

Predictive models for *Talaromyces marneffe* infection in HIV-infected patients using routinely collected data

Jianguang Hu

Xu Li

Tao Chen

Bingyu Liang

Wudi Wei

Dali Meng

Yu Lin

Minji Li

Peng Zhang

Wenjuan Jing

Gui Tang

Mingjia Lin

Zhongsheng Jiang (✉ jiangzs1111@sina.com)

Research

Keywords: *Talaromyces marneffe*, HIV infection, clinical predictive model, Random forest, logistic regression

Posted Date: January 2nd, 2020

DOI: <https://doi.org/10.21203/rs.2.19894/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objectives

Late diagnosis of *Talaromyces marneffe* (*T. marneffe*) in patients with HIV/AIDS infection is strongly associated with greater mortality. To date, effective predictive model for *T. marneffe* infection in clinical practice have not been established. We aimed to identify a non-culture-based method for rapid detection of *T. marneffe* infection in HIV/AIDS patients.

Methods

The prediction models were initially constructed using patients in a retrospective cohort study. We obtained demographics, clinical and laboratory data for each individual. Univariate comparisons, logistic regression, Random Forest (RF) analysis and receiver-operating characteristic curves (ROC) were used to identify and evaluate the predictive factors of *T. marneffe* infection status.

Results

HIV-infected patients with a baseline characterized by weight loss, typical skin lesions, peripheral or abdominal lymphadenopathy (POAL), hepatomegaly, splenomegaly, decrease lymphocyte count, abnormal aspartate aminotransferase (AST) level, higher AST to alanine aminotransferase (ALT) ratio index (AARI) level (>1) and lower (<50 cells/mL) CD4+ T-cell counts had an increased risk of *T. marneffe* infection. Skin lesions, POAL, AARI, AST level and CD4+ T-cell count resulted in good classifiers of *T. marneffe* infection by RF analysis. RF model had a relative high power [area under the ROC curve (AUC): 0.859] to predict *T. marneffe* infection in the present study. A new indicator combine AST level and AARI could increase the classification power of the model (AUC: 0.877).

Conclusion

Our data suggest that accurate assessments for *T. marneffe* infection can be obtained using routinely collected data of patients with HIV. The prediction model could used to identify HIV patients who currently have early stage of *T. marneffe* infection, which would be benefit to both patients and clinicians.

Background

Talaromyces marneffe (*T. marneffe*), previously named *Penicillium marneffe* [1, 2], is a pathogenic, thermal dimorphic fungus. Systemic mycosis and disseminated infection caused by *T. marneffe* is often found in patients with secondary immunodeficiency syndromes, especially in patients with AIDS [3], mainly popular in southeast Asia and southern China. The prevalence of systemic *T. marneffe* infection has grown up rapidly currently, in consistent with the increasing incidence of HIV infections [3]. The HIV epidemic has transformed *T. marneffe* from a rare infection to a leading AIDS-defining diagnosis in this region [4-7]. The proportion of HIV patients infected with *T. marneffe* is ~ 10% in China currently [3, 6], and

approximately 99% of *T. marneffeii* infections were reported from the southern regions, of which 43% were from Guangxi [6]. Thus, infection with *T. marneffeii* is common among HIV/ AIDS patients in Guangxi, ranking fourth among all AIDS-related complications with a 16.1% prevalence [8].

Previous studies have shown that the mortality rate of *T. marneffeii*-infected HIV/AIDS patients was significantly greater than that of HIV/AIDS patients without *T. marneffeii* infection [8, 9]. Late diagnosis is the main reason for the high fatality in HIV/AIDS patients co-infected with *T. marneffeii*, many cases of *T. marneffeii* infection persist undiagnosed until later stages [4, 5]. *T. marneffeii* infection has a considerably high mortality (50.6%-91%) if the disease was not diagnosed and treated in time [6, 10]. Antifungal treatment during the early stage of *T. marneffeii* infection can effectively control the disease [9, 11] and reduce death from *T. marneffeii* infection [6, 12].

Thus, early detection is critical for improving the prognosis of *T. marneffeii*. However, clinical experiences show that the diagnosis of *T. marneffeii* infection is difficult because its clinical manifestations may mimic other AIDS-related opportunistic infections [13, 14]. Microbiological culture is the gold standard method for diagnosis of *T. marneffeii*. However, this method requires prolonged 7–10 days of incubation time to isolate and identify the pathogen from clinical specimens, frequently resulting in the delay of appropriate antifungal therapy. In addition, the sensitivity of fungal culture from blood in HIV infected patients is about 76.7% [15]. Thus, development of non-culture-based method for rapid detection of *T. marneffeii* infection is urgently needed. Recently, a number of other diagnostic methods have been established including different polymerase chain reaction (PCR) methods [16–19], high-throughput sequencing of specimens [20], enzyme-linked immunosorbent assay (ELISA) [21] and Mp1p tests [22]. However, these costly approaches have not been applied in clinical practice with sensitivity ranging from 10–100% [23].

Notably, recent studies have indicated that accurate risk assessments for diseases progression and clinical outcomes could be obtained for patients using only data routinely collected in clinical practice [24–26]. There is barely effective predictive model for *T. marneffeii* infection in clinical practice right now. Therefore, the aim of this study was to analyse correlations among *T. marneffeii* infection and clinical parameters in large-scale HIV infected patients and to predict *T. marneffeii* infection status by using data that are routinely collected in clinical practice. This study has the potential to facilitate rapid diagnosis of *T. marneffeii* infection, thus initiating early therapeutic management of *T. marneffeii* in HIV infected patients possible.

