

Original Paper

Tongue Image–Based Diagnosis of Acute Respiratory Tract Infection Using Machine Learning: Algorithm Development and Validation

Qianzi Che^{1*}, PhD; Yuanming Leng^{2*}, MS; Wei Yang^{1*}, PhD; Xihao Cao³, MS; Zhongxia Wang⁴, PhD; Lizheng Liu⁵, PhD; Feibiao Xie⁶, PhD; Ruilin Wang⁴, PhD

¹Institute of Clinical Basic Medicine of Chinese Medicine, Chinese Academy of Traditional Chinese Medicine, Beijing, China

²Department of Biostatistics, School of Public Health, Boston University, Boston, MA, United States

³Department of Mathematics and Statistics, Boston University, Boston, MA, United States

⁴Department of Traditional Chinese Medicine for Liver Diseases, Fifth Medical Center of the Chinese People's Liberation Army General Hospital, Beijing, China

⁵Institute of Engineering and Applied Technology, Fudan University, Shanghai, China

⁶School of Mathematics and Statistics, Central South University, Hunan, China

*these authors contributed equally

Corresponding Author:

Ruilin Wang, PhD

Department of Traditional Chinese Medicine for Liver Diseases

Fifth Medical Center of the Chinese People's Liberation Army General Hospital

No. 100 West Fourth Ring Middle Road, Fengtai District

Beijing 1000039

China

Phone: 1 13811050593

Email: WRL7905@163.com

Abstract

Background: Human adenoviruses (HAdVs) and COVID-19 are prominent respiratory pathogens with overlapping clinical presentations, including fever, cough, and sore throat, posing significant diagnostic challenges without viral testing. Tongue image diagnosis, a noninvasive method used in traditional Chinese medicine, has shown correlations with specific respiratory infections, but its application remains underexplored in differentiating HAdVs from COVID-19. Advances in artificial intelligence offer opportunities to enhance tongue image analysis for more objective and accurate diagnostics.

Objective: This study aims to develop and validate artificial intelligence–based predictive models using tongue image features to differentiate COVID-19 from adenoviral respiratory infections, thereby improving diagnostic accuracy and integrating traditional diagnostic methods with modern medical technologies.

Methods: A total of 280 tongue images were collected from 58 patients with COVID-19, 84 patients with HAdVs, and 30 healthy controls. Deep learning methods were applied to extract tongue features, including color, coating, fissures, papillae, tooth marks, and granules. Four machine learning classifiers, logistic regression, random forest, gradient boosting model, and extreme gradient boosting, were developed to differentiate COVID-19 and HAdV infections. The key features identified by the machine learning algorithms were further visualized in a 2D space.

Results: Nine tongue features showed significant differences among groups (all $P < .05$), including coating color (red, green, and blue), presence of tooth marks, coating crack ratio, moisture level, texture directionality, roughness, and contrast. The extreme gradient boosting model achieved the highest diagnostic performance with an area under the receiver operating characteristic curve of 0.84 (95% CI 0.78-0.90) and an area under the precision-recall curve above 0.70. Shapley additive explanations analysis indicated tongue color, moisture, and texture as key contributors.

Conclusions: Our findings demonstrate the potential of tongue diagnosis in identifying pathogens responsible for acute respiratory tract infections at the time of admission. This approach holds significant clinical implications, offering the potential to reduce clinician workloads while improving diagnostic accuracy and the overall quality of medical care.

Keywords: tongue diagnosis; human adenoviruses; COVID-19; image feature extraction; machine learning; traditional Chinese medicine

Introduction

Human adenoviruses (HAdVs) are prominent respiratory pathogens affecting individuals of all ages, leading to acute upper and lower respiratory tract diseases, such as pneumonia and bronchitis [1]. Nearly half (45.7%) of acute respiratory tract infection (ARTI) outbreaks between 2009 and 2020 were attributed to HAdV-7 in China [2]. Since 2020, a temporary decline in HAdVs activity has been noted due to the rapid global spread of COVID-19 and subsequent public health measures [3,4]. As COVID-19 persists in its global circulation, a resurgence of HAdVs and other seasonal respiratory virus infections has occurred [5]. The clinical presentations of ARTI caused by HAdVs and COVID-19 are similar, often presenting as mild symptoms such as fever, rhinorrhea, cough, and sore throat, which poses a significant diagnostic challenge in the absence of viral testing. HAdVs are more prone than COVID-19 to causing acute respiratory infections in children, including pharyngitis, tonsillitis, pharyngoconjunctival fever, bronchitis, and pneumonia [6]. Moreover, recent studies indicate that mixed respiratory viral infections may lead to more severe disease outcomes than single infections [7]. Given the tendency of many patients with ARTI to self-medicate, it is essential to develop methods to differentiate between these common viral infections. Such differentiation could promote timely medical consultation and help reduce the transmission of adenovirus to children.

Tongue image diagnosis is a straightforward, noninvasive, and valuable diagnostic method used in traditional Chinese medicine (TCM), which assesses key features, such as the color, size, and shape of the tongue, along with the color, thickness, and moisture level of the tongue coating [8]. In respiratory infections, TCM theory associates specific tongue coatings with distinct syndromes [9]. Epidemiological studies have also identified distinctive tongue features in patients with acute respiratory infections. Patients with COVID-19 have been reported to present features, such as fissured tongue and strawberry tongue [10-13], whereas tongue signs in HAdV infections remain scarcely documented, limiting

comprehensive understanding. Given the convenience and cost-effectiveness of tongue imaging, its application for differentiating COVID-19 from HAdV infections shows promising potential.

Tongue diagnosis traditionally relies heavily on the experience and observational skills of TCM practitioners to interpret tongue features. However, advances in artificial intelligence (AI) technologies have enabled the extraction of tongue image characteristics, such as texture, color, and coating, facilitating more objective and intelligent tongue diagnosis. Recent studies have applied AI-assisted tongue image analysis in various diseases, including diabetes [14], fatty liver [15], non-small cell lung cancer [16], and gastric cancer [17], combining tongue image features with clinical indicators to develop exploratory disease risk prediction models. Similarly, Zhou et al [18] proposed an automatic multiview disease detection system using tongue images, achieving an average classification accuracy exceeding 95% for breast tumors, heart disease, fatty liver, and lung tumors. Given the limited research on AI-assisted tongue diagnosis for respiratory infections, there is a clear need to explore its potential in this domain. Therefore, this study aims to investigate the application of AI-based tongue image analysis to distinguish COVID-19 from adenoviral respiratory infections, with the goal of improving diagnostic accuracy and integrating traditional diagnostic methods with contemporary medical technologies.

Methods

Study Sample

Our study was conducted between November 2020 and January 2021 at the Department of Traditional Chinese Medicine Hepatology, which is in the Fifth Medical Center of the Chinese People's Liberation Army (PLA) General Hospital. The inclusion and exclusion criteria have been listed in [Textbox 1](#)

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Participants whose age is between 18 and 80 years old.
- Participants willing to participate in tongue image photography signed an informed consent form.

