International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, Delaware, USA
Volume 13– No. 1, August 2025 – www.ijais.org

Paediatrics Respiratory Diseases Diagnostic and Prediction System using Deep Learning Model

Oluwajana Kehinde Joseph
Department of Computer Science
Federal University of Technology, Akure

Adegun Iyanu Pelumi
Department of Computer Science
Federal University of Technology, Akure

Oluwadare Samuel Adebayo
Department of Computer Science
Federal University of Technology, Akure

ABSTRACT
The study presents a deep learning-based diagnostic and prediction method for paediatric respiratory illnesses that have a significant global impact on children's health, notably upper respiratory tract infections (URTI), chronic obstructive pulmonary disease (COPD), bronchiolitis, and pneumonia. The proposed framework combines Gated Recurrent Units (GRU) to describe sequential patterns in Mel-Frequency Cepstral Coefficients (MFCCs) and Convolutional Neural Networks (CNN) to capture local temporal features through the analysis of respiratory sound recordings. The model, which was trained on a labelled dataset, performed 84% of the time and showed good diagnostic and prediction abilities, particularly for cases of bronchiolitis (92% precision and recall) and healthy (100% precision and recall). The model's potential as an accurate and easily accessible tool for diagnosing and predicting paediatric respiratory diseases is proven by the results, despite a few misclassifications.

General Terms
Deep learning model, Paediatrics respiratory diseases, Diagnosis and prediction, Audio recordings, Respiratory sounds.

Keywords
Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), COPD, Healthy, Bronchiolitis, Pneumonia.

1. INTRODUCTION
Paediatric respiratory diseases, which range from common ailments like asthma to serious infections like pneumonia, represent a major healthcare burden worldwide, particularly in young children [1],[3]. These disorders typically cause recurring episodes that interrupt daily living and, in severe situations, need hospitalisation. Accurate and fast diagnosis is critical for providing optimal care and preventing consequences. Traditional diagnostic methods, such as clinical assessments, physical examinations, and diagnostic tests, frequently face challenges due to the overlapping symptoms of respiratory diseases, the reliance on subjective clinical judgement, and their resource-intensive nature [4],[19]. Furthermore, these methods may necessitate patient cooperation, which can be challenging to get in paediatric populations [22].

In recent years, advances in artificial intelligence (AI) and deep learning have emerged as revolutionary technologies for addressing diagnostic difficulties in healthcare. Deep learning models, which can analyse big and complicated datasets, have shown exceptional effectiveness in a variety of medical applications, including disease detection and screening [14],[15]. These technologies offer faster and more accurate diagnostics, better clinical decisions, and improved public health management.

Despite advances in AI, its application in paediatric respiratory disease detection remains relatively unexplored [2]. This study intends to close this gap by developing a diagnostic and prediction system using a hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) architecture. The system aims to automatically learn complex patterns in respiratory data, capture temporal correlations, and give interpretability in its decision-making process. By addressing the limitations of traditional diagnostic approaches and existing shallow machine learning models, the proposed system seeks to enhance diagnosis accuracy, eliminate misdiagnoses, and accelerate treatment initiation.

The study's aims include building and deploying a CNN-GRU model for detecting and predicting paediatric respiratory disorders, evaluating its performance using conventional metrics, and showing its potential to support earlier medical interventions. This study advances the field of paediatric healthcare by harnessing deep learning's strengths, with the goal of improving the quality of life for young patients and their families.

2. RELATED WORKS
Recent advances in deep learning and artificial intelligence have accelerated research into paediatric respiratory illness detection and prediction. Several studies have investigated novel strategies, emphasising both the potential and limitations of existing approaches.

[6] proposed a deep learning strategy based on Artificial Neural Networks (ANN) to predict paediatric asthma-related emergency department visits using Medicaid claims data. The ANN model beat traditional Lasso logistic regression by making use of its nonlinear properties. However, the model's reliance on Medicaid claims data prompted questions about its generalisation. Despite this constraint, the study demonstrated the benefits of deep learning in paediatric healthcare prediction.

[20] architecture reached a significant milestone by applying deep learning to respiratory sound recordings, achieving 98.5% accuracy on the ICBHI benchmark dataset. This finding emphasises the framework's strength and potential for practical use in detecting respiratory abnormalities and illnesses.

