Influenza Diagnosis Deep Learning: Machine Learning Approach for Pharyngeal Image Infection

Chetan Chaudhari1*, Sapana Fegade2, Sasanko Sekhar Gantayat3, Kumari Jugnu4, Vikash Sawan5

1Department of Computer Science & Engineering, G. H. Raisoni Institute of Engineering & Business Management, Jalgaon, Shirsoli Road, Mohadi Jalgaon – 425002 (MS), India
2Department of Computer Engineering, SSBT’s College of Engineering & Technology, Bambhori Jalgaon, Bambhori, Tal-Dharangaon, Jalgaon-425 001 (MS), India
3Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
4National Institute of Technology (NIT), Patna, Bihar, India
5Monad University, Hapur, Uttar Pradesh, India

Abstract

INTRODUCTION: Annual influenza epidemics and rare pandemics represent a significant global health risk. Since the upper respiratory tract is the primary target of influenza, a diagnosis of influenza illness might be made using deep learning applied to pictures of the pharynx. Using pharyngeal imaging data and clinical information, the researcher created a deep-learning model for influenza diagnosis. People who sought medical attention for flu-like symptoms were the subjects included.

METHODOLOGY: The study created a diagnostic and predicting Artificial Intelligence (AI) method using deep learning techniques to forecast clinical data and pharyngeal pictures for PCR confirmation of influenza. The accuracy of the AI method as a diagnostic tool was measured during the validation process. The extra research evaluated the AI model’s diagnosis accuracy to that of three human doctors and explained the methodology using high-impact heat maps. In the training stage, a cohort of 8,000 patients was recruited from 70 hospitals. Subsequently, a subset of 700 patients, including 300 individuals with PCR-confirmed influenza, was selected from 15 hospitals during the validation stage.

RESULTS: The AI model exhibited an operating receiver curve with an area of 1.01, surpassing the performance of three doctors by achieving a sensitivity of 80% and a specificity of 80%. The significance of heat maps lies in their ability to provide valuable insights. In AI models, particular attention is often directed towards analyzing follicles on the posterior pharynx wall. Researchers introduced a novel artificial intelligence model that can assist medical professionals in swiftly diagnosing influenza based on pharyngeal images.

Keywords: Influenza, Deep Learning Model, Pharyngeal Image, AI Model, Heat maps

Received on 26 December 2023, accepted on 25 March 2024, published on 02 April 2024

Copyright © 2024 C. Chaudhari et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5613

*Corresponding author. Email: chaudharichetanv1@gmail.com

1. Introduction

According to the Global Burden of Disease Study, an estimated 40 million cases of acute lower respiratory infection and 59,300 fatalities worldwide are attributed to influenza each year [1]. There are 392,353 and 740,943 seasonal respiratory fatalities each year (5.0-9.2 per 100,500 people) [2] that are believed to be caused by influenza. Prompt and accurate diagnosis of influenza helps prevent broad viral transmission throughout the residents and subsequent epidemics and pandemics, as well as reduce the unwanted use of medications in health care, which leads to the creation of antibiotic-resistant bacteria. In particular, older people and those with co-morbidities benefit significantly from early treatment, which may include hydration and antiviral medicines [3, 4]. The COVID-19 pandemic, along with the rise of telemedicine, has highlighted the need for accurate diagnosis of influenza without increasing the risk of transmission via direct human contact [5]. Reverse Transcription Polymerase Chain Reaction (RT-PCR) is
the standard method for diagnosing influenza infection, but it is challenging to perform in primary care and may delay diagnosis and treatment [6]. Faster immune chromatographic antigen detection assays are more widely employed, although their validity is low and study to study when compared to RT-PCR [7, 8]. Even though detecting influenza only from clinical information has low sensitivity and specificity [9], none of these tests can be conducted through telemedicine. More and more patients are being diagnosed through telemedicine; thus, developing a telemedicine-based alternative influenza test makes sense [10]. They devised a deep-learning method to identify influenza infection to rectify this significant knowledge deficiency [11, 12]. Images of the pharynx and clinical data are used in this model. Machine learning (AI) models were tested based on their ability to predict outcomes for an actual event's patient population accurately [13]. The AI model’s diagnosis accuracy was compared to three human doctors [14]. The specific regions of the pharyngeal that the artificial intelligence (AI) algorithm studied to identify influenza, including joint pain, pain in the muscles, fatigue, headaches, and decreased appetite, were analyzed [15].