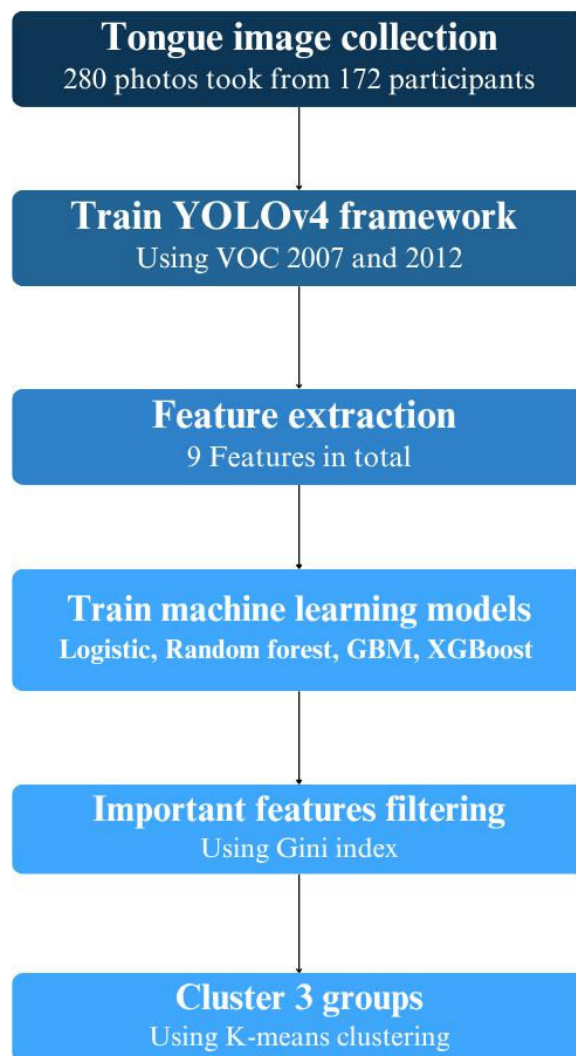
Exclusion criteria

- Individuals who were unable to describe their condition clearly due to mental factors or could not cooperate with the collection of tongue diagnosis images.
- Participants with severe acute complications, such as serious electrolyte imbalance and acidosis; patients with other serious internal diseases, such as tumors, immune system disorders, or hematologic diseases; and individuals taking medications, such as steroids, that affect glucose metabolism.
- Pregnant women and nursing mothers were excluded.

After applying the inclusion and exclusion criteria, 172 participants were included in the study. Of these, 58 were diagnosed with COVID-19, 84 with adenovirus, and 30 were assigned to the control group. Meanwhile, all the patients

had complete clinical data and demonstrated high compliance, while control group participants had no history of chronic or acute diseases in the past 3 months (Figure 1).

Figure 1. The flow diagram of the participant screening procedure. GBM: gradient boosting machines; XGBoost: extreme gradient boosting.



Outcome Definition

A confirmed case of COVID-19 is defined as a suspected case that yields a positive result on a real-time reverse-transcription polymerase chain reaction assay using respiratory specimens [19]. In this study, all patients with adenovirus infection met the diagnostic criteria outlined in the “Adenovirus Infection Diagnosis and Treatment Guidelines” issued in 2013 [20]. According to these guidelines, the criteria include (1) a real-time polymerase chain reaction test on throat swab samples that detects adenovirus-specific nucleic acids; (2) the presence of adenovirus-specific immunoglobulin M antibodies in the serum; and (3) a 4-fold or greater increase in adenovirus-specific immunoglobulin G antibodies in paired serum samples collected during the acute and recovery phases. The selection of diagnostic tests was based on clinical judgment. These tests were conducted on various specimen types, including nasopharyngeal swabs,

throat swabs, sputum samples, pleural effusion samples, and bronchoalveolar lavage fluid samples.

Tongue Image Collection

All tongue images were captured before the laboratory confirmation. Collecting tongue images was conducted either before a meal or 2 hours afterward. Before being photographed, participants were instructed not to eat any food or drink any colored beverages. During the examination, participants were asked to sit with their mouths open and their tongues extended, ensuring the tongue body was relaxed, the surface was flat, and the tip was drooping. A color correction card was held within the camera’s view during image capture to prevent external lighting from affecting the quality of the photograph. If a retake was necessary, the participants were advised to rest for 3-5 minutes before recollecting the image. The final choice for storage was the best-quality image. A digital diagnostic system was used to collect and analyze all tongue images.

Tongue Image–Based AI Deep Learning Models

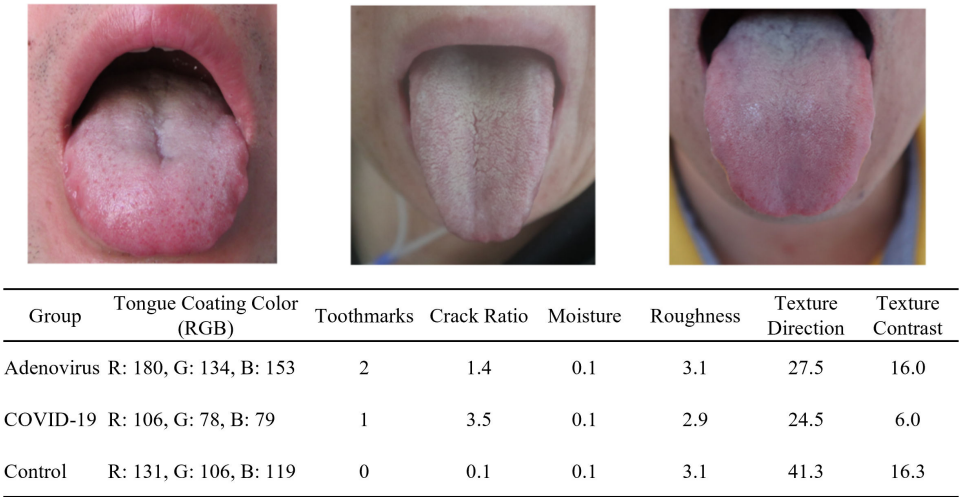
In this study, 4 key factors were primarily selected for detection: the tongue body, the coating, the teeth mark, and the fissures. To develop a target detection algorithm for tongue diagnosis images, the YOLOv4 object detection framework was used [21]. Through extensive experimentation with this algorithm, we used cross-validation to tune the convolutional neural network parameters and found the optimal combination of parameters [22]. We also increased the sample size and improved the model’s generalization using image preprocessing techniques, including random cropping, horizontal flipping, and color distortion [23]. This preprocessing yielded sample images with various tongue characteristics, including color, coating, fissures, prickles, tooth marks, and granules, categorized under data labels 1-20 (Multimedia Appendices 1 and 2). Expert TCM practitioners have reviewed all the extracted features to ensure that they possess clinical relevance and significance.

We used VOC 2007 and 2012 to pretrain the initial feature extractor due to its availability and general utility in learning low-level visual features. The training batch size was 32 images, and 2 NVIDIA 1080 8G GPUs were used. Each group was segmented individually for different tongue features, including tongue coating (color, shape, and moisture), fissures (cracks ratio), prickles, tooth marks, and granules (shape and reflectivity).

Tongue Image Feature Definition

The color of the tongue coating was calculated by taking the mean red, green, and blue values within a specified area, providing insights into the tongue’s health condition. Tooth marks were identified using the object detection algorithm, with a confidence threshold set at 0.6 to determine the presence and count of tooth marks. The proportion of cracks in the tongue coating was calculated by segmenting and merging areas where multiple fine cracks intersect, thus determining the area occupied by these cracks relative to the total area of the tongue coating. The moisture level of the tongue coating was assessed by identifying reflective areas where brightness exceeds a threshold value of 170, helping to evaluate the hydration status of the tongue. The texture features of the tongue coating were extracted using Tamura texture features, which focus on describing the textural characteristics of the tongue coating in terms of coarseness, contrast, directionality, and roughness [24]. In this study, coarseness refers to texture coarseness and is calculated based on the size of the elements. A larger element size or fewer repetitions of the element indicate a coarser texture. Contrast is primarily derived from the image’s grayscale, while directionality refers to the degree of alignment within the elements (Figure 2).

