

# Revolutionizing Infection Risk Scoring: An Ensemble “From Weak to Strong” Deduction Strategy and Enhanced Point-of-Care Testing Tools

Haoman Chen, Hui Ye, Fanxuan Chen, Yi Shi, Jiahe Li, Yuhang Li, Chaoyi Wei, Chengxi Zhang, Jinhao Jin, Quanyue Zhang, Shuo Pan, Dong Chen, Shichao Quan, Xiaowei Xu, Zongxing Yang, Lijun Xu, Yungui Zhang, Yi Wang, Yunlong Zhou, Ou Liu, Xiaoming Jiang, Liyu Liu, Qingzu He, Xiaoqu Hu, Jingye Pan, Jianwei Shuai, and Feifei Su\*


The COVID-19 pandemic exacerbates challenges faced by human immunodeficiency virus patients, who are at heightened risk for infection due to compromised immune systems. This study aims to develop a reliable, home-based point-of-care testing (POCT) tool for early screening of acquired immunodeficiency syndrome (AIDS) coinfecting with *Talaromyces marneffei* infection. Employing a “From weak to strong” deduction strategy for feature selection, data from 464 AIDS patients across four cohorts between February 5th, 2014, and January 8th, 2022, are analyzed. The top three features consistently observed are D-dimer, cluster of differentiation 4+, and aspartate transaminase. Based on these features, the simplest risk-scoring model is constructed, with the area under the receiver operating characteristic curve values of 0.91, 0.80, and 0.69 in the hold-out cohort, external cohort 1, and external cohort 2, respectively. This “From weak to strong” deduction strategy identifies advantageous clinical features, enabling the development of simplified clinical risk scores with multiple biomarkers. To facilitate practical implementation, enhanced POCT tools are introduced, specifically a strip with segmented testing capabilities that demonstrates sensitivity and strong correlation with clinical scoring models. Furthermore, an open-access website and a free Android mobile app are created to support community utilization. The findings underscore the effectiveness of the innovative deduction strategy and enhanced test strips, which enable bedside measurements without laboratory dependency.

## 1. Introduction

The COVID-19 pandemic has imposed unprecedented impediments on clinical diagnostics and therapeutic paradigms, attributable to its enhanced transmissibility and diverse symptomatology.<sup>[1]</sup> This scenario intensifies the quandary faced by patients with common infectious diseases competing for healthcare resources and also amplifies the risk of nosocomial infections.<sup>[2]</sup> Of particular note is the susceptibility of individuals infected with the human immunodeficiency virus (HIV), a cohort at high risk with impaired immune systems, who grapple with considerable threats and challenges in the context of the pandemic.<sup>[3–5]</sup> Consequently, the construction of reliable, portable point-of-care testing (POCT) devices for home use is paramount in mitigating infection risks within outpatient care settings.

However, the progress of POCT technologies relies on the presence of specific antigen–antibody markers, as demonstrated by COVID-19 and syphilis POCT assays.<sup>[6,7]</sup> Regrettably, not all infectious

H. Chen, H. Ye, D. Chen, F. Su  
Department of Infectious Diseases  
Wenzhou Central Hospital  
Wenzhou, Zhejiang 325000, China  
E-mail: feifeisuzs@163.com

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202300224>.

© 2023 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202300224

H. Chen, Y. Shi, J. Li, Y. Li, C. Wei, O. Liu, X. Jiang, L. Liu, J. Shuai  
Wenzhou Institute and Wenzhou Key Laboratory of Biophysics  
Wenzhou Institute  
University of Chinese Academy of Sciences  
Wenzhou, Zhejiang 325001, China

H. Chen, F. Chen, Y. Shi, J. Li, Y. Li, J. Jin, Q. Zhang, S. Pan  
Wenzhou Medical University  
Wenzhou, Zhejiang 325000, China

F. Chen, Q. Zhang, Y. Wang, Y. Zhou, J. Shuai  
School of Biomedical Engineering  
School of Ophthalmology and Optometry  
Eye Hospital  
Wenzhou Medical University  
Wenzhou, Zhejiang 325001, China

diseases exhibit easily identifiable, unambiguous antigen–antibody markers. In HIV-related infections caused by *Talaromyces marneffeii*, the quantity of relevant antigens in the bloodstream of HIV patients is strikingly low due to weakened immune function, thereby hindering the production of swift POCT screening tools.<sup>[8]</sup> Moreover, *Talaromyces marneffeii*, a dimorphic fungus, presents nonspecific clinical signs, often manifesting as rashes, respiratory symptoms, fever, anemia, and other systemic signals. Anticipating the onset of this disease is arduous, and its diagnosis requires prolonged pathogen cultivation, frequently resulting in diagnostic lags and increased

mortality rates.<sup>[8–10]</sup> A multitude of severe illnesses face similar obstacles, such as nonspecific clinical symptoms and challenges in obtaining diagnostic gold standards, as noted in sepsis, which typically employs the SOFA and APACHE IV scoring systems for screening and diagnosis.<sup>[10,11]</sup> As a result, the creation of a disease screening method that is not dependent on specific antigen–antibodies and is applicable to the early detection of critical illnesses, carries significant clinical importance and the possibility for broad application.

Z. Yang  
Department of Infectious Diseases  
Hangzhou Xixi Hospital  
Hangzhou, Zhejiang 320000, China

C. Zhang  
School of Materials Science and Engineering  
Shandong Jianzhu University  
Jinan, Shandong 250101, China

J. Jin, X. Xu, X. Hu  
The First Affiliated Hospital of Wenzhou Medical University  
Wenzhou, Zhejiang 325000, China

D. Chen, F. Su  
Wenzhou Key Laboratory for the Diagnosis and Treatment of Emerging and Re-emerging Infectious Diseases  
Wenzhou, Zhejiang 325000, China

S. Quan, J. Pan  
Department of Big Data in Health Science  
The First Affiliated Hospital of Wenzhou Medical University  
Wenzhou, Zhejiang 325100, China

S. Quan, J. Pan, J. Shuai  
Key Laboratory of Intelligent Treatment and Life Support for Critical Diseases of Zhejiang Province  
Wenzhou, Zhejiang 325000, China

S. Quan, J. Pan  
Zhejiang Engineering Research Center for Hospital Emergency and Process Digitization  
Wenzhou, Zhejiang 325000, China

L. Xu  
Department of Infectious Diseases  
The First Affiliated Hospital of Zhejiang University School of Medicine  
Hangzhou, Zhejiang 320000, China

Y. Zhang  
Yunnan Provincial Infectious Disease Hospital  
Kunming, Yunnan 650000, China

Y. Wang, Y. Zhou  
Engineering Research Center of Clinical Functional Materials and Diagnosis & Treatment Devices of Zhejiang Province  
Wenzhou Institute  
University of Chinese Academy of Sciences  
Wenzhou, Zhejiang 325001, China

Q. He  
Department of Physics, and Fujian Provincial Key Laboratory for Soft Functional Materials Research  
Xiamen University  
Xiamen, Fujian 361005, China

X. Hu  
Department of Thyroid and Breast Surgery  
The First Affiliated Hospital of Wenzhou Medical University  
Wenzhou, Zhejiang 325000, China

The core objective of devising an early screening tool for diseases that avoids reliance on specific antigen–antibodies is to identify a set of biomarkers that demonstrate high sensitivity and disease specificity. In recent times, machine learning techniques have seen extensive application in the field of biomarker identification.<sup>[12–16]</sup> Nonetheless, a large proportion of studies are devoid of sturdy guidelines for choosing the most suitable data method for their datasets, with selection outcomes often dependent on the performance of different methods on a single dataset. This introduces a degree of randomness into the final biomarker selection.<sup>[17,18]</sup> To lessen the irregularity in such research findings, our study suggests a more equitable feature selection process, subsequently quantifying the selected features for integration into POCT assays.

Present clinical point-of-care testing (POCT) strips primarily use colloidal gold assays, concentrating on specific antigen–antibody interactions, and are therefore designed for qualitative responses.<sup>[19]</sup> However, our ambition is to construct a disease-screening approach that overcomes dependence on specific antigen–antibody interactions, incorporating quantitative assessments of clinical scores and disease-related indices. Basic qualitative classification outcomes are insufficient to meet our needs. As such, we have devised a series of colloidal gold assay strips capable of semiquantitative measurements, allowing their smooth integration with clinical scoring systems.<sup>[17]</sup>

Employing patients with *Talaromyces marneffeii* infections within the HIV-infected population as a multicenter cohort study, our research suggests a selection strategy for nonspecific antigen–antibody disease biomarkers. Through implementing a “from weak to strong” feature selection approach, we have mitigated biases arising from different selection methods and innovatively applied the resultant selections to the fabrication of semiquantitative colloidal gold assay strips. This integration enables the combination of POCT detection and clinical scoring. Additionally, we have established a dedicated website and an efficient mobile app for AIDS patients and healthcare professionals to avail of early screening for *Talaromyces marneffeii* infections at no charge.