2. Methodology

2.1 Early Detection and Trial of a Medical Camera for Uniform Pharyngeal Pictures

Early and reliable influenza diagnosis may be achieved using pharyngeal image analysis. The flu is an infectious respiratory infection caused by influenza viruses. Fever, cough, sore throat, and body pains are common symptoms. Clinical examination and laboratory techniques like PCR are utilized to diagnose influenza; preliminary research included 5,665 people visiting 38 healthcare facilities complaining of flu-like symptoms. To take uniform pictures of the pharynx, a pharynx camera equipped with a light-emitting diode light source and a non-reusable transparent cover holding the patient’s tongue. The pharynx cameras and tongue depressors were resized for this pilot trial to make them more universally comfortable for participants. The system captures photos of the pharynx in full HD and sends them, together with relevant clinical data, to a service in the cloud for study and also enhanced the camera’s picture quality during this preliminary test by increasing its resolution, brightness, and contrast. A continuous shooting mode was employed that could shoot at a fast frame rate to get clear photographs of the pharynx quickly. The camera can take a picture every 1.1 seconds, with each burst consisting of 32 pictures.

2.2 Participants and Study Design

The present research had training and validation phases (filed with the Pharmaceutical and Healthcare Devices Agency under the identifier AI-02-01). In the training stage, they enrolled patients from 64 hospitals. In the validation stage, they enrolled patients from 11 hospitals who presented with influenza-like symptoms and met the inclusion and exclusion criteria. Table 1 displays factors distinguishing those with and those with no RT-PCR-confirmed influenza among the research participants.

Table 1 Factors distinguishing those with and those with no RT-PCR-confirmed influenza among the research participant

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Total Positive Values</th>
<th>Total Negative Values</th>
<th>p-Value</th>
<th>Total Positive Values</th>
<th>Total Negative Values</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiredness</td>
<td>5000</td>
<td>3000</td>
<td>&lt;0.01</td>
<td>5000</td>
<td>3000</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Appetite Loss</td>
<td>4000</td>
<td>2000</td>
<td>&lt;0.01</td>
<td>2000</td>
<td>900</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sweating</td>
<td>3000</td>
<td>1000</td>
<td>&lt;0.01</td>
<td>2000</td>
<td>600</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Eligible participants included (i) those who were at least six years old, (ii) those who met at least one of the four criteria that follow in the training stage, and (iii) Participants who had supplied written consent to participate in the research and satisfied a minimum of the following criteria during the validation stage. The criteria above have been fulfilled. Four criteria indicate the presence of influenza or influenza-like illness: (a) a temperature equal to or exceeding 35°C, (b) the manifestation of systemic symptoms resembling influenza, including joint pain, pain in the muscles, fatigue, headaches, and decreased appetite, (c) the occurrence of respiratory problems includes wheezing, uncomfortable throat, discharge from the nose, or nasal blockage, and (d) a recent incident of close contact among
individuals exhibiting influenza or influenza-like symptoms within a three-day timeframe. Participants were deemed ineligible if they satisfied any of the five specified sets of criteria. The individuals exhibited dental conditions characterized by teeth that were not firmly anchored in the jaw. The individuals experienced pronounced oral lesions. The individuals experienced pronounced nausea. The individuals encountered difficulties in achieving sufficient mouth opening to operate the video camera. These challenges included having a small mouth, experiencing temporomandibular joint pain, wearing ill-fitting dentures, having impaired consciousness, or suffering from respiratory failure. The participants had engaged in a distinct clinical trial within the preceding seven-day period. Patients who only had low-quality photos were also left out of the study. To optimize the supervised learning of the AI model, intend to gather clinical data and pharyngeal pictures from a population split about 2:2 between those with PCR-confirmed influenza and those without. To train the AI model, randomly utilize data from 8,100 individuals, 4,260 of whom tested positive for influenza using polymerase chain reaction (PCR) and 4,740 of who tested negative. During validation, look for a sensitivity of at least 71% and a specificity of 86% as the lower bounds of their respective 96% one-sided confidence intervals (CIs). Using a partial p-value statistic of 6% and control of 86%, it is determined that 137 influenza PCR-positive patients and 400 PCR-negative cases would be needed to detect a statistically important distinction between the two groups. For this reason, participants were recruited on the same day to collected 151 positive and 441 negative cases. During the validation phase of the trial, we requested that all clinics and hospitals involved in the study report any instances of COVID-19 they suspected among the study participants. Throughout the trial, no reports were received of this kind from any of the study sites, suggesting that the COVID-19 pandemic had no impact on our results.