Methods

Study Setting and Population

The prediction models were initially constructed using patients in a retrospective cohort study conducted at the People's Hospital of Liuzhou, a large-scale comprehensive tertiary hospital in Liuzhou and a large

treatment center for HIV/AIDS in the city. Individuals included in this study were more than 18 years of age, presence of HIV-1 infection, and had culture-proven *T. marneffei* infection from January 2015 to December 2018. Exclusion criteria were tumors; combined with tuberculosis; hepatitis A, B, C, D, E virus coinfection; liver damage caused by drugs, alcohol, autoimmune hepatitis active opportunistic infections; with evidence of cirrhosis or suspected liver cancer; pregnancy or lactation women; the presence of abnormal renal function, severe heart, lung, brain diseases. Figure 1 provides a flow chart with patient selection and enrolment. In a final step, we excluded 81 samples due to data collection problems, which resulted in a final cohort of 602 patients. The end point of the cohort is whether *T. marneffei* infection occurs. Therefore, patients were categorized into two groups: *T. marneffei* infection group (n = 343) or non-*T. marneffei* infection group (n = 259). HIV infection was determined by positive HIV ELISA and confirmatory western blot assays. *T. marneffei* infection was diagnosed by isolating *T. marneffei* from blood, skin scrapings, bone marrow, lymph node, and/or other body fluid samples according to standard culture techniques [3].

Data collection

Only data from the first *T. marneffei* admission were extracted from the medical records. Data that are routinely collected in clinical practice included the demographics, clinical characteristics (e.g. fever, cough and skin lesion), and associated laboratory examinations of the patients. We normalized the values of baseline parameters, by log₁₀ transformation of hemoglobin, white blood cell count (WBC), platelets, lymphocyte count, creatinine and urea values. Results of baseline labs abstracted also included CD4 T-cell count, pathological biopsy, aspartate aminotransferase (AST) and alanine aminotransferase (ALT). Using these values, AST to ALT ratio index (AARI) was calculated. Then, data from the medical records were independently doubleentered into a Microsoft Access 2013 database specially defined for this study to ensure data integrity.

Machine learning methods, predictive models construction and validation

Random forest (RF) analysis, a decision tree-based ensemble statistical method, was applied as multivariate statistics to identify the variables that best partitioned the overall study population according to *T. marneffei* infected status. RF is an ensemble learning method provides an unbiased selection of variables that make the largest contributions to the classification. The mean decrease in accuracy (MDA) variable was used to evaluate the RF interpretations, which estimates how much excluding (or permuting) each variable reduces the accuracy of the model during the out of bag error calculation phase.

Finally, the variables considered important discriminators of *T. marneffei* infection were selected, obtained from the multivariate comparisons ($P < 0.05$) or the RF analysis (with larger MDAs in the classification model). RF and logistic regression were used for analysis and further diagnosis models.

Receiver-operating characteristic (ROC) curves were generated to quantify how accurately the variables were able to discriminate between the groups. Models were evaluated by constructing the confusion matrix for test data. Model performance was compared using the area under the receiver operating characteristic curve (AUC) analysis and 95% confidence intervals (CI). For RF analysis, we replicated 500 times in order to obtain the mean and 95% CI for this AUC and validated each of these models on all the data within this cohort. In addition, sensitivity and specificity were also measured for each model.

Statistical analysis

A descriptive analysis of patients' characteristics was carried out using frequency tables for categorical variables and continuous variables. Data were presented in numbers (%) as appropriate. Normally distributed continuous variables were described as mean \pm standard deviation, nonnormally distributed continuous variables described as median and interquartile range (IQR). To find differences between the two groups, we compared their differences in demographic, clinical and laboratory characteristics at baseline using different methods. Univariate comparisons were performed using the nonparametric Mann–Whitney test or student's t-test for continuous variables and the Pearson's Chi-square test or Fisher's exact test for categorical variables. Variables with a p-value < 0.05 from the univariable analysis were then tested in multivariable models using forward stepwise procedures. A two-sided test at a p-value of < 0.05 was considered statistically significant. The data were analyzed using Statistical Package for the Social Sciences (SPSS) version 25.0 (SPSS Inc. Chicago, USA), the program 'R' (<http://cran.r-project.org>) and GraphPad Prism version 6.0 (GraphPad software, San Diego, California, USA).

Results

Baseline characteristics of the patients

During January 2015 to December 2018, 343 inpatients out of a total number of 602 eligible HIV-related hospitalization in Liuzhou People's Hospital presented with culture-proven *T. marneffei*. Table 1 shows demographic characteristics of the patients in this study. In total, the medium age of the patients was 46.7 years old with 436 (72.4%) male patients and 166 (27.6%) female patients. The infection has a sex bias with a ratio of 2.6:1 comparing male to female. 77.9% (469/602) were of Han nationality, 70.9% (427/602) were married, and 51.8% (312/602) were farmers. Overall, the route of HIV transmission mainly sexually acquired (587/602, 97.5%). Significant differences ($p < 0.05$) in baseline demographic characteristics were identified between *T. marneffei*-infected and uninfected patients in age, ethnic, and marital status upon admission (Table 1).

Table 1
The demographic and clinical features of the patients

Characteristics	Total (n = 602)	With T. marseffei infection (n = 343)	Without T. marseffei infection (n = 259)	p
Demographics				
Age, years (SD)	46.7 (15.0)	43.9 (14.6)	50.3 (14.8)	< 0.001
Gender, N (%)				0.318
Male	436 (72.4)	243 (70.8)	193 (74.5)	
Female	166 (27.6)	100 (29.2)	66 (25.5)	
Ethnic, N (%)				0.021
Han	469 (77.9)	255 (74.3)	214 (82.6)	0.015
Zhuang	113 (18.8)	72 (21)	41 (15.8)	0.108
Others	20 (3.3)	16 (4.7)	4 (1.5)	0.034
Occupation, N (%)				0.116
Farmer	312 (51.8)	167 (48.7)	145 (56)	0.076
Unemployed	136 (22.6)	78 (22.7)	58 (22.4)	0.920
Others	154 (25.6)	98 (28.6)	56 (21.6)	0.053
Marital status, N (%)				< 0.001
Married	427 (70.9)	219 (63.8)	208 (80.3)	< 0.001
Single	108 (17.9)	82 (23.9)	26 (10.0)	< 0.001
Other	67 (11.1)	42 (12.2)	25 (9.7)	0.317
HIV transmission group, N (%)				0.955
Sexual transmission	587 (97.5)	333 (97.1)	254 (98.1)	0.443
Injecting drug use	8 (1.3)	5 (1.5)	3 (1.2)	1.000
Mother-to-child transmission	6 (1.0)	4 (1.2)	2 (0.8)	0.704