## 2. Result

### 2.1. Patient Clinical Data from Three Medical Centers

A total of 464 patients diagnosed with AIDS were assessed, spanning from February 5th, 2014, to January 8th, 2022, across three medical centers in China, Figure S1, Supporting Information. These patients were methodically stratified into four distinct cohorts: a model development cohort ( $n = 318$ ) and a holdout validation cohort ( $n = 45$ ) from Wenzhou Central Hospital; an external cohort designated for model optimization ( $n = 49$ ) from Hangzhou Xixi Hospital (external cohort

1); and a final external validation cohort ( $n = 52$ ) from the First Affiliated Hospital of Zhejiang University School of Medicine (External Cohort 2). The development cohort's timeline encompassed the years 2014 through 2022, while the remaining cohorts spanned 2016 to 2021.

Patient selection and distribution among the cohorts were meticulously designed to optimize model performance and provide unbiased validation, ultimately bolstering the credibility of the study's conclusions. A set of four data distributions' heatmaps is elegantly depicted in Figure S2a–d, Supporting Information, while the precise data tables and intergroup statistical tests can be found in Table S1, Supporting Information. The interrelation among the attributes within each cohort is discernibly illustrated in Figure S2e, Supporting Information. The overall workflow and pipeline of the experiments are illustrated in Figure S3, Supporting Information.

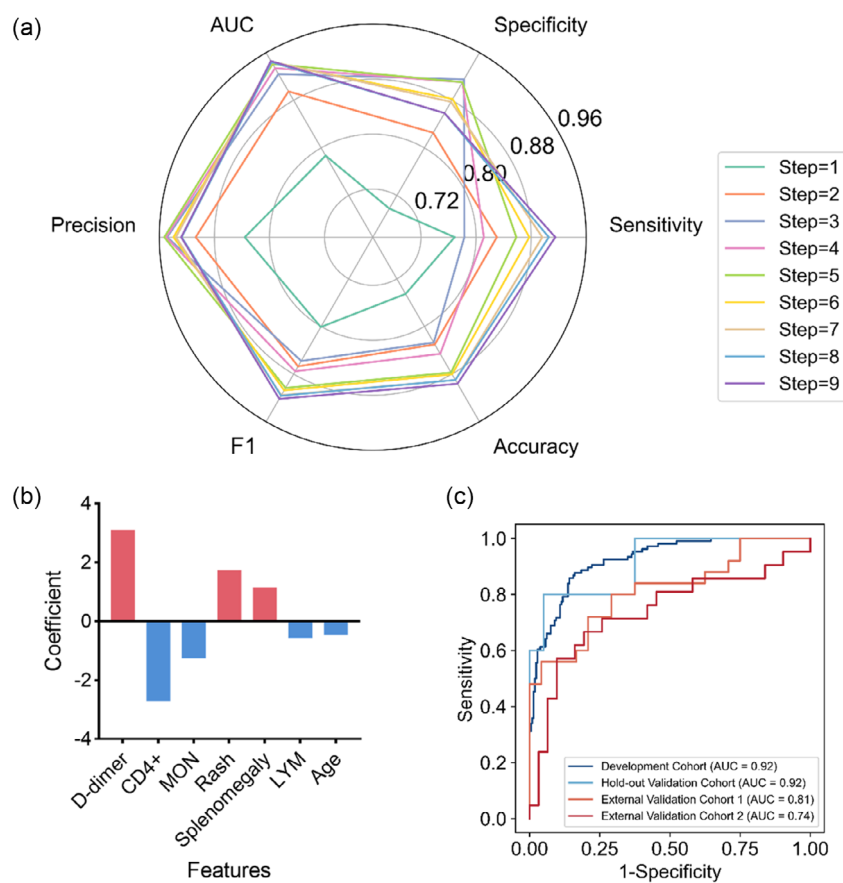
## 2.2. Classical Standardized Logistic Regression

After implementing Z-standardization, we conducted univariate regression analyses on all 32 features within the logistic

regression model to pinpoint variables exhibiting significant p-values. Subsequently, we carried out multivariate regression analyses on these variables utilizing a forward selection method. The performance at each stage of the stepwise forward logistic regression is aptly demonstrated in **Figure 1a**. Results from both univariate and multivariate regression analyses can be found in the Table S2, Supporting Information.

The multivariate regression analysis delineated seven key features, including rash (odds ratio (OR) = 5.712, 95% confidence interval (CI): 2.057–15.862), splenomegaly (OR = 3.179, 95% CI: 1.324–7.631), and D-dimer (OR = 22.458, 95% CI: 4.789–105.304), which were classified as risk factors for *Talaromyces marneffei* infection. Conversely, protective factors included age (OR = 0.635, 95% CI: 0.406–0.995), lymphocyte count (OR = 0.564, 95% CI: 0.332–0.958), monocyte count (OR = 0.287, 95% CI: 0.119–0.694), and CD4+ cell count (OR = 0.066, 95% CI: 0.014–0.317). The most salient risk factor was D-dimer, while the most potent protective factor was serum CD4+ cell count.

A ranking of characteristics, based on the optimal model's standardized logistic regression coefficients ( $\beta$ ), is vividly



**Figure 1.** Logistic regression experimental group visualization. a) The radar chart of the performance of each step of stepwise logistic regression. At step = 1, all metrics are at their lowest values. When the stepwise forward method automatically concludes at step = 9, the AUC, accuracy, F1, and sensitivity metrics all reach their peak values, which are 0.946, 0.896, 0.922, and 0.915, respectively. b) Feature ranking of  $\beta$  coefficients based on standardized logistic regression. The red and blue colors represent risk factors and protective factors, respectively, for predicting *Talaromyces marneffei* infection. Among the risk factors, the variable “D-dimer” has the highest  $\beta$  coefficient value of 3.112. Conversely, among the protective factors, the variable “CD4+” has the largest coefficient of  $-2.720$ . c) Four-cohort ROC curves of the optimal model based on standardized logistic regression.

portrayed in Figure 1b. The receiver operating characteristic (ROC) curves for the four cohorts, derived from the optimal standardized logistic regression model, are displayed in Figure 1c.

### 2.3. Shapley Additive Explanations and Permutation Importance of Machine Learning Models

The 12 machine learning models underwent a thorough process of 50 iterations of grid hyperparameter tuning. Based on the optimal parameters derived from this tuning, the models were subsequently trained on the primary cohort. The area under the curve (AUC) results for the 50 bootstrap resampling iterations are detailed in the supplementary material, specifically Table S3, Supporting Information, along with the optimal hyperparameter combinations for the models in Table S4, Supporting Information. The receiver operating characteristic (ROC) curve outcomes for the three test cohorts using the 12 machine learning models are provided in Figure S4, Supporting Information.

Utilizing the 12 optimally tuned machine learning models, we assessed and ranked the relative importance of all 32 predictor variables in deductive Sections 2 and 3. This evaluation employed the SHapley Additive exPlanations (SHAP) and permutation importance (PI) methodologies. The specific results for SHAP and PI can be inspected in Figures 2 and 3. Within each figure, subplot a presents the feature importance results for the 32 characteristics across the 12 machine learning models under both SHAP and PI frameworks; subplot b showcases the structural diagrams for SHAP and PI; subplot c illustrates the fluctuation of feature importance relative to the overall average level after adjusting the feature importance results of the 12 machine learning models under SHAP and PI frameworks by external cohort 1; and subplot d portrays the variations in the importance ranking of the final three features among the 32 features across the 12 machine learning models under both SHAP and PI frameworks.

### 2.4. From Weak to Strong Deduction and Consistency of the Sorting Results

In the standardized logistic analysis (Deduction I, Figure 4a), the uppermost seven features encompassed DD, CD4, rash, M, splenomegaly, L, and age, with respective scores of 3.112, 2.720, 1.743, 1.249, 1.156, 0.572, and 0.454. In the 12 machine learning models, we assessed and ranked the relative importance of all 32 predictor variables in deductive Sections 2 and 3 using SHAP and PI methods. In the SHAP analysis (Deduction II, Figure 4b), the top seven features included D-dimer, CD4+, rash, blood platelet (PLT), splenomegaly, aspartate transaminase (AST), and M, with scores of 873.84, 600.67, 480.20, 472.82, 449.11, 423.88, and 376.20. In the PI analysis (Deduction III, Figure 4c), the top seven features comprised D-dimer, CD4+, AST, rash, PLT, M, and splenomegaly, with respective scores of 799.39, 648.20, 432.93, 389.90, 370.55, 330.27, and 294.33.