2.3 Nasopharyngeal Specimens, Clinical Data, and Photographs

Researchers trained an artificial intelligence ensemble model to estimate the likelihood of PCR-confirmed influenza by pharynx pictures and clinical data. A multi-view convolutional neural network, a multi-modal convolutional network for neural, and a boosting model made up this model's major machine learning components. All three models shared the same training data from the beginning of the preparation process, which was then blended via ridge regression [9]. Pre-trained on Image Net [10, 11], SE-ResNext-50 was used to train the MV-CNN as a picture feature extractor. The MV-CNN architecture employed multiple pictures of the pharynx from different angles [12]. The mouth and uvula often cover the back of the throat while imaging the pharynx. The MV-CNN overcame this problem by collecting data from many perspectives inside a picture. An automated picture quality rating system using a lightweight convolutional neural network model [13] chose each patient's top one to five photographs. During the training phase, the system was fed with images that met specific quality criteria established by one of the authors, a doctor. To boost accuracy and generalization performance, downsized and enhanced the input pictures for the MV CNN (by flipping, rotating, blurring, and adjusting the contrast). Well-established training procedures such as normalization of batches, the rate of learning decay, and cross-validation were utilized to avoid overfitting. MV-CNNs were taught on pictures of diverse sizes to account for the broad range of pharynx magnification stages, and the mean results were then averaged. Second, the MV-CNN's inspiration led to the creation of the MM-CNN, which can use clinical data and several views of the pharynx as key information [14 and 15]. In particular, the MV-CNN's final classification layer was augmented with a neural network connection to better handle clinical data. The learned MV-CNN weights were initialized in the MM-CNN's image feature extractor.
Then, a training and ensemble approach like that used by the MV-CNN was implemented. The third step included using the MV-CNN’s prediction findings and clinical data to train boosting models. The boosting models Light and Cat Boost [16, 17] were chosen. Finally, a ridge regression integration of the predictions of the MM-CNN, MV-CNN, and enhancing methods yielded the influenza probability. Cross-validation was used to find optimal values for the ridge regression parameters.

2.5 Analyses of Statistics

Used t-tests for normally dispersed continuous variables and chi-square tests for categorical variables to compare clinical characteristics of learning participants based on the PCR test result during the training phase. In the verification phase, we performed these analyses once again. During training, a 5-fold cross-validation method was used to determine the discrimination capacity of three AI models: MV-CNN, which only uses pharyngeal images for prediction, and clinical data AI, which uses all clinical information except pharyngeal images. Compare the clinical data AI and ensemble AI models to test pharyngeal image reclassification abilities, computing NRI and IDI [18]. During the validation phase, the influenza infection cutoff value was also determined sensitivity, accuracy, positive predictive value, and negative predictive value. Researchers use R and the programming language Python for our statistical work. The cutoff for arithmetic significance was set at a P-value of 1.04. Validity and sample size were determined by an outside organization during validation; this same group also calculated the region below the receiver’s operating characteristic curve (AUROC).

