Received: 24 April 2026 | Revised: 27 July 2026 | Accepted: 15 September 2026 | Published online:10 November 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 00(00) 1–13
DOI: 10.47852/bonviewMEDIN620210119

BON VIEW PUBLISHING

Automated Pharyngitis Classification Using a Dual-Branch Hybrid Deep Learning Model with Explainable Artificial Intelligence and Retrieval-Augmented Generation-Based Clinical Recommendation

Mizanur Rahman[1], Fakrul Alam[2], Azharul Islam[3], Mazharul Islam[4], Sumon Rana[5], Firoz Kabir[6] and Irfan Sadiq Rahat[7,*]

[1] *Data Science and Artificial Intelligence, Campbellsville University, USA*

[2] *Science in Information Studies, Trine University, USA*

[3] *Department of Master of Science in Information System, Trine University, USA*

[4] *Science in Business Analytics, Trine University, USA*

[5] *School of Business, International American University, USA*

[6] *Information Technology, University of the Cumberlands, USA*

[7] *School of Computer Science and Engineering, VIT-AP University, India*

Abstract: Pharyngitis is a frequent inflammatory condition of the upper-airway system for which prompt preliminary evaluation can aid in the proper clinical evaluation and minimize treatment outside of that. However, the level of inflammation, the quality of illumination, the quality of the image, and the level of the observer's experience can affect the interpretation of the image. A research-stage framework for binary pharyngitis screening with 312 pharyngitis images classified as no pharyngitis and pharyngitis is presented. The proposed dual-branch hybrid deep learning model consists of a separable convolution-based local feature extraction branch and an attention-based global contextual learning and cross-scale feature fusion branch. It brings together the combined power of hybrid CNN/attention-based classification, explainable AI, and retrieval-augmented generation, along with a Qwen large language model, in a single screening workflow. The model performance was done by 5-fold cross-validation with leakage-safe and the result is 97.25% accuracy, 97.58% precision, 97.43% recall, and 97.50% F1 score. To explain the predictions, image regions were emphasized by explainability visualization, and the RAG/Qwen module produced structured recommendation reports using the predicted class and retrieved medical context. The framework can help with initial image screening, and clinician-reviewed decision support for telemedicine/triage environments. However, the limited sample size and absence of external multi-center and prospective testing by clinicians with the weathered data limits clinical generalizability. Thus, the proposed system is not intended as a diagnostic or a prescription tool for treatment, but as a decision support system in a research phase.

Keywords: pharyngitis classification, throat image analysis, deep learning, dual-branch hybrid model, explainable artificial intelligence

1. Introduction

Pharyngitis is the inflammation of the throat, often affecting the tissues of the throat and upper-airway structures [1]. It is one of the most common causes of sore throat and occurs in people of all ages – children as well as adults [2, 3]. It can be caused by viral infection, bacterial infection, recurrent inflammatory syndrome, or other irritative causes and is vital in terms of appropriate treatment, prevention of complications and unnecessary use of antibiotics for accurate diagnosis [4–6]. Diagnosis of pharyngitis is typically made during routine clinical practice using a combination of visual examination of the throat and the presence of

***Corresponding author:** Irfan Sadiq Rahat, School of Computer Science and Engineering, VIT-AP University, India. Email: rahat.21bce9985@vitapalum.ac.in

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

clinical symptoms including redness, swelling and pain, fever and presence of exudates [7]. These visual and clinical features can, however, be highly variable from person to person and can also be similar for both viral, bacterial, and non-infectious causes [8–11]. Throat appearance is also influenced by clinician experience, and this can result in variable interpretation of the appearance, especially during early stage and mild or ambiguous presentations. For this reason, objective and reliable diagnostic-support methods are necessary to enhance uniformity, minimize observer variability, and aid in the clinical decision-making process [12]. In the medical field, AI and deep learning have emerged as valuable tools for analyzing medical images and have demonstrated promising potential in automated disease classification [13, 14]. Complex visual pattern learning could take place in an unsupervised manner by deep learning models, potentially offering a standardized image-based assessment. Automated image analysis systems can enhance diagnostic consistency, minimize the subjective variation, and boost efficiency in clinical workflows [15–17]. These systems can process large volumes of images quickly and can offer objective support in the screening process for the clinician [18–20]. Automated image-based methods are especially relevant in the recognition of throat and upper-airway diseases, as visible features like inflammation, erythema, swelling, and exudates are vital diagnostic features [21, 22]. However, many deep learning models remain black boxes, in which the inner logic behind the decision-making process remains opaque [23]. This lack of explainability can decrease clinical trust and prevent its practical use in healthcare. Pharyngitis diagnosis and management is inconsistent in clinical practice as evidenced by clinical studies. The diagnosis, management, and decision support for pharyngitis are still important and challenging areas, as evidenced in clinical and artificial intelligence studies. According to clinical and AI research, there are several areas that are critical for pharyngitis, but difficult to achieve such as diagnosis, antimicrobial stewardship, image-based assessment, and interpretive decision support. Significant differences in diagnosis and treatment of group A streptococcal pharyngitis and the value of clinical scoring, rapid antigen tests, culture, and molecular diagnostics were documented by Luo et al. [24] and summarized by Mustafa & Ghaffari [25]. These concerns have been confirmed by more recent evidence. Curatola et al. [26] reviewed the diagnosis and management of contemporary streptococcal pharyngitis and noted that it is challenging to distinguish from clinical symptoms alone. At the same time, XAI is gaining traction in the medical imaging sector. Borys et al. [27] summarized SA methods applied to clinical images, while Champendal et al. [28] summarized the additional types of interpretability methods and the need for a clinically relevant evaluation. Structured decision rules have been shown to have a useful place in identifying patients at relatively low risk of having a streptococcal infection [29] and contemporary epidemiological patterns in group A streptococcal pharyngitis reported [30]. RGLMs have also recently been explored for use in health care, as discussed by Yang et al. [31] who talk about the potential of Retrieval-Augmented Generation for making generative AI more reliable, more traceable, and more current in health care. Differential diagnosis is also still crucial since recurrent pharyngitis is a feature of other syndromes like PFAPA, as illustrated by Wang et al. [32]. Bakhit et al. [33] also demonstrated that there is significant variation in the diagnostic value of the clinical prediction rules for the diagnosis of group A beta-hemolytic streptococci and Schmutzler et al. [34] examined myocarditis associated with streptococcal tonsillitis or pharyngitis. More recently, Melon et al. [35] focused on the use of retrieval graph for retrieval-augmented generation (RAG) in clinical decision support summaries, highlighting continued growing interest in retrieval-based medical LMs.

The novelty of this work is in the development of an integrated pharyngitis screening framework consisting of an image-based classification, explainable visual interpretation, and retrieval-augmented suggestion of the clinical recommendation within a unified workflow. The proposed framework aims to facilitate the entire screening process by introducing a dual-branch hybrid deep learning model for the classification of throat images into no pharyngitis or pharyngitis, explainable artificial intelligence to support the visual interpretation of the prediction result, and RAG and a Qwen large language model (LLM) to generate a structured recommendation report from the predicted class. This contribution is not a claim of fundamental algorithmic innovation nor clinical diagnosis, but rather a research-stage decision support framework for the initial evaluation of pharyngitis that integrates with hybrid CNN/attention classification, model interpretability, and context-aware recommendation generation. Thus, the work proposed is designed to increase transparency and usability of image-based screening for pharyngitis without removing the necessity for clinician review, external validation and medical expert verification prior to real-world clinical use.