Figure 2. Sample tongue images from the adenovirus (left), COVID-19 (middle), and control group (right) with extracted features.



Statistical Analysis

We obtained 280 tongue images from these participants, each contributing up to 7 images. Only images taken on the baseline day were retained for this study, and we took the average value of the extracted features in the images if participants had more than one image on the baseline day. There were 21 images with quality issues, including 11 unidentifiable images (all characteristic values were 0) and 10 unclear images (most characteristic values were 0). We applied a mean imputation technique to these images, averaging the image parameters based on outcome category

and symptom severity [25]. We conducted comparative analyses across 3 groups to examine variations in tongue features, including color, texture, moisture, and morphology.

The data were divided into 2 subsets, with 70% allocated for training and 30% reserved for internal validation, to address potential overfitting concerns. Advanced machine learning models were used to analyze tongue images and classify them into 3 categories: COVID-19, adenovirus, and control. We used several models in our analysis, including gradient boosting machines (GBM), random forest, logistic regression, and extreme gradient boosting (XGBoost). To fit each model, we applied 5-fold cross-validation on the

training set. We evaluated the models’ performance using 2 metrics: the area under the curve (AUC) and the area under the precision-recall curve (AUCPR) [26].

An importance variable rank was computed for each algorithm to identify the variables with the highest predictive power. In addition, we created Shapley additive explanations (SHAP) values that were calculated to indicate the contributing direction of the tongue image features [27]. Finally, we projected the selected variables of tongue images into a 2D space by a t-distributed stochastic neighbor embedding (t-SNE) plot to visually examine each group’s clustering distribution [28].

All statistical analyses were performed using R Studio version 4.2.0 (Posit, PBC) software. An independent samples 2-tailed *t* test was used to analyze the measurement data, such as age and identified tongue image features. The count data were analyzed using the chi-square test. Results with *P*<.05 were considered statistically significant.

Ethical Considerations

The Ethics Committee of the Fifth Medical Center of the Chinese People’s Liberation Army (PLA) General Hospital (2020074D) gave ethical approval on November 24, 2020.

Informed consent was waived by the committee, as the study involved retrospective analysis of anonymized data and did not include any identifiable personal information. All procedures were conducted in accordance with the ethical standards of the institutional and national research committees and with the principles outlined in the Declaration of Helsinki. We ensured that the privacy and confidentiality of all participants were strictly maintained; data were fully anonymized before analysis, and no individual-level identifiers were collected, stored, or reported. No compensation was provided to participants, as the study did not involve direct interaction with human participants.

Results

Characteristics of Patients With Each Pathogen

Our research involved 172 participants; 58 had COVID-19, 84 were diagnosed with adenovirus, and 30 were in the control group. Table 1 shows the main descriptive sociodemographic variables (sex and age). The average age in our sample was 23.65 (SD 9.61) years old (Table 1).

Table 1. Study sample and extracted feature characteristics.

Variables	Control (n=30)	COVID-19 (n=58)	Adenovirus (n=84)	<i>P</i> value
Gender				
male, n (%)	13 (43)	36 (62)	83 (99)	<.001
Age (years), mean (SD)	27.2 (8.0)	47.6(15.8)	23.7 (9.6)	<.001
Tongue coating, mean (SD)				
Red	119.0 (16.5)	109.8 (22.6)	149.2 (20.4)	<.001
Green	93.3 (13.8)	77.65 (21.2)	110.8 (20.7)	<.001
Blue	97.1 (19.7)	79.2 (22.0)	114.2 (25.4)	<.001
Toothmarks, n (%)				.049
0	20 (67)	20 (35)	33 (39)	
1	6 (20)	17 (29)	24 (29)	
2	4 (13)	21 (36)	27 (32)	
Crack ratio, mean (SD)	0.5 (1.0)	1.4 (2.4)	1.2 (2.0)	.09
Moisture, mean (SD)	0.1 (0.1)	0.1 (0.0)	0.1 (0.1)	<.001
Roughness, mean (SD)	3.2 (0.2)	3.1 (0.2)	3.3 (0.2)	<.001
Texture direction, mean (SD)	40.7 (19.1)	29.9 (17.7)	34.8 (19.1)	.04
Texture contrast, mean (SD)	17.1 (9.4)	12.5 (5.8)	24.6 (9.9)	<.001

Tongue Image Feature Extraction and Definition

We used transfer learning techniques using weights obtained from training on the VOC 2007 and VOC 2012 datasets, which enabled the model to improve its understanding of general object detection. After 60,000 iterations with a learning rate of 0.0001, the average loss was reduced to 0.2867 (SD 0.0462; Multimedia Appendix 3). Compared to the proposed fast recurrent neural network and faster region-based convolutional neural network, the Single Shot MultiBox Detector architecture demonstrates higher accuracy

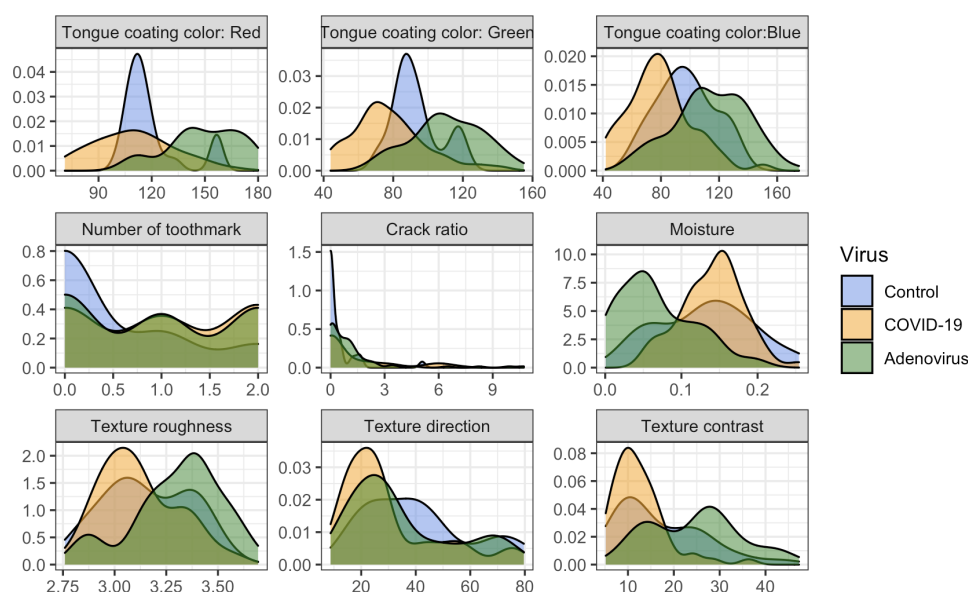
across 20 labeled datasets (Multimedia Appendix 4). After applying our feature extraction architecture, 9 features of tongue images were identified: tongue coating color values (red, green, and blue), the presence of tooth marks, tongue coating crack ratio, tongue coating moisture level, texture directionality, texture roughness, and texture contrast (Multimedia Appendix 5).