[15] proved the efficacy of deep learning algorithms in identifying common paediatric pulmonary disorders using Xray pictures, obtaining 92% accuracy in pneumonia identification, outperforming radiologists. However, the study's small dataset and narrow emphasis on paediatric

pulmonary disorders require additional research to improve real-world applicability. Similarly, [8] created a deep learning model for chest X-ray pictures that performed exceptionally well in paediatric pneumonia, increasing diagnostic accuracy, efficiency, and consistency.

[23] suggested a fine-grained diagnosis system that uses electronic health records (EHRs) to detect paediatric respiratory illnesses. The system attained 92.5% accuracy, proving its potential, while shortcomings in external validation and the study's scope were identified. [17] demonstrated a supervised multi-label classification framework for predicting 14 thoracic illnesses in paediatric chest radiographs, with an average AUC of 0.940. However, the study's dependence on a single dataset called into doubt its broader application.

[16] presented an ensemble of deep learning models for pneumonia identification that surpassed previous methods. The study emphasised the importance of validation across multiple therapeutic settings to prove broader value. [11] created a deep learning-based system for COVID-19 detection and pneumonia screening, which demonstrated good accuracy rates and represents a viable path for automated detection in clinical applications.

[9] investigated deep learning-based analysis of pulmonary auscultation audios and found encouraging results while raising concerns regarding generalisation and interpretability due to a lack of comparison study. [10] created a model for detecting wheezing in children based on real-world data, enhancing paediatric respiratory disease diagnosis and therapy. However, the study's dataset size was limited, and no comparisons with known methodologies were provided.

[18] suggested a model for detecting respiratory disorders that outperformed established methods but was limited by a small sample size and reliance on inpatient data. [13] used CNN based deep learning algorithms to diagnose lung diseases, with encouraging results. Despite its potential, the study has some drawbacks, including a small dataset, inherent biases, and limited generalisability.

[12] presented LDDNet, a deep learning system for identifying infectious lung illnesses that prioritises thorough optimisation and evaluation. [5] investigated a machine learning strategy for detecting paediatric pneumonia and demonstrated the Logistic Regression model's interpretability and effectiveness for medical practitioners.
Finally, [21] examined machine learning methods used for cough sound analysis, emphasising their utility in identifying childhood respiratory disorders. However, the study identified drawbacks such as input data unpredictability and a small number of articles analysed.

The reviewed studies collectively reveal remarkable improvements in using deep learning and machine learning approaches to diagnose and predict paediatric respiratory diseases. While the studies show excellent accuracy and robust methodology, common constraints such as small datasets, a lack of external validation, and limited generalisability highlights the need for additional research to develop these approaches for wider clinical use.

3. METHODOLOGY

This research presents a methodology for diagnosing and predicting paediatric respiratory diseases using a hybrid deep learning model that integrates Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU). The system utilizes audio recordings of respiratory sounds, with a structured workflow that includes data collection, preprocessing, feature extraction, data augmentation, model design, training, and evaluation.

3.1 Data Collection and Preprocessing

The Kaggle Respiratory Sound Dataset was employed, containing respiratory audio recordings and metadata, such as crackles, wheezes, and patient diagnoses (COPD, Bronchiolitis, Pneumonia, URTI, and Healthy). The data underwent preprocessing, including numerical encoding of disease labels and splitting into training, validation, and testing sets.

3.2 Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) were extracted from audio recordings as input features for the model. MFCCs, which represent the short-term power spectrum of sound, were computed using a sequence of transformations, including Fourier transforms, logarithmic power spectrum scaling, and Discrete Cosine Transforms. A total of 52 MFCC features were extracted per audio file. The derivation is represented as follows:

$$MFCC_{(n)} = \sum_{m=1}^{M-1} \log(X_m) \cos\left[\frac{n(m-0{\cdot}5)\pi}{M}\right] \quad 1$$

where X_m is the power spectrum of the audio signal, M is the number of mel bands, and n represents the MFCC order.

The input to the model is the MFCC features extracted from the audio data.

Let: $X \in \mathbb{R}^{TxF}$ 2

be the input MFCC features, where T is the number of time steps (or frames), and F is the number of MFCC features per frame.

