of grains that score low and high VIPs for some of the most important features to classify the data. Grey Level Variance measures the variance in grey-level intensity of the consecutive and adjacent blocks within the grains. In contrast, difference variance indicates the heterogeneity of texture inside the grains. In the three columns, the grains in the top and bottom rows are from different causative organisms, and the corresponding features values, which were very different, helped the discriminating process.

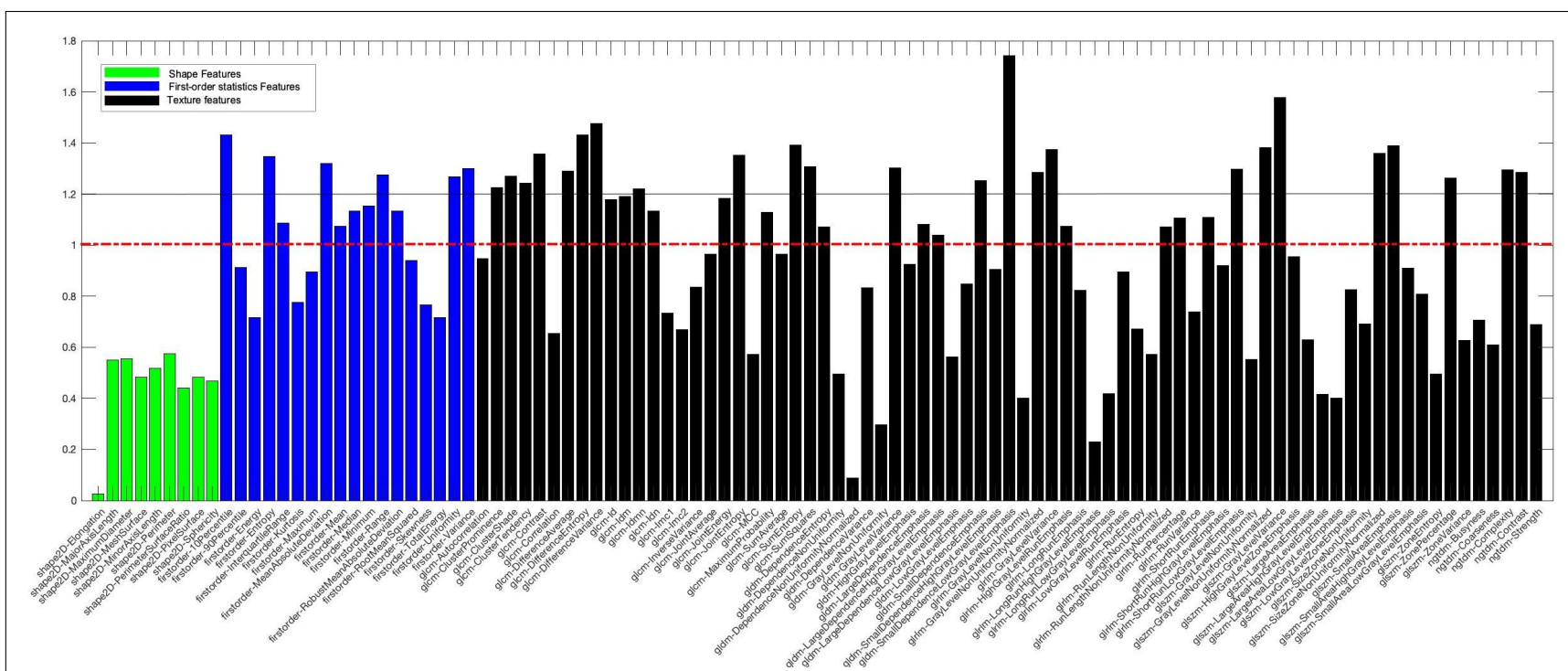


Fig 6. VIP of grains features: Variable importance scores of the PLS-DA model for discriminating mycetoma.