All tongue coating color values (red, green, and blue) differed significantly across groups (*P*<.001) (Table 1). Patients with adenovirus had the highest values, on average, especially in the red component, which may indicate more pronounced inflammation or other specific clinical features

related to adenovirus infections. The presence of tooth marks varied across groups, with the highest prevalence of 2 tooth marks in the COVID-19 group, with 21/58 (36%) participants showing this feature ($P=.049$). While the differences may not be statistically significant, both the COVID-19 and adenovirus groups exhibit a notably higher percentage of individuals with 2 tooth marks: 21/58 (36%) participants in the COVID-19 group and 27/84 (32%) participants in the adenovirus group. This is in contrast to the control group, where only 4/30 (13%) participants showed this characteristic ($P=.09$). This may be associated with other symptoms or manifestations of the illness. The moisture levels of tongue coating varied significantly among the different groups, with the adenovirus group exhibiting the lowest average moisture level ($P<.01$).

The adenovirus group shows considerably greater levels of tongue coating roughness compared to the COVID-19 and control groups ($P<.01$). In addition, both the adenovirus and COVID-19 groups exhibit significantly greater values in tongue texture direction than the control group, with the adenovirus group showing the highest mean value of 40.7 (SD 19.1). In terms of texture contrast, the COVID-19 group has the lowest average value at 12.5 (SD 5.8), while the adenovirus group reaches the highest average at 24.6 (SD 9.9), in contrast to the control group, which averages 17.1 (SD 9.4). The density plot suggested distinct patterns in tongue features among individuals with COVID-19, adenovirus, and those in the control group, demonstrating the potential of these features in differentiating respiratory viral infections (Figure 3).

Figure 3. Extracted tongue image feature density plot.



Diagnostic Performance Results

The tree-based and boosting models achieved over 70% AUCPR performance, with AUC exceeding 80% for general performance. The GBM achieved the highest AUC value of 0.888 and AUCPR value of 0.764. XGBoost came in second with an AUC value of 0.872 and an AUCPR value of 0.751, followed closely by random forest with an AUC value of 0.872 and an AUCPR value of 0.747. However, logistic regression showed a relatively worse performance, with an AUC value of 0.812 and an AUCPR value of 0.668 (Multimedia Appendix 6), and the confusion matrix of the test set is presented in Multimedia Appendix 7.

Explaining the Rationale Behind the Predicted Models

Based on SHAP values integrating 4 machine learning models, several factors contribute to diagnosing adenovirus

and COVID-19, including the color of tongue coating, moisture level, and texture direction. The feature importance plots indicate that the most significant variables are tongue color, moisture level, and texture direction (Multimedia Appendix 8). Specifically, a red tongue coating helps identify adenovirus cases, while a green tongue coating is beneficial for identifying COVID-19 cases (Figure 4). In addition, the t-SNE plot projects these variables into a 2D space and color codes each group. The distribution of patients with COVID-19 and adenovirus differs within the t-SNE space, showing some overlap, while the control group is between the 2 infectious cases (Figure 5).

Figure 4. Important features identified for distinguishing between COVID-19 (top), adenovirus (middle), and control (bottom) using SHAP. SHAP: Shapley additive explanations.

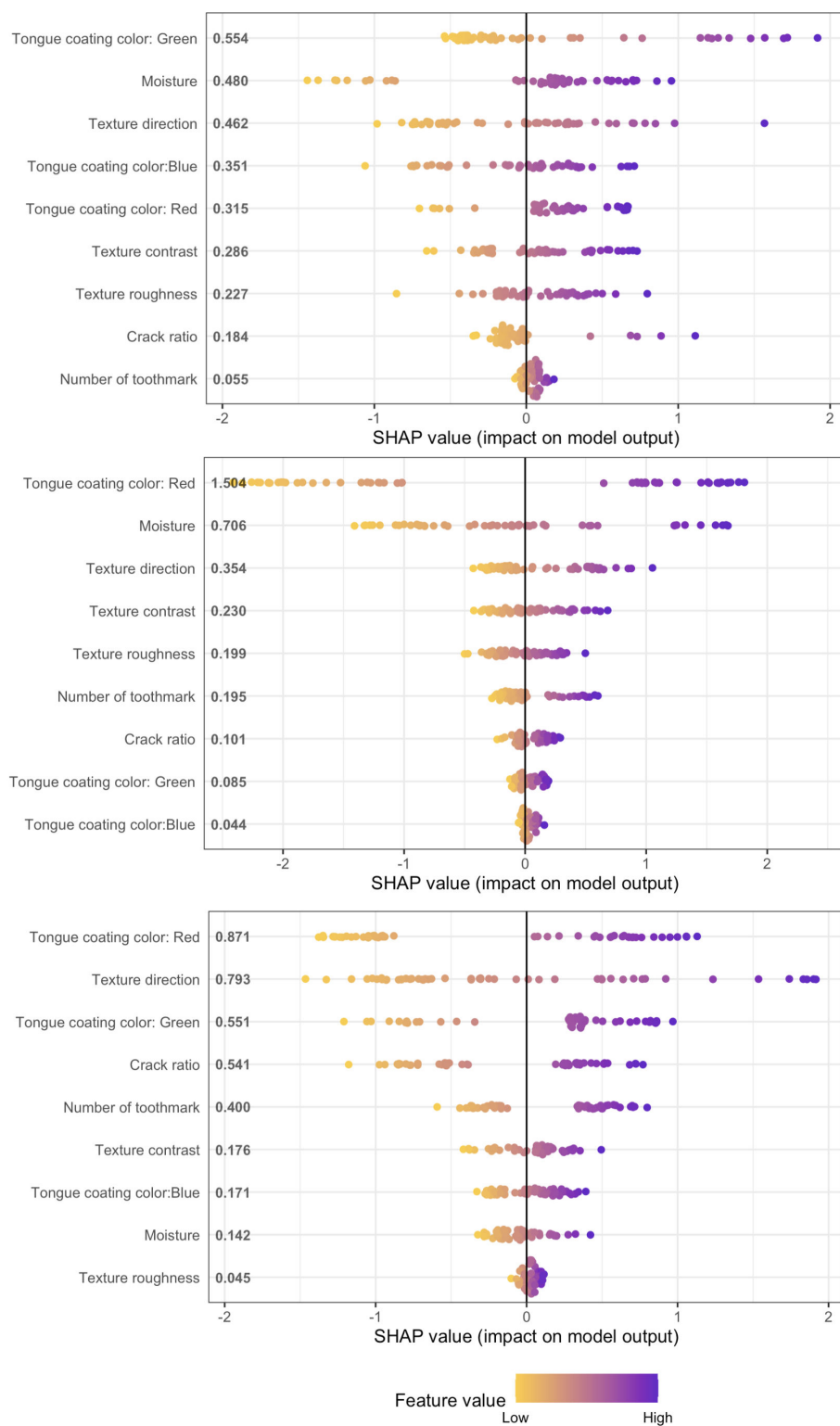
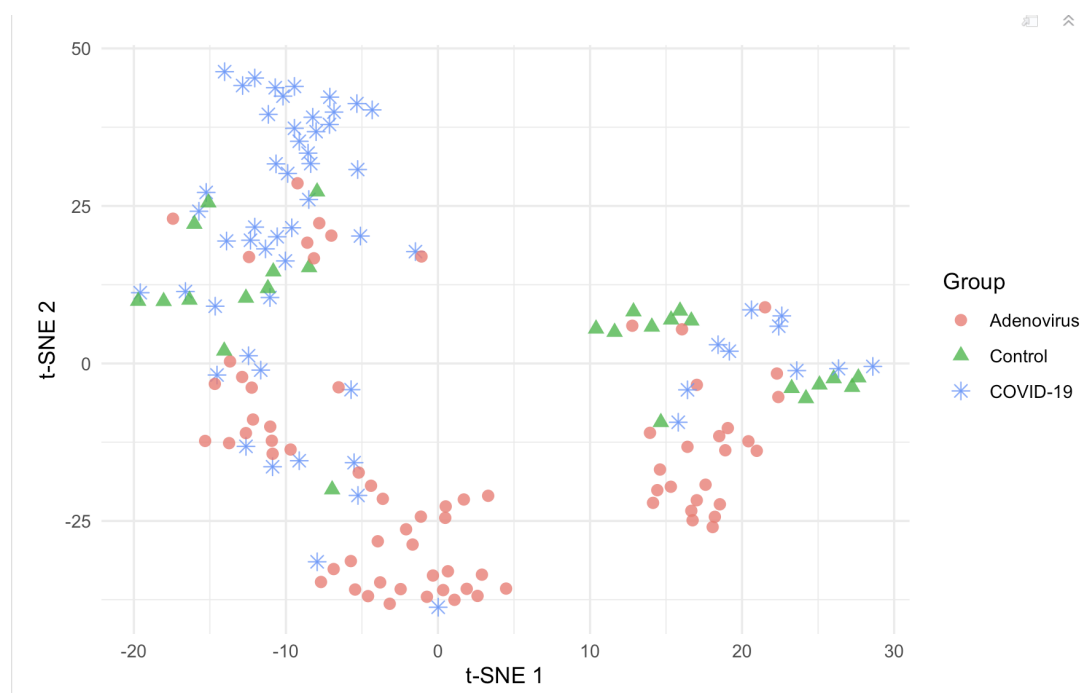