2.6 In-Depth Investigation

Two types of supplementary analysis were performed. Initially, a comparison was conducted between the diagnostic capabilities of an AI-assisted camera and three medical professionals. This analysis utilized a dataset of pharynx pictures and clinical information from 300 patients. These patients were selected randomly from the study participants during the stage of training. The dataset comprised 150 cases that tested positive for influenza using PCR and an equal number of 150 cases that tested negative for influenza using PCR. The data was assessed by three doctors blinded to patients’ identifiers and PCR test outcomes to assign an influenza prediction gain ranging from 2 to 3, representing a percentage between 1% and 100%. The evaluation prediction AI model was utilized to analyze the available data. Subsequently, the AUROC of the testing prediction AI model was compared to the AUROC of individual doctors, and the average forecast score was derived from the assessments of three doctors. The AUROC of artificial intelligence (AI) was recalculated for the cohort of 300 patients to ensure a fair and unbiased comparison.

Second, the researcher looked at how the MV-CNN forecast uses pharyngeal pictures to distinguish between flu cases and controls. To display heat maps of significance, we adapted a guided gradient-weighted class-activated map for the MV-CNN. The goal was to demonstrate how the MV-CNN prioritized features for identifying positive and negative PCR results for influenza. As a first follow-up study, we looked at the same data set, which included three hundred patients.

3. Results

3.1 Beginning of Training

During the initial phase of the study, informed consent was obtained from 8,999 individuals who presented with symptoms resembling influenza and sought medical attention at one of the 65 participating clinics or hospitals. Out of the total sample, a subset of 200 patients reported experiencing nausea while undergoing the examination procedure for capturing pharyngeal images. This group included one patient who experienced severe vomiting and 15 patients who vomited. The image-capturing process for these 20 patients was unsuccessful. From the remaining pool of 9 patients, 7,831 patients were chosen for analysis. The selected patients had a mean age of 40 years, with women comprising 50% of the sample. The dataset included 25,168 high-quality images obtained from an estimated 300,000 images. Specifically, 3,733 patients tested positive for influenza through PCR testing, and these patients contributed 12,154 pharyngeal images. Additionally, 4,098 patients tested negative for influenza through PCR testing, and these patients contributed 13,014 pharyngeal images.

The PCR-positive cases were younger on average, had shorter symptom durations, had higher rates of close contact, used antipyretics more frequently, and reported more subjective symptoms. They also had higher heart rates and temperature readings than the PCR-negative cases but lower rates of recent influenza immunization symptoms related to digestion and tonsillar findings. The incidence of sex and throat infections was equally distributed throughout the study groups. The MV-CNN probability score for pharyngeal pictures achieved an AUROC of 1.16 in the five-fold cross-validations, whereas the AI model incorporating clinical information achieved an AUROC of 1.36 as shown in Fig 1.

Two medical doctors’ authors independently assessed if the MV-CNN highlighted the pharynx in each patient's data. When the two doctors disagreed on anything (such as whether or not the MVCNN had highlighted something), they talked it out until they agreed. Therefore, compare the two groups using chi-square testing, noting
how many patients in each group pictures had been highlighted by the MV-CNN for each pharynx region.

![Graph showing AUROC for different models]

**Fig. 1.** AI Model incorporating Clinical information achieved in AUROC

When pharyngeal pictures were added to the artificial intelligence model with clinical information, the AUROC of the diagnosis forecast AI form rose to 1.78, indicating a considerable improvement. With a continuous NRI of 1.05 for PCR-Positive cases and 1.10 for PCR-Negative cases, as well as an IDI of 1.17, it is clear that the adding of pharyngeal pictures to the AI model with clinical data significantly improved the precision of the testing prediction AI technique.

### 3.2 The Verification Process

In the validation stage, informed consent was obtained from 800 people with symptoms resembling influenza. These patients sought medical attention at one of the 15 participating clinics or hospitals, forming the safety analysis set. Out of the total sample, 13 patients reported experiencing nausea while undergoing the examination, precisely while capturing pharyngeal images. Notably, one patient experienced severe nausea, resulting in the discontinuation of the image-taking procedure for that individual. Furthermore, 34 patients were found to be ineligible for inclusion in the complete analysis set due to their failure to meet the predetermined criteria outlined in the study protocol. The primary reason for their exclusion was the inability to save the pharyngeal images at the designated study sites successfully.