3.3 Data Augmentation

To enhance the robustness and generalizability of the model, data augmentation techniques were applied:

i. Noise Addition: It simulated background noise by adding random noise to the audio signals. The derivation is represented as follows:

$$data_{noise} = data + \alpha \cdot noise \quad 3$$

ii. Time Shifting: It adjusted the start time by shifting the audio forward or backward. The derivation is represented as follows:

$$shifed_{data} = np \cdot roll(data, shift_{value}) \quad 4$$

iii. Stretching: It altered the time dimension without changing the pitch. The derivation is represented as follows:

$$strected_{data} = librosa \cdot effects.time_{stretch(data,rate)} \quad 5$$

iv. Pitch Shifting: It modified the audio pitch to simulate different vocal characteristics. It is represented as follows:

$$pitch_{shifted_{data}} = librosa.effects.pitch_{shift(data,sr,n_{steps})} \quad 6$$

3.4 Model Architecture

The hybrid model combined CNN for spatial feature extraction and GRU for temporal sequence modelling as illustrated below in Figure 1.

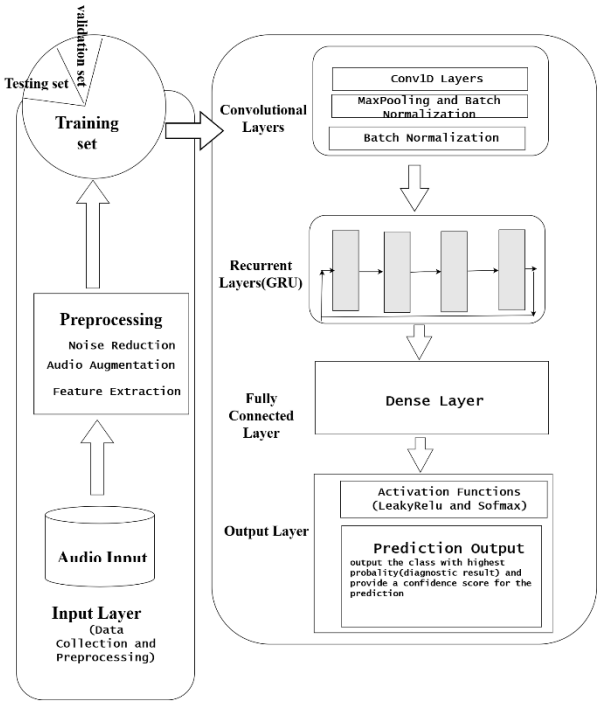

Figure 1: The architectural design for diagnosing and predicting paediatrics respiratory diseases.

3.4.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) layers were used for extracting spatial features from the MFCCs. These layers help detect local patterns in the MFCCs that corresponded to specific respiratory characteristics (wheezing, crackling).

The CNN architecture includes:

i. Conv1D Layers: It used Conv1D layers to detect local patterns in MFCCs corresponding to respiratory features. The derivation is represented as follows:

$$H_t = ReLU(W_c * X_t + b_c) \quad 7$$

where:

H_t is the output of the convolutional layer at time step t ,

W_c is the convolutional filter,

$*$ represents the convolution operation,

X_t is the MFCC input at time t,

b_c is the bias term.

ii. ReLU Activation: It is a non-linear activation function that ensured non-linearity in the model. The derivation is represented as follows:

$$ReLU(x) = max(x\ , 0) \quad 8$$

iii. MaxPooling: It reduced the dimensionality of the feature maps and retained the most important features while reducing the computational cost. The derivation is represented as follows:

$$P_t = \max\ (H_t) \quad 9$$

where:

P_t is the pooled feature at time step t.

iv. Batch Normalisation: It normalised the output of the previous layer, speeding up training and improving model stability. The derivation is represented as follows:

$$\widehat{P_t} = \frac{P_t - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad 10$$

where:

μ and σ^2 are the mean and variance of the pooled features,

ϵ is a small constant to prevent division by zero.

3.4.2 Gated Recurrent Unit (GRU)

It modelled temporal dependencies in sequential data, capturing how respiratory sounds evolve over time.