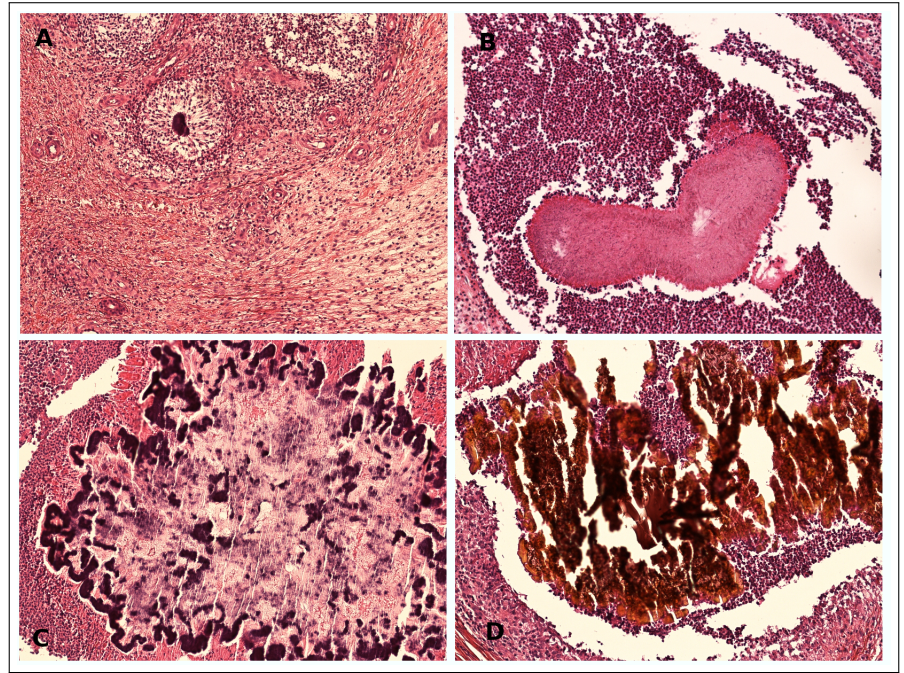


Fig 7. Diameter of the grains. The first and the second rows represents the grains that have the shortest and longest diameter, respectively. (A, C): actinomycetoma, and (B, D): eumycetoma.