Figure 5. T-distributed stochastic neighbor embedding plot visualizing sample distribution across three groups: adenovirus (circle), control (triangle), and COVID-19 (asterisk), highlighting distinct clustering patterns among the conditions. t-SNE: t-distributed stochastic neighbor embedding.



Discussion

In this study, we developed and validated an AI-based prediction model to analyze tongue images for the differential diagnosis of respiratory infections caused by COVID-19 and human adenoviruses. Our models demonstrated high diagnostic accuracy, with key tongue image features, such as color, coating, moisture, and texture serving as significant discriminators. These findings highlight the potential of noninvasive tongue image analysis as an effective and cost-efficient diagnostic tool, which may facilitate early and accurate differentiation of respiratory viral infections and support clinical decision-making.

In recent years, the concurrent prevalence of respiratory infectious diseases, such as COVID-19 and HAdVs, has made it challenging to establish a differential diagnosis without pathogenic testing, thereby impacting the efficacy of the treatment [29]. Tongue image diagnosis has emerged as an effective, noninvasive method for auxiliary diagnosis that can be conducted in various settings, catering to the global demands of primary health care systems [30]. In this study, we developed machine learning models that use tongue images to predict acute respiratory tract infection (ARTI) pathogens. The accuracy of these models exceeded 80%, with feature importance plots revealing tongue color, moisture level, and tongue texture direction as pivotal variables. We then projected the tongue image features of the 3 groups of participants into a lower-dimensional space by t-SNE plot. The resulting clusters of COVID-19 and adenovirus were notably distinct, suggesting that the tongue image parameters we extracted contribute significantly to the diagnosis of ARIs. This noninvasive, convenient, and rapid approach has the potential to mitigate unnecessary diagnostic procedures and reduce health care costs.

The constructability of a risk warning model using tongue diagnosis data has been validated in research on metabolic diseases and cancers [16,31]. Our study is pioneering in applying AI deep learning techniques to investigate the diagnostic value of tongue images in acute respiratory tract infection diagnosis and can be integrated into clinical practice to aid decision-making and alleviate physician burden. Previous research by Mai and Krauthammer [32] aimed to predict common respiratory viruses in the United States by combining natural language processing tools with machine learning techniques. However, the model's prediction performance of HAdVs was moderate, with an area under the receiver operating characteristic curve (AUROC) of 0.53 [32]. Chang et al [33] used an XGBoost model, incorporating demographic, physical examination, laboratory, and vital sign data, to predict HAdVs in hospitalized children with respiratory symptoms, achieving an AUROC of 0.82. Our model exhibited superior performance, achieving an AUROC of 0.87, following the inclusion of various features in its development. In contrast to models established in previous studies, the model developed in this paper benefited from the usage of noninvasive tongue image data, obviating the need for laboratory examination indicators. By using the YOLOv4 object detection framework for image feature extraction and integrating ensemble learning algorithms (ie, GBM and XGBoost), we have demonstrated the clinical utility of tongue images. All 4 models exhibited discriminative validity over tongue images, indicating that tongue images can serve as a reliable tool for ARTI diagnosis and are robust across different AI deep learning model types. Notably, tree-based models, such as GBM, XGBoost, and random forest, outperform logistic regression due to their ability to capture complex nonlinear relationships and interactions among features. Tongue image-derived features often display nonlinear patterns that contradict the linear

assumptions of logistic regression. In contrast, tree-based models are nonparametric and resistant to multicollinearity, making them ideal for complex datasets derived from image analysis.

Previous research has validated the use of objective tongue image acquisition equipment, methods, and data analysis techniques [34]. Research on AI in TCM tongue diagnosis has predominantly focused on standardizing tongue diagnosis to minimize human errors [34]. Xu et al [35] developed a multitask joint learning model for segmenting and classifying tongue images, using a deep neural network to optimally extract tongue image features. Meng et al [36] proposed a novel feature extraction framework, termed constrained high dispersal neural networks, to extract unbiased features and minimize human labor in TCM tongue image diagnosis. This study used the YOLOv4 framework, which excelled in extracting high-level image features, thereby enhancing the model's capacity to identify intricate patterns and subtle variations [21]. The model was initially pretrained on the VOC 2007 and VOC 2012, which are effective for capturing general low-level visual features but lack domain-specific characteristics relevant to medical imagery, such as tongue images. To overcome the challenge of transferring features from natural datasets to specialized domains, domain-specific fine-tuning was performed using collected samples and synthetic data. Although the limited sample size may restrict the model's ability to capture complex patterns in tongue imagery, the fine-tuned feature extractor achieved approximately 95% classification accuracy across 20 defined groups, demonstrating effective adaptation. This integration, coupled with the fusion of ensemble learning algorithms, harnesses the complementary strengths of clinical and image-derived features, effectively addressing the limitations of each modality in isolation [37]. Consequently, the model gains enhanced discriminative power, leading to more accurate predictions of ARTI.

The analysis of feature importance in machine learning models (GBM, XGBoost, and random forest) revealed that a red tongue coat is most indicative of HAdVs and control group, while the green color is more suggestive of COVID-19 cases. Furthermore, statistical analysis of the tongue images of the 3 groups showed that the COVID-19 group had a higher moisture level and lower contrast rate, suggesting that the patients with COVID-19 are more likely to have a thicker and greasier tongue coating compared to the other 2 groups. Our findings align with the TCM theory. According to TCM, COVID-19 is classified as a "cold-dampness" disease, characterized by a tongue that is pale red or dark and a thick, white greasy coating. HAdVs are considered "warm" diseases in TCM, typically presenting with red and dry tongues. The thickness and dryness of the tongue coating can also indicate the severity of the disease and the extent of fluid damage. Epidemiological studies conducted in China, the United Kingdom, and Ukraine have yielded analogous findings; the predominant tongue colors observed were pale pink and dark red, with the most frequently encountered tongue coating being thin and greasy [38-40]. Therefore, a patient presenting with acute respiratory symptoms and a red

tongue devoid of a thick, greasy coating is more likely to have an HAdVs infection. Such observations could guide health care professionals in suggesting additional diagnostic tests and preventive measures, like isolation or improved hygiene, to avert transmission to susceptible household members. Considering its noninvasive nature and rapid assessment capability, AI-assisted tongue image analysis could be particularly valuable as a screening or triage tool in primary care clinics and emergency departments, where timely differentiation of respiratory infections is essential. The advancement of large language models and multimodal AI has facilitated the development of intelligent diagnostic systems that enhance tongue image interpretation and provide real-time decision support. These technologies not only assist clinicians in prioritizing patients for confirmatory testing and early intervention, optimizing health care resource allocation but also enable remote patient monitoring and telemedicine applications.

