I.J. Image, Graphics and Signal Processing, **2025, 2, 47-61**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2025.02.03

A Novel Approach for Enhancing COVID-19 Diagnosis Accuracy through Graph Neural Networks Using Respiratory Sound Data

Nagaraju Sonti*
Department of ECE, Vasireddy Venkatadri Institute of Technology, Namburu, Pedakakani, Guntur, Andhra Pradesh
Email: nagarajusonti@gmail.com
ORCID iD: https://orcid.org/0000-0003-3336-6802
*Corresponding Author

Rukmini M. S. S.
Department of ECE, Vignan's Foundation for Science, Technology & Research, Vadlamudi, Andhra Pradesh
Email: mssrukmini@gmail.com
ORCID iD: https://orcid.org/0009-0006-8283-2020

Venkatesh Munagala
Department of ECE, Vasireddy Venkatadri Institute of Technology, Namburu, Pedakakani, Guntur, Andhra Pradesh
Email:Venkatesh.munagala@gmail.com
ORCID iD: https://orcid.org/0009-0009-8314-8898

Received: 31 January, 2024; Revised: 20 October, 2024; Accepted: 12 January, 2025; Published: 08 April, 2025

Abstract: This research presents a groundbreaking method using graph neural networks (GNN) for the accurate identification of COVID-19 through the analysis of respiratory sounds. The method utilizes advanced signal processing and machine learning techniques, including Fast Fourier Transforms (FFTs), Mel-spectrograms, and GNN methodology. FFTs are used as a preprocessing step to convert raw respiratory sound signals into frequency-domain representations, enhancing signal quality and isolating informative acoustic patterns. Mel-spectrograms are used to extract essential feature vectors for diagnostic classification, enhancing the model's ability to discern subtle patterns indicative of COVID-19 infection.

The GNN methodology feeds preprocessed audio features into a graph neural network architecture, which excels at capturing complex relationships and dependencies within data by modeling them as graphs. In this context, respiratory sound data is represented as a graph, with nodes corresponding to specific audio features and edges representing relationships between them. The GNN effectively learns to propagate information across the graph, enabling it to identify meaningful patterns indicative of COVID-19 infection. The research findings show that GNN surpasses convolutional neural network (CNN) in terms of accuracy, precision, recall, and F1 score, indicating significant progress in the application of GNN in medical diagnostics. The study provides a comprehensive examination of the possibilities of using advanced neural network techniques to transform disease detection and diagnosis, with a validation accuracy of up to 97% under rigorous constraints.

Index Terms: Convolutional Neural Network, COVID-19, Graph Neural Networks, Fast Fourier Transform, Mel-Spectrogram

1. Introduction

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, was officially declared a global pandemic by the World Health Organization (WHO) in March 2020 [1]. Since its onset, the virus has significantly impacted global healthcare systems, leading to millions of infections and challenging medical infrastructures. Current diagnostic methods, such as RT-PCR (Reverse Transcription Polymerase Chain Reaction) and antigen tests, though highly accurate, are invasive, time-consuming, and require specialized personnel [2]. These limitations make them unsuitable for large-scale, rapid testing, particularly in resource-constrained environments.

There is an increasing demand for non-invasive, scalable diagnostic tools that can be easily administered and deployed across different environments. Respiratory sound analysis offers a promising alternative for diagnosing COVID-19 by detecting abnormalities in coughs, breath sounds, and other respiratory patterns that may be indicative of infection [3]. Respiratory infections, including COVID-19, often affect the lungs and airways, leading to detectable acoustic changes in respiratory sounds, which may not be perceptible to the human ear but can be captured using advanced sound processing and machine learning techniques.

This paper introduces a novel approach that utilizes Graph Neural Networks (GNN) to analyze respiratory sounds for the detection of COVID-19. GNNs are particularly effective at modeling complex relationships between sound features, making them more suitable for this task than traditional methods like Convolutional Neural Networks (CNNs), which may struggle to process structured data of this kind. By employing Mel-spectrograms and Fast Fourier Transforms (FFT) to preprocess and extract key features from respiratory sounds, our model capitalizes on GNN's capability to capture subtle, graph-structured acoustic patterns unique to COVID-19 patients.

One of the major advantages of this method is its non-invasive nature, which eliminates the discomfort associated with traditional swab-based tests. This encourages more frequent testing and reduces the risk of viral transmission to healthcare workers during sample collection. In contrast to RT-PCR, which is considered the gold standard but can be slow and resource-intensive, this GNN-based approach offers a rapid, contactless, and scalable alternative [2]. Sound data can be easily collected using common devices such as smartphones or digital recorders, making it suitable for large-scale screening, even in remote or resource-constrained areas.

The dataset used in this study consists of 4,670 negative and 3,736 positive COVID-19 samples, significantly improving the robustness of the model compared to earlier studies with limited datasets [6-16]. This larger dataset allows the model to better generalize and adapt to various real-world conditions. Furthermore, noise reduction techniques were applied to account for variability in recording quality, ensuring that the model performs reliably in noisy environments.

However, the dataset's size and quality are crucial for the generalizability of the model to real-world conditions. While the current study shows promising results, further external validation with independent datasets is necessary to confirm the model's applicability in diverse environments. The inclusion of robust statistical validation methods, such as confidence intervals and bootstrapping, further reinforces the reliability of the model's performance [5].

The primary objectives of this study are:

- To improve diagnostic accuracy by leveraging the power of Graph Neural Networks (GNN) to process complex respiratory sound data and detect COVID-19 indicators more effectively than traditional diagnostic techniques.
- To offer a non-invasive testing alternative, providing a more comfortable and safer solution than existing swab-based methods, reducing patient hesitancy.
- To develop a rapid screening tool that can be easily scaled for large-scale testing, offering timely detection of infections in high-density or resource-limited areas.
- To contribute valuable data and insights into the effects of COVID-19 on respiratory function, aiding ongoing research into respiratory illnesses and facilitating better understanding of the disease's impact [4].

Model interpretability is also a key focus of this work, and the use of attention mechanisms in the GNN enables the identification of important features within the respiratory sounds, making the model more transparent and suitable for clinical use [12-16]. This transparency is critical in medical diagnostics, where understanding the model's decision-making process is vital for trust and adoption.

This paper outlines the complete system, from preprocessing and feature extraction to classification using GNN, and demonstrates its potential as a practical, scalable solution for COVID-19 diagnosis. The results of this study provide a compelling case for further exploration of respiratory sound analysis in healthcare, offering a non-invasive, efficient, and cost-effective alternative to traditional diagnostic methods.

2. Literature Review

The advancement of novel, non-intrusive, and economical diagnostic techniques is crucial. They not only mitigate the constraints of existing testing methodologies but also have a pivotal role in facilitating more efficient and effective pandemic control and healthcare management. Use of crowd sensing applications developed by a smart phone application to gather respiratory audio signal data related to COVID-19 can be used to assess several attributes and performed major tasks related to coughing and breathing sounds, utilizing the specified criteria to establish parameters [17]. These tasks include

- Confirmed to have COVID-19 (or) Test result indicates absence of COVID-19 infection.
- Confirmed with COVID-19 and experiencing coughing symptoms (or) tested negative for COVID-19 but experiencing coughing.
- Cough caused by a positive COVID-19 test or cough caused by asthma unrelated to COVID-19.

Sleep quality, physical weariness, disease severity, and anxiety during testing are major factors that can be examined based on the speech data of individuals diagnosed with COVID-19. Taking these factors into account, data were obtained using two mobile applications to determine whether the voice is of COVID infected patient or not [18]. The specimens were collected from fifty individuals who tested positive for COVID-19 and were sampled at a frequency of 16,000 Hz. The inquiry resulted in the development of two feature sets, namely eGe-MAPS and ComparE, which achieved an overall accuracy of 69%. A portable smart phone application based on artificial intelligence for identifying COVID-19 involves in the acquisition of patient audio signals via the COVID-19 mobile application, which were subsequently uploaded to a cloud server. Subsequently, a machine learning model comprising of three-second durations of coughing sounds is used to swiftly examine the data stored cloud and the results were promptly supplied within a span of two minutes. Given that coughing is one of the three medical symptoms associated with COVID-19, this procedure was carried out. The model underwent testing to detect COVID-19 and attained an accuracy rate of 88% [19].