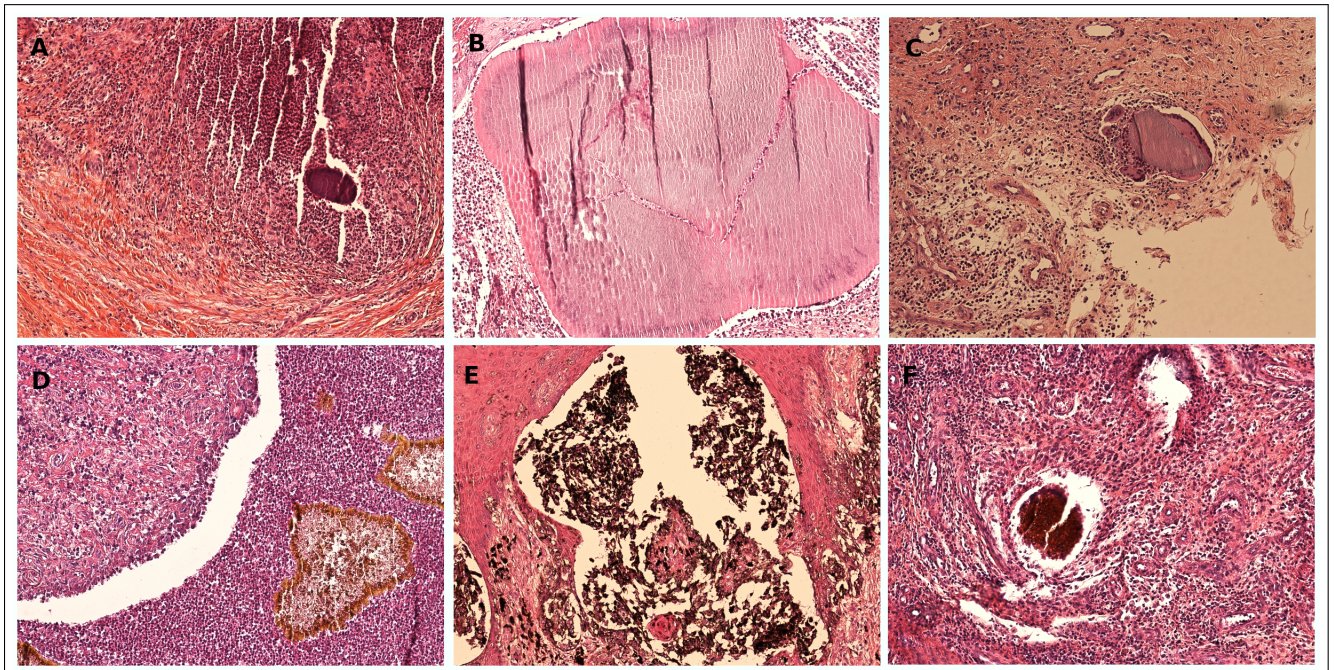


Fig 8. Grains which score greatest and smallest score for selected VIPs. First and second rows show lowest and highest score, respectively. (A, D): Difference variance, (B, E): Small dependence low-gray-level emphasis (peak feature), and (C, F): Grey-level variance.

208 Predicted values for the two classes are given in Fig.9. A threshold of 0.1 is shown

as the horizontal red dashed line and apply for samples classification. This threshold was calculated using Bayes' theorem. A sample is labelled as BM if it scores value greater than 0.1 and FM otherwise. With such a threshold, only one sample of BM was misclassified, while a few FM were misclassified. It indicated that while most samples were correctly classified, a few of eumycetoma samples presented similar radiomics features compared to actinomycetoma.

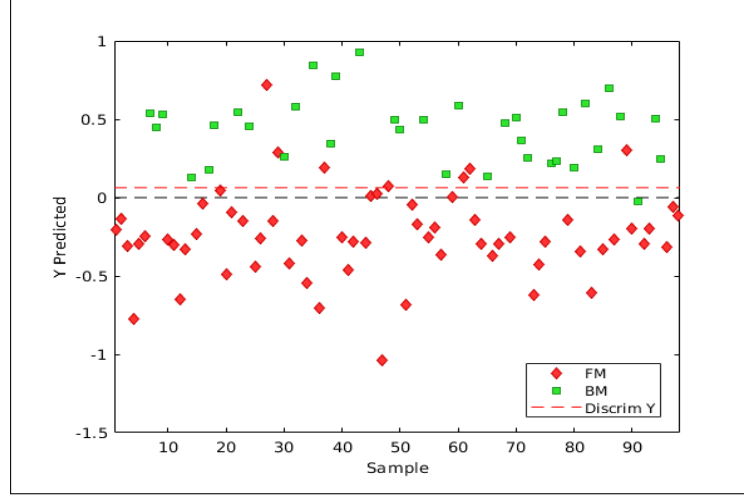


Fig 9. Predictions of the Samples: Estimated class values for samples for discrimination between eumycetoma and actinomycetoma

Evaluation of the model

The evaluation results given in Table 3 illustrates that the base on our experiments, the model is reliable and robust, with similar results obtained between training and validation. The value of sensitivity and specificity are good estimates of the model, and they are compatible with prediction in Fig.9.

Table 3. Estimated metrics for the model.

	Training set	Validation set
Sensitivity	0.948	0.971
Specificity	0.885	0.889
Accuracy	92%	91.8%
AUC	-	0.9683
Matthew's correlation	0.844	0.836

The estimated threshold for classes discrimination is indicated by a vertical red line in Fig 10B, with the intersection points corresponding to the sensitivity and specificity of the model. ROC curve in Fig 10A displays the sensitivity and the specificity of the model for a similar distributed set of samples with different classification thresholds. A very high AUC of 0.9683 was achieved on our data. The permutation test was used to analyse the reliability of the proposed model and the existence of a real association between grains features and their classes.