Although advances have been made in the automated classification of pharyngitis and throat disease, some features are missing from the existing systems. Current systems are primarily geared towards classification accuracy without offering structured clinical recommendations or treatment related advice. But when making a clinical decision, it's not enough to recognize the class of a disease; there is a need to know what action is required next, how to treat symptoms, what further action might be required, and when medical evaluation is needed. Furthermore, few explanations available can make it hard for a clinician to decide whether a prediction made by the model is due to clinically relevant areas in the throat or irrelevant image artifacts. As such, integrated frameworks for artificial intelligence that can not only accurately classify images but also provide an explanation of the image's concepts and make clinically relevant decisions are needed. The aim of this study is to tackle these issues by proposing a unified system for automatic pharyngitis classification and clinical suggestions. Then, a separable convolution-based feature extraction module and the attention-based feature learning module are integrated by using a dual-branch hybrid deep learning model to recognize throat image as no pharyngitis and pharyngitis.

Second, the model is evaluated using leakage-safe 5-fold cross-validation on a dataset of 312 throat images to ensure reliable and unbiased performance assessment. Third, explainable artificial intelligence is incorporated through feature visualization and interpretability methods to highlight image regions that contribute to the model prediction. Fourth, a LLM-based clinical recommendation system is integrated to provide structured treatment guidance. RAG is used to retrieve relevant medical context and generate disease-specific recommendations based on the predicted class. Finally, the proposed framework supports automated screening, explainable diagnosis, and clinical decision support, thereby improving the practical usability of artificial intelligence in pharyngitis assessment.

2. Materials and Methods

2.1. Dataset description

The main dataset employed for model development was Pharyngitis Images Dataset which was available from the Kaggle

dataset repository. The dataset consists of a total of 312 RGB images of the throat with each image belonging to one of two classes, no pharyngitis or pharyngitis, and is available at https://www.kaggle.com/datasets/bebeaa/pharyngitis-images-dataset. The images are of different appearances of the throat: different color, texture of the mucosa, redness, inflammation, and swelling of the throat. These differences in appearance are associated with the visual heterogeneity found in screening for pharyngitis with images and render discrimination difficult, especially when the inflammatory signs are mild or appear in a subtle way in the image. The images and classes labels in the public dataset were used as the main data for training and internal cross-validation of the proposed model. An independent publicly available dataset – PGUPharyngitis – was used for external evaluation of model performance. This dataset is made available in Figshare and may be identified by the DOI 10.6084/m9.figshare.28163513. PGUPharyngitis has 742 optical throat photographs, among which the bacterial and nonbacterial pharyngitis images are 430 and 312, respectively. Images were taken by Samsung Galaxy S21 Ultra and Xiaomi Redmi 8 Pro smartphones, and multiple physicians annotated each image, with the final labels selected by majority voting. The data were gathered from October 2023 to May 2024 in two geographic areas in Iran, and the age range of participants was 0-100 years. There were two sets of data used in experiments for different purposes. The Kaggle Pharyngitis Images Dataset was used for model development and internal evaluation while PGUPharyngitis was held out as an external test set for evaluation of the model and was not used for model training or hyperparameter optimization. The two label spaces were not directly comparable because the primary dataset classified the subjects as no pharyngitis or pharyngitis, while the PGUPharyngitis dataset classified the subjects as bacterial or nonbacterial pharyngitis; caution was taken in interpreting results from the external analysis. The detailed processing, augmentation, and leakage-safe 5-fold cross-validation procedures for the primary dataset are described in the following steps. Detailed preprocessing, augmentation, and leakage-safe five-fold cross-validation procedures for the primary dataset are described as follows.

2.2. Data preprocessing and cross-validation

As a preprocessing step, all the images from the throat were resized to 224 × 224 pixels. The resizing process guarantees that all samples have the same input size for the deep learning architecture, which guarantees uniformity in processing. By normalizing the size of the image, the model can learn the features more effectively, as it will not be affected by the image resolution. Moreover, resizing decreases the computation complexity and memory consumption and retains vital visual features relevant to the classification of pharyngitis, including color, texture, and structural information. Image normalization was also applied to enhance the stability of training and learning efficiency. Normalization ensures that the distributions of pixel values are uniform across images, facilitating the model to learn meaningful representations of its features. The following process helps with faster convergence when training and also helps in preventing instability due to different intensity levels of the images. The training data were augmented with more data to further improve the model's generalizability. The benefit of augmentation is to generate more variations of the existing images, so that the model can learn more robust features. Typical augmentation techniques included rotation, flipping, and brightness variation, which simulate natural variations observed in clinical imaging. This helps to decrease overfitting because the model doesn't "memorize" individual image patterns. Importantly, augmentation was only done to the training data, not to the validation and test data, to ensure a fair evaluation. To evaluate the model performance, the safe 5-fold cross-validation method was used. To train and test the dataset, the data were split into five separate folds with 80% of the data used for training and 20% for testing in each fold. The folds were turned over so that each sample was tested once. The preprocessing and augmentation were done independently in each training fold, and the validation and test folds were not changed. This leakage-safe process guarantees there is no overlap between training and test sample and offers unbiased and reliable performance assessment.

3. Proposed Methodology

3.1. Overview of the proposed model

To effectively deal with the differences in the visual properties of pharyngitis images in the throat, a dual-branch hybrid model for automated classification of pharyngitis is proposed. To effectively capture the differences in the visual properties of pharyngitis images in the throat, a dual-branch hybrid model for automated classification of pharyngitis is proposed. The model aims to capture both local visual features and global contextual information, both of which are vital for good differentiation between no pharyngitis and pharyngitis class. The architecture has two parallel branches that work simultaneously on the same image to extract the features. The first branch is based on separable convolution layers and channel attention mechanisms, which capture local spatial features like texture, color variation, and structural patterns. The focus of this branch is on the fine grain visual details that are important in detecting inflammation-related changes. The second branch applies multi-scale transformer attention to learn the global contextual information and long range spatial relationships. This branch allows the model to learn more general patterns and relationships in the images, not necessarily captured by the convolution. A feature fusion mechanism is then implemented to fuse the feature representations obtained from both branches into one single representation. These fused features are forwarded to classification layers, and final prediction is generated for the classes No Pharyngitis and Pharyngitis. An overall architecture is shown in Figure 1. The hybrid design enables to leverage the complementary information provided by the convolution and attention-based learning, which helps to achieve accurate and reliable classification of pharyngitis.