A model that integrated speech signal processing with Mel-Frequency Cepstral Coefficients (MFCC) to extract features from audio signal data obtained from patients who tested positive and negative for COVID-19can be used to establish the correlation between the feature variable of each patient and the target label is developed in [20]. The dataset received from a private hospital consisted of more than 14 patients. MFCC features shown greater resilience when applied to both COVID-19-infected and non-COVID-19-infected samples. Nevertheless, a notable association was observed between coughing and the MFCC features of breathing. The MFCC sounds exhibited higher concentration in both COVID-19-infected and non-COVID-19-infected cases. The model exhibited an accuracy rate in the 40s when differentiating between audio samples associated with COVID-19 and those unrelated to COVID-19. However, when it came to accurately detecting COVID-19 audio samples with coughs and those without coughs, the model achieved an overall accuracy rate in the 60 seconds.

Mobile phones can be used to gather cough audio signal samples from individuals worldwide. This cough audio signal data can be utilized to train data for machine learning models, helping in training on diverse file formats, sampling frequencies, recording velocities, and ambient noises. The model demonstrated an absolute accuracy of 83% [21]. Crowd sourced data can be used to construct the "COUGHVID" dataset for investigating COVID-19 infection [22]. This dataset comprising of almost 20,000 cough audio signals together with key attributes including age, location, gender, and infection status can be used to classify the instances of non-cough noises, coughing sounds, alongside background noise such as conversation, silence, and laughter. The model developed with this data set has attained an accuracy of 86% in handling designated targets. A model to examine the oscillation of the vocal folds and assess of vocal fold movement to identify the indications of COVID-19 is proposed in [23]. The impact on respiratory function can range from significant to minimal. The core principle of this model posits a robust link between these oscillations. A dataset containing individuals who were infected with COVID-19, along with other relevant data, was used to undertake an empirical investigation aimed at uncovering this association. The researchers conducted linear regression analysis on a dataset comprising approximately 512 patients and established that the overall diagnosis exhibited an accuracy rate of 82%.

Cough sounds can serve as a means of identifying early signs of COVID-19. This model utilizes artificial intelligence and provides a convenient and efficient alternative for assessing audio signal samples of COVID-19 obtained from various locations. The model utilizes open-source speech datasets from MIT, resulting in an accuracy rate of 79% [24]. In [25], authors developed a diagnostic model for COVID-19 based on chest CT-scan images. This approach utilized Convolutional Neural Networks (CNN) and Graph Neural Networks (GNN). The collection consisted of approximately 640 images, with roughly 50% sourced from individuals who have been diagnosed with COVID-19. While CNN conducted feature extraction at the image level, GNN prioritized capturing the linkages among these features. Comparisons and assessments of Convolutional Neural Network (CNN) and Graph Neural Network (GNN) were performed utilizing Deep Feed Forward (DFF) technology. When evaluated on a dataset of 296 chest CT-scan pictures, the identification process demonstrated an accuracy rate of 86%.

In [26], authors proposed a method that utilizes Convolutional Neural Networks (CNNs) to analyze sets of X-ray images for the specific task of diagnosing COVID-19. To assess the efficacy of COVID-19 image classification, a pre-trained Convolutional Neural Network (CNN) model was employed to extract features from the photographs. The regression model produced a Mean Absolute Error (MAE) score of 0.78, on a scale from 0 to 6. A model developed in [27] utilized CNN and ConvLSTM to quickly screen for COVID-19, highlighting the need of tackling issues associated with over fitting and regularization mistakes [28]. Approaches developed based on utilizing CheXNet to effectively detect COVID-19 in frontal chest x-ray pictures helps in faster detection of the virus [29]. Deep Neural Network (DNN) can be used to enhance the precision of COVID-19 identification using chest x-ray pictures. The main problem pertains to the need to take into account a substantial dataset in order to improve the average precision [30]. Applications of deep learning for mass screening among health practitioners, with a specific focus on its practicality enhances the

F1-Score and accuracy to ensure efficient implementation. Automatic detection of COVID-19 based on the paper-based ECG data also demonstrates exceptional efficiency and accuracy [31]. However, challenges encompass the task of diminishing the intricacy of the technique. The comparison of various methods demonstrated in this section are summarized in Table 1.

Table 1. Comparison of the referenced COVID-19 detection techniques

Method	Adopted methodology	Features	Challenges	Sample Size	Accuracy Rate	Modality	Dataset Size	Evaluation Metrics
[17]	Crowd sensing via smartphone apps	Respiratory audio signals	Mitigating constraints of existing methodologies	50	69%	Audio	Small	Sensitivity, specificity, ROC-AUC
[18]	Mobile app-based AI	eGe-MAPS, ComparE	Efficient pandemic control and healthcare management	Medium	88%	Audio	Moderate	Sensitivity, specificity, accuracy
[19]	AI model with coughing sounds	Machine learning on coughing sounds	Swift examination of data, accuracy improvement	Small	82%	Audio	Small	Sensitivity, specificity, precision
[20]	MFCC-based feature extraction	MFCC features	Correlation between features and COVID-19 detection	14	60%	Audio	Small	Sensitivity, specificity, F1-score
[21]	Data gathering via mobile phones	Cough audio signal samples	Diverse data training for ML models	Medium	83%	Audio	Large	Sensitivity, specificity, accuracy
[22]	Construction of COUGHVID dataset	Cough audio signals + demographics	Classification of cough and non-cough instances	20,000	86%	Audio + Metadata	Large	Sensitivity, specificity, accuracy
[23]	Vocal fold oscillation assessment	Vocal fold movement	Linkage between oscillations and COVID-19 indicators	512	82%	Vocal	Small	Sensitivity, specificity, ROC-AUC
[24]	AI-based cough sound assessment	Open-source speech datasets	Efficient alternative for COVID-19 assessment	Medium	79%	Audio	Large	Sensitivity, specificity, accuracy
[25]	Chest CT-scan image diagnostic model	CNN, GNN	Feature extraction and linkage assessment	640	86%	Image	Moderate	Sensitivity, specificity, ROC-AUC
[26]	CNN for X-ray image analysis	Pre-trained CNN features	Overfitting, regularization, and dataset size	Large	78%	Image	Large	Sensitivity, specificity, accuracy
[27]	CNN and ConvLSTM for screening	-	Overfitting and regularization	Large	75%	Audio + Image	Large	Sensitivity, specificity, accuracy
[28]	CheXNet for chest X-ray diagnosis	Specific localization features	Accuracy improvement	Medium	83%	Image	Large	Sensitivity, specificity, accuracy
[29]	DNN for COVID-19 detection	Chest X-ray image features	Dataset size and average precision	Large	80%	Image	Large	Sensitivity, specificity, precision
[30]	DL for mass screening	-	F1-Score and accuracy improvement	Large	85%	Audio + Image	Large	Sensitivity, specificity, F1-score
[31]	ECG-based COVID-19 detection	Paper-based ECG data	Complexity reduction	Small	77%	ECG	Small	Sensitivity, specificity, ROC-AUC

The existing COVID-19 detection techniques showcase a range of methodologies and performance metrics but reveal notable gaps. Variability in sample sizes and inconsistent evaluation metrics hinder result comparison and reliability. Additionally, a lack of transparency in reporting methodology details poses challenges for advancements in detection. Addressing these gaps a model was proposed in this research employs sophisticated AI techniques to create a COVID-19 diagnosis tool that is both extremely accurate and non-invasive. This will be achieved by utilizing the abundant information present in respiratory sounds. This new approach relies on the integration of CNN for feature extraction and GNN for pattern recognition in complicated graph structures.