The permuted models and original model are compared for the goodness of fit (R^2) and prediction (Q^2) and tested for over-fitting. The results of the permutation test are given in Fig.11. The horizontal axis represents the correlation coefficient between the actual

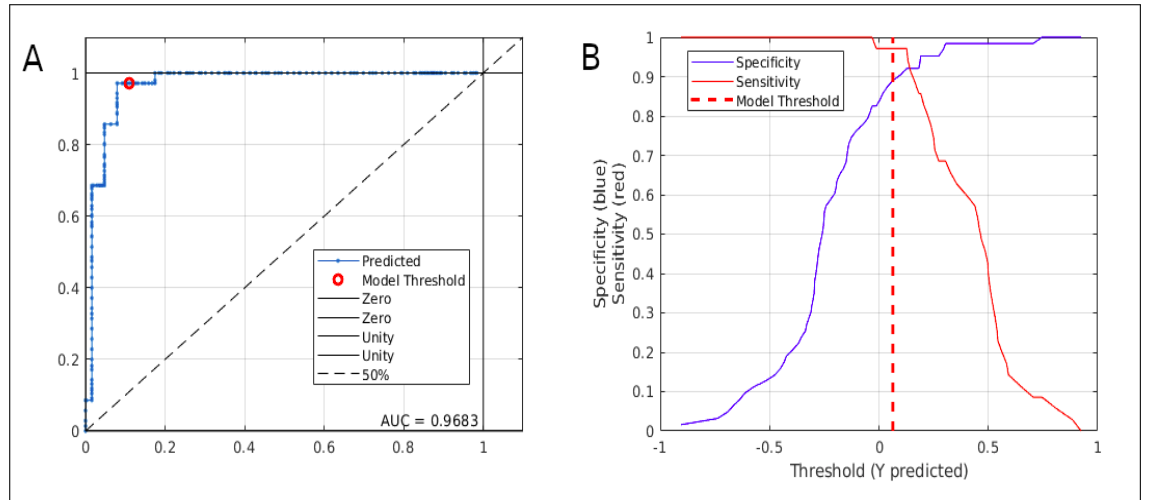


Fig 10. Receiver Operating Characteristic curves (ROC) and threshold plots.
(A):Predicted ROC,(B):Modelled Threshold

230 classes and the permuted classes, while the vertical axis translates the standardised sum
231 of squares that measures the deviation of each model away from the mean value and
232 is expressed in standard deviations. Small values of the sum of squares mean that the
233 model differs considerably from the mean of error, and since the permutation shuffles the
234 samples classes, there should be a great distance from the mean of error. Furthermore,
235 permuting classes might leave a small correlation between original and permuted classes;
236 hence, the right side of the plot represent the original model. The actual model scores
237 significantly higher value compared to permuted models indicating several standard
238 deviations away from the mean of error [40]. The permutation plot complies with the
239 classical validation criteria [41], and it suggests that the proposed model is not over-fitted
240 and that it is unlikely for the relation between classes and features to be random.

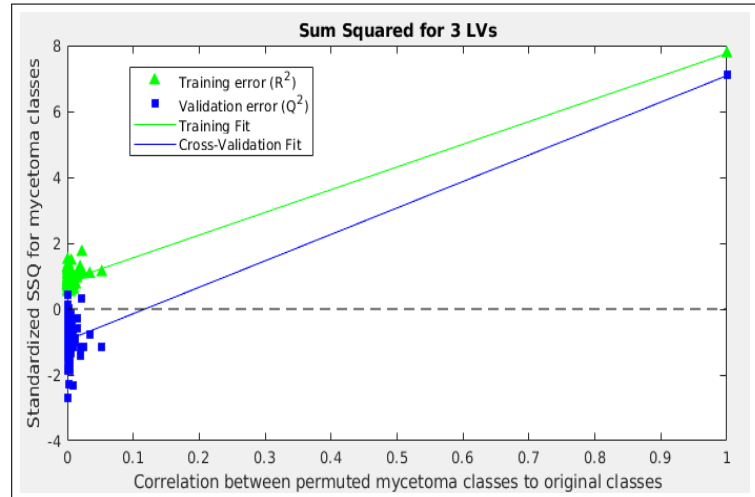


Fig 11. Permutation Test

241 With the objective of assessing the model on external data set, the secondary data set
242 was used to evaluate the robustness of the model with images from new patients that
243 have distinct acquisition parameters and slides preparation techniques. Fig.12 shows

the classification of test samples. The proposed model achieves 0.892 accuracy, which is quite similar to validation results, confirming the robustness of our approach.