The GRU layers used update and reset gates for efficient sequence learning, with stacked layers for deeper temporal feature extraction. it is presented as follows:

a. Update Gate

The update gate controlled how much of the previous hidden state was retained; it is presented as follows:

$$z_t = \sigma(W_z[h_{t-1}\ , \widehat{P_t}\ ]) \quad 11$$

where:

z_t is the update gate value,

W_z is the weight matrix for the update gate,

h_{t-1} is the hidden state from the previous time step,

$\widehat{P_t}$ is the current input from the CNN,

σ is the sigmoid activation function.

b. Reset Gate

The reset gate controlled how much of the past information that was forgotten, as represented as follows:

$$r_t = \sigma(W_r[h_{t-1}\ , \widehat{P_t}\ ]) \quad 12$$

where:

r_t is the reset gate value,

W_r is the weight matrix for the reset gate.

The candidate hidden state $\widehat{h_t}$ computed as follows:

c. Candidate Hidden State

The candidate hidden state $\widehat{h_t}$ computed as follows:

$$\widehat{h_t} = \tanh\ (W[r_t \odot h_{t-1}, \widehat{P_t}) \quad 13$$

where:

$\widehat{h_t}$ is the candidate hidden state,

$\odot$ represents element-wise multiplication,

W is the weight matrix.

d. Final Hidden State

The final hidden state at time t is a combination of the previous hidden state and the candidate hidden state, which was controlled by the update gate:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \widehat{h_t} \quad 14$$

3.4.3 Dense Layers and Softmax Output:

The Dense layers refined features extracted by GRU using Leaky ReLU activation.

The Dense Layers with Leaky ReLU activation was used to process the output from GRU as computed as follows:

$$o_t = LeakyReLU(W_d . h_t + b_d) \quad 15$$

where:

o_t is the output of the dense layer,

W_d and b_d are the weights and biases of the dense layer.

A softmax output layer produced probability distributions across five disease classes.

The probability distribution over the classes is represented as follows:

$$\widehat{y_t} = \text{softmax}(W_o . o_t + b_o) \quad 16$$

where:

$\widehat{y_t}$is the probability distribution over the classes at time step t,

$W_o$and $b_o$are the weights and biases of the output layer.

The softmax function is defined as follows:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad 17$$

where $x_i$is the input to the softmax layer for the i-th class.

3.5 Training and Evaluation

The model was trained using the categorical cross-entropy loss function and optimized with the Adam optimizer. The derivate is computed as follows:

$$\text{Loss} = -\sum_i^C y_i \log(\hat{y}_1) \quad 18$$

where C is the number of classes, $y_i$is the true label for class i, and $\hat{y}_1$is the predicted probability for class i.

Early stopping and model checkpointing were employed to prevent overfitting. The performance was assessed using the following metrics:

i. Accuracy: Proportion of correctly classified instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN} \quad 19$$

where:
TP (True Positives) correctly predicted positive instances.
TN (True Negatives) correctly predicted negative instances.
FP (False Positives) incorrectly predicted positive instances.
FN (False Negatives) incorrectly predicted negative instances.

ii. Precision: Accuracy of positive predictions. It indicated the model's accuracy in predicting positive instances.

$$Precision = \frac{TP}{TP+FP} \quad 20$$

where:
TP: True Positives
FP: False Positives

iii. Recall: Ability to identify positive instances. It measured the model's ability to identify positive cases.

$$Recall = \frac{TP}{TP+FN} \quad 21$$

where:
TP: True Positives
FN: False Negatives

iv. F1-Score: Harmonic mean of precision and recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad 22$$

v. Confusion Matrix: Displayed correct and incorrect predictions for each class. it is represented as follows:

$$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \quad 23$$

3.6 Diagnostic and Predictive Tasks

The system performed two primary tasks:

a. Diagnostic Task: It classified respiratory conditions into predefined categories (COPD, Bronchiolitis, Pneumonia, URTI, and Healthy).

b. Predictive Task: It leveraged GRU's temporal modelling to forecast the progression of respiratory conditions based on sequential data.