Characteristics	Total (n = 602)	With T. marneffei infection (n = 343)	Without T. marneffei infection (n = 259)	p
Other	1 (0.2)	1 (0.3)	0 (0.0)	1.000
Symptoms				
Fever, N (%)	473 (78.6)	300 (87.5)	173 (66.8)	< 0.001
Weight loss, N (%)	405 (67.3)	253 (73.8)	152 (58.7)	< 0.001
POAL, N (%)	386 (64.1)	275 (80.2)	111 (42.9)	< 0.001
Cough, N (%)	251 (41.7)	149 (43.4)	102 (39.4)	0.317
Hepatomegaly, N (%)	172 (28.6)	151 (44.0)	21 (8.1)	< 0.001
Splenomegaly, N (%)	165 (27.4)	136 (39.7)	29 (11.2)	< 0.001
Skin lesions, N (%)				
Typical skin lesions	145 (24.1)	140 (40.8)	5 (1.9)	< 0.001
Non-typical skin lesions	23 (3.8)	17 (5.0)	6 (2.3)	0.094
No Skin lesions	434 (72.1)	186 (54.2)	248 (95.8)	< 0.001

The common clinical manifestations among all patients were fever (473/602, 78.6%), followed by weight loss (405/602, 67.3%), peripheral or abdominal lymphadenopathy (POAL) (386/47, 64.1%), cough (251/602, 41.7%), hepatosplenomegaly (172/602, 28.6%), splenomegaly (165/602, 27.4%), and typical skin lesions (145/602, 24.1%). Table 1 shows the comparison of clinical datas between patients with and without T. marneffei infection in this study. Clearly, patients infected with T. marneffei were more likely to have fever, weight loss, POAL, splenomegaly, hepatomegaly and typical skin lesions at baseline.

Data from laboratory examinations are shown in Table 2. In the study population, laboratory findings differed between groups; patients infected with T. marneffei exhibited decreased hemoglobin, WBC, platelets, lymphocyte count, but AST, ALT and AARI were significantly higher in T. marneffei infected patients than those in non-T. marneffei infected patients.

Table 2
Laboratory test results of the patients

Laboratory findings	Total (n = 602)	With T. marneffeii infection (n = 343)	Without T. marneffeii infection (n = 259)	p
Hemoglobin, g/L (IQR)	91 (76–107)	91 (76–107)	100 (81–117)	< 0.001
WBC, cells/ μ L (IQR)	4.40 (2.97–6.70)	4.23 (2.90–6.04)	4.80 (3.20–4.81)	0.004
Platelets, cells/ μ L (IQR)	152.00 (77.75–231.50)	111.00 (54.00–197.00)	200.00 (132.00–270.00)	< 0.001
Lymphocyte count (IQR)	0.49 (0.29–0.83)	0.38 (0.24–0.61)	0.66 (0.38–1.10)	< 0.001
ALT, U/L (IQR)	29 (19–53)	35 (22–63)	25 (14–40)	< 0.001
AST, U/L (IQR)	56.5 (31–125.5)	89 (50–159)	34 (23–55)	< 0.001
AARI (IQR)	1.91 (1.25–2.95)	2.34 (1.63–3.80)	1.50 (1.04–2.10)	< 0.001
CD4 + T-cell count, cells/mL (IQR)	16 (9–35)	12 (7–25)	24 (12–50)	< 0.001
Creatinine, μ mol/L (IQR)	68.0 (55.0–85.0)	68.3 (55.0–85.0)	68.0 (55.0–85.9)	0.952
Urea, mmol/L (IQR)	4.70 (3.30–6.60)	4.77 (3.21–6.60)	4.37 (3.50–6.50)	0.390

Time to treatment initiation and outcome of T. marneffeii infected patients

T. marneffeii was successfully isolated from clinical specimens of the patients. For patients infected with T. marneffeii, the interval between onset of T. marneffeii symptoms and the time of admission to our hospital was 30 days (IQR:14–39), and the median interval from the time of admission to initiation of antifungal therapy was 3 days (IQR:1–6). The median duration of antifungal treatment in the hospital was 12 days (IQR:6–16). The outcome of the most majority of patients at discharge was clinical improvement (234/343, 68.22%), 9.33% patients (32/343) died after admission, and 19.24% (66/343) were discharged early due to request from the family members, or referred to other hospitals.

Factors associated with *T. marneffeii* infection

Baseline variables shown significantly higher risk of *T. marneffeii* infection were further included in the multivariate analysis. Results of the logistic regression are displayed in Table 3, which showed that *T. marneffeii* infection significantly increased with weight loss [odds ratio (OR) and 95% confidence interval (CI): 1.79 (1.10–2.91)]. Importantly, typical skin lesions had a stronger association with the probability of *T. marneffeii* infection [OR:19.27; 95% CI (7.17–51.78)]. Patients with POAL [OR: 2.86; 95% CI (1.78–4.61)], hepatomegaly [OR: 2.57; 95% CI (1.34–4.95)], or splenomegaly [OR:1.96; 95% CI (1.08–3.57)], had a higher risk of *T. marneffeii* infection. Moreover, patients with decreased lymphocyte count, abnormal AST level, higher AARI level (> 1) or lower (< 50 cells/mL) CD4 + T-cell count also had an increased risk of *T. marneffeii* infection, with an OR of 1.84 (95% CI: 1.03–3.27), 4.29 (95% CI: 2.65–6.94), 2.54 (95% CI:1.27–5.07) and 9.48 (95% CI:2.12–42.37) respectively. Therefore, the results suggest that these baseline variables favoured *T. marneffeii* infection.