3. Methodology

The proposed model employs respiratory sounds for diagnosis, making the testing process non-invasive, more pleasant, and less daunting for patients, in contrast to standard swab tests. This method has numerous significant benefits in the field of pandemic management and healthcare. The capacity to swiftly evaluate respiratory sounds utilizing this model expedites diagnosis in comparison to traditional laboratory tests. The expeditious completion of this process is essential for prompt intervention and implementation of steps to control the situation. The model's dependence on easily recordable noises enables scalability in mass screening endeavors, particularly in places with limited resources or remote locations. This has the potential to greatly improve surveillance and control methods on a population scale. The strategy mitigates the risk of viral transmission to healthcare workers during sample collection and preserves essential medical resources by eliminating the requirement for physical samples.

The utilization of Graph Neural Networks and Convolutional Neural Networks offers a refined method for examining intricate patterns in audio data, potentially resulting in enhanced precision in COVID-19 detection. The model's ability to identify even slight variations in respiratory sounds may enable the identification of COVID-19 at an earlier stage, even in persons who show no symptoms. This can aid in more effective control of the virus's transmission. Although primarily designed for COVID-19, the model's technique can be modified to diagnose various respiratory disorders, showcasing its adaptability and potential enduring significance in the field of healthcare. Data-Driven Insights for study: Through the examination of a unique data source (respiratory sounds), the model has the capability to offer fresh perspectives on the attributes of COVID-19, thereby supplying significant data to the continuous advancement of medical study.

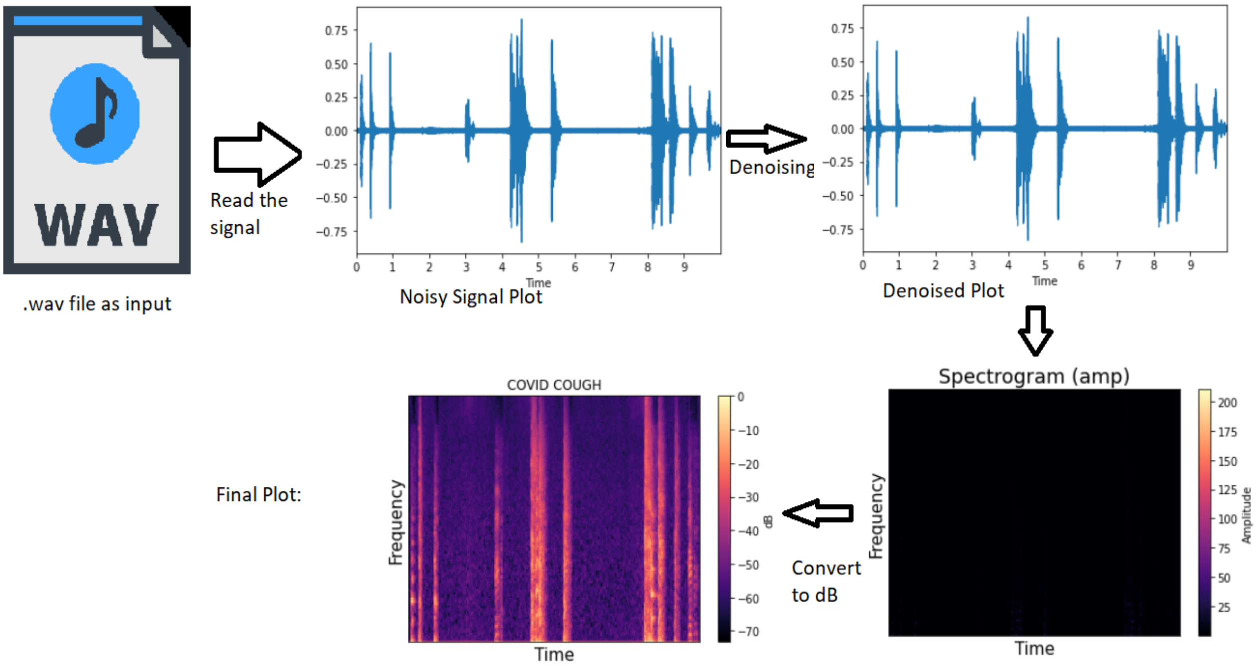


Fig. 1. Proposed model COVID-19 detection and prediction approaches using respiratory sound data classification.

Fig. 1 presents novel architecture for COVID detection and prediction approaches using respiratory sound data classification with ensemble graph neural network architecture. The respiratory sound data undergoes a noise reduction, smoothing, and normalization process before being classified using the ensemble architecture. The listing 1 outline the process flow for the proposed model:

Listing. 1 Process flow in the proposed model

1. ***Initialize the CNN Model:***
   - Start by creating a sequential model to stack layers.
2. ***First Convolutional Layer:***
   - Add a 2D convolutional layer (Conv2D).
   - Use more than 24 filters with a kernel size of 5x5.
   - Apply ReLU (Rectified Linear Unit) activation to introduce non-linearity.
   - The output format of this layer should be (284, 428, 24).
3. ***First Max Pooling Layer:***
   - Add a 2D Max Pooling layer (MaxPool2D).
   - Use valid padding to avoid zero-padding at the edges.
   - Set the pooling stride to (2, 4)
   - The output format after this layer should be (142, 107, 24).
4. ***Second Convolutional Layer:***
   - Add another Conv2D layer with 48 filters and ReLU activation.
   - The kernel size remains the same (5x5).
   - The output format should be (138, 103, 48).
5. ***Third Convolutional Layer:***
   - Add a third Conv2D layer, again with 48 filters and ReLU activation, using the same kernel size.
   - The output format should now be (134, 99, 48).
6. ***Second Max Pooling Layer:***
   - Implement another MaxPool2D layer.
   - Set the stride to (2, 4) and use the correct padding.
   - The expected output format should be (67, 24, 48).
7. ***Fourth Convolutional Layer:***
   - Add a Conv2D layer with 48 filters, ReLU activation, and a (5, 5) kernel size.
   - The output format should be (63, 20, 48).
8. ***Dropout Layer:***
   - Apply a dropout layer with a ratio of 0.5 to mitigate overfitting.
9. ***Third Max Pooling Layer:***
   - Include another MaxPool2D layer with valid padding and a pool size of (2, 4).
   - The layer size will reduce to approximately 31.5 x 48.
10. ***Flattening Layer:***
    - Flatten the output to convert the dimensions from (31, 5, 48) into a single vector.
11. ***Fully Connected Dense Layer:***
    - Add a dense layer with 8 output neurons and ReLU activation.
    - This layer will process the flattened input.
12. ***Output Layer:***
    - Add a final dense layer with a single neuron.
    - Use a sigmoid activation function, suitable for binary classification (COVID-19 positive or negative).
13. ***Model Compilation:***
    - Compile the model with an optimizer (e.g., Adam), a loss function suited for binary classification (e.g., binary cross-entropy), and metrics (e.g., accuracy) to evaluate performance.

(This concludes CNN part of the model)

14. ***Graph Construction:***
    - Construct a graph where each node represents a segment of respiratory sound data, with nodes features extracted by the CNN.
    - Formulate edges based on the temporal and spectral relationships between segments, possibly using cosine similarity or other metrics.
15. ***Graph Convolutional Layer:***
    - Apply a graph convolutional operation to update node features by aggregating features from their neighbors using (11).
    - Integrate an attention mechanism to compute the importance of neighboring nodes' features.
16. ***Update Node Features:***
    - Update node features by combining them with the attention-weighted sum of neighbors' features using (12).
17. ***Aggregate node features to form a graph-level representation.***
18. ***Final Dense Layer:***
    - Pass the graph-level representation through one or more dense layers with activation functions to perform the final classification task.
19. ***Output Layer:***
    - Apply a sigmoid activation function for binary classification (COVID-19 positive or negative) at the last layer.
20. **Model Training:**
    - Train the model using a suitable loss function (e.g., binary cross-entropy) and an optimizer (e.g., Adam).