3.7 Data Flow Design and Flowchart design for paediatrics respiratory diseases diagnostic and prediction system

i. Audio Input → MFCC Feature Extraction: The respiratory sound recordings were converted into MFCC features.

ii. Data Augmentation: Various augmentation techniques were applied to expand the dataset.

iii. CNN Layers: The augmented MFCCs were processed by the CNN layers to extract spatial features.

iv. GRU Layers: The output of the CNN was fed into the GRU layers, which captured the temporal patterns in the data.

v. Dense Layers → Output Layer: The GRU output was processed by the dense layers, and the SoftMax output layer generated the final class probabilities.

vi. Evaluation Metrics: The model's predictions were compared with the ground truth to evaluate its performance.

The system processes data through sequential stages as illustrated below in Figure 2

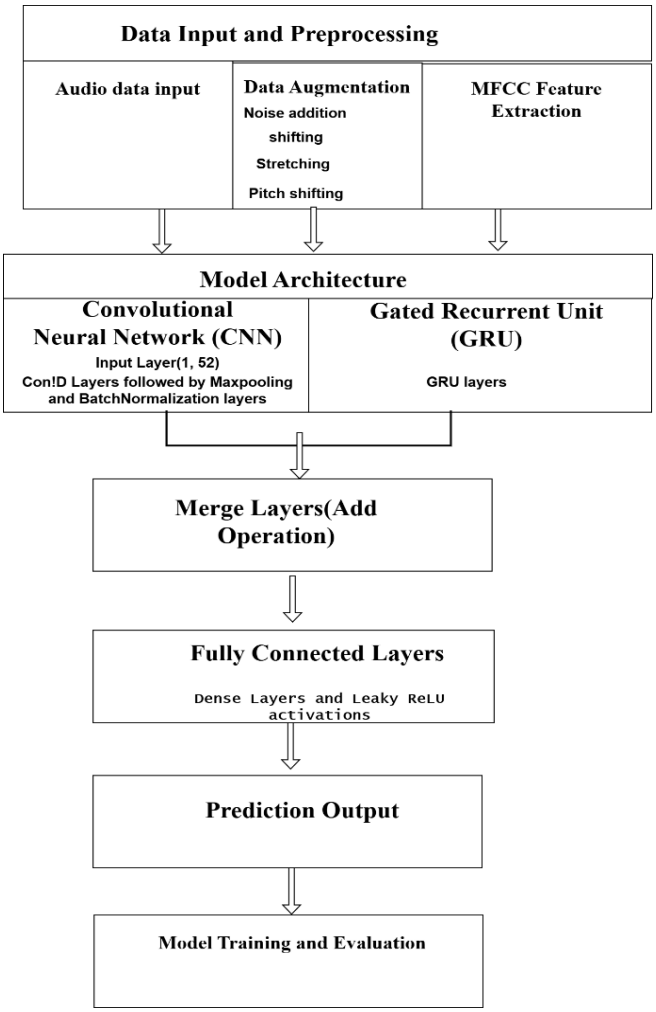

Figure 2: The system design for the paediatrics respiratory disease diagnosis and prediction model using deep learning

3.7.1 Connecting the Flowchart

The flowchart consists of the following sequential steps, as shown Figure in 3, and illustrates how the audio input processes through various stages to produce a final output and evaluate the model's performance as follows:

1. Audio input → 2. MFCC extraction → 3. Data augmentation → 4. CNN layers → 5. GRU layers → 6. Dense and softmax layers → 7. Evaluation.