Hypothetically, the SARS-CoV-2 virus may induce alterations in the expression levels of genes coding for apoptosis and necroptosis of epithelial cells [41,42], resulting in the accumulation of oral epithelial cells and increasing tongue coating thickness. However, these proposed mechanisms remain preliminary and require further experimental validation to clarify their roles in oral manifestations of COVID-19. Supporting the clinical relevance of tongue changes, Wang et al [43] found a correlation between tongue coating thickness in patients with COVID-19 and levels of white blood cells as well as the neutrophil-to-lymphocyte ratio. Conversely, in patients with fever who tested negative for COVID-19, the presence of slimy or greasy tongue fur was associated with the level of C-reactive protein [43]. In addition, studies have indicated that greasy tongue fur is associated with higher blood fibrinogen levels in patients with stroke and with increased activity of glossal epithelial cells and vascular permeability in rodent models [44,45].

It is important to acknowledge several limitations of this study that may impact the generalizability of its findings. Although our predictive models demonstrate the potential of tongue image features for the differential diagnosis of acute respiratory tract infections, their accuracy remains modest, with area under the precision-recall curve (AUPRC) values below 80% across all models (Multimedia Appendix 7). Visualization of the data using t-SNE projected the 3 groups into generally distinct but overlapping clusters in 2D feature space, with considerable overlap, especially between the COVID-19 and adenovirus groups. This overlap is further reflected in our confusion matrix analyses, which consistently show misclassifications between these 2 groups across all classification models. Such overlap between the COVID-19 and adenovirus groups observed in the t-SNE projection may be partly attributable to the similarity of tongue image features shared by these 2 patient populations. In addition, the dimensionality reduction inherent in t-SNE can lead to loss of critical discriminative information, further exacerbating this overlap. Together with the limited sample size used for model development and validation, as well as the exclusion of additional clinical information, such as patient

symptoms and other relevant features, these factors collectively contribute to the modest accuracy of our predictive models. Furthermore, the COVID-19 pandemic's operational constraints necessitated data collection at a single tertiary hospital. Nevertheless, as a national referral center serving Beijing and surrounding provinces, our institution treats patients from geographically diverse regions, which may partially mitigate concerns regarding population representativeness. Future studies should incorporate larger and more diverse cohorts, combined with the integration of comprehensive clinical data, which are warranted to improve the discriminative power and generalizability of the models.

Conclusion

This study illustrates the utility of AI in helping clinicians identify potential pathogens in ARTI at the time of admission.

The interpretability and clinical relevance of our models suggest that they may help reduce unnecessary medical costs and diagnostic procedures while maintaining diagnostic accuracy. Moreover, by alleviating clinicians' workloads, this approach has the potential to enhance the overall quality of medical care. Looking forward, further research is warranted to externally validate these predictive models in independent cohorts and to expand the sample size to improve generalizability. In addition, incorporating clinical symptoms and other patient-specific features could further refine model performance and support more comprehensive decision-making in respiratory infection diagnosis.

Acknowledgments

We would like to acknowledge Dr. Wei Yang as a co-corresponding author. Correspondent to: Prof. YANG Wei, Department of Medical Statistics, Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China. yangyxq@ruc.edu.cn; Ruilin Wang, PhD, The Fifth Medical Center of PLA General Hospital, Beijing, China, WRL7905@163.com.

This work was supported by the National Key Research and Development Program (2023YFC3503404), Science and Technology Innovation Project of the China Academy of Chinese Medical Sciences (CI2023C066YLL, CI2021B003, CI2021A04706).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to data-sharing agreements with the participating hospitals but are available from the corresponding author on reasonable request.

Authors' Contributions

QC and YL conceptualized and designed the study, as well as drafted the main manuscript text. ZW, LL and RW collected the tongue image data and curated the dataset. FX developed and implemented the machine learning models. XC contributed to data visualization and statistical analysis. WY and RW supervised the overall study and provided critical revisions to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tongue image segmentation and feature extraction parameters for 20 labels.

[\[DOCX File \(Microsoft Word File\), 13 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Parameters for each layer in the proposed architecture.

[\[DOCX File \(Microsoft Word File\), 13 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Iteration loss plot. We iterated 40,000 times with an initial learning rate of 0.001, then reduced it to 0.0001 and continued until 60,000 iterations. Finally, we set the learning rate to 0.00001 and extended the iteration to 70,000.

[\[PNG File \(Portable Network Graphics File\), 75 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Accuracy comparison among fast recurrent neural network (blue), faster region-based convolutional neural network (orange), and Proposed Single Shot MultiBox Detector (green).

[\[PNG File \(Portable Network Graphics File\), 32 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Identified tongue image variables.

[\[DOCX File \(Microsoft Word File\), 13 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Model assessment matrices for 4 classifiers and 95% CI.

[\[DOCX File \(Microsoft Word File\), 13 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

Confusion matrix for machine learning model.

[\[PNG File \(Portable Network Graphics File\), 202 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

Mean Shapley additive explanations value over the random forest, gradient boosting machines, extreme gradient boosting. A longer bar indicates the variable is of higher importance in model performance. GMB: Gradient boosting machines.

[\[PNG File \(Portable Network Graphics File\), 160 KB-Multimedia Appendix 8\]](#)