3.2. Multi-scale transformer attention branch of the proposed pharyngitis classification model

The multi-scale transformer attention branch aims to obtain the global context features, which will be complementary to the local context features extracted by the convolution branch. Transformer attention allows for the model to focus on various parts of the image and helps it understand the relationships between them. This branch starts with multi scale convolutions, which extract features from different scales. The multi-scale features are able to detect both fine scale visual pattern (localized redness) and the larger scale structural pattern (overall inflammation regions). This ability enhances the ability to detect fine and wide range of visual characteristics, relevant to the classification of pharyngitis. After multi-scale feature extraction, the branch adds a local window attention mechanism to pay attention to the significant spatial

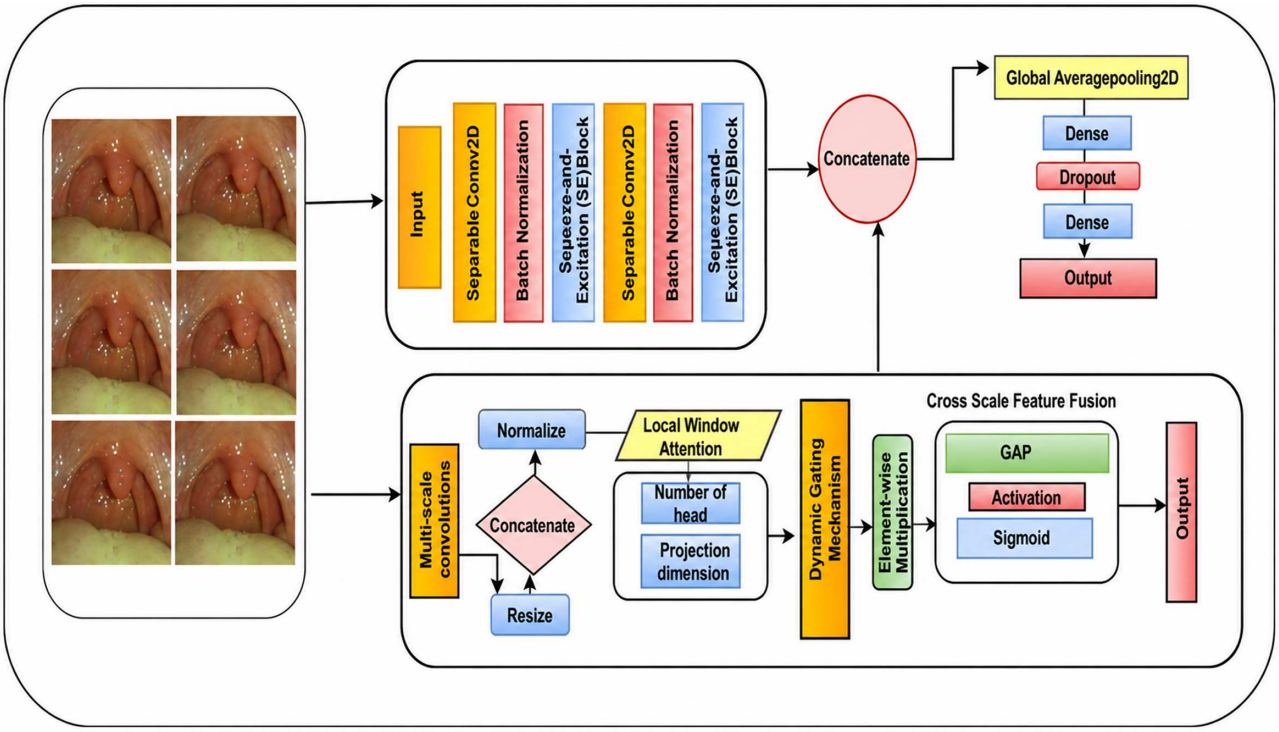


Figure 1. Architecture of the proposed dual-branch hybrid deep learning model for automated pharyngitis classification

region. The local window attention can highlight the important local region of the image while retaining the spatial structure of the image. This attention mechanism is able to highlight the relationship between adjacent areas, which promotes the contextual understanding of the throat image. This allows the model to more accurately detect the significant visual information from the variations in the background. A dynamic gating mechanism is used to further improve the feature representation. Dynamic gating mechanism is used to assign importance weights to the feature channels, enhancing the important features while canceling out less important ones. This process enhances the quality and relevance of the extracted features. Finally, a cross-scale feature fusion step is carried out to fuse features from various scales. Cross-scale feature fusion means combining the features from multiple resolution levels, resulting in a more robust and complete feature representation. This branch helps the model to absorb the contextual information from the globe and helps to improve the performance of the model in the classification of pharyngitis.

3.3. Feature fusion and classification

The features from the separable convolution branch and the multi-scale transformer attention branch are fused by feature concatenation. The output of each branch is a $7 \times 7 \times 256$ dimensional feature tensor, and it concatenates along the channel dimension to fuse the outputs. This operation returns a feature tensor of size $7 \times 7 \times 512$, which is a fused operation. Concatenation process: local spatial features are extracted using convolution, and global contextual features are learned using transformer attention. This fused representation enhances feature completeness and offers a more holistic perspective of visual patterns pertinent to classification of pharyngitis. After feature fusion, the fused feature tensor is passed to global average pooling. This is an operation that reduces the size of the vector from $7 \times 7 \times 512$ to 1×512. Global average pooling is a method that eliminates a lot of parameters while retaining vital feature information and enhancing generalization performance. This produces the feature vector that feeds into a fully connected dense (256) layer with rectified linear unit (ReLU) activation function. This layer is trained with discriminative feature representations to improve the classification ability. Dropout (0.40) is used in order to avoid overfitting and to enhance model robustness, as well as generalization. Finally, a Dense layer with 2 neurons and Softmax activation function creates probability scores for the two output classes – no pharyngitis and pharyngitis. This classification module yields the end classification and allows for the correct classification of pharyngitis.

3.4. Training configuration

All the experiments were run on a high performance workstation for efficient training and evaluation of the proposed model. It was powered by an Intel Core i9 8 Core 12th Generation processor, 32 GB of DDR5 RAM for data processing and model calculation. For graphics acceleration, the system featured an NVIDIA RTX 3080 graphics card with 10GB of VRAM, supporting efficient deep learning training. The system also had 1 TB of NVMe SSD storage, providing high-speed data access and loading of the models. We used Python 3.10, PyTorch 2.*x*, and CUDA 12.*x* for deep learning implementation and GPU acceleration, respectively (Table 1). The operating system used was Windows 11 64 bit. The model was trained with the Adam optimizer with a starting learning rate of 0.001 to optimize the gradient. ReduceLROnPlateau was used to adapt the learning rate during training, with learning rate reduction factor 0.1, patience 5 epochs, and the minimum learning rate of 0.000001. This model was trained for 100 epochs, and early stopping was used to prevent overfitting by stopping training after the validation performance

did not improve anymore after 10 epochs. For stability of gradient update, and efficient use of the GPU resources, a batch size of 32 was chosen (Table 2). A categorical cross-entropy loss function was used, which is suitable for binary classification with softmax output and the measures of prediction error between the predicted and true labels. These training settings guarantee stable convergence, enhance classification performance, and offer reliable pharyngitis classification.

Table 1. Hardware and software environment used for experiments

System	Configuration
CPU	Intel Core i9 (12th Gen), 8-core
Memory (RAM)	32 GB DDR5
GPU	NVIDIA RTX 3080 (10 GB VRAM)
Storage	1 TB NVMe SSD
Operating System	Windows 11 (64-bit)
Framework	Python 3.10, PyTorch 2.*x*, CUDA 12.*x*

Table 2. Training characteristics of the proposed model

Parameter	Description
Model Name	(Proposed Model)
Input Data Format	Pre-processed
Image Size	224 × 224 pixels
Batch Size	32
Epochs	100
Learning Rate	0.001 (with decay using Reduce LROnPlateau)
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Regularization Techniques	Batch Normalization, Dropout (0.40)

3.5. Disease prediction and Query Formulation

The first step in the LLM workflow for treatment recommendation is the automated classification of pharyngitis from the input throat image using a trained dual-branch hybrid model. The image passes through the feature extraction branches and classification layers where appropriate features of the image are extracted and analyzed. The final layer in the classification network uses the Softmax activation function to output the probability scores of each classification. From these probabilities, the model assigns the ultimate predicted class by picking the highest probability class. The output class is no pharyngitis or pharyngitis, which is an automatic result of the disease screening. This prediction will give a structured and objective description of the image input and will be the most important input to the next clinical recommendation system. After predicting the disease, the query formulation engine produces a structured clinical query based upon the predicted class information. This query is automatically generated by the query formulation engine which includes both the predicted disease label and disease contextually relevant information related to the disease condition and classification result. The query is structured in a textual form. A structured query to retrieve medical knowledge that is relevant to the query, optimized for semantic search, and able to identify the clinical information relevant for the disease. The generated query is then fed to the RAG system, which uses the query to retrieve the medical knowledge. This helps to achieve context-aware clinical recommendation and disease-specific recommendation generation. This step makes the classification of pharyngitis more directly linked with clinical recommendation which will facilitate automated clinical decision support and make the proposed system more useful in practice.