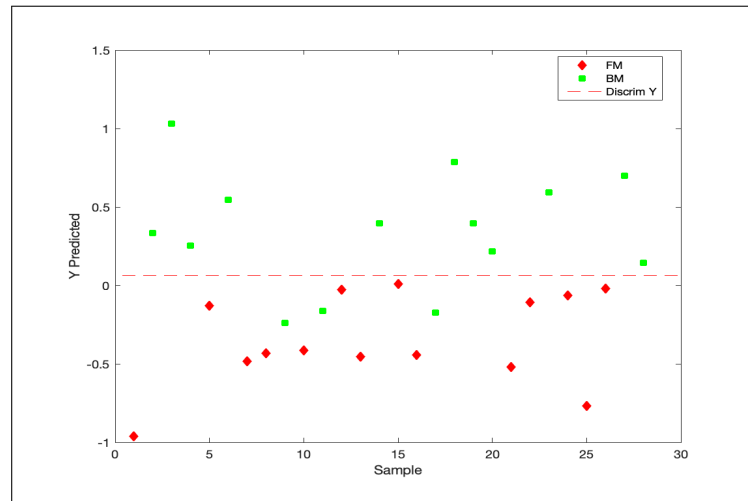


Fig 12. Evaluation of the model. Prediction of images with different acquisition and preparation methods from the training data set.

Discussion

Proper mycetoma management and treatment require accurate identification of mycetoma causative organisms [8,20]. Currently, histopathological techniques seem to be the optimal methods for identification of mycetoma organisms in terms of performance, cost and time [6,9,18]. However, the ability of pathologists to discriminant between eumycetoma and actinomycetoma is restricted by their knowledge about the microscopic appearance of the organisms. Furthermore, some organisms look very similar to each other [6,19,20]. Hence, the judgment is vulnerable to false results. Therefore, in this work, we pioneered a computational method to differentiate between eumycetoma and actinomycetoma from grains quantitative features in histopathological microscopic images.

In this study, we use the histopathological microscopic images from surgical biopsies from 55 randomly chosen patients from the MRC. The imaging protocol was set to uniform parameters with a digital optical microscope. As images acquisition and samples preparation method are feasible in many clinical centres that have histopathology departments, we used this technique in the present study.

The performance of our method was evaluated in terms of sensitivity, specificity, accuracy, AUC and MCC. The proposed model achieved an accuracy of 91.8%, sensitivity of 0.971, and specificity of 0.889. The obtained results were in line with the reported results from trained expert pathologist from the MRC [18]. The 92% accuracy of expert pathologists indicates that on our data set and the proposed model can be as efficient as an expert. MCC metric measures the statistical accuracy taking into account the different sizes of classes. It scored 0.836, which indicates a skilful model. To evaluate the robustness of the model, AUC is computed by aggregating the performance of the model across different thresholds. AUC of 0.968 indicated that the proposed model has excellent discrimination ability. AUC and MCC values were homogeneous, which again showed the ability of the model to separate the classes. Furthermore, the permutation test assesses the risk of not predicating mycetoma class for a new sample. The proposed model exhibited to be reliable and robust. Therefore, the proposed model found to be