Table 3
Risk factors associated with *T. marneffei* infection (logistic regression model)

Characteristics	OR (95% CI)	p
Weight loss		
Yes	1.79 (1.10–2.91)	0.019
No	Reference	
Skin lesions		
Typical skin lesions	19.27 (7.17–51.78)	< 0.001
Non-typical skin lesions	1.89 (0.62–5.70)	0.265
No Skin lesions	Reference	
POAL		
Yes	2.86 (1.78–4.61)	< 0.001
No	Reference	
Hepatomegaly		
Yes	2.57 (1.34–4.95)	0.005
No	Reference	
Splenomegaly		
Yes	1.96 (1.08–3.57)	0.027
No	Reference	
Lymphocyte count		
Decrease	1.84 (1.03–3.27)	0.039
Increase	0.23 (0.05–0.97)	0.045
Normal	Reference	
AST level		
Abnormal	4.29 (2.65–6.94)	< 0.001
Normal	Reference	
AARI		
> 1	2.54 (1.27–5.07)	0.008

Characteristics	OR (95% CI)	p
≤ 1	Reference	
CD4 + T-cell count		
< 50	9.48 (2.12–42.37)	0.003
50–99	4.91 (0.96–25.11)	0.056
≥ 100	Reference	

Predictive models by machine learning method

RF analysis was used as a multivariate method to evaluate the classificatory value of *T. marneffei* infection determinations in baseline characteristics (Fig. 2). The variable importance graph for the model predicting is shown in Fig. 2A. RF algorithm confirmed that skin lesions, AARI, POAL, AST, and CD4 + T-cell count were good classifiers in a ranked list of classification scheme (Fig. 2A). Moreover, it is important to highlight that these variable were also the most significant variable either in the univariate test or logistic regression, which suggested that they are powerful indicators of *T. marneffei* infection over time.

Assessing and comparing models performance

Finally, we evaluated the potential clinical usefulness of our highlighted candidates. Accordingly, logistic regression and RF algorithm were further used to construct predictive models with the involvement of parameters. Sensitivity and specificity of each method are shown in Table 4, and ROC curves are plotted in Fig. 2B.

Table 4
Specificity and sensitivity of classification and AUC of predictive modules based on two methods.

Methods	Variable	AUC(95% CI)	Sensitivity	Specificity
RF	Model 1	0.859 (0.807–0.912)	78.2%	81.6%
	Model 2	0.877 (0.809–0.904)	83.2%	77.6%
Logistic regression	Model A	0.575 (0.528–0.621)	74.1%	40.9%
	Model B	0.712 (0.672–0.753)	45.9%	95.7%
	Model C	0.687 (0.643–0.731)	80.3%	57.2%
	Model D	0.681 (0.639–0.724)	44.4%	91.8%
	Model E	0.642 (0.598–0.686)	39.7%	88.7%
	Model F	0.623 (0.577–0.669)	83.5%	40.5%
	Model G	0.721 (0.678–0.764)	82.4%	61.9%
	Model H	0.589 (0.542–0.636)	94.1%	23.7%
	Model I	0.583 (0.536–0.630)	90.6%	25.3%

We compared the performance of the logistic regression models and RF models for *T. marneffei* infection with AUC analysis and 95% CI (Table 4). For logistic regression, our results showed that the AUC of each variable was less than 0.8 (Fig. 3, Model B-J). Additionally, we carried out validations by conducting Module 1 and Module 2 of RF method. The model 1 displayed an AUC value of 0.859 of patients with 78.2% of sensitivity and 81.6% specificity (Fig. 2B) (Table 4). Moreover, given that AST level and AARI are significant alterations of *T. marneffei* infection, AST level and AARI combine to form a new indicator. Then, to refine the predictive model, we evaluate the ability of the new indicator to correctly classify *T. marneffei* infection in this study. As expected, the AUC value increased by 0.018, specificity increased by 5.0% and it did improve the percentage of classification (Fig. 2D) (Table 4).

Discussion

Early and accurate diagnosis of *T. marneffei* infection is crucial to initiate the proper management and treatment in clinical practice, especially for HIV infected patients.

Accurate risk assessments for *T. marneffei* infection among HIV patient represent the major strength of our study. For predictive models, two modules were conducted separately to find the most appropriate model in this study. We suggest RF as a machine learning method to aid in prediction and diagnosis for binary classification of *T. marneffei* infection status, which has a relative higher power of 0.877 compare to logistic regression models. Due to the reliance on classical forms of statistical analysis that restrict the

number of predictor variables, logistic regression models primarily limited to baseline data and may have only moderate accuracy in *T. marneffei* risk prediction. RF analysis was used to build prediction models because this method is able to incorporate many predictor variables without compromising the accuracy of the risk prediction [27]. The sensitivity of the our models is comparable to the microbiological culture (78.2% ~ 83.2% vs 76.7%) [28]. Notably, the novel method with only clinical parameters was quite valuable, indicating that our model for *T. marneffei* infection prediction and early diagnosis is possible.

Our findings reveal that patients with typical skin lesions had a stronger association with the probability of *T. marneffei* infection in the univariate and multivariate analysis. Moreover, RF analysis also revealed skin lesions as most important indicator in the prediction models. Thus, the presence of skin lesions may associated with a more rapid to antifungal initiation and shorter mean duration of antifungal treatment [4]. Interestingly, previous studies have found that typical skin lesions were present in 71%-83% of patients infected with *T. marneffei* [4, 17, 28], but this rate was only 40.8% in our study. This unexpected finding may be explained by different reasons. In our study, the median interval from the time of admission to initiation of antifungal therapy was only 3 days. Thus, early diagnosis and timely antifungal therapy for *T. marneffei* may favoured immunological recovery and lower incidence rate of skin lesions. On the other hand, patients with typical skin lesions may also diagnosed and treated in primary hospitals. In the absence of skin lesions, the differential diagnoses for *T. marneffei* among AIDS-associated febrile illness still challenging, epidemiological data and specimen examination are essential to make a diagnosis for patients without skin lesions [29].

In addition, we observed that patients with increased AST and AARI levels also had higher risk of *T. marneffei* infection. RF analysis revealed increased AST and AARI as important variables in the classification scheme (Fig. 2), which suggested that they are powerful indicators of *T. marneffei* infection. As important serum biomarkers for accessing liver injury, ALT and AST are separately sensitive to the damage of hepatocyte membrane and hepatocyte mitochondria [30, 31]. In our study, the proportion of increased AST and ALT accounted for 75% and 63% of the *T. marneffei*-infected population, respectively, which were similar to the related studies [6]. Clinically, *T. marneffei* is characterized by fungal invasion of multiple body organ systems, as it can proliferates in macrophages and disseminates via the reticuloendothelial system [3]. Thus, liver is vulnerable to *T. marneffei* attack. Previous studies have also shown that liver biopsy of HIV/AIDS patients in making a diagnosis of hepatic *T. marneffei* and that there are different patterns of pathologic damage [32]. Thus, the presence of increased AST and AARI levels may be favoring accuracy and early diagnosis of *T. marneffei*.