We executed a comprehensive evaluation of the three independent deductions, revealing that D-dimer, CD4+, and rash consistently ranked 1st, 2nd, 3rd, or 4th across the three deductions, thereby, exhibiting high importance regardless of the data fitting method employed. For instance, D-dimer accounted for 28.3%, 23.8%, and 24.5% of the total proportion of the top seven features

in standardized logistic regression, SHAP, and PI, respectively, as depicted in Figure 4d–f. D-dimer, CD4+, and rash together constituted 68.8%, 28.8%, and 27.7% of the overall feature importance in standardized logistic regression, SHAP, and PI, respectively, as illustrated in Figure 4g.

Consequently, we selected these three features to streamline our model. We reimplemented logistic regression using these three features in the development cohort and tested it on the validation, external cohort 1, and external cohort 2 datasets. The acquired AUC values were 0.883, 0.822, and 0.728, signifying that the discriminative prowess of this streamlined model was not markedly inferior to the 32-feature model. The specific results can be examined in Figure 4h. Therefore, developing the most streamlined scoring system employing these three features represents a robust approach.

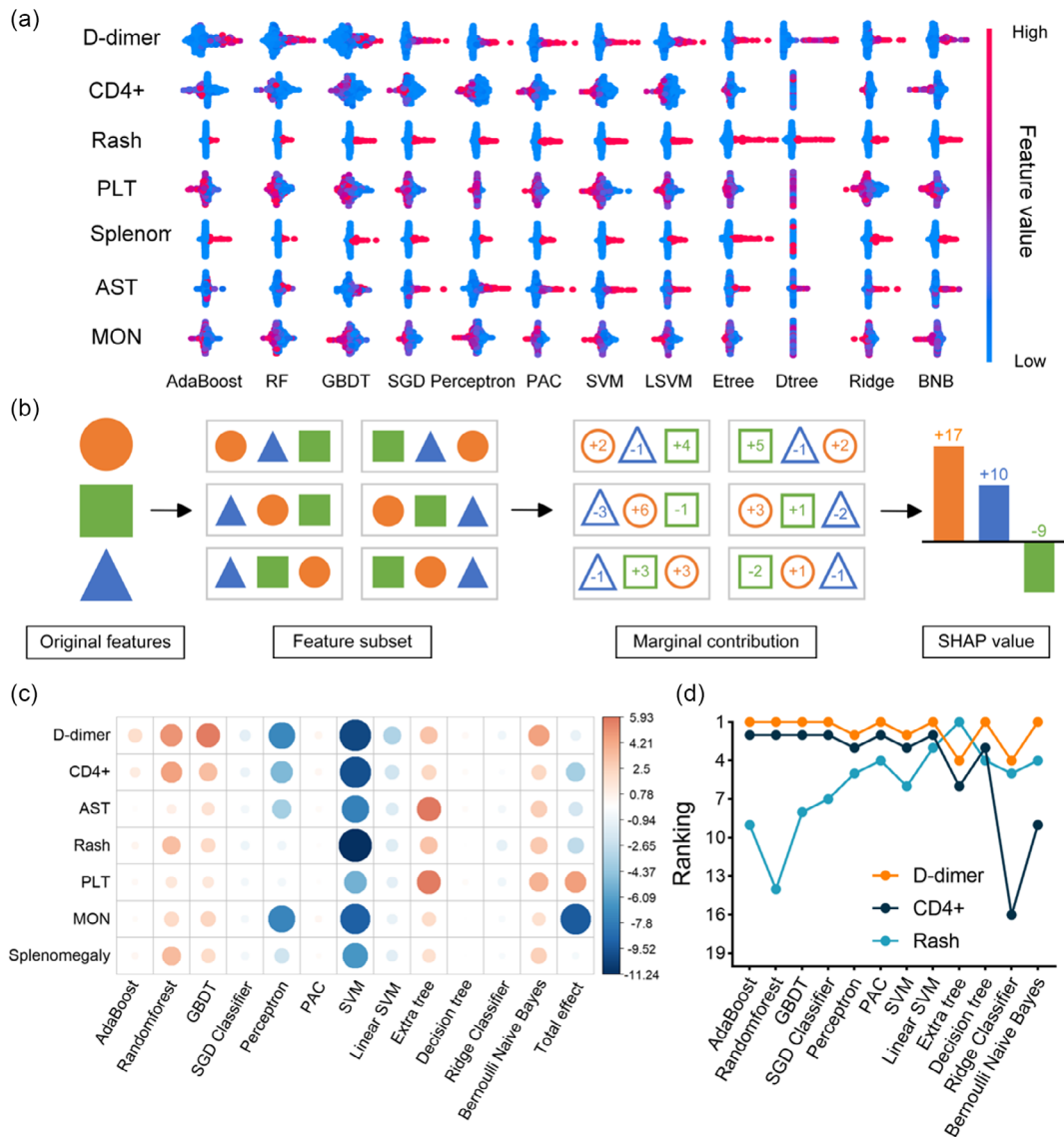
### 2.5. The Establishment of the Simplest Scoring System and the Evaluation of System Performance

To devise a robust risk scoring system founded on the initial three characteristics, we harnessed the prowess of logistic regression models' beta coefficients. Abiding by well-established clinical guidelines, we demarcated scoring thresholds for D-dimer, CD4+ cells, and rash, adhering to these stipulations: D-dimer  $< 0.5 \text{ mg L}^{-1}$  yields 0 points, D-dimer  $> 0.5 \text{ mg L}^{-1}$  awards 1 point, D-dimer  $> 1 \text{ mg L}^{-1}$  grants 2 points, D-dimer  $> 2 \text{ mg L}^{-1}$  accords 3 points; CD4  $> 200 \times 10^6 \text{ L}^{-1}$  merits 0 points, CD4  $> 100 \times 10^6 \text{ L}^{-1}$  secures 2 points, CD4  $> 50 \times 10^6 \text{ L}^{-1}$  earns 4 points, CD4  $< 50 \times 10^6 \text{ L}^{-1}$  garners 6 points; rash's absence scores 0 points, while its presence bestows 2 points.<sup>[20–23]</sup> Our risk scoring system manifested admirable discriminatory acumen across diverse datasets, boasting AUC values of 0.91, 0.80, and 0.69 in the hold-out cohort, external cohort 1, and external cohort 2, respectively, as delineated in Figure 5h.

Furthermore, we conducted a meticulous examination of all threshold points to ascertain the model's precision, sensitivity, and specificity. These findings can be scrutinized in Figure 5a–e, while the data are accessible in Table S5, Supporting Information. To validate our scoring system's distinctiveness within the data, we implemented unsupervised clustering through the K-means algorithm on the entire cohort of 464 patients, opting for the optimal cluster number  $k = 2$  as recommended by the elbow plot. Our analysis unveiled statistically significant disparities between the scores of the autonomously clustered group. Detailed outcomes can be perused in the f and g subplots within Figure 5, and the elbow plot results are available in Figure S5, Supporting Information.

### 2.6. The Reformation of POCT Test Strips

Our ingeniously devised POCT test strip reflects concentrations of D-dimer and CD4+ through the number of red T-lines displayed on each strip, while the presence or absence of a rash can be ascertained via visual examination. This POCT test strip empowers clinicians and home-based patients to directly obtain a risk score for *Talaromyces marneffe* infection. Patients amassing scores of 7 or above are deemed high-risk candidates for *Talaromyces marneffe* infection. This design expedites prompt