strongly significant since the achieved results are comparable to expert classification analysis performed at the MRC. The proposed method is objective and reproducible and can reduce the need for highly specialised pathologists for diagnosis in endemic areas. In the literature, several studies proposed size, mostly diameter, and the border of the grains as characteristic features of the mycetoma grains [4, 6, 9, 20]. Eumycetoma has the largest grains, while actinomycetoma grains vary from small to medium. The grains of both mycetoma types can be rounded or oval, while actinomycetoma has irregular borders, vermiform shape, or multi-lobed shape. Based on our results, there are some overlapping features between the causative organisms, and it is inconvenient to describe the grains in terms of size or shape for identification of the causative organism. The VIPs results concluded that textural features are the most dominant and powerful features for differential diagnosis. The top textural features were difference variance, small dependence low-grey-level emphasis, grey-level variance, and complexity. These features illustrate that eumycetoma tends to have a non-uniform and complex pattern within grains, and it is usually composed of connected blocks that are less homogeneous. On the other hand, actinomycetoma grains are compact with a simple or regular pattern. These results are of particular interest due to the reported fact, that eumycetoma grains known to be harder with a coarse texture and tend to be fractured [6, 9, 10, 20]. Hence, we believe that our proposed model provides quantitative features which are quite connected to the qualitative features in use by expert pathologists. In the light of the aforementioned results, we can conclude that the discriminating features of mycetoma types depend on the variation in texture for the adjacent regions within grains. In other words, the homogeneity of textures and the tendency for closer blocks to have similar spatial variation.

The results of this work were drawn from a well-defined set of images using uniform image acquisition parameters. However, it was interesting to test the model on different images from new patients with distinct acquisition parameters and slides preparation techniques and to evaluate the robustness of the model. We observed a good classification of this data set. It is noticeable that the misclassified samples belong to actinomycetoma (*Nocardia*), which is uncommonly seen in Africa [3] and which was not included in the training/validation data set. Certainly, the variability of the data set affects the performance of the model, so this limitation should be investigated carefully in future studies considering the distribution of species.

Considering the various subtypes of actinomycetoma and the mycetoma retrospective study [18], it can be inferred that mycetoma can be classified into more than two types. Thus, the proposed approach might be extended to consider more types according to the individual causative organisms. Clearly, bigger sample size is needed to train the model. Despite this, we tested this assumption, and the preliminary result with an accuracy of 80.6% is encouraging. The literature reports that smaller grains for two of the tested classes (*Actinomadura madurae* and *Actinomadura pelletierii*) are similar [6, 18, 20]. This strengthens the fact that the classification tasks for these types are challenging and the accuracy we obtained was not odd. Hence, we believe that the limited number of images for these classes strongly affects the performance of the extended model. Accordingly, we suggest increasing the number of images of the different classes for the extended model in upcoming works.

Tackling to develop an automated diagnostic model for mycetoma histopathological diagnosis, as another natural perspective, the proposed classification model could be integrated with a segmentation technique. The promising discrimination results fulfilled in this study are encouraging for expanding the model in the future.

In conclusion, the different effective diagnostic tools used for mycetoma diagnosis require expert personnel and good set up. In this work, we introduced a novel, simple, and low-cost computational method that could be integrated into routine histopathological

326 diagnosis procedures. The proposed method uses radiomics in conjunction with PLS-
327 DA to effectively discriminate between actinomycetoma and eumycetoma with 91.8%
328 accuracy and robustness to samples preparation techniques. This could reduce the need
329 for expert pathologists in non-specialised clinical centres to perform the histological
330 analysis.

331 Acknowledgments

332 We are grateful to Dr.Cécila Rousselot-Denis for assistance with the preparation of the
333 histology slides, Mr.Emmanuel Edwar Siddig and Miss Omnia Babekir Abdallah who
334 were in charge of samples collection.