3.6. Knowledge retrieval using vector database

The second step of the LLM-based treatment recommendation workflow is to retrieve medical knowledge for the pharyngitis classification result, as shown in Figure 2. This retrieval component is based on the general idea of RAG, where external knowledge is retrieved and fed into the language model for more context-aware and grounded response generation. The knowledge retrieval and LLM-based clinical summarization techniques used in ClinicGraphRAG have been extended, in which graph-based retrieval and LLM was employed to create clinical decision support summaries with high reliability [35]. The proposed framework includes a vector database for knowledge retrieval, which allows for semantic search, not just keyword matching. The structured query is first mapped to a numerical embedding representation based on the predicted class. This embedding represents the meaning of the query and enables comparison to stored medical knowledge entries, based on conceptual similarity. Thus, more effective retrieval of the relevant medical information related to the anticipated screening result for pharyngitis is possible than by simple keyword matching. The generated query embedding is then used to do similarity search in Qdrant vector store. For this database, the entries of medical knowledge are stored as vector representations and similarity comparison is done between the query embedding and the stored embeddings to find the most relevant ones. Retrieved data might contain descriptions of diseases, typical symptoms, general management considerations, and follow-up tracking for the forecasted class. The retrieved contextual snippets and metadata are then fed into the Qwen LLM as reference input to the RAG technique. This retrieval step is used to link pharyngitis classification based on images to external medical knowledge, and also to facilitate the generation of a structured recommendation report. The retrieved material and generated recommendation has been provided for research purposes to aid in making decisions and are not intended for clinical use without the review of medical experts.

3.7. Treatment recommendation using large language model

The medical context information retrieved in the last stage of the treatment recommendation workflow is passed to the LLM to generate treatment recommendations. This stage converts the result of the classification of pharyngitis into clinically useful output to produce structured treatment guidance. The LLM classifies the expected class (no pharyngitis and pharyngitis) and queries the medical knowledge to provide suitable suggestions. Such integration facilitates automatic clinical decisions support based on image predictive and utilization of medically meaningful guidance in a more useful manner. A domain constrained Qwen LLM is used to ensure safe and reliable recommendation generation. The

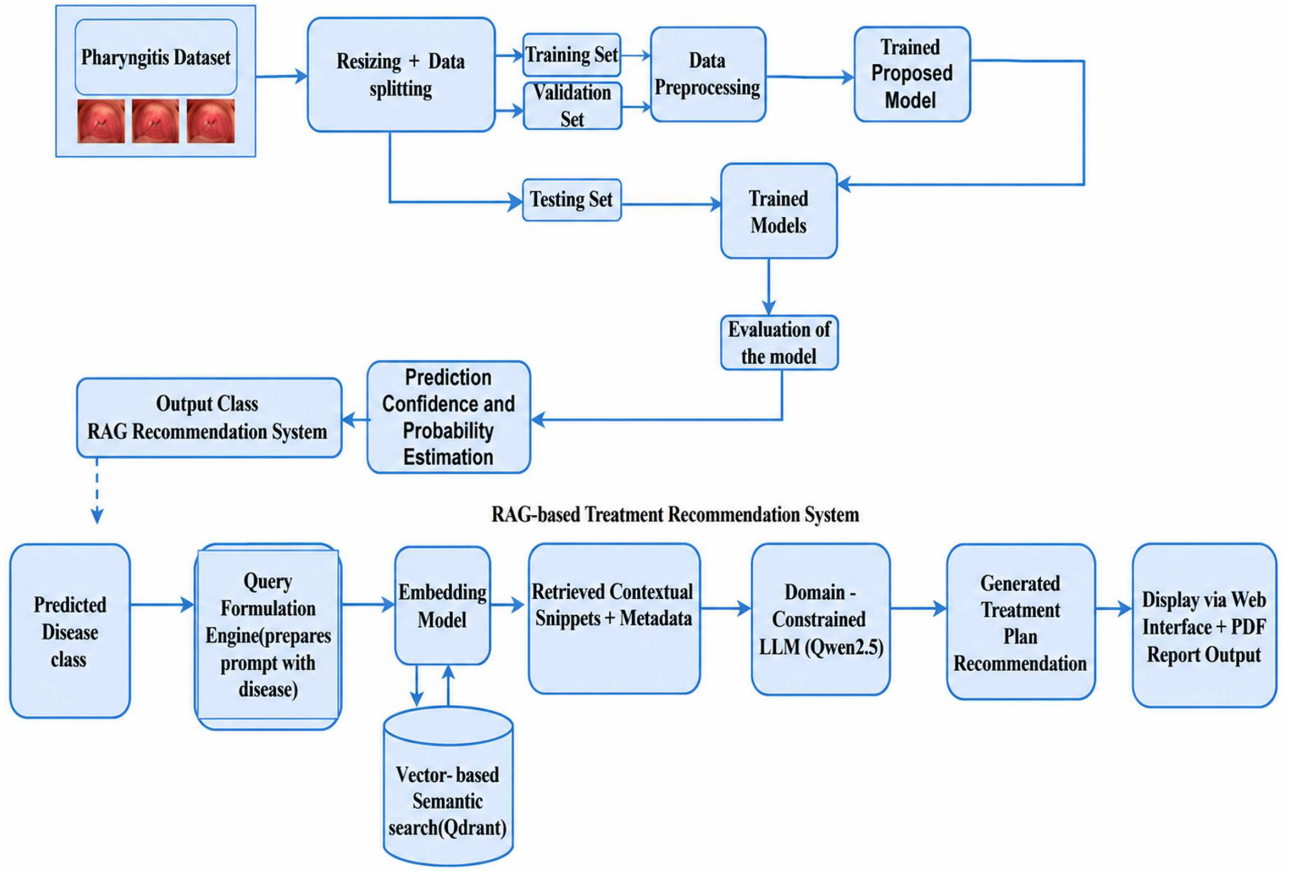


Figure 2. Overall workflow of the proposed pharyngitis classification and clinical recommendation system integrating deep learning, explainable artificial intelligence, retrieval-augmented generation, and the Qwen large language model

domain constrained configuration forces the model to work on medical domain knowledge to avoid unrelated and irrelevant outputs. This constraint helps to ensure reliability and add to the safety of the responses produced. The retrieved medical context from the vector database will be given to the Qwen LLM as reference data, so that the model can produce medical responses based on the verified information. The following features of this RAG approach make it possible to generate context-aware responses and medically relevant and contextually specific recommendations that reflect the predicted disease condition. The output generated is displayed in the form of a clinical report, showing the predicted condition and treatment recommendations. The clinical report is a document that includes organized data, including symptom management suggestions and follow-up recommendations. This format enhances interpretability and decision support that is interpretable. The clinical report can help the healthcare professionals to understand the automated prediction and suggested course of action. This step takes classification output to actionable clinical recommendation, and improves the overall performance of the pharyngitis classification system by providing automated clinical decision support.