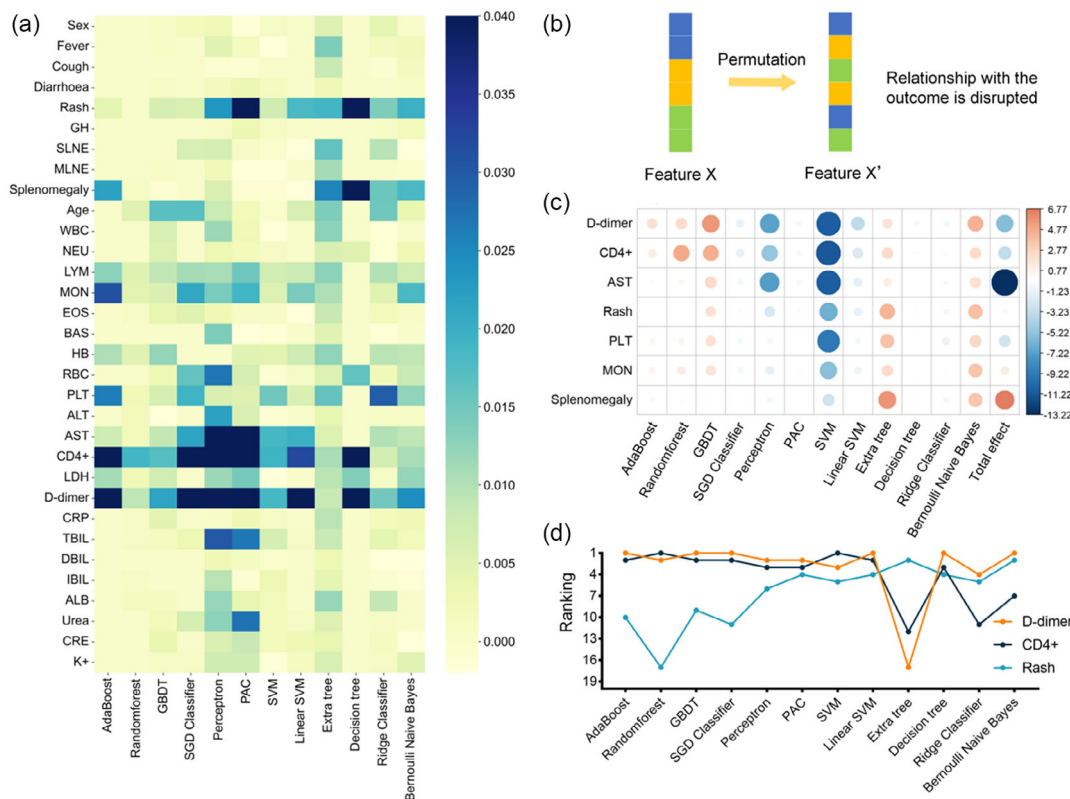
**Figure 2.** Shapley additive explanations experimental group visualization. a) The heat map of the feature importance results of 32 features under SHAP interpretation in 12 machine learning models. Colored bars indicate the varying magnitudes of feature values. We present the results for the top seven features, as ranked by their cumulative SHAP values, from 12 machine learning models. b) The schematic diagram of SHAP. c) The comparative chart of fluctuations of SHAP results after correction by external cohort 1. Comparison of SHAP results for 12 machine learning models before and after calibration on external cohort 1. In comparison to the average AUC for each model on external cohort 1, blue signifies a relative decrease in model scores after calibration, while red indicates a relative increase in model scores after calibration. The larger the circle, the greater the degree of change. d) The ranking changes of the final three main features on 12 machine learning models under the SHAP explanation.

bedside and community-based testing. The application protocol of the test suite, along with the completed prototype experiments, is depicted in **Figure 6**.

Figure 6a elucidates the workflow of the POCT test strip's clinical application, Figure 6b exhibits the colorimetric scoring interpretation methodology for the test strip, Figure 6c unveils the fundamental principle of the POCT test strip, and Figure 6d,e portrays the intermediary procedures and outcomes of our test strip fabrication. The specific chemical manufacturing process can be accessed in the supplementary material.

## 2.7. The Establishment of a Free Internet Platform

In a bid to aid clinicians and patients in making well-informed diagnostic assessments, a gratuitous open platform has been meticulously constructed, which can be accessed at [www.aoids-diagnosis.com](http://www.aoids-diagnosis.com), as exhibited in **Figure 7b**. This considerably designed website enables both physicians and patients to input relevant clinical data, from which it subsequently yields reliable forecasts. Concurrently, we have crafted a mobile application for the Android platform, ensuring that users can access the system



**Figure 3.** Permutation importance of experimental group visualization. a) The heat map of the feature importance results of 32 features under PI interpretation in 12 machine learning models. b) The schematic diagram of PI. c) The comparative chart of fluctuations of PI results after correction by external cohort 1. Comparison of PI results for 12 machine learning models before and after calibration on external cohort 1. In comparison to the average AUC for each model on external cohort 1, blue signifies a relative decrease in model scores after calibration, while red indicates a relative increase in model scores after calibration. The larger the circle, the greater the degree of change. d) The ranking changes of the final three main features on 12 machine learning models under the PI explanation.

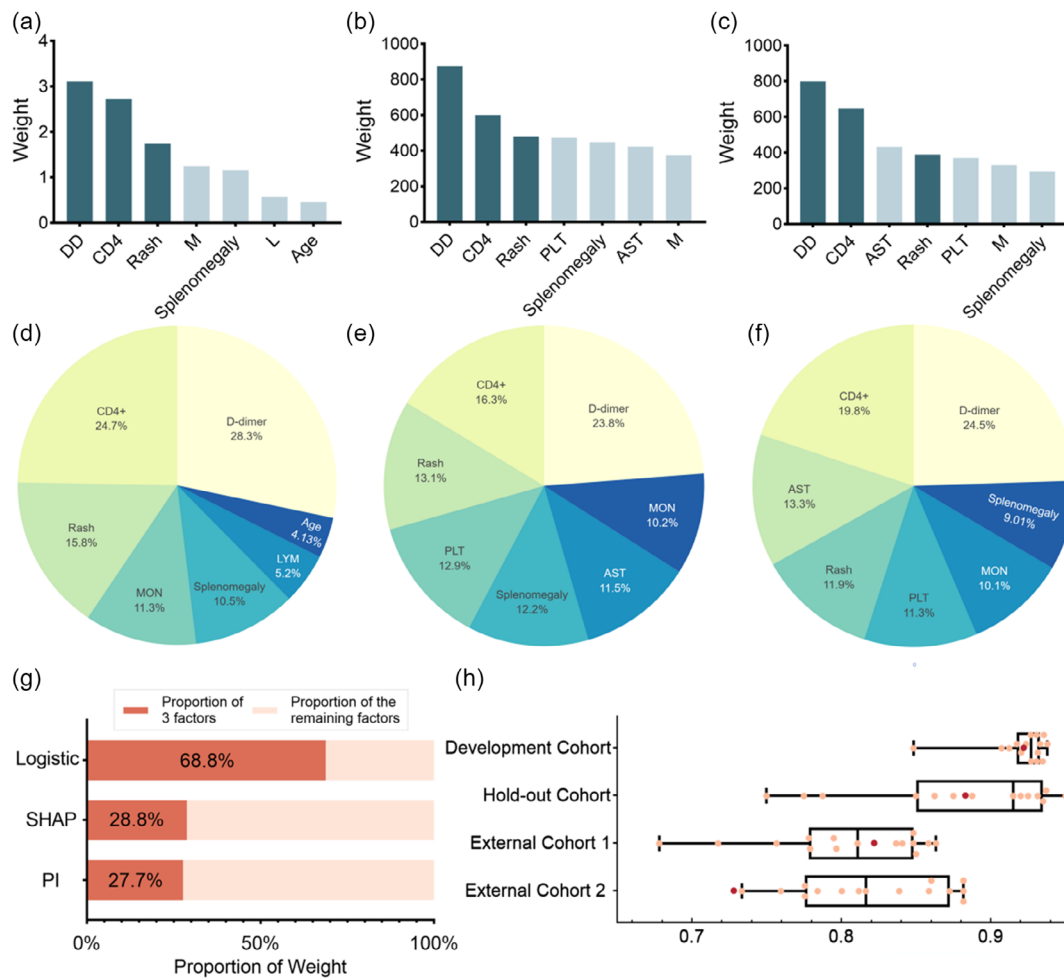
irrespective of location, as demonstrated in Figure 7c. Figure 7a outlines the exhaustive development process of the website, application, and accompanying software.

### 3. Discussion

Dimorphic fungi, distinguished by their capacity to effectuate morphological transitions in response to environmental stimuli, are essential facilitators of adaptation throughout their life cycles.<sup>[24]</sup> They transform from filamentous structures to yeast-like forms within host organisms, thus initiating pathogenic infections. The complex morphological transformations exhibited by fungi such as *Histoplasma*, *Talaromyces marneffeii*, and *Sporothrix*, pose substantial challenges to the host immune system. This complexity not only complicates the diagnosis and treatment of infections but also significantly hinders accurate and efficient therapeutic intervention.<sup>[24]</sup> *Talaromyces marneffeii* epitomizes this group of fungi, demonstrating intricate morphological changes, immunological hurdles, diverse infection sites, and a broad range of infection severity, leading to a myriad of atypical clinical manifestations that obstruct precise diagnosis and effective therapy.<sup>[8]</sup>

In the present investigation, we employed an amalgamation of conventional statistical methods, machine learning techniques, advanced biomarker screening, clinical scoring, and point-of-care testing (POCT) approaches to formulate an exhaustive early screening tool development process. This process was specifically designed for AIDS patients during the COVID-19 pandemic, with the goal of facilitating community and home-based screening.