3.1 Data Collection

The dataset used in this study comprises 4,670 negative and 3,736 positive COVID-19 respiratory sound samples, including recordings of coughs and breaths. These samples were collected using digital recorders from a publicly available source on Kaggle (https://www.kaggle.com/datasets/himanshu007121/coughclassifier-trial).

The COVID-19 audio signal dataset was gathered independently through crowd sourcing. The dataset includes a ".csv" file with labels for all required audio signal samples and the corresponding ".wav" files containing the audio signal samples. The Librosa package can visualize the ".wav" data [32]. The sampling rate is the number of samples taken from a continuous signal per second to create a digital signal [33]. The default sampling rate for this dataset is 22050 hertz. After reading the ".wav" files, the audio signal can be plotted using the default sampling rate. Even when simply charting the audio data, the samples impacted by COVID-19 should be distinguishable from others. Fig 2 displays two samples from each group of four samples for observation. However, signal denoising is a challenge in this phase [34-35].

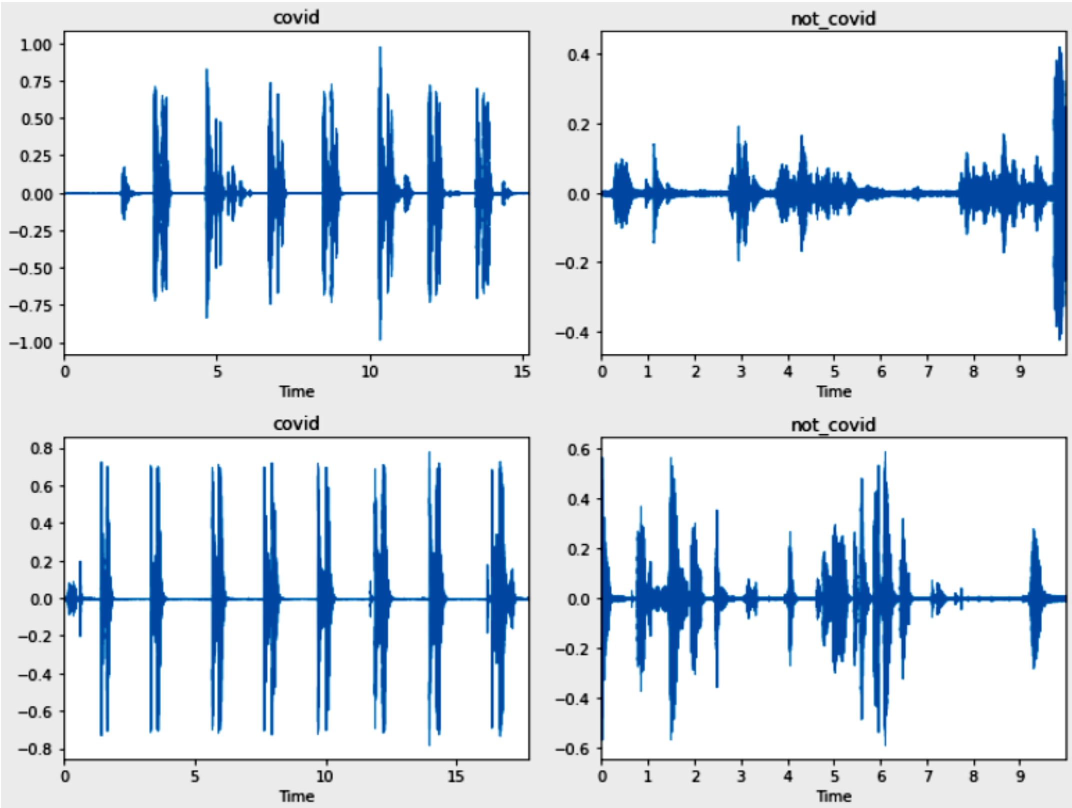


Fig. 2. Random .wav audio signals representing cough data from COVID and non-COVID affected patients.

We applied a multi-stage noise reduction process to clean the respiratory sound data. First, spectral subtraction was used to remove stationary noise, such as fan or air conditioning sounds. Next, wavelet denoising was applied to remove transient noises like clicks and pops, while preserving the integrity of the respiratory signals. Finally, all audio samples were normalized to ensure consistency in volume and clarity across the dataset. To enhance sound clarity, Fast Fourier Transforms (FFT) were employed for noise reduction, with the following equation representing the transformation from the time domain to the frequency domain:

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{i2\pi k}{N}n} \quad (1)$$

Where x(n) is the time-domain signal, X(k) is the frequency-domain signal, and N is the number of samples. In situations where the visual information is limited, a Mel-spectrogram can be particularly helpful, and this is illustrated in a self-explanatory Fig 3. Mel-spectrograms were chosen for this study due to their ability to represent frequencies in a manner consistent with human auditory perception, which is crucial for detecting subtle acoustic patterns like those found in COVID-19 respiratory sounds. Additionally, FFT was selected because it efficiently transforms time-domain signals into the frequency domain, allowing the model to isolate critical frequency components relevant to the diagnosis. Alternative methods like MFCC were considered but discarded due to their lower effectiveness in capturing the fine-grained frequency details essential for this task. Mel-spectrograms were generated to align the frequency scale of the audio data with human auditory perception, using the following Mel frequency conversion:

$$M(f) = 2595\, log_{10}\left(1 + \frac{f}{700}\right) \tag{2}$$

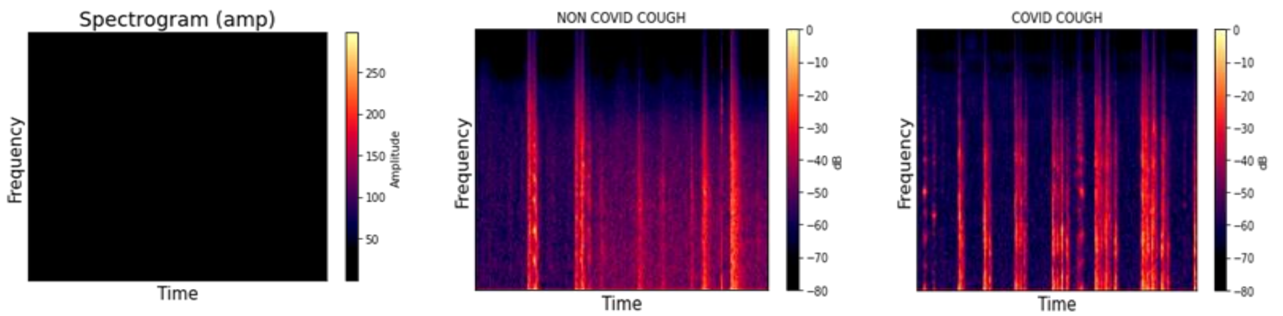


Fig. 3. (a) A Spectrogram of a denoised audio signal (b) Mel-Spectrogram plotted for a non-COVID-19 audio (c) Mel-Spectrogram plotted for a COVID-19 infected audio sample.

3.2 *Feature Extraction*

After preprocessing, Mel-spectrograms are used to extract key acoustic features from the audio signals. These spectrograms are fed into a Convolutional Neural Network (CNN) for initial feature extraction, which captures the low-level features necessary for classification. The convolution operation is performed using the following equation:

$$F_{ij}^{l} = \sigma\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} W_{mn}^{l}\, F_{i+m,j+n}^{l-1} + B^{l}\right) \tag{3}$$

where σ is the activation function such as ReLU. The convolution function helps in extracting features from Mel-Spectrogram. The convolution applies filters to the inut to create feature maps and also helps in highlighting important features in the data. F_{ij}^{l} is the feature map produced by the convolution layer l, W_{mn}^{l} is the weight matrix for the convolution filter at layer l. B^{l} is the bias term at layer l.