4. Results and Discussion

Leakage-safe 5 fold cross-validation was used to validate the classification performance of the proposed dual-branch hybrid model. The model had a training accuracy of 97.68 per cent and a cross-validation accuracy of 97.25 percent. In addition, the model obtained a precision of 97.58 percent, recall of 97.43 percent, and F1 score of 97.50 percent. The results showed good classification performance for classification of pharyngitis (Table 3). High precision was achieved, meaning the model has a low number of false positive predictions; high recall was achieved, meaning the model is good at correctly identifying pharyngitis cases. The F1 score is high, indicating a good balance between precision and recall, which means the classification results are reliable. In addition, the small error between the accuracy of the train and the cross-validation suggests that the model has a good generalization capability, and that it is not overfitting. A comparison of the performances with baseline models is shown in Table 3. The proposed model gave the best performance among all the tested models. The Focal Transformer gave 97.03 percent accuracy and F1 score of 96.62 percent whereas ConvNeXt achieved an accuracy of 95.87 percent and F1 score of 95.36 percent. The accuracy of the CSWin Transformer is 94.18 percent and the F1 score is 93.52 percent, while GhostNetV2 is 89.68 percent and 89.63 percent, respectively. The proposed model has an accuracy and F1 score greater than all baseline models. The improvement can be credited to the hybrid feature extraction design which combines the two types of feature, convolution-based and attention-based, and the combination of both produces a better representation. The results obtained validate the effectiveness of the proposed architecture while strong pharyngitis classification capability is demonstrated by the leakage-safe cross-validation for evaluating the results.

The training and validation loss and accuracy curves of the proposed dual-branch hybrid model are shown in Figure 3. The loss has been gradually reduced from around 0.90 to 0.23 while

Table 3. Performance comparison (precision, recall, and F1-score) of MFN and baseline models under 5-fold cross-validation

Model	Train accuracy (%)	CV Val Acc (%)	Precision (%)	Recall (%)	F1-score (%)
Proposed Model	97.68	97.25	97.58	97.43	97.50
Focal Transformer	97.24	97.03	96.78	96.47	96.62
ConvNeXt	96.76	95.87	95.60	95.12	95.36
CSWin Transformer	94.85	94.18	93.88	93.17	93.52
GhostNetV2	90.17	89.68	90.16	89.11	89.63

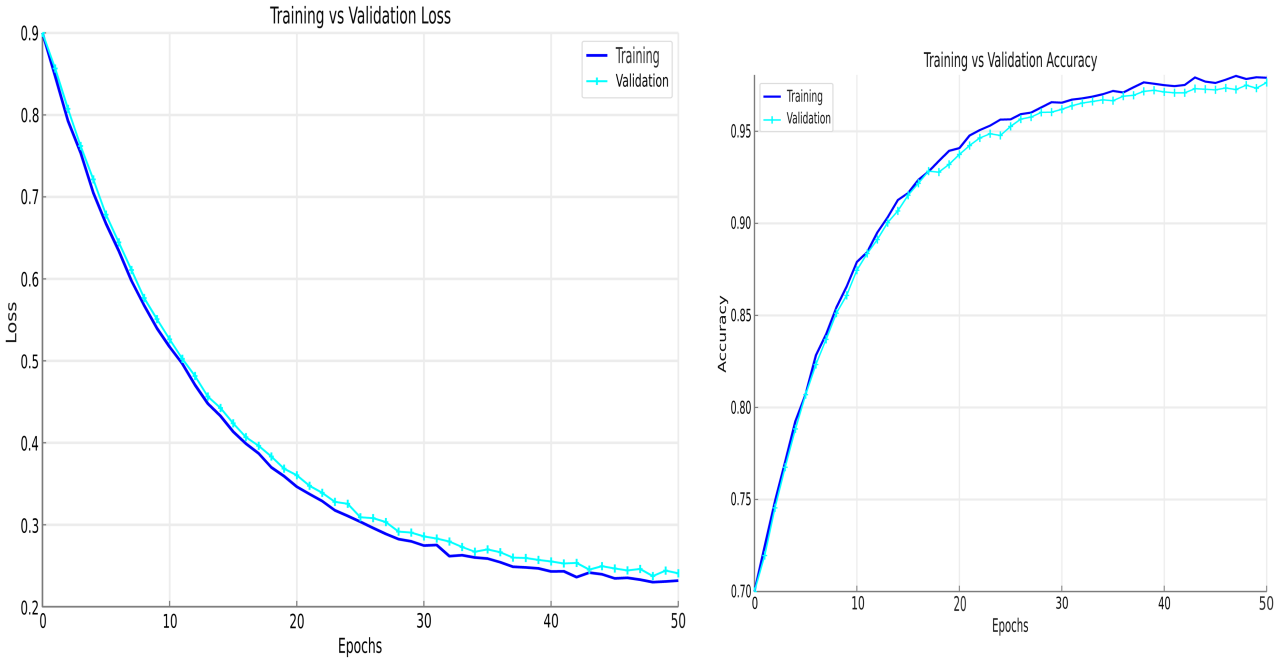


Figure 3. Training and validation accuracy and loss curves of the proposed model during pharyngitis classification training used in: training performance analysis

training and 0.24 while validation converging to a satisfactory level. The accuracy rose from approximately 0.70 to 0.98 for training, and 0.97 for validation, which is similar to the accuracy achieved in the cross-validation of 97.25%. The similarity between the training and validation curve implies stable learning, good generalization and less overfitting when it comes to the classification of pharyngitis. The confusion matrix of the proposed model for the classification of pharyngitis shows the prediction performance analysis of the two classes in detail. A confusion matrix is provided to indicate the correct classification of the samples according to the output of the model. The confusion matrix is shown below, in which the 168 samples of no pharyngitis were classified correctly as such, whereas the 17 samples of no pharyngitis were misclassified as pharyngitis. Furthermore, 124 samples of Pharyngitis were correctly classified as pharyngitis while 3 samples of Pharyngitis were misclassified as no pharyngitis. The true negatives are 168, the false positives are 17, the true positives are 124, and the false negatives are 3. From the confusion matrix, it is seen that most of the samples are accurately classified and only a few misclassifications are noted. The low false negative value of 3, means that the system has a high ability to detect Pharyngitis cases.

4.1. External evaluation

An independent external evaluation was then performed on an independent external dataset, the PGUPharyngitis dataset, which comprises of 742 throat images that are labeled for the classification of bacterial versus nonbacterial pharyngitis, made available publicly. All external images were resized to 224 × 224 pixels and tested with the same image-only pipeline, with no training or hyperparameter tuning of the model from the external dataset. Table 4 summarizes the data for the external data set: class distribution, imaging devices, annotation protocol, and demographic coverage. Table 5 shows the comparative performance of the proposed model and baseline architectures on this independent dataset. The best overall performance was obtained by the proposed model, which achieved an accuracy of 95.15%, precision of 96.93%, recall of 94.42%, F1 score of 95.68%, AUC of 0.9722, and specificity of 94.55%. The results show that the proposed dual-branch hybrid architecture kept excellent discriminative ability on a separated dataset of pharyngitis cases and benefitted from a competitive performance when compared to several transformer-based baselines, using fewer parameters.