335 References

- 336 1. WHO: Mycetoma; Accessed January 2020. [https://www.who.int/buruli/
337 mycetoma/en/](https://www.who.int/buruli/mycetoma/en/).
- 338 2. Fahal A. Mycetoma: a thorn in the flesh. Transactions of the Royal Society of
339 Tropical Medicine and Hygiene. 2004;98(1):3–11.
- 340 3. van de Sande WW. Global burden of human mycetoma: a systematic review and
341 meta-analysis. PLoS neglected tropical diseases. 2013;7(11):e2550.
- 342 4. Relhan V, Mahajan K, Agarwal P, Garg VK. Mycetoma: an update. Indian
343 journal of dermatology. 2017;62(4):332.
- 344 5. Fahal A, Mahgoub ES, Hassan AME, Abdel-Rahman ME. Mycetoma in the
345 Sudan: an update from the mycetoma research centre, University of Khartoum,
346 Sudan. PLoS neglected tropical diseases. 2015;9(3):e0003679.
- 347 6. Fahal A. Mycetoma—clinicopathological monograph, Khartoum University Press.
348 Khartoum, Sudan. 2006.
- 349 7. A F. Bailey and Love's Short Practice of Surgery;.
- 350 8. van de Sande WW, Fahal AH, Goodfellow M, Welsh O, Zijlstra E, et al. The
351 mycetoma knowledge gap: identification of research priorities. PLoS neglected
352 tropical diseases. 2014;8(3):e2667.
- 353 9. Ahmed AA, van de Sande W, Fahal AH. Mycetoma laboratory diagnosis. PLoS
354 neglected tropical diseases. 2017;11(8):e0005638.
- 355 10. Emmanuel P, Dumre SP, John S, Karbwang J, Hirayama K. Mycetoma: a
356 clinical dilemma in resource limited settings. Annals of clinical microbiology and
357 antimicrobials. 2018;17(1):35.
- 358 11. Efared B, Tahiri L, Boubacar MS, Atsam-Ebang G, Hammas N, Chbani L, et al.
359 Mycetoma in a non-endemic area: a diagnostic challenge. BMC clinical pathology.
360 2017;17(1):1.
- 361 12. van de Sande WW, Fahal AH, Goodfellow M, Welsh O, Zijlstra EE, et al. Merits
362 and pitfalls of currently used diagnostic tools in mycetoma. PLoS neglected
363 tropical diseases. 2014;8(7):e2918.
- 364 13. Abd El Bagi ME. New radiographic classification of bone involvement in pedal
365 mycetoma. American Journal of Roentgenology. 2003;180(3):665–668.

- 366 14. El-Bagi MEA, Fahal AH. Mycetoma revisited. *Saudi Med J*. 2009;30(4):529–533.
- 367 15. Fahal A, Sheik H, Homeida M, Arabi Y, Mahgoub E. Ultrasonographic imaging
368 of mycetoma. *British journal of surgery*. 1997;84(8):1120–1122.
- 369 16. El Shamy M, Fahal A, Shakir M, Homeida M. New MRI grading system for the
370 diagnosis and management of mycetoma. *Transactions of the Royal Society of*
371 *Tropical Medicine and Hygiene*. 2012;106(12):738–742.
- 372 17. Alam K, Maheshwari V, Bhargava S, Jain A, Fatima U, ul Haq E. Histological
373 diagnosis of madura foot (mycetoma): a must for definitive treatment. *Journal of*
374 *global infectious diseases*. 2009;1(1):64.
- 375 18. Siddig EE, Mhmoud NA, Bakhet SM, Abdallah OB, Mekki SO, El Dawi NI,
376 et al. The Accuracy of Histopathological and Cytopathological Techniques in
377 the Identification of the Mycetoma Causative Agents. *PLoS neglected tropical*
378 *diseases*. 2019;13(8):e0007056.
- 379 19. van de Sande W, Fahal A, Ahmed SA, Serrano JA, Bonifaz A, Zijlstra E, et al.
380 Closing the mycetoma knowledge gap. *Medical mycology*. 2018;56(suppl_1):S153–
381 S164.
- 382 20. Siddig E, Fahal A. Histopathological Approach in Diagnosis of Mycetoma Causative
383 Agents: A Mini Review. *J Cytol Histol*. 2017;8:466.
- 384 21. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics:
385 the process and the challenges. *Magnetic resonance imaging*. 2012;30(9):1234–1248.
- 386 22. Avanzo M, Stancanella J, El Naqa I. Beyond imaging: The promise of radiomics.
387 *Physica Medica*. 2017;38:122–139.
- 388 23. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton
389 P, et al. Radiomics: extracting more information from medical images using
390 advanced feature analysis. *European journal of cancer*. 2012;48(4):441–446.
- 391 24. Limkin E, Sun R, Dercle L, Zacharaki E, Robert C, Reuzé S, et al. Promises and
392 challenges for the implementation of computational medical imaging (radiomics)
393 in oncology. *Annals of Oncology*. 2017;28(6):1191–1206.
- 394 25. Lee LC, Liong CY, Jemain AA. Partial least squares-discriminant analysis (PLS-
395 DA) for classification of high-dimensional (HD) data: a review of contemporary
396 practice strategies and knowledge gaps. *Analyst*. 2018;143(15):3526–3539.
- 397 26. Brereton RG, Lloyd GR. Partial least squares discriminant analysis: taking the
398 magic away. *Journal of Chemometrics*. 2014;28(4):213–225.
- 399 27. Kvalheim OM. Interpretation of partial least squares regression models by means
400 of target projection and selectivity ratio plots. *Journal of Chemometrics*. 2010;24(7-
401 8):496–504.
- 402 28. McCann MT, Ozolek JA, Castro CA, Parvin B, Kovacevic J. Automated histology
403 analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*.
404 2014;32(1):78–87.
- 405 29. Komura D, Ishikawa S. Machine learning methods for histopathological image
406 analysis. *Computational and structural biotechnology journal*. 2018;16:34–42.