References

1. Lynch JP 3rd, Fishbein M, Echavarria M. Adenovirus. *Semin Respir Crit Care Med*. Aug 2011;32(4):494-511. [doi: [10.1055/s-0031-1283287](#)] [Medline: [21858752](#)]
2. Liu MC, Xu Q, Li TT, et al. Prevalence of human infection with respiratory adenovirus in China: a systematic review and meta-analysis. *PLoS Negl Trop Dis*. Feb 2023;17(2):e0011151. [doi: [10.1371/journal.pntd.0011151](#)] [Medline: [36812245](#)]
3. Del Riccio M, Caini S, Bonaccorsi G, et al. Global analysis of respiratory viral circulation and timing of epidemics in the pre-COVID-19 and COVID-19 pandemic eras, based on data from the Global Influenza Surveillance and Response System (GISRS). *Int J Infect Dis*. Jul 2024;144:107052. [doi: [10.1016/j.ijid.2024.107052](#)] [Medline: [38636684](#)]
4. Chow EJ, Uyeki TM, Chu HY. The effects of the COVID-19 pandemic on community respiratory virus activity. *Nat Rev Microbiol*. Mar 2023;21(3):195-210. [doi: [10.1038/s41579-022-00807-9](#)] [Medline: [36253478](#)]
5. Cho HJ, Rhee JE, Kang D, et al. Epidemiology of respiratory viruses in Korean children before and after the COVID-19 pandemic: a prospective study from national surveillance system. *J Korean Med Sci*. May 20, 2024;39(19):e171. [doi: [10.3346/jkms.2024.39.e171](#)] [Medline: [38769924](#)]
6. Wen S, Lin Z, Zhang Y, et al. The epidemiology, molecular, and clinical of human adenoviruses in children hospitalized with acute respiratory infections. *Front Microbiol*. 2021;12:629971. [doi: [10.3389/fmicb.2021.629971](#)] [Medline: [33664719](#)]
7. Babawale PI, Guerrero-Plata A. Respiratory viral coinfections: insights into epidemiology, immune response, pathology, and clinical outcomes. *Pathogens*. Apr 12, 2024;13(4):316. [doi: [10.3390/pathogens13040316](#)] [Medline: [38668271](#)]
8. Wang WY, Zhou H, Wang YF, Sang BS, Liu L. Current policies and measures on the development of traditional Chinese medicine in China. *Pharmacol Res*. Jan 2021;163:105187. [doi: [10.1016/j.phrs.2020.105187](#)] [Medline: [32916255](#)]
9. Li J, Chen Y, Yu X, Xie Y, Li X, Diagnosis and Treatment Guideline for Chinese Medicine on Acute Trachea-Bronchitis working team, Respiratory Disease Branch of China Association of Chinese Medicine, Respiratory Disease Branch of China Medical Association of Minorities. Diagnosis and treatment guideline for Chinese medicine on acute trachea-bronchitis. *J Evidence Based Medicine*. Dec 2021;14(4):333-345. [doi: [10.1111/jebm.12460](#)]
10. Iranmanesh B, Khalili M, Amiri R, Zartab H, Aflatoonian M. Oral manifestations of COVID-19 disease: A review article. *Dermatol Ther*. Jan 2021;34(1):e14578. [doi: [10.1111/dth.14578](#)] [Medline: [33236823](#)]
11. Jones VG, Mills M, Suarez D, et al. COVID-19 and Kawasaki Disease: Novel Virus and Novel Case. *Hosp Pediatr*. Jun 2020;10(6):537-540. [doi: [10.1542/hpeds.2020-0123](#)] [Medline: [32265235](#)]
12. Chiotos K, Bassiri H, Behrens EM, et al. Multisystem Inflammatory Syndrome in Children During the Coronavirus 2019 Pandemic: A Case Series. *J Pediatric Infect Dis Soc*. Jul 13, 2020;9(3):393-398. [doi: [10.1093/jpids/piaa069](#)] [Medline: [32463092](#)]
13. Taşlıdere B, Mehmetaj L, Özcan AB, Gülen B, Taşlıdere N. Melkersson-Rosenthal Syndrome Induced by COVID-19. *Am J Emerg Med*. Mar 2021;41(262):262. [doi: [10.1016/j.ajem.2020.08.018](#)] [Medline: [32829989](#)]
14. Li J, Chen Q, Hu X, et al. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. *Int J Med Inform*. May 2021;149:104429. [doi: [10.1016/j.ijmedinf.2021.104429](#)] [Medline: [33647600](#)]
15. Jiang T, Guo XJ, Tu LP, et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD. *Comput Biol Med*. Aug 2021;135:104622. [doi: [10.1016/j.compbiomed.2021.104622](#)] [Medline: [34242868](#)]
16. Shi Y, Guo D, Chun Y, et al. A lung cancer risk warning model based on tongue images. *Front Physiol*. 2023;14:1154294. [doi: [10.3389/fphys.2023.1154294](#)] [Medline: [37324390](#)]

17. Yuan L, Yang L, Zhang S, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. *EClinicalMedicine*. Mar 2023;57:101834. [doi: [10.1016/j.eclinm.2023.101834](https://doi.org/10.1016/j.eclinm.2023.101834)] [Medline: [36825238](https://pubmed.ncbi.nlm.nih.gov/36825238/)]
18. Zhou J, Zhang Q, Zhang B. An automatic multi-view disease detection system via Collective Deep Region-based Feature Representation. *Future Gener Comput Syst*. Feb 2021;115:59-75. [doi: [10.1016/j.future.2020.08.038](https://doi.org/10.1016/j.future.2020.08.038)]
19. Dutta D, Naiyer S, Mansuri S, et al. COVID-19 diagnosis: a comprehensive review of the RT-qPCR method for detection of SARS-CoV-2. *Diagnostics (Basel)*. Jun 20, 2022;12(6):1503. [doi: [10.3390/diagnostics12061503](https://doi.org/10.3390/diagnostics12061503)] [Medline: [35741313](https://pubmed.ncbi.nlm.nih.gov/35741313/)]
20. Matthes-Martin S, Feuchtinger T, Shaw PJ, Matthes-Martin S, et al. European guidelines for diagnosis and treatment of adenovirus infection in leukemia and stem cell transplantation: summary of ECIL-4 (2011). *Transpl Infect Dis*. Dec 2012;14(6):555-563. [doi: [10.1111/tid.12022](https://doi.org/10.1111/tid.12022)] [Medline: [23146063](https://pubmed.ncbi.nlm.nih.gov/23146063/)]
21. Bochkovskiy A, Wang CY, Liao HYM. Yolov4: optimal speed and accuracy of object detection. *arXiv*. Preprint posted online on Apr 23, 2020. *arXiv:2004.10934*
22. O'Shea K. An introduction to convolutional neural networks. *arXiv*. Preprint posted online on Dec 2, 2015. [doi: [10.48550/arXiv.1511.08458](https://doi.org/10.48550/arXiv.1511.08458)]
23. Hao R, Namdar K, Liu L, Haider MA, Khalvati F. A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted MRI using convolutional neural networks. *J Digit Imaging*. Aug 2021;34(4):862-876. [doi: [10.1007/s10278-021-00478-7](https://doi.org/10.1007/s10278-021-00478-7)] [Medline: [34254200](https://pubmed.ncbi.nlm.nih.gov/34254200/)]
24. Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. *IEEE Trans Syst, Man, Cybern*. 1978;8(6):460-473. [doi: [10.1109/TSMC.1978.4309999](https://doi.org/10.1109/TSMC.1978.4309999)]
25. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. Oct 2006;59(10):1087-1091. [doi: [10.1016/j.jclinepi.2006.01.014](https://doi.org/10.1016/j.jclinepi.2006.01.014)] [Medline: [16980149](https://pubmed.ncbi.nlm.nih.gov/16980149/)]
26. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. *Learning from Imbalanced Data*. Springer; 2018.
27. Marcilio WE, Eler DM, editors. *From explanations to feature selection: assessing SHAP values as feature selection mechanism*. Presented at: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); Nov 7-10, 2020; Porto de Galinhas, Brazil.
28. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res*. 2014;15(1):3221-3245. [doi: [10.5555/2627435.2697068](https://doi.org/10.5555/2627435.2697068)]
29. Krumbein H, Kümmel LS, Fragkou PC, et al. Respiratory viral co-infections in patients with COVID-19 and associated outcomes: a systematic review and meta-analysis. *Rev Med Virol*. Jan 2023;33(1):e2365. [doi: [10.1002/rmv.2365](https://doi.org/10.1002/rmv.2365)] [Medline: [35686619](https://pubmed.ncbi.nlm.nih.gov/35686619/)]
30. Wen G, Ma J, Hu Y, Li H, Jiang L. Grouping attributes zero-shot learning for tongue constitution recognition. *Artif Intell Med*. Sep 2020;109:101951. [doi: [10.1016/j.artmed.2020.101951](https://doi.org/10.1016/j.artmed.2020.101951)] [Medline: [34756217](https://pubmed.ncbi.nlm.nih.gov/34756217/)]
31. Li J, Yuan P, Hu X, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. *J Biomed Inform*. Mar 2021;115:103693. [doi: [10.1016/j.jbi.2021.103693](https://doi.org/10.1016/j.jbi.2021.103693)] [Medline: [33540076](https://pubmed.ncbi.nlm.nih.gov/33540076/)]
32. Mai MV, Krauthammer M. Controlling testing volume for respiratory viruses using machine learning and text mining. *AMIA Annu Symp Proc*. 2017;2016:1910-1919. [Medline: [28269950](https://pubmed.ncbi.nlm.nih.gov/28269950/)]
33. Chang TH, Liu YC, Lin SR, et al. Clinical characteristics of hospitalized children with community-acquired pneumonia and respiratory infections: using machine learning approaches to support pathogen prediction at admission. *J Microbiol Immunol Infect*. Aug 2023;56(4):772-781. [doi: [10.1016/j.jmii.2023.04.011](https://doi.org/10.1016/j.jmii.2023.04.011)] [Medline: [37246060](https://pubmed.ncbi.nlm.nih.gov/37246060/)]
34. Tania MH, Lwin K, Hossain MA. Advances in automated tongue diagnosis techniques. *Integr Med Res*. Mar 2019;8(1):42-56. [doi: [10.1016/j.imr.2018.03.001](https://doi.org/10.1016/j.imr.2018.03.001)] [Medline: [30949431](https://pubmed.ncbi.nlm.nih.gov/30949431/)]
35. Xu Q, Zeng Y, Tang W, et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. *IEEE J Biomed Health Inform*. Sep 2020;24(9):2481-2489. [doi: [10.1109/JBHI.2020.2986376](https://doi.org/10.1109/JBHI.2020.2986376)] [Medline: [32310809](https://pubmed.ncbi.nlm.nih.gov/32310809/)]
36. Meng D, Cao G, Duan Y, et al. Tongue images classification based on constrained high dispersal network. *Evid Based Complement Alternat Med*. 2017;2017:7452427. [doi: [10.1155/2017/7452427](https://doi.org/10.1155/2017/7452427)] [Medline: [28465706](https://pubmed.ncbi.nlm.nih.gov/28465706/)]
37. Li B, Chen H, Lin X, Duan H. Multimodal learning system integrating electronic medical records and hysteroscopic images for reproductive outcome prediction and risk stratification of endometrial injury: a multicenter diagnostic study. *Int J Surg*. Jun 1, 2024;110(6):3237-3248. [doi: [10.1097/JS9.0000000000001241](https://doi.org/10.1097/JS9.0000000000001241)] [Medline: [38935827](https://pubmed.ncbi.nlm.nih.gov/38935827/)]
38. Horzov L, Goncharuk-Khomyn M, Hema-Bahyna N, Yurzhenko A, Melnyk V. Analysis of tongue color-associated features among patients with PCR-confirmed COVID-19 infection in Ukraine. *Pesqui Bras Odontopediatria Clín Integr*. 2021;21. [doi: [10.1590/pboci.2021.109](https://doi.org/10.1590/pboci.2021.109)]