4.2. Explainability and LLM report analysis

The integrated Qwen LLM recommendation system generates a clinical report as shown in Figure 4. The report contains the forecasted disease state, disease treatment recommendation, and follow-up guidance. Also, the report contains a clinical safety disclaimer to provide appropriate interpretation of the automated output. Depending on the classification result, the structured

Table 4. Characteristics of the independent PGUPharyngitis dataset used for external evaluation

Attribute	Details
Dataset name	PGUPharyngitis
DOI	10.6084/m9.figshare.28163513
Repository	Figshare (Springer Nature)
Published	2025
License	CC BY 4.0
Total images	742
Classes	2 (Bacterial / Nonbacterial)
Bacterial	430 / 742 (57.9%)
Nonbacterial	312 / 742 (42.1%)
Image type	Optical throat photography
Capture device	Samsung Galaxy S21 Ultra; Xiaomi Redmi 8 Pro
Original resolution	Variable (cropped; not standardized)
Input resolution used	224 × 224 pixels
Sex distribution	Male: 383 (51.6%); Female: 359 (48.4%)
Age range	0–100 years (multi-age cohort)
Annotators per image	4–9 physicians (mean: 6)
Labeling protocol	Majority vote
Collection period	October 2023 – May 2024
Collection sites	Two cities, Iran (warm coastal and cold mountainous regions)
Modality used	Images only (clinical symptoms excluded)
Validation split	742 images (hold-out external test set)

Table 5. External evaluation performance comparison on the independent PGUPharyngitis dataset

Method	Acc (%)	Prec (%)	Recall (%)	F1-score (%)	AUC	Spec (%)	Params (*M*)
GhostNetV2	88.41	91.55	88.14	89.07	0.9124	88.78	6.2
CSWin Transformer	91.24	93.76	90.93	91.25	0.9387	91.67	23.4
ConvNeXt	92.32	94.51	92.09	93.51	0.9509	92.63	28.6
Focal Transformer	93.67	95.49	93.49	94.70	0.9622	93.91	28.9
Proposed Model	95.15	96.93	94.42	95.68	0.9722	94.55	15.3

report will include medically relevant information related to no pharyngitis or pharyngitis. To include the retrieved medical knowledge, we use retrieval-augmented generation, allowing us to achieve context-aware recommendation generation. The integration facilitates the delivery of clinically useful information from the classification output. Structured clinical report of medical guidance that can help healthcare professionals with decision-making. The classification, explainability, and recommendation not only brings transparency but also boosts the clinical trust and enables better practical usability of the proposed framework.

The proposed model is visualized using feature maps in Figure 4, which gives an insight into the internal feature extraction process for the classification of pharyngitis. The feature map shows the most significant parts of the image of the throat that are involved in the classification decision. The model targets clinically relevant features, as shown by the high activation regions in anatomical areas that are relevant to the disease. In case of pharyngitis, the activation is limited to areas of apparent inflammation, redness, and tonsillar swollen areas. These areas correspond with clinical signs of pharyngitis seen on physical exam. The feature map illustrates the model's ability to focus and detect pathological areas without picking up irrelevant background areas. This visualization shows that the features learned by the model are consistent with the internal decision-making process of the model and that the learned features correspond to the clinical characteristics. Consequently, the feature map facilitates the interpretation of the model and enables classification of pharyngitis in a way that is explainable (with visual evidence of the model reasoning).

The Grad-CAM heatmap is displayed in Figure 5 along with the image of the throat overlaid on the heatmap. Highlighted areas are those that played an important role in the prediction of the model, such as parts of the tonsillar region, posterior pharyngeal wall and adjacent inflammatory mucosa. Clinically plausible because erythema and swelling, or exudative appearance are considered in the examination of the throat. But these visual markers are not specific to a particular cause of pharyngitis and can be present in bacterial, viral, inflammatory, and non-infectious diseases. Grad-CAM is a coarse localization method, that only provides a coarse, post hoc explanation of the location of the model's decision, and should not be interpreted as a causal explanation or a lesion boundary. It is important to note that an anatomically relevant region of activation in itself does not prove the clinical accuracy of the prediction, since image contrast,



Figure 4. LLM-generated clinical report showing predicted disease condition and treatment recommendation using retrieval-augmented generation and Qwen large language model

cropping, illumination etc. may all be correlated with the region of activation and cause it to appear in a different area. In this study, quantitative localization in clinician annotated regions and independent evaluation by otolaryngologists was not performed. Thus, the Grad-CAM visualization should be used as a "transparency" tool to look at how a model is behaving, but not as absolute evidence of either its correctness or reliability in clinical practice.

4.3. Potential domain shift and robustness under real-world imaging conditions

The proposed model had an accuracy of 97.25% in internal cross-validation on PGUPharyngitis dataset and an accuracy of 95.15% on the same dataset. This performance degradation might be aligned with differences between the datasets, but domain shift wasn't explicitly measured. The cross-dataset performance may be affected by differences in the model of the smartphone, lighting, flash power, white balance, resolution of the image, compression, cropping, distance between the patient and the camera, angle of the camera, position of the tongue, saliva reflections, patient characteristics, annotation protocols, or diagnostic criteria. In addition, the internal dataset classifies no pharyngitis as no pharyngitis, while PGUPharyngitis classifies it as either bacterial or nonbacterial. The classification targets cannot be directly equated with each other so the external analysis must be treated with care and the handling process of the labels must be well explained. There is some initial cross-dataset evidence of promise, but this does not guarantee generalizability. Throat photographs of real individuals may also have sub-optimal lighting, motion blur, uneven lighting, or occlusion by the tongue or dental structures, may not show the entire pharyngeal area, may not be taken from the correct angle or distance, may be compressed, and may vary depending on the camera type. The rotation, flipping, and augmentation in brightness during training only limitedly simulate variations in acquisition but does not set the robustness



Figure 5. Feature map and overlay visualization highlighting model attention regions for explainable pharyngitis classification

under uncontrolled clinical/telemedicine conditions. It is therefore important to take into account the development of automated image quality assessment, the minimum visibility and sharpness standards, rejection of poor quality images, uncertainty-based abstention and inability to assess output, standardized instructions for image acquisition, and prospective testing with acquired images in normal practice.

4.4. Reliability and maintenance of the RAG knowledge base

The trustworthiness of the report generated by RAG will also rely on the pertinence of the retrieved information and the management of the knowledge base. In this prototype, the structured query is used to fetch contextual information from the Qdrant vector database and the fetched content is provided to the Qwen LLM to generate a report. This process might enhance the context grounding, but retrieval does not ensure factual accuracy, clinical fullness, or freedom from errors associated with the synthesis of language models. The output can still be affected by irrelevant retrieval, missing evidence, conflicting recommendations and outdated clinical guidance. The knowledge base needs to be limited to authoritative, verifiable, medical sources with established source, publication and update dates, as well as version information, prior to being used in clinical practice. It would be necessary to have periodic medical examination, to replace old guidance and re-index/re-embed as a result of source changes. Retrieved-source identifiers, versions of the knowledge base, and generated reports should also be kept for audit, and supporting sources should be visible to the clinician who is reviewing the generated report. The output generated should not be prescriptive and should not include recommendations for antibiotics or medications without supporting evidence and should be reviewed by humans before making impact on patient care.