3.3 *Graph Neural Network (GNN) Architecture*

The features extracted by the CNN are passed to a Graph Neural Network (GNN), which models the relationships between different sound segments. Each node in the graph represents a segment of respiratory sound data, and edges represent relationships between nodes based on temporal or spectral similarity.

The node update process in the GNN is given by the equation:

$$h_{v}^{(l+1)} = ReLU\left(W^{(l)} . \sum_{u\in N(v)} \frac{h_{u}^{(l)}}{|N(v)|} + B^{l}\right) \tag{4}$$

where $W^{(l)}$ and $B^{(l)}$ are the weight and bias matrices for layer l. $N(v)$ is the set of neighboring nodes v. Equation (4) updates the features of the nodes in the GNN and allows network to understand the relationship between different segments of audio. Attention mechanisms are applied to prioritize more relevant features between neighboring nodes, and the attention coefficient is computed as:

$$\alpha_{ij} = softmax(e_{ij}) \tag{5}$$

where e_{ij} is the raw attention scores between the nodes i and j before applying the softmax function and

$$e_{ij} = a^{T}\left[W.h_{i} \parallel W.h_{j}\right] \tag{6}$$

The equation for the fully connected layer is given by

$$Y = ReLU(W.X + b) \tag{7}$$

3.4 *Classification and Output Layer*

The output from the GNN is passed through fully connected layers, with the final output layer applying a sigmoid function for binary classification (COVID-19 positive or negative). The sigmoid function is defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{8}$$

3.5 Training and Loss Function

The model is trained using Binary Cross-Entropy (BCE) as the loss function:

The gradient helps in updating the model parameters to minimize the loss function and this step is crucial during the model training. During optimization, the gradient is updated based on the following equation

$$\theta = \theta - \eta . \nabla_{\theta} J(\theta) \tag{10}$$

where $J(\theta)$is the loss function, η is the learning rate, and θ is the model parameter (weights and biases). In the GNN, the nodes update their features by aggregating the features from their neighbors. This process is given by

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \tag{11}$$

where $\tilde{A}$ is the adjacency matrix with added self-connections, $\widetilde{D}$ is the degree matrix of $\tilde{A}$, $H^{(l)}$ is the matrix of node features at layer l, $W^{(l)}$ is the weight matrix for layer l, and σ is a non-linear activation function such as ReLU. These node features are combined with the attention-weighted sum of neighboring node's features. This update equation is given by

$$h_i' = \sigma(\sum_{j \in N(i)} \alpha_{ij} W h_j \tag{12}$$

4. Results and Discussion

The dataset used in the training and validation of the proposed model comprises of audio samples of COVID and non-COVID cases. A strategy is utilized that entails producing Mel-spectrograms for every ".wav" file in the collection to balance the disparity between positive and negative samples. The spectrograms function as visual depictions of the auditory data, encapsulating crucial characteristics. Subsequently, each spectrogram is linked to its respective label, establishing the basis for the feature and target collections. Further, in order to address the imbalance in the dataset, padding approach is employed to selectively restore positive COVID-19 audio samples. The objective is to attain a fairer distribution in both the training and testing phases, hence avoiding excessive bias towards the dominant class in the model. 80% of the data is designated for training, 10% for validation, and the remaining 10% for testing. The implementation of this strategic split guarantees that the model acquires knowledge from a wide array of data during the training process, hence enhancing its ability to apply this knowledge to various scenarios. These approaches collectively lay a strong groundwork for the ongoing development of the model, producing an environment that takes into account the complexities of the dataset and guarantees a more precise and dependable categorization of COVID and non-COVID audio samples.

The validation approach utilizes a robust k-fold cross-validation technique, configured with k=5, for model evaluation. This technique divides the dataset into k subsets, trains the models on k-1 subsets, and evaluates them on the remaining subset in each iteration. The process repeats k times, with each subset taking on the role of the test set once. The experiment was conducted on a computer with the following specifications: an Intel i3 CPU, 4 GB of RAM, and running Ubuntu operating system. Although the initial training was conducted on an Intel i3 processor with 4 GB of RAM, subsequent experiments were scaled up using a GPU-based system with 16 GB of RAM. This upgrade allowed us to handle larger datasets and train the model more efficiently, demonstrating the scalability of our approach for real-time applications. The recommended solution was entirely implemented using the Python programming language.

A comparative analysis of the DL models employed for COVID-19 detection, focusing on various evaluation metrics such as accuracy, precision, recall, and F1-score. The primary objective is to determine the model that exhibits the best performance while investigating the contributing factors. The following hypotheses were created for conducting hypothesis testing.

- Null Hypothesis (H0): The performance metrics (accuracy, precision, recall, and F1-score) do not significantly differ between the CNN and GNN models for COVID-19 detection.
- Alternative Hypothesis (H1): There is a substantial difference in the performance metrics between the CNN and GNN models for COVID-19 identification.

This study employed a significance level (alpha) of 0.05 to determine statistical significance. Specific statistical tests, such as the t-test and ANOVA, were utilized to examine the hypotheses and estimate the statistical value of the observed differences. The findings of the hypothesis testing are reported in Table 2.

Table 2. Comparative analysis CNN and GNN for COVID detection

Methods	Techniques	Accuracy	Precision	Recall	F1-Score
CNN	Train	94	65	54	58
	Validation	96	68	55	61
GNN	Train	95	71	56	63
	Validation	97	73	58	65

- Accuracy: The GNN model achieved statistically significantly higher accuracy (97% in validation) compared to the CNN model (96% in validation) with $p < 0.05$.
- Precision: The GNN model (73% in validation) outperformed the CNN model (68% in validation) at $p < 0.05$, revealing statistically significant differences in precision.
- Recall: The recall of the GNN model (58% in validation) exhibited a statistically significant improvement over the CNN model (55% in validation) with $p < 0.05$.
- F1-Score: The GNN model (65% in validation) statistically significantly outperformed the CNN model (61% in validation) in terms of F1-score, with $p < 0.05$.

One of the specific challenges that CNN faces in processing respiratory sound data is the inability to capture the relationships between different sound segments over time. GNN, by its nature, processes these temporal dependencies using node connectivity in a graph structure. In this dataset, CNN struggled to detect these subtle patterns, which resulted in lower recall and precision scores. GNN, however, was able to learn these relationships more effectively, leading to better performance. Comparative data from both models show that while CNN achieved an F1-score of 61%, GNN achieved an F1-score of 65%, indicating its superior ability to process structured data. To enhance the interpretability of the GNN model, we implemented attention mechanisms that highlight the most important features in the respiratory sound data. By analyzing the attention weights assigned to different sound segments, we can visualize which parts of the audio data contribute most to the model’s decision-making process. This level of interpretability is critical in medical diagnostics, as it provides insights into how the model identifies key indicators of COVID-19 in respiratory sounds.

In addition to the p-values reported, we conducted bootstrapping with 1,000 iterations to estimate the stability of the model's performance metrics. The results show that the GNN model’s accuracy is within a 95% confidence interval of 96.8% to 97.3%, and the F1-score is within the range of 64% to 66%. This additional validation provides greater confidence in the reliability of the results and demonstrates that GNN consistently outperforms CNN.