Similar to the results from others studies, *T. marneffei*-infected HIV/AIDS patients are characteristics with low CD4 + T cell count, POAL, hepatomegaly, splenomegaly, and weight loss in our study [4, 6, 15]. Among patients with lower (< 50 cells/mL) CD4 + T-cell counts, the body's immunity could not maintained at a certain level and clinical *T. marneffei* symptoms (e.g. POAL, hepatosplenomegaly and weight loss) may present earlier in the disease course, which may involve more rapid diagnosis and appropriate antifungal therapy. Therefore, in the absence of skin lesions, clinicians should also be alert to mixed infection with *T.*

marneffeii especially in endemic areas for the patients in a low CD4 + T cell count who present with symptoms of POAL, hepatosplenomegaly, and weight loss, and do not respond to the empirical or directed therapies.

Several limitations are to be noted including the lack of more diverse external cohort for further verification and the exclusion patients with T. marneffeii latent infections, which may reduce the effectiveness of our model. More samples and studies are highly required based on our models, which may substitute method into a practical clinic protocol. Although the diagnosis of T. marneffeii still needs to be confirmed by fungal culture, the model could assess and find high-risk patients of T. marneffeii infection before later stages, guiding clinicians in early early diagnosis and proper treatment.

Conclusions

In conclusion, we have successfully developed and validated accurate prediction models for T. marneffeii infection status among HIV individuals. More importantly, it is the first time to construct an effective model to diagnose and predict T. marneffeii infection in HIV-infected patients by using these clinical parameters, which can help to develop machine learning methods for clinical diagnosis. As a commonly screening method, the model construction based on only data routinely collected in clinical practice also facilitates application of our model, which is more convenient practice in some primary health care units with poor access to blood culture and other diagnostic laboratory. If validated among larger populations, application of our risk prediction models in clinical practice can help accurately tailor monitoring post-T. marneffeii infected individuals.

Abbreviations

Talaromyces marneffeii, T. marneffeii; Random forest, RF; receiver-operating characteristic curves, ROC; area under the ROC curve, AUC; peripheral or abdominal lymphadenopathy, POAL; AST to ALT ratio index, AARI; polymerase chain reaction, PCR; enzyme-linked immunosorbent assay, ELISA; white blood cell count, WBC; aspartate aminotransferase, AST; alanine aminotransferase, ALT; mean decrease in accuracy, MDA; confidence intervals, CI; interquartile range, IQR; odds ratio, OR.

Declarations

Acknowledgements

We thank all the physicians and patients who took part in this study.

Authors' contributions

ZSJ, BYL, MJL designed the study and provided the correlative knowledge. TC, DLM, YL collected the data, including data extraction. WDW, PZ, WJJ, GT cleaned and analyzed the data. JGH, XL drafted of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by The National Science and Technology Major Project of China During the 13th Five-Year Plan Period (2018ZX10302104), National Natural Science Foundation of China (NSFC, 81803295, 81760602), the committee of Health of Guangxi Zhuang autonomous region (Z20190034), Open project of Guangxi Key Laboratory of AIDS Research (gklapt201904).

Ethics approval and consent to participate

This study was approved by the Human Research Ethics Committee of the People's Hospital of Liuzhou (Ethical Review No. 2019-02-01). Written informed consent was obtained from each participant prior to their enrollment.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Publisher's Note

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

References

- [1] Duong T A. Infection due to *Penicillium marneffei*, an emerging pathogen: review of 155 reported cases[J]. *Clin Infect Dis*, 1996, 23(1): 125-130.
- [2] Piehl M R, Kaplan R L, Haber M H. Disseminated penicilliosis in a patient with acquired immunodeficiency syndrome[J]. *Arch Pathol Lab Med*, 1988, 112(12): 1262-1264.
- [3] Vanittanakom N, Cooper C R, Jr., Fisher M C, Sirisanthana T. *Penicillium marneffei* infection and recent advances in the epidemiology and molecular biology aspects[J]. *Clin Microbiol Rev*, 2006, 19(1): 95-110.
- [4] Le T, Wolbers M, Chi N H, Quang V M, Chinh N T, Lan N P, et al. Epidemiology, seasonality, and predictors of outcome of AIDS-associated *Penicillium marneffei* infection in Ho Chi Minh City, Viet Nam[J]. *Clin Infect Dis*, 2011, 52(7): 945-952.