5. Limitations and Future Work

The proposed structure is shown to be performing well for automated classification of pharyngitis, but there are some limitations that need to be taken into consideration. First, the number of images in this dataset is 312, and the leakage-safe cross-validation was used, but the robustness and representation ability of the model may also be better with increasing the number of images. Second, the present framework only uses throat images for the classification. However, the use of image-based analysis alone cannot give all the information that is necessary to make a diagnosis and use of other clinical data, like the patients' symptoms or history or laboratory findings, may still improve the diagnostic accuracy and clinical relevance. Third, external validation with independent data sets was not done; multi-center validation could be useful to evaluate the generalizability of the proposed model in other clinical settings. Besides that, the quality and completeness of the knowledge retrieved from RAG is crucial to the recommendations generated by the Qwen LLM. Thus, it is necessary

to validate the clinical experts before implementing it in a real clinical context to guarantee safety and appropriateness.

These limitations will be tackled, and the scope of the proposed system will be expanded in future. Larger datasets and a greater variety of datasets will be taken into consideration for robustness and generalization. The framework can also be expanded to include categorization of other throat diseases, other than pharyngitis categorization. The system will be validated clinically to assess it in real-life healthcare settings. Real time use will be enabled by integration with clinical decision support platforms. To make more improvements in RAG, the following developments will take place: Expert validated knowledge bases will be developed to increase the reliability of recommendations. Furthermore, multimodal information such as clinical symptoms and patient data could be added to the future work that will enhance the diagnostic support. These advancements will be expected to increase the overall system reliability and usability in the clinic.

6. Conclusion

The presented study incorporated a research-stage framework for automated binary pharyngitis screening with an image of the throat, a dual-branch hybrid deep learning model, explainable artificial intelligence, and clinical recommendation supporting via RAG. The designed architecture integrates two types of visual information through two stages: 1) the local feature extraction is achieved by applying separable convolution and 2) the contextual information is acquired by the attention mechanism. The framework aimed to transcend the mere classification task and establish an interpretable decision support process for initial pharyngitis evaluation by incorporating more than just image classification, encompassing image visualization and a Qwen LLM for recommendations. The proposed model was tested on a set of 312 throat images which were classified into no pharyngitis and pharyngitis using leakage-safe 5-fold cross-validation. The model achieved an accuracy of 97.25%, precision of 97.58%, recall of 97.43%, and F1-score of 97.50%. The results show that the combination of the two branches can successfully classify local and global features in the images, and that this dual-branch model can be used to learn discriminative features from the images that are relevant to the pharyngitis screening problem, in the dataset evaluated. Use of leakage-safe cross-validation further enhances the reliability of the internal evaluation, as there is less risk of leakage between train and test information. The framework not only classifies the images but also provides explainable artificial intelligence for the purpose of identifying image areas that are important for the model prediction. This visualization aspect enhances transparency by enabling users to check if the model includes clinically relevant areas in the throat. The RAG module then enhances the system by using the Qwen LLM to retrieve the relevant medical context and generate a constructed informational recommendation report. This integrated design can aid clinician-reviewed telemedicine triage or initial screening workflows, when well validated. But these results need to be taken with a grain of salt. The study was conducted with a small single public dataset, and there was no external multi-center validation or prospective clinician-reviewed testing. The proposed system will NOT make clinical diagnosis, bacterial or viral etiology determination, antibiotic prescription or be a substitute for physician judgment. The recommendations also need to be validated by medical experts in order to be used clinically by the RAG. Future research is needed to apply a larger and more diverse set of data, validate externally across clinical environments, assess in a prospective fashion with healthcare workers, and review the recommendation module by experts. The system is not yet validated as a clinical diagnostic system; it can be regarded as a simulation-based medical AI research system until the process of validation is complete.

Ethical Statement

This study involved secondary analysis of publicly available human throat photographs from the Pharyngitis Images Dataset and PGUPharyngitis. The authors did not recruit participants, collect new patient images, perform clinical interventions, or conduct animal experiments.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

We confirm that the image datasets described in this study are publicly available in the following repositories: 1. Pharyngitis Images Dataset Kaggle: https://www.kaggle.com/datasets/bebeaa/pharyngitis-images-dataset. 2. PGUPharyngitis Figshare: https://doi.org/10.6084/m9.figshare.28163513.

Author Contribution Statement

MD Mizanur Rahman: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Md Fakrul Alam:** Validation, Formal analysis, Data curation. Md Azharul Islam: Methodology, Validation, Formal analysis. **Mazharul Islam:** Data curation, Writing – review & editing. **Md Sumon Rana:** Formal analysis, Writing – review & editing. **Md Firoz Kabir:** Validation, Formal analysis, Writing – review & editing. **Irfan Sadiq Rahat:** Software, Writing – original draft, Supervision.

References

[1] Anselmi, F., Dusser, P., & Kone-Paut, I. (2025). Periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome in children—from pathogenesis to treatment strategies: A comprehensive review. *Pediatric Drugs*, *27*(5), 575–592. https://doi.org/10.1007/s40272-025-00699-1

[2] Mihrete, K., Awoke, T., Melake, A., Sisay, T., & Abera, B. (2026). Bacterial pharyngitis, antimicrobial susceptibility, and associated factors among children at comprehensive specialized hospitals, Northwest Ethiopia. *BMC Infectious Diseases*, *26*(1), 235. https://doi.org/10.1186/s12879-025-12427-8

[3] Pellegrino, R., Timitilli, E., Verga, M. C., Guarino, A., Iacono, I. D., Scotese, I., . . . , & Chiappini, E. (2023). Acute pharyngitis in children and adults: Descriptive comparison of current recommendations from national and international guidelines and future perspectives. *European Journal of Pediatrics*, *182*(12), 5259–5273. https://doi.org/10.1007/s00431-023-05211-w

[4] Hossain, M. A., Rahman, M. A., Hossain, M. S., Karmakar, C., Rahman, M. S., Shekhor, K. C., . . . , & Basit, A. (2026). PharynFuseFormer supports image based pharyngitis screening from throat photographs using transformer CNN feature fusion. *Discover Computing*, *29*(1), 311. https://doi.org/10.1007/s10791-026-10220-y

[5] Hamilton, J. L., & McCrea II, L. (2024). Streptococcal pharyngitis: Rapid evidence review. *American Family Physician*, *109*(4), 343–349.

[6] Fajriyani, Aliyah., I Gede Pasek Suta, Wijaya., & Bimantoro, Fitri. (2025). Classification of throat disease using CNNs: EfficientNetB0 and ResNet50. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, *9*(2), 144–153. https://doi.org/10.29303/jcosine.v9i2.661

[7] Chng, S. Y., Tern, P. J. W., Kan, M. R. X., & Cheng, L. T.-E. (2024). Deep learning model and its application for the diagnosis of exudative pharyngitis. *Healthcare Informatics Research*, *30*(1), 42–48. https://doi.org/10.4258/hir.2024.30.1.42

[8] Meghla, H. M., Rahim, A., Shehan, M. S. H., Melon, M. M. H., Shahiduzzaman, M., Alam, M. F., . . . , & Tesfaye, A. (2026). SeAttnFusionNet for pharyngitis versus no pharyngitis classification from RGB throat images with grad-CAM-based interpretability. *Discover Analytics*, *4*(1), 22. https://doi.org/10.1007/s44257-026-00073-6

[9] Shojaei, N., Rostami, H., Barzegar, M., Farzaneh, S. S., Farrar, Z., Alimohammadi, M., . . . , & Gonbadi, L. (2025). A publicly available pharyngitis dataset and baseline evaluations for bacterial or nonbacterial classification. *Scientific Data*, *12*(1), 1418. https://doi.org/10.1038/s41597-025-05780-5