39. Wen Z, Min C, Ding S, et al. Tongue and pulse features of 668 asymptomatic patients infected with the severe acute respiratory syndrome coronavirus 2 omicron variant in Shanghai. *J Tradit Chin Med*. Dec 2022;42(6):1006-1011. [doi: [10.19852/j.cnki.jtcm.20220922.004](https://doi.org/10.19852/j.cnki.jtcm.20220922.004)] [Medline: [36378060](https://pubmed.ncbi.nlm.nih.gov/36378060/)]
40. D J. Diagnosis and management of TCM on infected or suspected patients with Covid-19-48 cases study in UK. *JCV*. 2021;01(4). URL: <https://www.scienceworldpublishing.org/journals/journal-of-corona-virus-/current-issue#articlepage> [doi: [10.47690/JCV.2021.1403](https://doi.org/10.47690/JCV.2021.1403)]
41. Ravindra NG, Alfajaro MM, Gasque V, et al. Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. *PLoS Biol*. Mar 2021;19(3):e3001143. [doi: [10.1371/journal.pbio.3001143](https://doi.org/10.1371/journal.pbio.3001143)] [Medline: [33730024](https://pubmed.ncbi.nlm.nih.gov/33730024/)]
42. Zhu Z, Shi J, Li L, Wang J, Zhao Y, Ma H. Therapy targets SARS-CoV-2 infection-induced cell death. *Front Immunol*. 2022;13:870216. [doi: [10.3389/fimmu.2022.870216](https://doi.org/10.3389/fimmu.2022.870216)] [Medline: [35655782](https://pubmed.ncbi.nlm.nih.gov/35655782/)]
43. Wang ZC, Cai XH, Chan J. Tongue coating in COVID-19 patients: a case-control study. *medRxiv*. Preprint posted online on Apr 15, 2024. [doi: [10.1101/2022.03.14.22272342](https://doi.org/10.1101/2022.03.14.22272342)]
44. Gao L, Liu P, Song J xian, et al. Study on the correlation of tongue manifestation with fibrinogen and neutrophil in acute cerebral infarction patients. *Chin J Integr Med*. Dec 2012;18(12):942-945. [doi: [10.1007/s11655-012-1298-y](https://doi.org/10.1007/s11655-012-1298-y)] [Medline: [23239003](https://pubmed.ncbi.nlm.nih.gov/23239003/)]
45. Qi WJ, Zhang MM, Wang H, Wen Y, Wang BE, Zhang SW. Research on the relationship between thick greasy tongue fur formation and vascular endothelial cell permeability with the protein expression of zonula occludens-1. *Chin J Integr Med*. Jul 2011;17(7):510-516. [doi: [10.1007/s11655-011-0784-1](https://doi.org/10.1007/s11655-011-0784-1)] [Medline: [21725876](https://pubmed.ncbi.nlm.nih.gov/21725876/)]

Abbreviations

AI: artificial intelligence
ARTI: acute respiratory tract infection
AUC: area under the curve
AUCPR: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
GBM: gradient boosting machine
HAdV: human adenovirus
PLA: People's Liberation Army
RNN: R
SHAP: Shapley additive explanations
t-SNE: t-distributed stochastic neighbor embedding
TCM: traditional Chinese medicine
XGBoost: extreme gradient boosting

Edited by Arriel Benis; peer-reviewed by Adekunle Adeoye, Kudirat Abidemi Obisesan Olawuni; submitted 18.03.2025; final revised version received 24.06.2025; accepted 24.06.2025; published 25.08.2025

Please cite as:

Che Q, Leng Y, Yang W, Cao X, Wang Z, Liu L, Xie F, Wang R
Tongue Image-Based Diagnosis of Acute Respiratory Tract Infection Using Machine Learning: Algorithm Development and Validation
JMIR Med Inform 2025;13:e74102
 URL: <https://medinform.jmir.org/2025/1/e74102>
 doi: [10.2196/74102](https://doi.org/10.2196/74102)

© Qianzi Che, Yuanming Leng, Wei Yang, Xihao Cao, Zhongxia Wang, Lizheng Liu, Feibiao Xie, Ruilin Wang. Originally published in *JMIR Medical Informatics* (<https://medinform.jmir.org>), 25.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.