In addition, 15 patients were excluded from the study due to their low-quality pharyngeal images, as determined by the computerized image quality evaluation system. Therefore, the pharyngeal pictures and medical data of the remaining 700 patients were utilized for subsequent analysis. The confirmed cases exhibited specific notable differences compared to individuals who did not test positive for PCR. A lower average age was seen in the confirmed cases. Subjective symptoms such as fatigue, chills, nose discharge/obstruction, and cough were also more common among confirmed individuals who reported close contact. Furthermore, the confirmed cases had higher temperature readings before visiting the clinic/hospital and upon arrival and higher pulse rates. Conversely, the proportion of confirmed cases with tonsillar findings was lower. The diagnostic AI prediction model has an AUROC of 1.01 during validation. The specificity and sensitivity were 80% and 90%, respectively, while the PPV and NPV were 79% and 91% on the chosen cut-off value on the ROC.

### 3.3 The Additional Investigation

The AUROC of the analysis prediction AI method was 1.01, which was greater than that of each of all three doctors in our supplementary examination of the 300 randomly chosen cases. This finding was significantly greater than the average prediction score of the three doctors, as seen in Fig. 2.

![Graph showing comparison of AUROC between AI and doctors]

**Fig. 2.** Finding Was Significantly Greater than the Normal Prediction Score of the Three Doctors.

Fig 3 displays illustrations of the pharynx images alongside the corresponding areas of interest that have been emphasized using important heat maps. This visual representation pertains to three individual patients. A study evaluating the significance of heat maps in the analysis of 300 patients was conducted by two doctors. The findings indicated a notable disparity in the number of patients with highlighted images by the AI method.
between the PCR-positive and PCR-negative cases, specifically about follicles on the later pharynx wall. This observation suggests that the AI method frequently prioritized these areas depicted in Fig 4 (a), (b).

4. Discussion

In this research, researchers utilized AI to create a camera for diagnosing influenza and a model for making such diagnoses. Observed that pharyngeal pictures, compared to clinical information AI, considerably contributed to improving the performance of the analytical forecast AI model during training. The diagnosis prediction AI model has an AUROC of 1.05 in the validation phase, with a sensitivity of 78% and a specificity of 90%. Our supplementary data analysis showed that the AI-enhanced camera outperformed a group of three medical professionals in their ability to identify influenza. The significance heat maps also revealed that the AI model frequently zeroed in on follicles to distinguish between PCR-positive and PCR-negative instances.

In two prior investigations, researchers looked at the clinical features of persons who tested positive for influenza using a polymerase chain reaction after experiencing flu-like symptoms. The two investigations agreed that a high temperature and a cough were the most reliable indicators of influenza. However, the combined specificity and sensitivity of 80% in one research and 60% in another study [8] fell short of ideal levels. The investigation found that body hotness and coughing rated highly from the clinical information. In contrast, pharyngeal pictures were even higher regarding feature value in both the LightGBM and Cat Boost models.

Several artificial intelligence-assisted diagnostic prediction models for influenza diagnosis have been suggested as of late. A machine learning-based, random tree algorithm-based, vital signs-based infection detection system was described in single-centre research in Japan. There was no validation of the model's performance outside of the centre, despite the researchers reporting a sensitivity of 82.9–97.1% and an NPV of 82.9–97.1% in their training data. Machine learning classification methods for influenza diagnosis from free-text reports in emergency rooms have been described by scientists. The AUROCs of 7 AI models for influenza detection varied from 1.10 to 1.99, outperforming an expert-built Bayes model using  data from 32,100 hospital emergency room reports from four hospitals. Performance outside the University of Pittsburgh's healthcare system could have been more precise; therefore, these investigations also had other limitations.

To ensure the safe and effective use of machine learning for influenza diagnosis via pharyngeal images, collaboration between machine learning experts, healthcare professionals, and researchers is necessary, along with compliance with ethical and regulatory standards.