[10] Tu, H., Chen, M., Zhu, P., Hong, J., & He, L. (2026). Global research trends and knowledge gaps in chronic pharyngitis: A bibliometric analysis (2007–2025). *European Archives of Oto-Rhino-Laryngology*, *283*(7), 4171–4186. https://doi.org/10.1007/s00405-026-10135-w

[11] Leung, A. K. C., Lam, J. M., Barankin, B., Leong, K. F., & Hon, K. L. (2024). Group A β-hemolytic streptococcal pharyngitis: An updated review. *Current Pediatric Reviews*, *21*(1), 2–17. https://doi.org/10.2174/1573396320666230726145436

[12] Van Der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis*, *79*, 102470. https://doi.org/10.1016/j.media.2022.102470

[13] Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S. W., Karthikesalingam, A., King, D., . . . , & Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *npj Digital Medicine*, *4*(1), 65. https://doi.org/10.1038/s41746-021-00438-z

[14] Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine*, *28*(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

[15] Mylonas, I. (2025). Pharyngitis. In In *I. Mylonas, Sexually transmitted diseases* (pp. 147–152). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78204-6_10

[16] Li, Z., Huang, J., & Hu, Z. (2019). Screening and diagnosis of chronic pharyngitis based on deep learning. *International Journal of Environmental Research and Public Health*, *16*(10), 1688. https://doi.org/10.3390/ijerph16101688

[17] Coronel, E. J. D., Mababangloob, M. N., & Balbin, J. R. (2021). Detection of outer throat infection using deep convolutional neural network. In *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 1–6. https://doi.org/10.1109/HNICEM54116.2021.9731949

[18] Wijayanto, I., Irawati, I. D., Fahrozi, F., & Hadiyoso, S. (2022). Computer aided detection system for pharyngitis based on convolutional neural network. In *2022 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*, 1–4. https://doi.org/10.1109/APWiMob56856.2022.10014256

[19] Tohit, N., & Shamugam, M. (2025). Acute pharyngitis. In Y. Swarna Nantha (Ed.), *Screen, diagnose, manage and monitor* (pp. 787–792). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-7640-5_89

[20] Swathi, M., & Regunathan, R. (2024). A multi-fusion approach to classify pharyngitis, tonsillitis and oral cancer with iterative relief feature weighting. *Alexandria Engineering Journal*, *104*, 359–370. https://doi.org/10.1016/j.aej.2024.06.027

[21] Jeng, P.-H., Yang, C.-Y., Huang, T.-R., Kuo, C.-F., & Liu, S.-C. (2024). Harnessing AI for precision tonsillitis diagnosis: A revolutionary approach in endoscopic analysis. *European Archives of Oto-Rhino-Laryngology*, *281*(12), 6555–6563. https://doi.org/10.1007/s00405-024-08938-w

[22] Samsir, S., Daulay, N. K., Harahap, S. Z., Zalmi, W. F., Sari, A. N., Nasution, F. A., & Watrianthos, R. (2022). Implementation learning vector quantization using neural network for classification of ear, nose and throat disease. *Journal of Physics: Conference Series*, *2394*(1), 012016. https://doi.org/10.1088/1742-6596/2394/1/012016

[23] Muhammad, D., & Bendechache, M. (2024). Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis. *Computational and Structural Biotechnology Journal*, *24*, 542–560. https://doi.org/10.1016/j.csbj.2024.08.005

[24] Luo, R., Sickler, J., Vahidnia, F., Lee, Y.-C., Frogner, B., & Thompson, M. (2019). Diagnosis and management of group A streptococcal pharyngitis in the United States, 2011–2015. *BMC Infectious Diseases*, *19*(1), 193. https://doi.org/10.1186/s12879-019-3835-4

[25] Mustafa, Z., & Ghaffari, M. (2020). Diagnostic methods, clinical guidelines, and antibiotic treatment for group A streptococcal pharyngitis: A narrative review. *Frontiers in Cellular and Infection Microbiology*, *10*, 563627. https://doi.org/10.3389/fcimb.2020.563627

[26] Curatola, A., Di Sarno, L., Massese, M., Caroselli, A., Gatto, A., & Chiaretti, A. (2024). Diagnosis and treatment of group A streptococcal pharyngitis in children. *Pediatric Annals*, *53*(6). https://doi.org/10.3928/19382359-20240407-08

[27] Borys, K., Schmitt, Y. A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C. M., & Nensa, F. (2023). Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches. *European Journal of Radiology*, *162*, 110787. https://doi.org/10.1016/j.ejrad.2023.110787

[28] Champendal, M., Müller, H., Prior, J. O., & Sá Dos Reis, C. (2023). A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging. *European Journal of Radiology*, *169*, 111159. https://doi.org/10.1016/j.ejrad.2023.111159

[29] Shapiro, D. J., Barak-Corren, Y., Neuman, M. I., Mandl, K. D., Harper, M. B., & Fine, A. M. (2020). Identifying patients at lowest risk for streptococcal pharyngitis: A national validation study. *The Journal of Pediatrics*, *220*, 132–138.e2. https://doi.org/10.1016/j.jpeds.2020.01.030

[30] Kline, M. C., Kissler, S. M., Whittles, L. K., Barnett, M. L., & Grad, Y. H. (2024). Spatiotemporal trends in group A

streptococcal pharyngitis in the United States. *Clinical Infectious Diseases*, *78*(5), 1345–1351. https://doi.org/10.1093/cid/ciae083

[31] Yang, R., Ning, Y., Keppo, E., Liu, M., Hong, C., Bitterman, D. S., . . . , & Liu, N. (2025). Retrieval-augmented generation for generative artificial intelligence in health care. *npj Health Systems*, *2*(1), 2. https://doi.org/10.1038/s44401-024-00004-1

[32] Wang, A., Manthiram, K., Dedeoglu, F., & Licameli, G. R. (2021). Periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) syndrome: A review. *World Journal of Otorhinolaryngology - Head and Neck Surgery*, *7*(3), 166–173. https://doi.org/10.1016/j.wjorl.2021.05.004

[33] Bakhit, M., Gamage, S. K., Atkins, T., Glasziou, P., Hoffmann, T., Jones, M., & Sanders, S. (2024). Diagnostic performance of clinical prediction rules to detect group A beta-haemolytic streptococci in people with acute pharyngitis: A systematic review. *Public Health*, *227*, 219–227. https://doi.org/10.1016/j.puhe.2023.12.004

[34] Schmutzler, L., Mirna, M., Hoppe, U. C., & Lichtenauer, M. (2022). From streptococcal pharyngitis/tonsillitis to myocarditis: A systematic review. *Journal of Cardiovascular Development and Disease*, *9*(6), 170. https://doi.org/10.3390/jcdd9060170

[35] Melon, M. M. H., Islam, M. M., Rahman, M. M., Miah, N., Rahman, M. M., Rabby, M. F., . . . , & Shahbazi, M. (2026). ClinicGraphRAG knowledge graph and LLM retrieval-augmented generation for reliable clinical decision support summaries. *Discover Artificial Intelligence*, *6*(1), 833. https://doi.org/10.1007/s44163-026-01492-w

How to Cite: Rahman, M., Alam, F., Islam, A., Islam, M., Rana, S., Kabir, F., & Rahat, I. S. (2026). Automated Pharyngitis Classification Using a Dual-Branch Hybrid Deep Learning Model with Explainable Artificial Intelligence and Retrieval-Augmented Generation-Based Clinical Recommendation. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN620210119