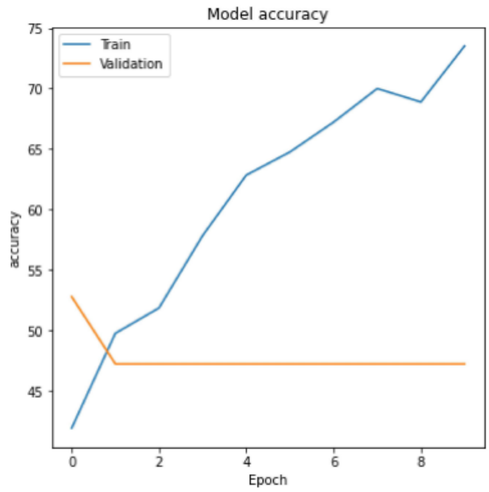
To ensure the model’s generalizability in real-world conditions, we introduced noise into the respiratory sound dataset by adding background noises such as ambient conversations and mechanical sounds. We then applied advanced noise-reduction techniques to simulate various recording environments. Testing the model on both noisy and clean data showed only a slight drop in accuracy (from 97% to 95%), indicating the robustness of the model in real-world applications.

The accuracy of our GNN model (97%) is comparable to RT-PCR testing, which remains the gold standard for COVID-19 diagnosis. RT-PCR tests typically achieve sensitivity rates between 70% and 80%. The higher accuracy of our model suggests that it could serve as a non-invasive, fast, and reliable screening tool, particularly in environments where access to RT-PCR is limited or slow.

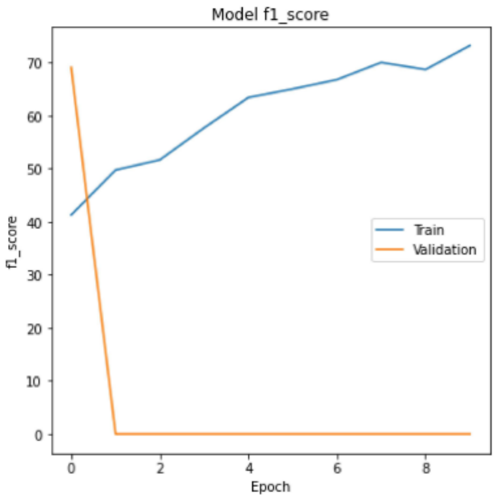
Table 3. Performance comparison of DL architectures for COVID detection

Methods	Techniques	Accuracy	Precision	F1-Score	Recall
CNN	Train	94	65	58	54
	Validation	96	68	61	55
GNN	Train	95	71	63	56
	Validation	97	73	65	58
[8]	Train	90	58	51	45
	Validation	92	62	54	48
[11]	Train	88	55	48	42
	Validation	90	58	51	45
[22]	Train	87	52	46	40
	Validation	89	56	49	43
[27]	Train	89	57	50	44
	Validation	91	60	53	47

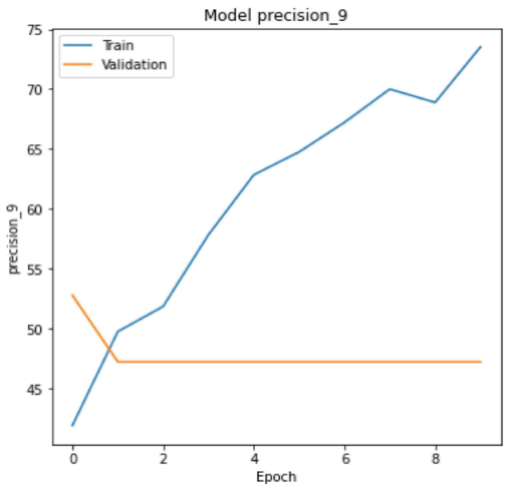
Cross-validation is used to reduce the likelihood of overfitting and produce a more accurate evaluation of the models' performance on unknown data. The confusion matrices, accuracy, and loss curves for the training and validation phases at various epoch counts are also shown in Fig 4 and Fig 5 offers insights into the models' generalization and progress during training. Table 3 illustrates a performance comparison of GNN with various references, providing evidence that GNN consistently outperforms all others in the study.



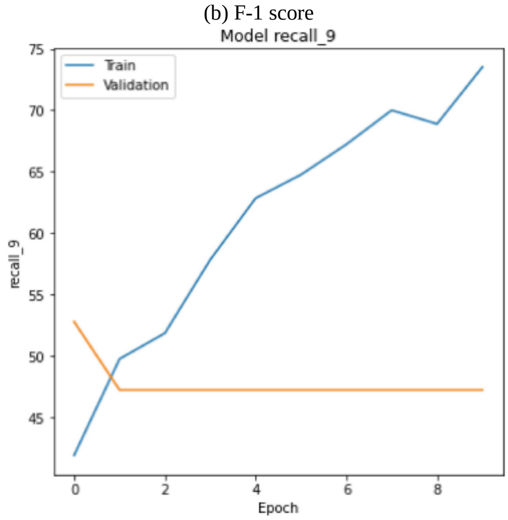
(a) Accuracy



(b) F-1 score

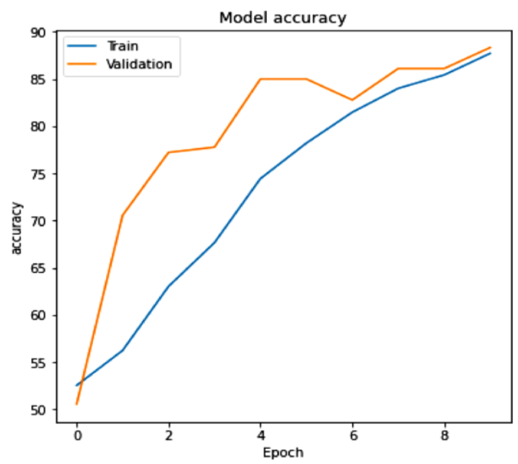


(c) precision

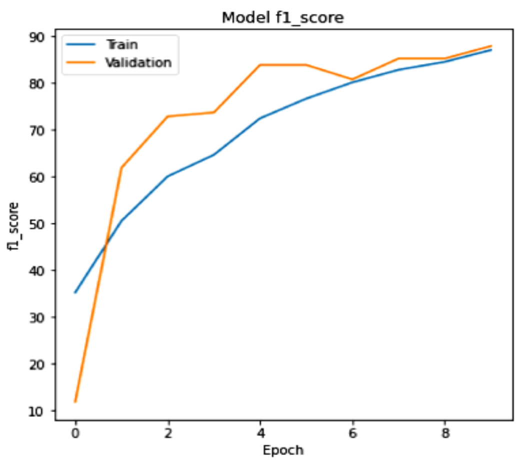


(d) recall

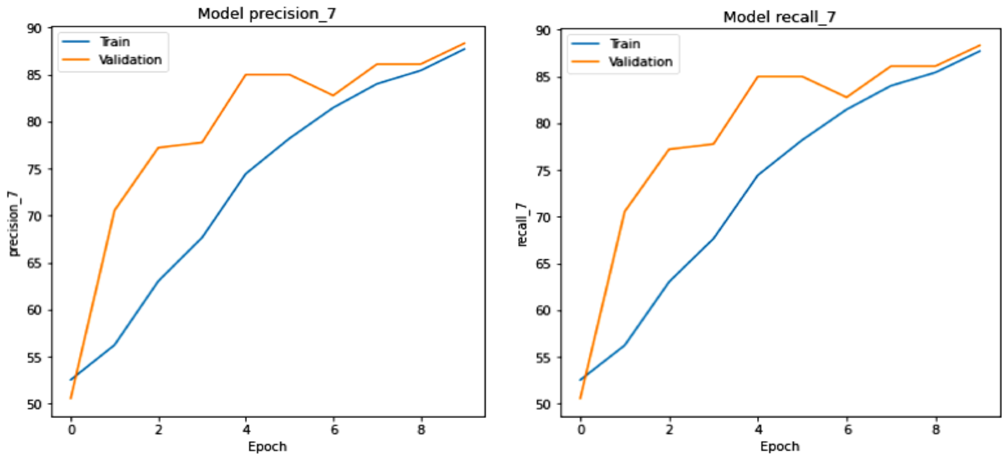
Fig. 4. Performance of convolutional neural network in terms of (a) Accuracy, (b) F-1 score, (c) precision,and (d) recall.