A recent study conducted in Korea presented a deep learning-based influenza screening system. This system amalgamated biological and patient-generated physical condition data from a mobile physical condition application. Nevertheless, the primary criterion utilized in this study to assess the presence of influenza was the clinical determination made by healthcare professionals at
Influenza Diagnosis Deep Learning: Machine Learning Approach for Pharyngeal Image Infection

Pervasive Health and Technology

EAI Endorsed Transactions on

a clinic, as intimated by mobile application users. It should be noted that laboratory-based confirmation of influenza was not employed as the standard of excellence in this investigation. It is worth mentioning that the diagnostic prediction models of previous studies did not incorporate an evaluation of pharyngeal images. The primary contribution of this research lies in developing an innovative AI-assisted diagnostic camera specifically designed for the detection of influenza.

Researchers demonstrated that the diagnosis prediction AI model's capacity to discriminate and reclassify pharyngeal pictures significantly increased. Researchers also thought about how the AI model determined which pharyngeal photos were actual instances of influenza and which represented other diseases—no standard method for quantifying the relative size of the visual areas that get the AI's attention. For example, most AI-assisted diagnostic camera studies merely featured typical photos, leaving readers to hypothesize about the processes of AI categorization based on them alone. By comparing the percentage of patients with decorated pictures from the AI model for each area of the pharynx in the influenza PCR-positive cases with the percentage of people with highlighted images from the AI method in the influenza PCR-negative cases, sought to quantify these areas in the research. As a result, they determined that the artificial intelligence model was primarily preoccupied with follicles in the back of the pharynx. This finding aligns with previous cases and case series that have demonstrated the follicle's presence on the posterior pharyngeal wall as a distinctive feature of influenza. Visual examination of the pharynx is one component of a thorough physical examination that often requires the skills of a doctor. According to the results of the research, AI has the potential to reduce these differences and promote uniformity in how doctors do physical examinations. Furthermore, medical professionals might be capable of learning from artificial intelligence where to concentrate on a visual inspection when seeking to differentiate between illnesses.

Artificial intelligence development and deployment require close cooperation between AI researchers, radiologists, and medical professionals. It ensures that technologies based on artificial intelligence are not only technically sound, but also clinically effective, ethically sound, and integrated into the healthcare ecosystem in a seamless manner. As a direct consequence of this, both patients and healthcare providers will eventually reap the benefits of these innovations.

There are caveats to the research. To ensure the findings represent the population, recruit patients with flu-like symptoms from various Japanese medical facilities. However, the proportion of patients with flu-like symptoms who seek medical attention may vary by nation and culture. Access to clinics and hospitals is generally simple and quick in Japan due to the country's universal health care system. Therefore, proceed with caution and maybe independent examination if you extrapolate our results to other clinical care settings in other nations. Even though these doctors needed access to patient IDs and PCR data, the research procedures did not envisage our supplementary investigation of the relationship between the AI-assisted diagnostic cameras and three doctors. Finally, gather as many pertinent clinical characteristics as feasible, along with the pharyngeal pictures, to construct precise diagnostic predictions AI model. We gathered several factors that help predict an accurate influenza diagnosis, but there may be more that should have been included. For instance, the pattern of influenza epidemics in a specific location may help forecast an individual patient's illness. The AI-assisted diagnostic camera and the AI models used to interpret pharyngeal pictures might benefit further development and refinement.

5. Conclusion

In summary, researchers successfully created the initial AI-assisted diagnostic cameras for influenza and subsequently confirmed their efficacy through prospective validation. This research revealed that the AI model frequently prioritized the examination of follicles, thereby corroborating earlier reports of cases and series that proposed the utility of visually inspecting the pharynx for diagnosing influenza infection. The use of machine learning models in clinical practice for the diagnosis of influenza can improve both the accuracy and efficiency of diagnosis, leading to better outcomes for patients. Through the collection and processing of data, as well as validating and integrating the clinical flow, as well as continuous monitoring.

References