- [5] Wu T C, Chan J W, Ng C K, Tsang D N, Lee M P, Li P C. Clinical presentations and outcomes of *Penicillium marneffei* infections: a series from 1994 to 2004[J]. *Hong Kong Med J*, 2008, 14(2): 103-109.
- [6] Hu Y, Zhang J, Li X, Yang Y, Zhang Y, Ma J, et al. *Penicillium marneffei* infection: an emerging disease in mainland China[J]. *Mycopathologia*, 2013, 175(1-2): 57-67.
- [7] Armstrong-James D, Meintjes G, Brown G D. A neglected epidemic: fungal infections in HIV/AIDS[J]. *Trends Microbiol*, 2014, 22(3): 120-127.
- [8] Jiang J, Meng S, Huang S, Ruan Y, Lu X, Li J Z, et al. Effects of *Talaromyces marneffei* infection on mortality of HIV/AIDS patients in southern China: a retrospective cohort study[J]. *Clin Microbiol Infect*, 2019, 25(2): 233-241.
- [9] Nor-Hayati S, Sahlawati M, Suresh-Kumar C, Lee K C. A retrospective review on successful management of *Penicillium marneffei* infections in patients with advanced HIV in Hospital Sungai Buloh[J]. *Med J Malaysia*, 2012, 67(1): 66-70.
- [10] Liu B, Ping F U. Research Progress of Penicilliosis Marneffei[J]. *Journal of Dermatology & Venereology*, 2010.
- [11] Wong S Y, Wong K F. *Penicillium marneffei* Infection in AIDS[J]. *Patholog Res Int*, 2011, 2011: 764293.
- [12] Le T, Kinh N V, Cuc N T K, Tung N L N, Lam N T, Thuy P T T, et al. A Trial of Itraconazole or Amphotericin B for HIV-Associated Talaromycosis[J]. *N Engl J Med*, 2017, 376(24): 2329-2340.
- [13] Desakorn V, Smith M D, Walsh A L, Simpson A J, Sahassananda D, Rajanuwong A, et al. Diagnosis of *Penicillium marneffei* infection by quantitation of urinary antigen by using an enzyme immunoassay[J]. *J Clin Microbiol*, 1999, 37(1): 117-121.
- [14] Chaiyaroj S C, Chawengkirttikul R, Sirisinha S, Watkins P, Srinoulprasert Y. Antigen detection assay for identification of *Penicillium marneffei* infection[J]. *J Clin Microbiol*, 2003, 41(1): 432-434.
- [15] Kawila R, Chaiwarith R, Supparatpinyo K. Clinical and laboratory characteristics of penicilliosis marneffei among patients with and without HIV infection in Northern Thailand: a retrospective study[J]. *BMC Infect Dis*, 2013, 13: 464.
- [16] Elias N A, Cuestas M L, Sandoval M, Poblete G, Lopez-Daneri G, Jewtuchowicz V, et al. Rapid identification of *Histoplasma capsulatum* directly from cultures by multiplex PCR[J]. *Mycopathologia*, 2012, 174(5-6): 451-456.
- [17] Hien H T A, Thanh T T, Thu N T M, Nguyen A, Thanh N T, Lan N P H, et al. Development and evaluation of a real-time polymerase chain reaction assay for the rapid detection of *Talaromyces marneffei* MP1 gene in human plasma[J]. *Mycoses*, 2016, 59(12): 773-780.

- [18] Pongpom M, Sirisanthana T, Vanittanakom N. Application of nested PCR to detect *Penicillium marneffei* in serum samples[J]. *Med Mycol*, 2009, 47(5): 549-553.
- [19] Sun J, Najafzadeh M J, Zhang J, Vicente V A, Xi L, De Hoog G S. Molecular identification of *Penicillium marneffei* using rolling circle amplification[J]. *Mycoses*, 2011, 54(6): e751-759.
- [20] Zhu Y M, Ai J W, Xu B, Cui P, Cheng Q, Wu H, et al. Rapid and precise diagnosis of disseminated *T.marneffei* infection assisted by high-throughput sequencing of multifarious specimens in a HIV-negative patient: a case report[J]. *BMC Infect Dis*, 2018, 18(1): 379.
- [21] Prakit K, Nosanchuk J D, Pruksaphon K, Vanittanakom N, Youngchim S. A novel inhibition ELISA for the detection and monitoring of *Penicillium marneffei* antigen in human serum[J]. *Eur J Clin Microbiol Infect Dis*, 2016, 35(4): 647-656.
- [22] Cao L, Chen D L, Lee C, Chan C M, Chan K M, Vanittanakom N, et al. Detection of specific antibodies to an antigenic mannoprotein for diagnosis of *Penicillium marneffei* penicilliosis[J]. *J Clin Microbiol*, 1998, 36(10): 3028-3031.
- [23] Ning C, Lai J, Wei W, Zhou B, Huang J, Jiang J, et al. Accuracy of rapid diagnosis of *Talaromyces marneffei*: A systematic review and meta-analysis[J]. *PLoS One*, 2018, 13(4): e0195569.
- [24] Rodriguez-Gallego E, Gomez J, Pacheco Y M, Peraire J, Vilades C, Beltran-Debon R, et al. A baseline metabolomic signature is associated with immunological CD4+ T-cell recovery after 36 months of antiretroviral therapy in HIV-infected patients[J]. *AIDS*, 2018, 32(5): 565-573.
- [25] Rodriguez-Gallego E, Gomez J, Domingo P, Ferrando-Martinez S, Peraire J, Vilades C, et al. Circulating metabolomic profile can predict dyslipidemia in HIV patients undergoing antiretroviral therapy[J]. *Atherosclerosis*, 2018, 273: 28-36.
- [26] Konerman M A, Zhang Y, Zhu J, Higgins P D, Lok A S, Waljee A K. Improvement of predictive models of risk of disease progression in chronic hepatitis C by incorporating longitudinal data[J]. *Hepatology*, 2015, 61(6): 1832-1841.
- [27] Wei R, Wang J, Wang X, Xie G, Wang Y, Zhang H, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning[J]. *EBioMedicine*, 2018, 35: 124-132.
- [28] Larsson M, Nguyen L H, Wertheim H F, Dao T T, Taylor W, Horby P, et al. Clinical characteristics and outcome of *Penicillium marneffei* infection among HIV-infected patients in northern Vietnam[J]. *AIDS Res Ther*, 2012, 9(1): 24.
- [29] Cooper C R, Jr., Mcginnis M R. Pathology of *Penicillium marneffei*. An emerging acquired immunodeficiency syndrome-related pathogen[J]. *Arch Pathol Lab Med*, 1997, 121(8): 798-804.

- [30] Hou J, Wang G, Wang F, Cheng J, Ren H, Zhuang H, et al. Guideline of Prevention and Treatment for Chronic Hepatitis B (2015 Update)[J]. J Clin Transl Hepatol, 2017, 5(4): 297-318.
- [31] Li W, Zhao J, Zou Z, Liu Y, Li B, Sun Y, et al. Analysis of hepatitis B virus intrahepatic covalently closed circular DNA and serum viral markers in treatment-naive patients with acute and chronic HBV infection[J]. PLoS One, 2014, 9(2): e89046.
- [32] Yousukh A, Jutavijittum P, Pisetpongsa P, Chitapanarux T, Thongsawat S, Senba M, et al. Clinicopathologic study of hepatic *Penicillium marneffei* in Northern Thailand[J]. Arch Pathol Lab Med, 2004, 128(2): 191-194.