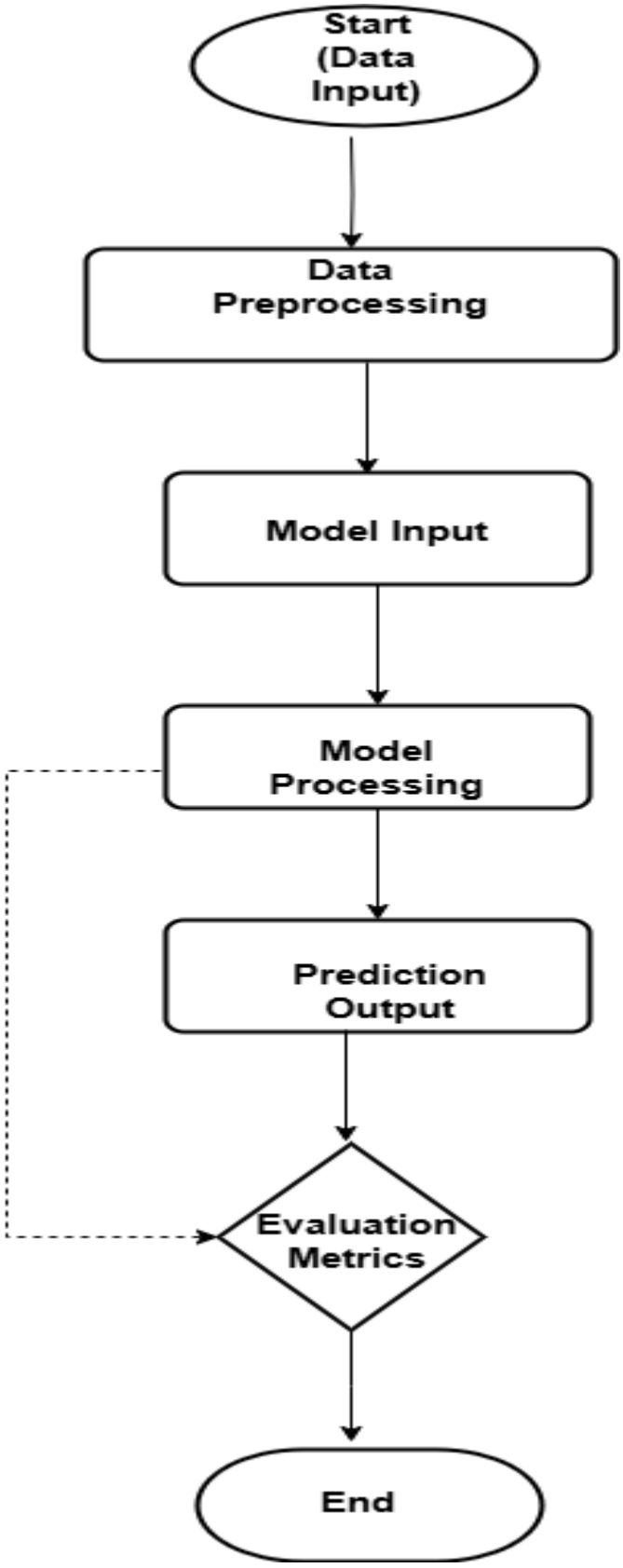

Figure 3: The flowchart for the paediatrics respiratory disease diagnosis and prediction model using deep learning

The methodology combines advanced signal processing techniques and deep learning architectures to deliver a robust system for diagnosing and predicting paediatric respiratory diseases, addressing challenges such as data variability, temporal complexity, and interpretability.

4. IMPLEMENTATION RESULTS AND DISCUSSION

4.1 Implementation specification of the system

The paediatric respiratory disease diagnostic and prediction system was developed to classify conditions such as COPD, Bronchiolitis, Pneumonia, URTI, and Healthy cases. The system utilized Python as the programming language and several deep learning libraries, including TensorFlow, Keras, and Librosa, for model building, training, and audio feature extraction. Libraries like Numpy and Pandas facilitated data handling, while Seaborn and Matplotlib were used for visualization. Scikit-learn supported dataset splitting and evaluation metrics. This combination of tools ensured a robust pipeline for analysing respiratory sounds using Mel-Frequency Cepstral Coefficients (MFCCs), a key feature in distinguishing between various conditions.

The implementation required a hardware setup with GPU support for faster training and at least 8GB of RAM to process audio data effectively. Software requirements included Python 3.7+, TensorFlow 2.x, and Librosa for audio analysis. This setup provided the computational efficiency and flexibility needed to train and test the deep learning model, making it capable of addressing the complexities of diagnosing paediatric respiratory diseases with high accuracy.

4.2 Dataset Description

The Kaggle Respiratory Sound Database comprises 920 annotated recordings from 126 patients, spanning children, adults, and the elderly, totalling 5.5 hours of respiratory sounds. These recordings, ranging from 10 to 90 seconds in length, include 6898 respiratory cycles, with 1864 containing crackles, 886 with wheezes, and 506 exhibiting both. The dataset captures clean and noisy respiratory sounds, simulating real-life conditions. Digital stethoscopes and advanced recording techniques were employed to gather this data, which is essential for understanding respiratory health and diagnosing conditions like asthma, pneumonia, and chronic obstructive pulmonary disease (COPD).