- 407 30. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al.
408 Computational radiomics system to decode the radiographic phenotype. *Cancer*
409 *research*. 2017;77(21):e104–e107.
- 410 31. dos Santos VHJ, Ramos AS, Pires JP, Engelmann PdM, Lourega RV, Ketzer
411 JM, et al. Discriminant analysis of biodiesel fuel blends based on combined data
412 from Fourier Transform Infrared Spectroscopy and stable carbon isotope analysis.
413 *Chemometrics and Intelligent Laboratory Systems*. 2017;161:70–78.
- 414 32. Ballabio D, Consonni V. Classification tools in chemistry. Part 1: linear models.
415 *PLS-DA. Analytical Methods*. 2013;5(16):3790–3798.
- 416 33. Ramos PM, Ruisánchez I. Data fusion and dual-domain classification analysis of
417 pigments studied in works of art. *Analytica chimica acta*. 2006;558(1-2):274–282.
- 418 34. Pérez NF, Ferré J, Boqué R. Calculation of the reliability of classification in dis-
419 criminant partial least-squares binary classification. *Chemometrics and Intelligent*
420 *Laboratory Systems*. 2009;95(2):122–128.
- 421 35. Wise BM, Gallagher N, Bro R, Shaver J, Windig W, Koch RS. *PLS Toolbox 4.0*;
422 2007.
- 423 36. Esbensen KH, Geladi P. Principles of proper validation: use and abuse of re-
424 sampling for validation. *Journal of Chemometrics*. 2010;24(3-4):168–187.
- 425 37. Villa JE, Quiñones NR, Fantinatti-Garboggini F, Poppi RJ. Fast discrimination of
426 bacteria using a filter paper-based SERS platform and PLS-DA with uncertainty
427 estimation. *Analytical and bioanalytical chemistry*. 2019;411(3):705–713.
- 428 38. Pereira LS, Lisboa FL, Neto JC, Valladão FN, Sena MM. Screening method for
429 rapid classification of psychoactive substances in illicit tablets using mid infrared
430 spectroscopy and PLS-DA. *Forensic science international*. 2018;288:227–235.
- 431 39. Botelho BG, Reis N, Oliveira LS, Sena MM. Development and analytical validation
432 of a screening method for simultaneous detection of five adulterants in raw milk
433 using mid-infrared spectroscopy and PLS-DA. *Food chemistry*. 2015;181:31–37.
- 434 40. Ahn JK, Kim S, Kim J, Hwang J, Kim KH, Cha HS. A comparative metabolomic
435 evaluation of Behcet’s disease with arthritis and seronegative arthritis using
436 synovial fluid. *PloS one*. 2015;10(8):e0135856.
- 437 41. Musharraf SG, Siddiqui AJ, Shamsi T, Choudhary MI, Rahman Au. Serum
438 metabonomics of acute leukemia using nuclear magnetic resonance spectroscopy.
439 *Scientific reports*. 2016;6:30693.