Figures

Fig. 1

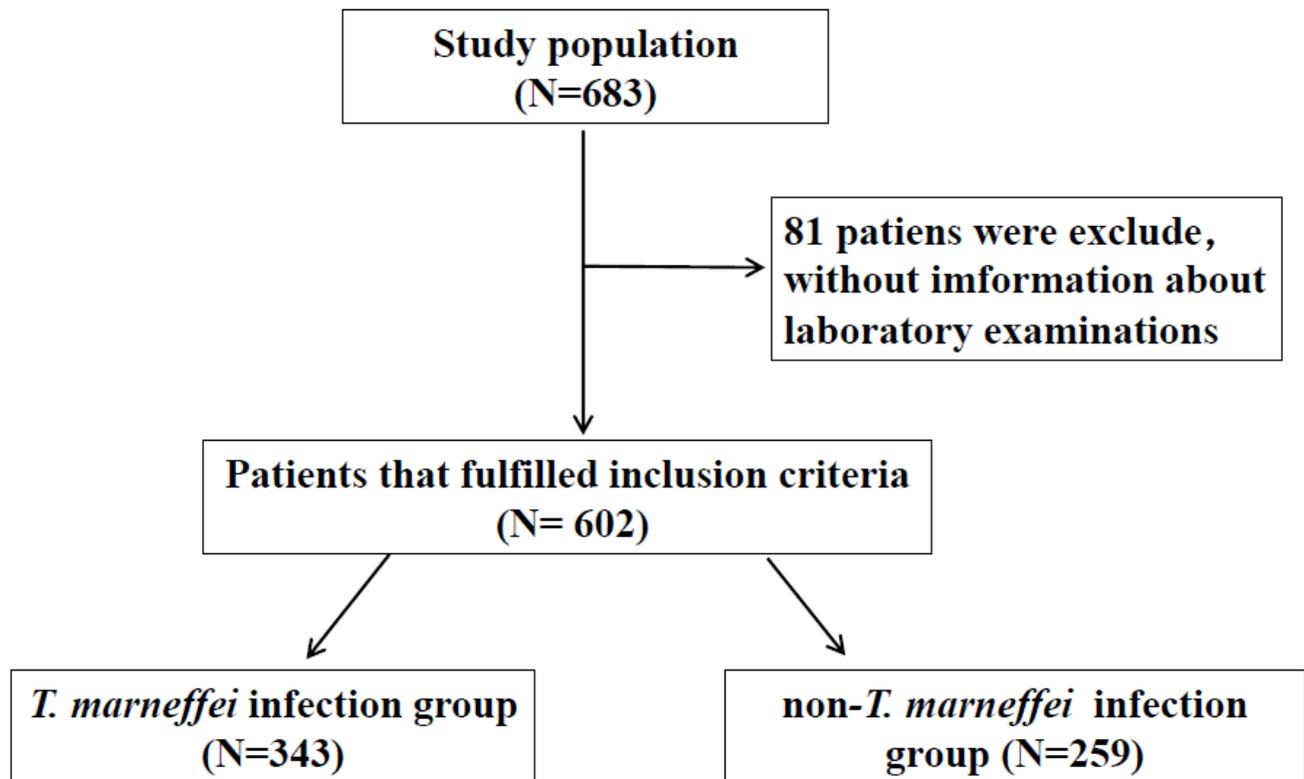


Figure 2

Chart of the patients included in the study.