In clinical settings, when procuring gold standard pathogen cultures is unfeasible, the identification of pertinent biomarkers for infection screening becomes paramount. However, differentiating high-diagnostic-value features from a multitude of complex and variable characteristics is a significant challenge. Predominant feature selection methods include logistic regression and machine learning.<sup>[25,26]</sup> Logistic regression, given its superior interpretability, is frequently utilized in the creation of clinical diagnostic models. Nonetheless, its linear fitting model might overlook nonlinear and fundamental latent features.<sup>[27]</sup> In contrast, machine learning provides flexibility and adaptability in feature selection, exploring optimal fits through a diverse array of algorithms.<sup>[25]</sup> However, during the development of machine learning models, feature weights for a particular disease often exhibit variability depending on the algorithms



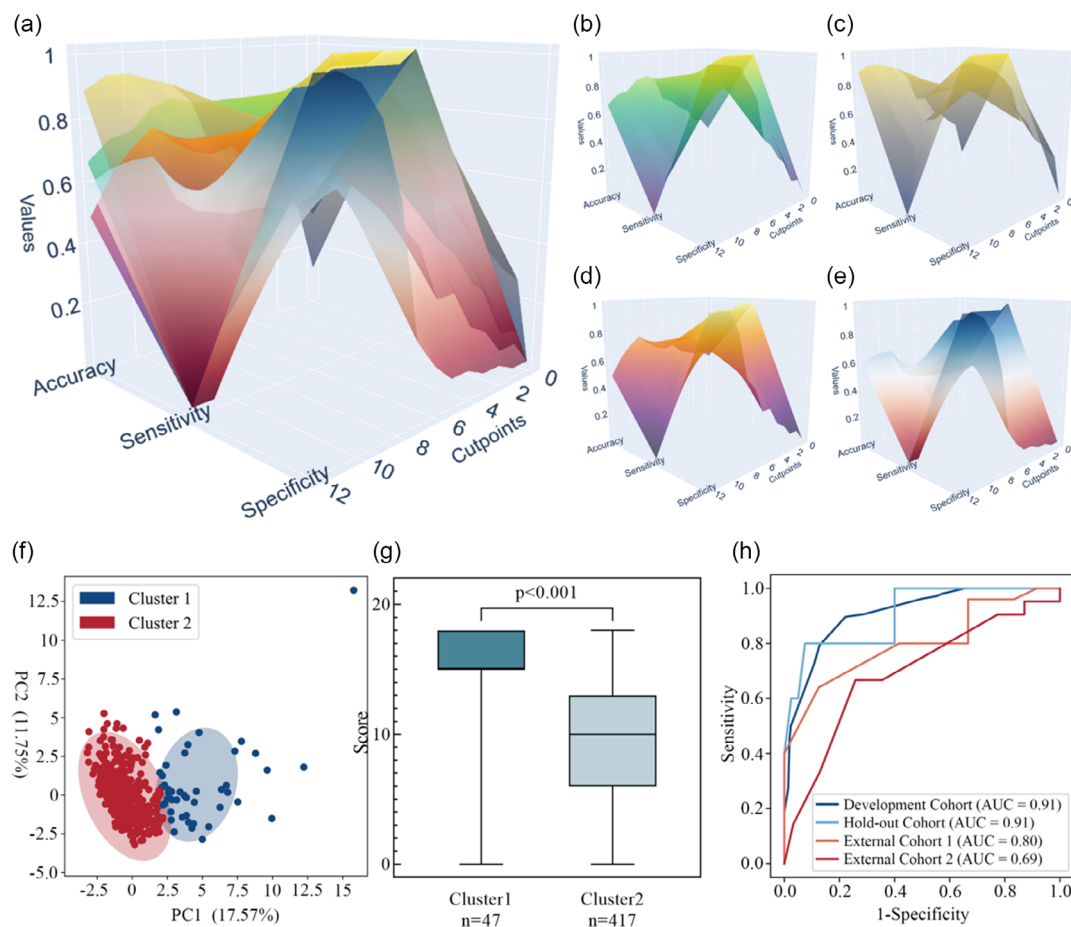
**Figure 4.** From Weak to Strong Deduction Group visualization. a) Histogram of weight for seven features with weak Deduction I (Standardized Logistic Regression). b) Histogram of weight for first seven features with weak Deduction II (Shapley additive explanations). c) Histogram of weight for first seven features with weak Deduction III (permutation importance). d) Pie chart of proportion of weight for seven features with weak Deduction I (Standardized Logistic Regression). e) Pie chart of proportion of weight for first seven features with weak Deduction I (Standardized Logistic Regression). f) Pie chart of proportion of weight for first seven features with weak Deduction I (Standardized Logistic Regression). g) Bar chart of the percentage distribution of three factors and the remaining factors. h) Boxplots of the AUC of machine learning algorithms and logistic regression. The red dot denotes the AUC of the three features logical regression. The AUC of the logistic regression with three features demonstrates considerable competitiveness across four distinct cohorts.

utilized. Moreover, different feature weight calculation methods within a single algorithm can lead to divergent rankings of feature importance. This inconsistency necessitates attention in current research, as the lack of standardization and distinct outcomes arising from varying algorithms and feature importance ranking methods compromise the reliability of these models.

To overcome these challenges, we propose an integrative strategy that combines logistic regression with multiple machine learning models, allowing for the ranking of feature importance from multiple perspectives. Our method analyzes feature importance through three separate aspects: standardized logistic regression coefficients, SHapley Additive exPlanations (SHAP), and permutation importance (PI). This “from weak to strong” inductive process ultimately produces highly consistent results.

In logistic regression, the regression coefficients indicate the extent to which a feature influences the dependent variable. However, direct comparisons between unique features’ regression coefficients are confounded by unit effects.<sup>[28,29]</sup> To mitigate this issue, we utilize standardized regression coefficients, derived from standardizing both the features and the dependent variable. This standardization process eliminates disparities in magnitude and scale, thereby facilitating the examination of weight magnitudes across various variables.<sup>[26]</sup>

Furthermore, specific machine learning models typically possess their own feature weight analysis methods, such as the coefficients of Support Vector Machine (SVM) class models and feature importance of tree models like Random Forest. Nonetheless, these methods are primarily designed for internal model analysis and prove inadequate for comparisons between distinct algorithms. Directly comparing various algorithm



**Figure 5.** Simplest scoring system experimental group visualization. a) 3D visualization of sensitivity, specificity, and accuracy on four cohorts with different cut-off values. b) 3D visualization of sensitivity, specificity, and accuracy on development cohort with different cut-off values. c) 3D visualization of sensitivity, specificity, and accuracy on hold-out cohort with different cut-off values. d) 3D visualization of sensitivity, specificity, and accuracy on external cohort1 with different cut-off values. e) 3D visualization of sensitivity, specificity, and accuracy on external cohort 2 with different cut-off values. f) K-means plot. The patients were unsupervised and clustered based on 32 features. The number of clusters  $K=2$ , as recommended by the elbow graph, was used. g) Boxplots of the scores for the two clusters. Cluster 1 has statistically significantly higher scores than cluster 2 ( $n = 464$ , Mann-Whitney  $U$  test,  $p < 0.001$ ). h) AUC of the scoring system in the development cohort, hold-out cohort, external cohort 1, and external cohort 2.

models using disparate weight analysis methods risks introducing biases into the feature importance analysis.<sup>[20,21,29]</sup> To rectify this issue, we employ two external perspective feature evaluation techniques: SHAP and permutation importance. The SHAP framework serves as an interpretable tool for machine learning model predictions, ascertaining the contributions of individual input features to prediction outcomes based on the Shapley value concept from cooperative game theory and providing an inclusive explanation of the model's decision-making process.<sup>[29,30]</sup> Permutation importance, a model-agnostic method, calculates feature importance by assessing the extent to which performance scores decrease when features are randomly permuted.<sup>[31]</sup> By implementing these two independent feature evaluation methods, we attain a unified perspective for evaluating feature importance across a multitude of models.