The dataset is divided into five respiratory condition classes: COPD, Bronchiolitis, Pneumonia, Upper Respiratory Tract Infection (URTI), and Healthy cases. Each class represents unique symptoms and acoustic characteristics—COPD involves airflow limitations with wheezing, Bronchiolitis affects infants with chest congestion, Pneumonia causes lung inflammation with breathing difficulties, URTI involves nasal congestion and sneezing, while Healthy cases exhibit normal respiratory sounds. This comprehensive dataset provides a valuable resource for applying deep learning techniques to automate the diagnosis and prediction of paediatric and adult respiratory diseases

4.3 Loading and Analysing Audio Files.

The audio files were loaded and analysed using the `librosa` library, which converted respiratory recordings into a consistent format (e.g., 16 kHz sampling rate) and normalized signals for uniformity. The filenames were structured to include patient IDs, which were extracted and linked to diagnoses from a CSV file, resulting in a merged dataset containing filenames, patient IDs, and diseases. This dataset facilitated feature extraction, exploratory analysis, and model development, with COPD identified as the most common condition, highlighting class imbalance.

Table 1: A small subset of data with two columns: pid (patient ID) and disease

	Pid	Disease
0	162	COPD
1	193	COPD
2	138	COPD
3	207	COPD
4	176	COPD
5	151	COPD
6	215	Bronchiectasis
7	210	URTI
8	156	COPD
9	140	Pneumonia

Figure 4 illustrates the distribution of disease categories in the dataset. COPD is the most prevalent, with over 60 instances, followed by Healthy samples (around 25) and URTI (approximately 15). Bronchiectasis, Pneumonia, and Bronchiolitis each have 5–10 samples, while Asthma and LRTI are rare, with fewer than 5 instances. This significant class imbalance could impact the model's performance, favouring well-represented categories like COPD while underperforming on rarer conditions.

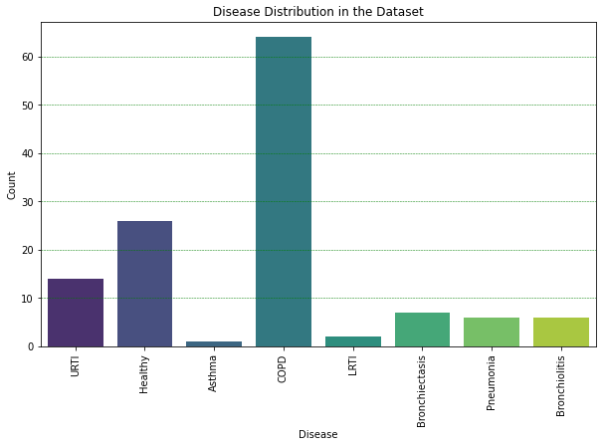

Figure 4: The disease distribution in the dataset.

4.4 Feature Extraction Using MFCCs and Data Augmentation

MFCCs were extracted using librosa to capture key acoustic features like pitch and intensity, creating fixed-length feature vectors for each audio sample. Data augmentation techniques,

including noise addition, time shifting, time stretching (1.2x and 0.8x), and pitch shifting (+3 semitones), were applied to enhance dataset diversity and improve model generalization to real-world variability in respiratory sounds.

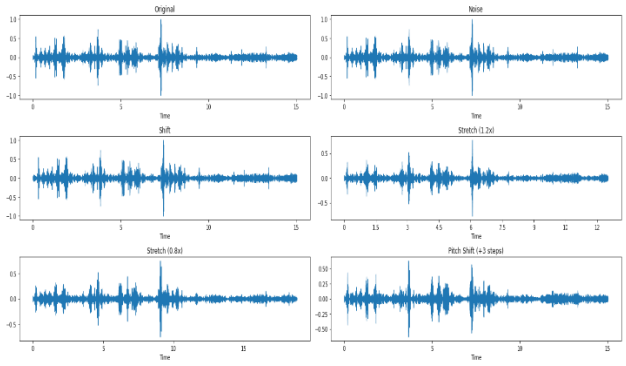

Figure 5: The MFCC waveform plots.

Figure 5 illustrates MFCC spectrograms for original and augmented audio data, representing audio in the frequency domain. The x-axis indicates time, the y-axis shows frequency bands, and colour intensity represents energy levels, with lighter areas signifying higher energy.