Fig. 2

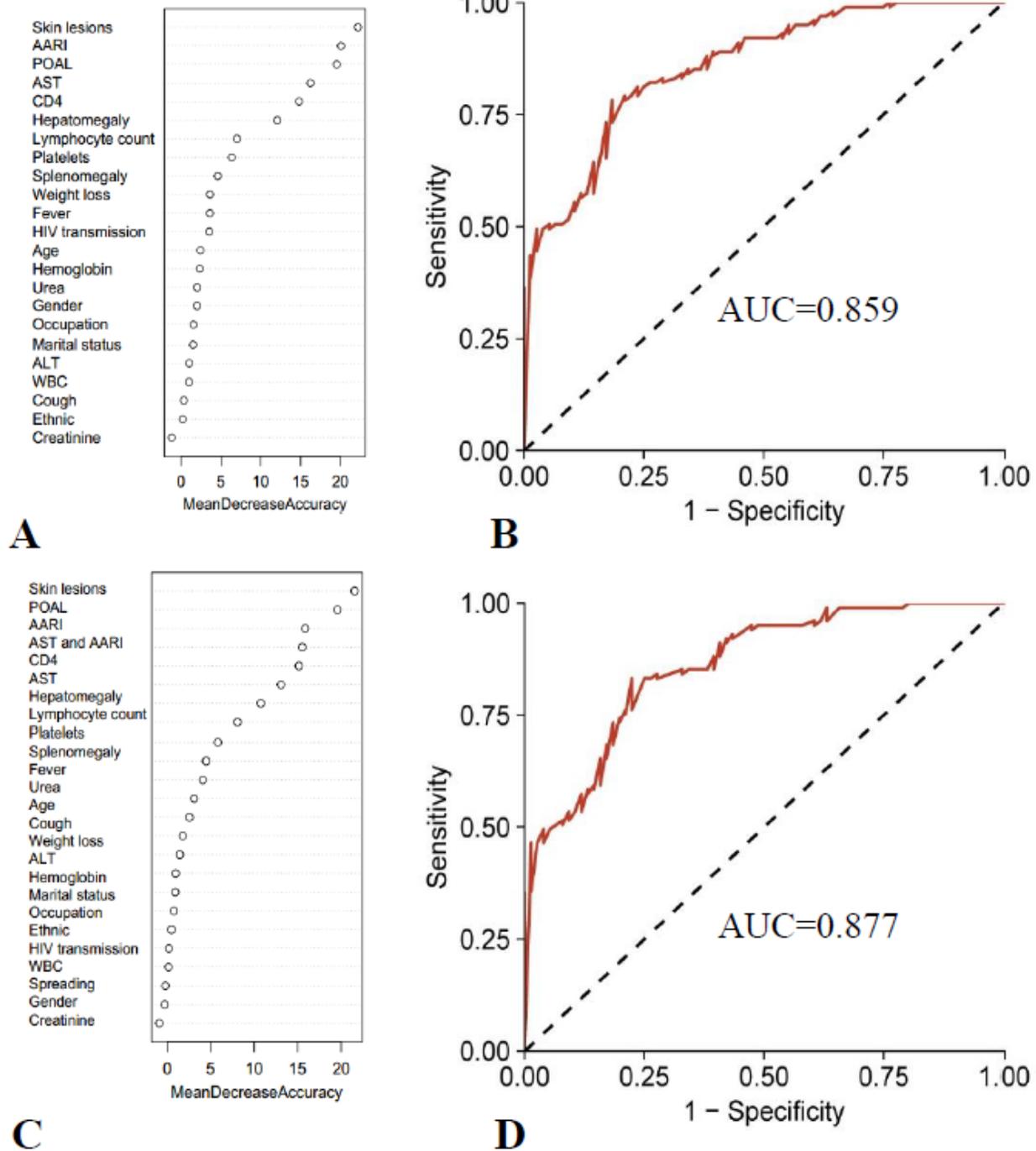


Figure 4

RF variable importance plots and ROC curves to classify patients according to baseline characteristics. (A, C) Variable importance graph for predicting *T. marneffei* infection. The ranked list of variables mean the importance of each variable in classifying data. The variables are ordered top-to-bottom as most-to-least important in classifying between *T. marneffei* infected patients and non- *T. marneffei* infected

patients. (B, D) Using ROC curves and 95% confidence interval, we assessed models that could accurately predict *T. marneffei* infection according to different clinical parameters.

Fig. 3

ROC Curve

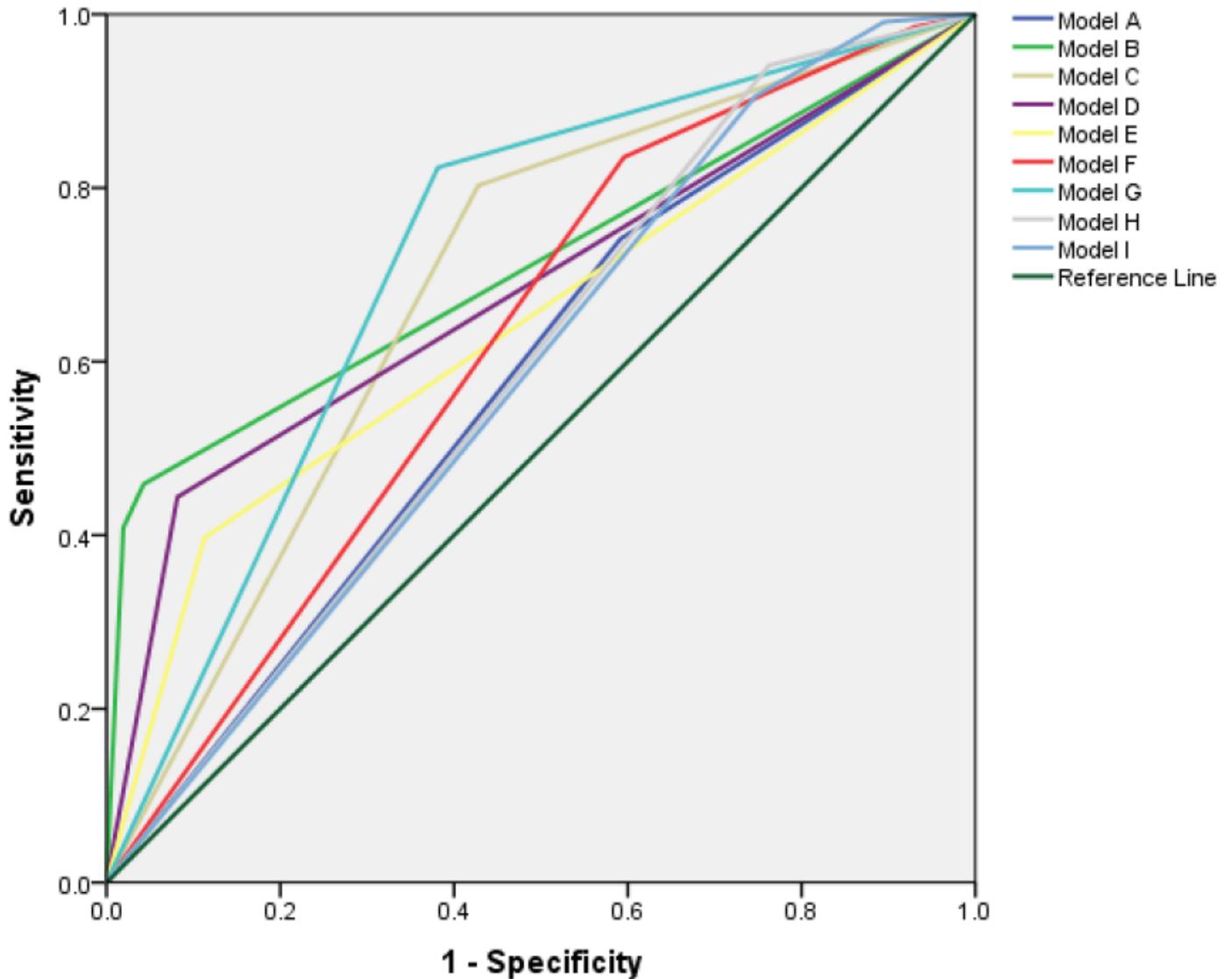


Figure 6

ROC curve of each variable for use in the distinguish *T. marneffei* infection of HIV patients. Using ROC curves, we assessed the ability of 9 models to accurately predict *T. marneffei* infection according to different parameters obtained in the logistic regression. Model A-I was respectively constructed with weight loss, skin lesions, POAL, hepatomegaly, splenomegaly, lymphocyte count, AST level, AARI and CD4+ T-cell count.