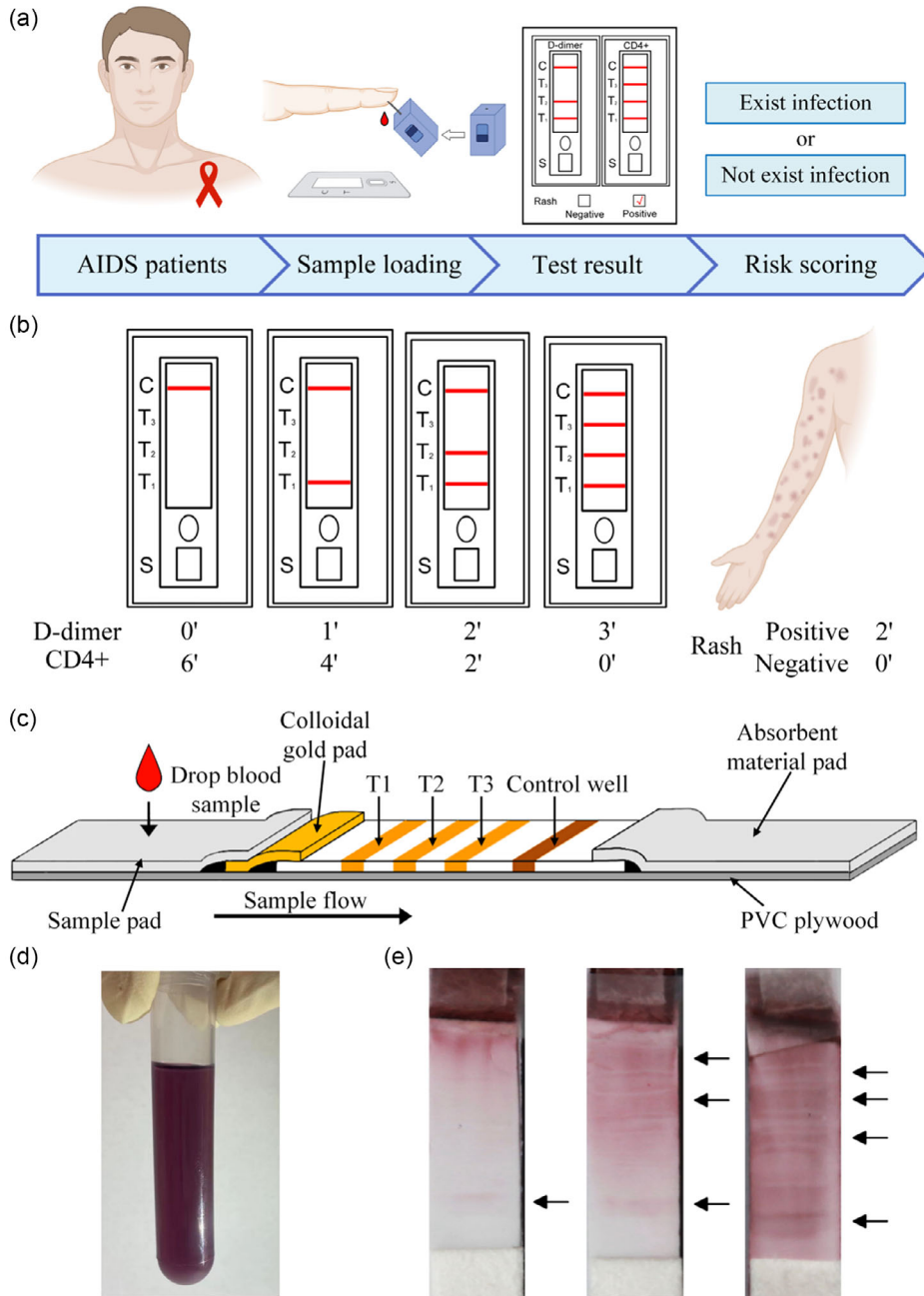
The combination of classical statistics, machine learning algorithms, and the “from weak to strong” inductive process resulted in findings that highlight D-dimer, CD4 + T-cell count, and rash

as the most significant features in both development and validation sets. These three features received weight contributions of 68.8%, 28.8%, and 27.7% from standardized, PI, and SHAP evaluation methods, respectively. Ultimately, the simplest scoring system attained 0.91, 0.80, and 0.69 AUROC in hold-out cohort, external cohort 1, and external cohort 2, thus corroborating the reliability and stability of our feature selection approach.

CD4 cells perform a pivotal role within the human immune system. Gradual or irregular decreases in CD4 cells among HIV-infected individuals denote serious immune system impairment.<sup>[21]</sup> When CD4 cell counts fall below  $200 \text{ cells mm}^{-3}$ , numerous opportunistic infections or tumors can arise, particularly in HIV-infected patients with CD4 cell counts below  $100 \text{ cells mm}^{-3}$ , where *T. marneffeii* activation becomes highly probable.<sup>[20]</sup>

In addition, *T. marneffeii* frequently compromises blood vessels, creating fungal emboli, and disseminating throughout the body. D-dimer, a product of cross-linked fibrin clot





**Figure 6.** POCT test strips experimental group visualization. a) Flowchart of the procedure for using the POCT strip. b) Method for calculating the POCT strip scores. When only a distinct purple band appears on the control line (C line), D-dimer can be read as 0 points, and CD4 can be read as 6 points. When both the control line (C line) and the test line (T line) display distinct purple bands simultaneously, D-dimer can be read as 1 point, and CD4 can be read as 4 points. Proceeding in this manner, the scores for D-dimer and CD4 can be determined. The presence or absence of the rash will be adjudged as 2 points and 0 points, respectively. c) Schematic diagram of the POCT strip. d) A sample of colloidal gold solution. e) Results of the POCT strip. We present the colloidal gold test strip samples produced, categorized as: one control line (C line); one control line (C line) and two test lines (T lines); one control line (C line) and three test lines (T lines).

dissolution by plasmin, primarily reflects fibrinolytic function. It can assist in predicting thrombosis, infection, tissue necrosis, and AIDS-related complications, and to a certain degree, disease progression and mortality rates.<sup>[28,32]</sup> Simultaneously, *T. marneffei* infection can provoke inflammatory responses in the skin and soft tissues, resulting in the development of rashes. Rashes often

present as erythema, papules, blisters, or ulcers and may be accompanied by itching or pain. In cases of AIDS coinfecting with *T. marneffei*, the presence of a rash may suggest localized infection dissemination and deep tissue involvement. Therefore, monitoring rash occurrence is vital for screening AIDS patients coinfecting with *T. marneffei*.<sup>[21]</sup>



**Figure 7.** Web and app development result diagram. a) The workflow of creating the web and app page. b) Display of our free web platform. Building upon our preliminary infection screening model, we have devised a complimentary web-based platform to facilitate early detection of *Talaromyces marneffei* infection among medical professionals and patients alike. c) Display of our free app platform. To overcome constraints related to time, location, and equipment, our team has developed a versatile mobile application. This application is specifically designed for seamless integration and use within the Android operating system.

Based on this feature consensus, to circumvent complex model computations, the study transforms the features into a clinical scoring system, enabling calculation by healthcare professionals and promoting early screening indicator dissemination. However, as the serum markers incorporated in this model necessitate laboratory testing, the required facilities and testing times preclude rapid home testing for HIV patients. By determining D-dimer and CD4 cell concentrations, we convert existing coronavirus POCT test strips from qualitative to semiquantitative testing. A scoring model directly related to semiquantitative testing estimates the risk of *T. marneffei* infection in HIV patients based on test outcomes. POCT test strip-displayed scores can indicate whether an individual is part of a high-risk group for *T. marneffei* infection. This semiquantitative POCT test strip, in conjunction with clinical scoring incorporating serological markers (e.g., Alvarado score, International Prognostic Index (IPI), Systemic Inflammatory Response Syndrome (SIRS) score, etc.), enables swift disease risk

assessment without dependence on laboratory testing. This innovative development holds the potential for extension to other studies.

Despite the promising results of this study, several limitations must be acknowledged. The model, founded on a retrospective cohort, specifically targets HIV-infected patients in China, potentially resulting in variations in disease prevalence and clinical characteristics in other countries. Further research is required to establish the generalizability of our findings to diverse populations and to validate the model's performance in a prospective setting.

In conclusion, our research presents a novel risk assessment system, developed through a feature selection strategy termed "from weak to strong" deduction. The performance and generalizability of our model were evaluated using multicenter cohorts, revealing promising outcomes for the early diagnosis of opportunistic *Talaromyces marneffei* infections in AIDS patients. By incorporating logistic regression and machine

learning models, our methodology streamlines biomarker screening “from weak to strong” and facilitates the development of a composite biomarker scoring system. Ultimately, to foster community outreach, we efficiently integrated the clinical scoring system into semiquantitative test strips, offering expansive potential applications across a range of clinical assessment domains. If the scores comprise segmented serological biomarkers, such test strips can be produced.

#### 4. Experimental Section

**Ethical Principles and Group Access Criteria:** The current investigation adheres to the stipulations of the Declaration of Helsinki, obtaining ethical committee consent from Wenzhou Central Hospital (L2021-03-082), Hangzhou Xixi Hospital (2020 Ethics Approval No. 34), and the First Affiliated Hospital of Zhejiang University School of Medicine (2021-II T-599). All participating investigators signed agreements to maintain the confidentiality of medical information, thereby ensuring patient privacy. Personal data were anonymized through established deidentification protocols. The study’s flowchart, detailed in the supplementary materials, demarcates the inclusion and exclusion criteria for participant selection and provides a comprehensive guide for the research process.