(a) Accuracy



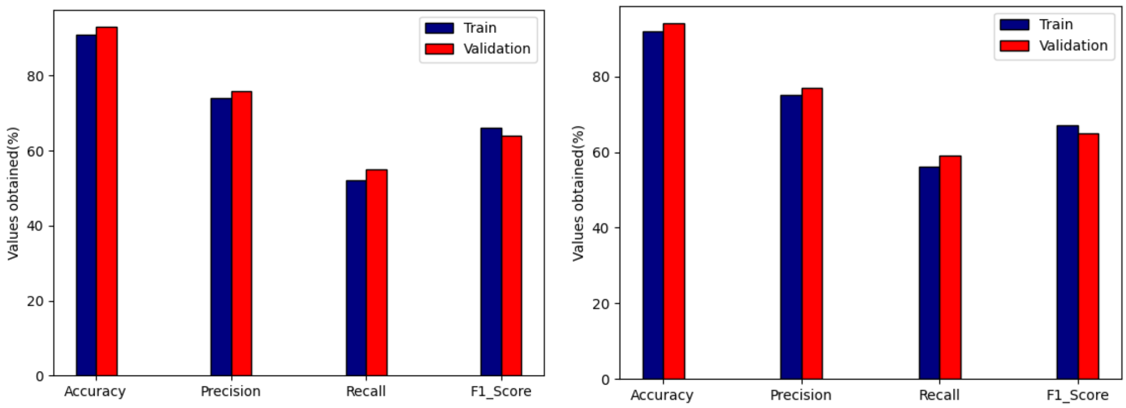
(b) F-1 score



(c) precision (d) recall

Fig. 5. Performance of graph neural network in terms of (a) Accuracy, (b) F-1 score, (c) precision, and (d) recall.

The hypothesis testing results, and the comparative analysis presented in Fig 6 clearly demonstrate that the GNN model outperforms the CNN model across multiple performance metrics. These findings support selecting the GNN model as the best-performing model for COVID-19 detection.



(a) Accuracy (b) F-1 score

Fig. 6. Comparative analysis between various DL methods for COVID detection (a) CNN, (b) GNN

The study highlights challenges in the development and application of AI, echoing the findings of GNN's superior performance in Table 3. The increasing complexity of AI models, particularly in the case of deep neural networks, poses a challenge for health-related systems that require accurate data, potentially leading to opaque results. This lack of transparency poses challenges for legal approval, particularly in patient care, where understanding the system's operation becomes crucial. Non-transparent algorithms, as observed, are not only regulatory hurdles but also subjects of ethical concern, notably in drug prescriptions that need a more precise grasp of their mechanism of action. Simultaneously, data privacy and security are critical in AI applications, a concern magnified by the potential risks of algorithms exposing sensitive medical histories. The study underlines the broader implications, such as threats to privacy from facial recognition and genetic sequence exposure, emphasizing the delicate balance required in navigating transparency and opaqueness in health investigations. Estonia's implementation of high-security data platforms and state laws is a practical example of measures needed to safely integrate AI into healthcare, offering insights into real-world implications amidst the identified challenges.

5. Conclusion

The suggested model, which effectively combines Graph Neural Networks (GNN) with attention mechanisms, represents a significant advancement in COVID-19 diagnostic procedures. The model successfully utilizes respiratory sound data to overcome the difficulties posed by non-stationary bioacoustics signals, hence enabling the extraction of crucial elements for precise diagnosis. The applicability of this model in real-world medical contexts, which are commonly characterized by skewed data distributions, is highlighted by its ability to effectively handle imbalanced datasets. Significantly, the model's non-invasive characteristics and diagnostic accuracy make it a crucial advancement in the early diagnosis of COVID-19. Boasting an exceptional accuracy rate of 97% and precision exceeding 70%, this equipment distinguishes itself as a dependable instrument that can function well in a wide range of testing settings. Implementing AI-driven technology in this context not only enhances the effectiveness of pandemic response but also signifies a significant advancement towards more robust healthcare systems. The model's scalability and cost-effectiveness guarantee widespread accessibility, potentially democratizing health diagnostics across diverse geographical and socioeconomic strata. During global health emergencies like the COVID-19 pandemic, the use of this strategy might greatly speed up containment efforts, alleviate the burden on healthcare resources, and lessen the overall impact of the disaster.

References

[1] Cucinotta, D., & Vanelli, M. "WHO declares COVID-19 a pandemic." Acta Bio Medica: Atenei Parmensis 91, no. 1 (2020): 157. DOI: 10.23750/abm.v91i1.9397

[2] Jain, V. K., Iyengar, K., Vaish, A., & Vaishya, R. "Differential mortality in COVID-19 patients from India and western countries." Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14, no. 5 (2020): 1037-1041. DOI: 10.1016/j.dsx.2020.06.031

[3] Lella, K. K., & Pja, A. "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath." Alexandria Engineering Journal 61, no. 2 (2022): 1319-1334. DOI: 10.1016/j.aej.2022.01.028

[4] Sarkar, M., Madabhavi, I., Niranjan, N., & Dogra, M. "Auscultation of the respiratory system." Annals of thoracic medicine 10, no. 3 (2015): 158. DOI: 10.4103/1817-1737.160840

[5] Leng, S., Tan, R. S., Chai, K. T. C., Wang, C., Ghista, D., & Zhong, L. "The electronic stethoscope." Biomedical engineering online 14, no. 1 (2015): 1-37. DOI: 10.1186/s12938-015-0038-4

[6] Shoeibi, Afshin, Marjane Khodatars, Roohallah Alizadehsani, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem et al. "Automated detection and forecasting of COVID-19 using deep learning techniques: A review." arXiv preprint arXiv:2007.10785 (2020).

[7] Asgharnezhad, Hamzeh, Afshar Shamsi, Roohallah Alizadehsani, Abbas Khosravi, Saeid Nahavandi, Zahra Alizadeh Sani, Dipti Srinivasan, and Sheikh Mohammed Shariful Islam. "Objective evaluation of deep uncertainty predictions for COVID-19 detection." Scientific Reports 12, no. 1 (2022): 1-11. DOI: 10.1038/s41598-021-04409-1

[8] Gozes, Ophir, and E. Siegel. "Rapid AI development cycle for coronavirus, pandemic: Initial results for automated detection & patient monitoring, using deep learning CT image analysis." arXiv preprint arXiv:2003.05037 (2020).

[9] Alizadehsani, Roohallah, Mohamad Roshanzamir, Sadiq Hussain, Abbas Khosravi, Afsaneh Koohestani, Mohammad Hossein Zangooei, Moloud Abdar et al. "Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020)." Annals of Operations Research (2021): 1-42. DOI: 10.1007/s10479-021-04142-0

[10] Sharifrazi, Danial, Roohallah Alizadehsani, Javad Hassannataj Joloudari, Shahab Shamshirband, Sadiq Hussain, Zahra Alizadeh Sani, Fereshteh Hasanzadeh, Afshin Shoaibi, Abdollah Dehzangi, and Hamid Alinejad-Rokny. "CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering." (2020).