**Diagnostic Criteria and Clinical Data Acquisition:** Diagnoses of AIDS were in strict accordance with the 2021 edition of the World Health Organization (WHO) AIDS Diagnosis and Treatment Guidelines. Classifications and diagnoses of HIV opportunistic infections were performed following the recommendations of the Centers for Disease Control and Prevention (CDC). Patients in each cohort were stratified based on the presence or absence of a confirmed *Talaromyces marneffe* infection. The experimental group consisted of individuals with positive *Talaromyces marneffe* culture results. The standards for data management are outlined in the supplementary materials. Clinical data were systematically collected from the inpatient diagnosis and treatment system databases of each participating medical center.

A standard numerical form was designed to consolidate data regarding patient characteristics, including demographic factors (age and sex), clinical manifestations (fever, cough, rash, gastrointestinal hemorrhage), laboratory parameters, and radiological features. The data, collected from each medical center, were systematically entered into the standard form.

**Statistical Analysis and Standardized Logistic Regression Coefficients:** Data processing was conducted using Microsoft Excel 2019. To preserve data integrity, missing values for continuous variables in both positive and negative patients in the development cohort were filled using imputation via the series mean method in Statistical Package for the Social Sciences (SPSS) (version 24.0, IBM). The compositional differences of variables between the experimental and control groups were investigated using the Mann–Whitney *U* test for continuous variables and the Pearson chi-squared test for categorical variables. Univariate and multivariate logistic regression analyses were executed in SPSS, utilizing forward selection for feature selection and the likelihood ratio test for model fit assessment. Two-tailed *p*-values were computed, with a statistical significance threshold of 0.05. Z-standardization was implemented to normalize discrepancies in dimensions and scales, thereby facilitating the comparison of diverse variables. The standardized regression coefficients, or beta coefficients, were generated by the logistic regression equation.

**Machine Learning Models and Feature Importance Ranking Methodologies:** To address potential bias from dataset-specific characteristics during feature selection with a single machine learning algorithm, 12 representative classifiers were integrated. Additionally, two external evaluation methods for assessing feature importance were adopted, thus providing a balanced evaluation of different machine learning selection outcomes.

The 12 classifiers comprised AdaBoost Classifier, Gradient Boosting Classifier, Random Forest Classifier, Decision Tree Classifier, Extra Tree Classifier, Stochastic Gradient Descent Classifier (SGDC), Passive Aggressive Classifier (PAC), Perceptron, Support Vector Classifier

(SVC), Linear Support Vector Classifier (Linear-SVC), Bernoulli Naive Bayes Classifier, and Ridge Classifier. All algorithms were implemented using the Scikit-learn package (version 1.2.0) in Python 3.9.13. A five-fold cross-validation technique was applied to assess the model’s performance. This cross-validation procedure was conducted 50 times to ensure robust estimates of model performance, with the random seed set to 1. A grid search method was employed to determine the optimal hyperparameter combination for the model. The area under the receiver operating characteristic curve (AUROC) of the cross-validated model served as the objective for hyperparameter optimization.

Two external feature importance ranking methods were employed: Shapley additive explanations (SHAP) and permutation importance (PI). SHAP, a method grounded in cooperative game theory, provided a measure of a feature’s relative influence over the entire prediction process, represented by the mean SHAP value. Conversely, PI measured the contribution of a feature to the model prediction error by randomly permuting one feature while maintaining the order of all others.

To address the issue of differing scales across machine learning models, MinMaxScaler was applied to the SHAP and PI outputs of all models, constraining the range for each feature between 0 and 100. Scores for each feature across the models were subsequently computed and accumulated to derive an overall ranking for each feature. However, all information sources in this process originated solely from the development cohort and did not improve the model’s performance on other cohort datasets. Therefore, a calibration method employing the AUROC performance of an independent external validation dataset (Xixi Hospital) was proposed. The original feature scores were multiplied by the AUROC index of the external dataset to derive the adjusted scores. This method uses the external dataset’s AUROC to calibrate the model weights derived from the development cohort, thereby enhancing the generalizability of the final model.

**From Weak to Strong Deduction and Feature Selection:** To circumvent limitations and biases associated with feature selection via a single data-fitting method, a “From weak to strong” deduction-based feature selection strategy was devised, ensuring the stability of the final feature selection results across different fitting methods and data distributions. This strategy integrated feature selection outcomes from three independent derivation methods to obtain highly reliable and generalizable representative features. Initially, standardized regression coefficients were employed to rank feature importance within the logistic models. Subsequently, PI was utilized to rank feature importance within the machine learning models. Lastly, SHAP was employed to rank feature importance within the machine learning models. The ranking outcomes from these three methods were analyzed, deriving higher levels of evidence, and informing feature selection.

**Simplified Model Construction:** A streamlined clinical scoring model was devised based on the stable features identified through the selection process. To generate scores, each feature’s beta coefficient value was divided by the smallest constant and rounded to the nearest whole number. This scoring methodology was then employed to predict the risk of *Talaromyces marneffe* infection and validated within an external validation cohort at the First Affiliated Hospital of Zhejiang University.

**Chemicals and Materials:** The absorbent pad used is an AN3 absorbent pad (0.6 mm thick), and both the sample pad and gold label pad consist of glass fiber membrane number 8,975 (0.5 mm thick). These materials were procured from Hangzhou Bulus Trading Co., Ltd (www.bulus.com.cn). The nitrocellulose membrane used is a pure nitrocellulose blotting membrane P/N 66 485 (0.22 μm thick), purchased from Nanyang Ruitai Biotechnology Co., Ltd.

**Colloidal Gold Solution Preparation:** An aqueous solution of chloroauric acid (HAuCl<sub>4</sub>) at 0.01% concentration was prepared using 200 mL of water, heated to boiling, and maintained for 3 min. An equimolar amount of 1% trisodium citrate was rapidly added to the solution, and the mixture was heated and stirred for an additional 30 min. Upon turning wine-red, the solution was purified, concentrated, and the gold nanoparticle solution was obtained through multiple centrifugation steps.

**Colloidal Gold Modification with Antibodies:** The colloidal gold solution was adjusted to pH 9.0 using a 0.2 mol L<sup>-1</sup> Na<sub>2</sub>CO<sub>3</sub> solution, and 1 mL was transferred into a centrifuge tube. Concurrently, 0.1 mL of a

200  $\mu\text{g mL}^{-1}$  antibody solution was added to the tube, and the mixture was agitated for 40 min. Subsequently, a 6% BSA solution was added to achieve a 1% concentration, and the mixture was shaken for an additional 20 min. The solution was then centrifuged at 12 000 rpm for 10 min, the supernatant was discarded, and the concentrate was resuspended in 0.1 mL of 0.01 mol L<sup>-1</sup> Tris-HCl solution (containing 1% BSA and 1% sucrose). This resuspension produced the immunocolloidal gold solution.

**Strip Preparation:** The sample pad, gold label pad, and nitrocellulose membrane were pretreated by soaking in a solution of 1% BSA, 2% glucose, pH 7.2, 0.1 mol L<sup>-1</sup> PBS buffer for 12 h. They were then dried in a 37 °C incubator for 12 h and set aside.

The colloidal gold semiquantitative test strip components were sequentially adhered to a PVC board (7 cm × 0.5 cm): the sample pad (2 cm × 0.5 cm), the antibody colloidal gold pad (1 cm × 0.5 cm), the nitrocellulose membrane (2 cm × 0.5 cm), and the absorbent filter paper (2.5 cm × 0.5 cm).

An antigen solution was prepared by combining 50  $\mu\text{L}$  of antigen (Native Human D-Dimer protein ab35949-100ug) with 450  $\mu\text{L}$  of 1% PBS to achieve a 1% concentration of Solution A. Solution B was prepared by combining 100  $\mu\text{L}$  of Solution A with 300  $\mu\text{L}$  of 1% PBS solution to create a 1.25% concentration. T1, T2, and T3 lines were drawn on the pretreated nitrocellulose membrane using 0.1  $\mu\text{L}$  of Solution A, maintaining a 3 mm interval between each line. The quality control line was drawn using 0.1  $\mu\text{L}$  of secondary antibody (AffiniPure Rabbit Anti-Mouse IgG (H + L) (min X Hu Sr Prot) 315-005-045), positioned 3 mm after T3. After completing the lines, the membrane was dried in a 37 °C incubator for 2 h. This process was repeated three times to produce three identical test strips.

**Web Platform and Mobile Application Development:** The construction of a comprehensive web-based platform was undertaken to facilitate the utilization of the clinical scoring model. Advanced technologies and methodologies were utilized, ensuring the platform's functionality, performance, and security. The structure, esthetic, and interactivity of the platform were designed using HyperText Markup Language 5 (HTML5), CSS3, and JavaScript. Asynchronous JavaScript and XML (AJAX) techniques were implemented, enabling real-time communication between client-side and server-side components.

The management of data storage and processing was achieved through PHP and MySQL. Optimization strategies were incorporated to ensure the performance of the website. Adherence to web security best practices was maintained throughout the development process. Cross-browser compatibility testing was conducted to ensure a consistent user experience across various platforms.

Simultaneously, the development of the Clinical Scoring Model Android application was undertaken. The application was designed using Java, Android Jetpack, and the Gradle toolkit, which resulted in an efficient, visually appealing, and user-friendly interface. Components such as Responsive Design, LiveData, and ViewModel were integrated to enhance accessibility and handle user input effectively.

Communication with the back-end server was facilitated through HTTP client libraries via a RESTful API. Server-side data storage, processing, and retrieval were managed effectively using PHP and MySQL. Performance optimization strategies were implemented, and Android security best practices were adhered to throughout the development process. Compatibility testing was conducted across various Android versions and devices to ensure a consistent user experience. The culmination of this comprehensive and methodical approach was the creation of a user-friendly platform designed to accurately predict the risk of *Talaromyces mameffei* infection.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

H.C., H.Y., and F.C. contributed equally to this work. This work is supported by the STI2030-Major Projects (2021ZD0201900), the National

Natural Science Foundation of China (grant no. 12090052), the Wenzhou Key Laboratory of Diagnosis and Treatment of Emerging and Recurrent Infectious Diseases (grant no. WKJ-ZJ-2138), the Key Laboratory of Intelligent Treatment and Life Support for Critical Diseases of Zhejiang Province (grant no. 2021E10016), and the Construction and management of a sub-placement care unit under epidemic conditions (grant no. Y20220505).

## Conflict of Interest

Haoman Chen and Fanxuan Chen pursued their master's degree programs under the mentorship of Jianwei Shuai and Xiaoqu Hu. Chaoyi Wei pursued a Ph.D. program with Jianwei Shuai and Xiaoming Jiang. Hui Ye was member of Prof. Feifei Su's team. Yi Shi, Jiahe Li, and Yuhang Li were members of Jianwei Shuai and Xiaoqu Hu's team. Yungui Zhang, Zongxing Yang, Lijun Xu, and Feifei Su were research peers and collaborated on this work. Ou Liu and Xiaoming Jiang were members of Jianwei Shuai's lab.

## Author Contributions

J.S., F.S., J.P., H.C., Y.L., C.W., F.C, J. L., and Y.S. conceived the study; J.L., F.C., Y.L., Y.S., Y.W., and H.C. conducted the data science experiments; F.C. and H.C. performed the POCT strip experiments; J.S. provided guidance and methodology for the experiments. F.S., H.Y., Y.L., J.L., Y.S., F.C., and H.C. collected and reconciled the clinical data from Wenzhou Central Hospital; L.X. led the data collection and data verification team at The First Affiliated Hospital of Zhejiang University School of Medicine; Z.Y. headed the data collection and data verification team at Hangzhou Xixi Hospital; H.C., F.C., Y.S., Y.L., and J.L. drafted the manuscript; J.S., O.L., X.J., Q.H., L.L., X.H., J.P., S.Q., and D.C. directed the review team. All authors participated in the article revision process. H.C. coordinated the student team and lab collaborative work throughout the project. F.S., J.S., and J.P. ensured the quality of all work.

## Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## Keywords

AIDS opportunistic infections, feature selection strategy, machine learning, point-of-care testing, test strips, risk scoring system

Received: May 5, 2023

Revised: July 26, 2023

Published online:

- [1] V. J. McGowan, C. Bamba, *Lancet Public Health* **2022**, *7*, e966.
- [2] A. Lal, N. A. Erond, D. L. Heymann, G. Gitahi, R. Yates, *Lancet* **2021**, *397*, 61.
- [3] World Health Organization (WHO), <https://www.who.int/publications/i/item/9789240031593>.
- [4] H. Lin, J. Shuai, *New J. Phys.* **2010**, *12*, 043051.
- [5] Y. Mo, B. Ren, W. Yang, J. Shuai, *Physica A* **2014**, *399*, 31.
- [6] Y. Zhou, Y. Wu, L. Ding, X. Huang, Y. Xiong, *TrAC, Trends Anal. Chem.* **2021**, *145*, 116452.
- [7] H. Young, A. Moyes, L. Seagar, A. McMillan, *J. Clin. Microbiol.* **1998**, *36*, 913.

- [8] C. Cao, L. Xi, V. Chaturvedi, *Mycopathologia* **2019**, *184*, 709.
- [9] B. Yang, J. Wang, H. Jiang, H. Lin, Z. Ou, A. Ullah, Y. Hua, J. Chen, X. Lin, X. Hu, *Front. Microbiol.* **2021**, *11*, 603183.
- [10] T. Le, M. Wolbers, N. H. Chi, V. M. Quang, N. T. Chinh, N. P. Huong Lan, P. S. Lam, M. J. Kozal, C. M. Shikuma, J. N. Day, *Clin. Infect. Dis.* **2011**, *52*, 945.
- [11] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, L. G. Thijs, *Intensive Care Med.* **1996**, *22*, 707.
- [12] K.-H. Yu, A. L. Beam, I. S. Kohane, *Nat. Biomed. Eng.* **2018**, *2*, 719.
- [13] Z. Obermeyer, E. J. Emanuel, *N. Engl. J. Med.* **2016**, *375*, 1216.
- [14] J. Zhao, J. Sun, S. C. Shuai, Q. Zhao, J. Shuai, *Briefings Bioinf.* **2023**, *24*, bbac527.
- [15] W. Wang, L. Zhang, J. Sun, Q. Zhao, J. Shuai, *Briefings Bioinf.* **2022**, *23*, bbac463.
- [16] X. Li, P. Zhang, Z. Yin, F. Xu, Z.-H. Yang, J. Jin, J. Qu, Z. Liu, H. Qi, C. Yao, *Research* **2022**, *2022*, 9838341.
- [17] T. Chen, X. Li, Y. Li, E. Xia, Y. Qin, S. Liang, F. Xu, D. Liang, C. Zeng, Z. Liu, *Am. J. Kidney Dis.* **2019**, *74*, 300.
- [18] J. Zhang, J. Xu, X. Hu, Q. Chen, L. Tu, J. Huang, J. Cui, *BioMed Res. Int.* **2017**, *2017*, 7961494.
- [19] B. Kuswandi, *Nanomaterials in Diagnostic Tools and Devices*, Elsevier Publishing Group, Amsterdam, The Netherlands **2020**, pp. 439–477.
- [20] J. Sun, W. Sun, Y. Tang, R. Zhang, L. Liu, Y. Shen, J. Wang, J. Chen, T. Qi, Z. Wang, *BMC Infect. Dis.* **2021**, *21*, 514.
- [21] A. H. Limper, A. Adenis, T. Le, T. S. Harrison, *Lancet Infect. Dis.* **2017**, *17*, e334.
- [22] E. Leeansyah, D. F. Malone, D. D. Anthony, J. K. Sandberg, *Curr. Opin. HIV AIDS* **2013**, *8*, 117.
- [23] E. Hamlyn, S. Fidler, W. Stöhr, D. A. Cooper, G. Tambussi, M. Schechter, J. M. Miro, M. McClure, J. Weber, A. Babiker, *AIDS* **2014**, *28*, 869.
- [24] A. Sil, A. Andrianopoulos, *Cold Spring Harbor Perspect. Med.* **2015**, *5*, a019794.
- [25] G. Handelman, H. Kok, R. Chandra, A. Razavi, M. Lee, H. Asadi, *J. Intern. Med.* **2018**, *284*, 603.
- [26] M. M. Engelhard, J. A. Oliver, R. Henao, M. Hallyburton, L. E. Carin, C. Conklin, F. J. McClernon, *JAMA Netw. Open* **2019**, *2*, e197939.
- [27] M. Fernandes, S. M. Vieira, F. Leite, C. Palos, S. Finkelstein, J. M. Sousa, *Artif. Intell. Med.* **2020**, *102*, 101762.
- [28] J. S. Cramer, Tinbergen Institute Working Paper No. 2002-119/4, **2002**, <https://doi.org/10.2139/ssrn.360300>.
- [29] Y. Nohara, K. Matsumoto, H. Soejima, N. Nakashima, *Comput. Methods Prog. Biomed.* **2022**, *214*, 106584.
- [30] S. M. Lundberg, S.-I. Lee, *Advances in Neural Information Processing Systems*, MIT Press, Massachusetts, USA **2017**, p. 30.
- [31] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, *Bioinformatics* **2010**, *26*, 1340.
- [32] M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255.