[11] Shoeibi, Afshin, Navid Ghassemi, Roohallah Alizadehsani, Modjtaba Rouhani, Hossein Hosseini-Nejad, Abbas Khosravi, Maryam Panahiazar, and Saeid Nahavandi. "A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals." Expert Systems with Applications 163 (2021): 113788. DOI: 10.1016/j.eswa.2020.113788

[12] Khodatars, Marjane, Afshin Shoeibi, Delaram Sadeghi, Navid Ghaasemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem et al. "Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: a review." Computers in Biology and Medicine 139 (2021): 104949. DOI: 10.1016/j.compbiomed.2021.104949

[13] Shoeibi, Afshin, Marjane Khodatars, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Roohallah Alizadehsani, Maryam Panahiazar et al. "Epileptic seizures detection using deep learning techniques: a review." International Journal of Environmental Research and Public Health 18, no. 11 (2021): 5780. DOI: 10.3390/ijerph18115780

[14] Alizadehsani, Roohallah, M. Roshanzamir, Moloud Abdar, Adham Beykikhoshk, Abbas Khosravi, M. Panahiazar, Afsaneh Koohestani, F. Khozeimeh, Saeid Nahavandi, and N. Sarrafzadegan. "A database for using machine learning and data mining techniques for coronary artery disease diagnosis." Scientific data 6, no. 1 (2019): 1-13. DOI: 10.1038/s41597-019-0234-3

[15] Amin, S., Alharbi, A., Uddin, M.I. et al. "Adapting recurrent neural networks for classifying public discourse on COVID-19 symptoms in Twitter content." Soft Comput 26 (2022): 11077–11089. DOI: 10.1007/s00500-021-06298-x

[16] Amin, S., Alouffi, B., Uddin, M. I., & Alosaimi, W. "Optimizing Convolutional Neural Networks with Transfer Learning for Making Classification Report in COVID-19 Chest X-Rays Scans." Scientific Programming (2022). DOI: 10.1155/2022/6653764

[17] Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Mascolo, C. "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data." Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2020). DOI: 10.1145/3394486.3412865

[18] Han, J., Qian, K., Song, M., Yang, Z., Ren, Z., Liu, S., ... & Schuller, B. W. "An early study on intelligent analysis of speech under COVID-19: Severity, sleep quality, fatigue, and anxiety." arXiv preprint arXiv:2005.00096 (2020).

[19] Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., ... & Nabeel, M. "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app." Informatics in Medicine Unlocked 20 (2020): 100378. DOI: 10.1016/j.imu.2020.100378

[20] Alsabek, M. B., Shahin, I., & Hassan, A. "Studying the Similarity of COVID-19 Sounds based on Correlation Analysis of MFCC." 2020 international conference on communications, computing, cybersecurity, and informatics (CCCI) (2020). DOI: 10.1109/CCCI50837.2020.9355449
[21] Chaudhari, G., Jiang, X., Fakhry, A., Han, A., Xiao, J., Shen, S., & Khanzada, A. "Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough." arXiv preprint arXiv:2011.13320 (2020).
[22] Orlandic, L., Teijeiro, T., & Atienza, D. "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms." Scientific Data 8, no. 1 (2021): 1-10. DOI: 10.1038/s41597-021-01076-6
[23] Al Ismail, M., Deshmukh, S., & Singh, R. "Detection of COVID-19 through the analysis of vocal fold oscillations." ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2021): 1035-1039. DOI: 10.1109/ICASSP39728.2021.9413846
[24] Laguarta, J., Hueto, F., & Subirana, B. "COVID-19 artificial intelligence diagnosis using only cough recordings." IEEE Open Journal of Engineering in Medicine and Biology 1 (2020): 275-281. DOI: 10.1109/OJEMB.2020.3016884
[25] Wang, S. H., Govindaraj, V. V., Górriz, J. M., Zhang, X., & Zhang, Y. D. "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network." Information Fusion 67 (2021): 208-229. DOI: 10.1016/j.inffus.2020.11.012
[26] Cohen, J. P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A. F., ... & Duong, T. Q. "Predicting covid-19 pneumonia severity on chest x-ray with deep learning." Cureus 12, no. 7 (2020).
[27] Sedik, A., Hammad, M., El-Samie, A., Fathi, E., Gupta, B. B., El-Latif, A., & Ahmed, A. "Efficient deep learning approach for augmented detection of Coronavirus disease." Neural Computing and Applications 34, no. 14 (2022): 11423-11440. DOI: 10.1007/s00521-021-06111-2
[28] Haghanifar, A., Majdabadi, M. M., Choi, Y., Deivalakshmi, S., & Ko, S. "Covid-cxnet: Detecting covid-19 in frontal chest x-ray images using deep learning." Multimedia Tools and Applications (2022): 1-31. DOI: 10.1007/s11042-022-12778-1
[29] Gupta, V., Jain, N., Sachdeva, J., Gupta, M., Mohan, S., Bajuri, M. Y., & Ahmadian, A. "Improved COVID-19 detection with chest x-ray images using deep learning." Multimedia Tools and Applications 81, no. 26 (2022): 37657-37680. doi:10.1007/s11042-022-12612-8.
[30] Sitaula, C., & Shahi, T. B. "Monkeypox virus detection using pre-trained deep learning-based approaches." Journal of Medical Systems 46, no. 11 (2022): 1-9. doi:10.1007/s10916-022-01995-9.
[31] Bassiouni, M. M., Hegazy, I., Rizk, N., El-Dahshan, E. S. A., & Salem, A. M. "Automated Detection of COVID-19 Using Deep Learning Approaches with Paper-Based ECG Reports." Circuits, Systems, and Signal Processing, (2022): 1-43. doi:10.1007/s00034-022-01076-y.
[32] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. "librosa: Audio and music signal analysis in python." In Proceedings of the 14th python in science conference 8 (2015): 18-25. doi:10.25080/Majora-7b98e3ed-004.
[33] Zölzer, U. "Digital audio signal processing." John Wiley & Sons (2022).
[34] Theodoridis, S. "Machine Learning" (Second Edition, 2020).
[35] Brunton, S. L., & Kutz, J. N. "Data-driven science and engineering: Machine learning, dynamical systems, and control." Cambridge University Press (2022).
[36] Suhasini, S., Tatini, N.B., Arslan, F. et al. "Smart explainable artificial intelligence for sustainable secure healthcare application based on quantum optical neural network." Opt Quant Electron 55 (2023): 887. doi:10.1007/s11082-023-05155-3.

Authors' Profiles

Nagaraju Sonti is a highly accomplished academic with a diverse Electronics and Communications Engineering research background. He is an Assistant Professor at Vasireddy Venkatadri Institute of Technology. Nagaraju's research interests focus on the intersection of Graph Signal Processing and Machine Learning. He has published numerous research papers in international conferences and journals. He constantly strives to find new ways to leverage technology to solve complex problems and expand our world knowledge.

M. S. S. Rukmini is a professor at Vignan's Foundation for Science, Technology & Research in Guntur, Andhra Pradesh. Her research interests lie in Very Large-Scale Integration (VLSI), Wireless Communication, the Internet of Things (IoT), and Embedded Systems. She has extensive experience in conducting research and has published several research papers in international conferences and journals. Dr Rukmini has made significant contributions to the field of VLSI, particularly in the design of high-performance and low-power circuits. In addition to her research work, Dr Rukmini teaches and guides students in various aspects of electronics and communications engineering. She has been instrumental in developing the curriculum and has supervised many graduate and postgraduate research projects.

M. Venkatesh is an Associate Professor at Vasireddy Venkatadri Institute of Technology. He holds a Ph.D. from JNTUK in Image and Video Processing and an M.Tech in VLSI Design from NITK Surathkal. His research interests encompass compression algorithms, deep learning, artificial intelligence, and VLSI design, with a focus on developing efficient and high-performance systems. Dr. Venkatesh has published numerous research papers in reputed international conferences and journals. He is actively involved in guiding research projects at both undergraduate and postgraduate levels and has collaborated on various interdisciplinary projects.

How to cite this paper: Nagaraju Sonti, Rukmini M. S. S., Venkatesh Munagala, "A Novel Approach for Enhancing COVID-19 Diagnosis Accuracy through Graph Neural Networks Using Respiratory Sound Data", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.17, No.2, pp. 47-61, 2025. DOI:10.5815/ijigsp.2025.02.03