Each augmentation affects the spectrogram differently: noise introduces variability in low-energy bands; shifting displaces the audio along the time axis; time stretching (1.2x and 0.8x) compresses or elongates the frequency patterns; and pitch shifting (+3 semitones) raises the frequencies without altering timing. These augmentations enhance the model's ability to generalize by exposing it to diverse acoustic variations.

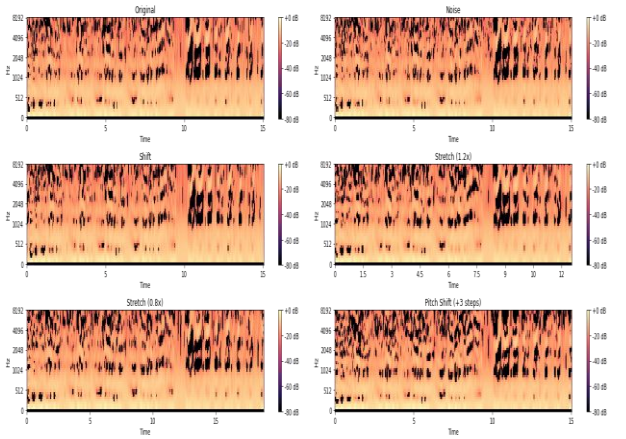

Figure 6: The MFCC Spectrograms.

4.5 Label Encoding

Label encoding was used to convert categorical disease labels (COPD, Bronchiolitis, Pneumonia, URTI, Healthy) into numerical representations. One-hot encoding was applied, assigning each label a unique binary vector, enabling the model to process multi-class classification effectively.

```
# Convert to numpy arrays
mfccs_features = np.array(x_mfccs)
labels = np.array(Y_data)

# Check the shapes of both arrays
print("MFCCs Features Shape:", mfccs_features.shape)
print("Labels Shape:", labels.shape)
```

```
MFCCs Features Shape: (915, 52)
Labels Shape: (915, 5)
```

Figure 7: The code snippet (encoding disease labels into a one-hot vector format).

Figure 7 demonstrates the one-hot encoding process, where categorical disease labels (COPD, Bronchiolitis, Pneumonia, URTI, Healthy) are mapped to unique binary vectors. For instance, 'COPD' is represented as [1, 0, 0, 0, 0], while 'Healthy' is [0, 0, 0, 0, 1]. The encoded labels are stored in a float64 array (`Y data`) with a shape of (915, 5), corresponding to 915 samples and 5 disease categories. This encoding format is essential for multi-class classification, enabling the model to process and predict disease categories effectively.

4.6 Addressing Class Imbalance

To address class imbalance, the implementation used oversampling to duplicate samples from underrepresented classes and applied additional augmentations like noise and pitch-shifting more frequently to minority classes such as Bronchiolitis and Pneumonia. A bar chart visualized the dataset distribution, showing Pneumonia (222) and Healthy (210) as the most represented categories, while URTI (138) had the lowest count. Bronchiolitis (174) and COPD (171) had moderate representation. These techniques ensured a more balanced dataset, enabling the model to learn effectively across all categories

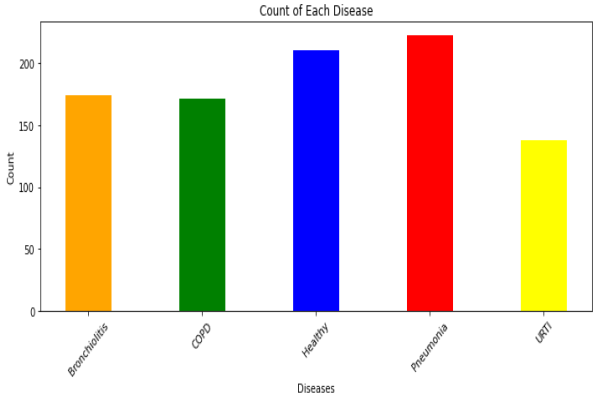

Figure 8: The distribution of disease labels after data augmentation.

4.7 Dataset Splitting and Sampling

The dataset was pre-processed and split into training (82.5%), validation (17.5%), and test (7.5%) sets using stratified sampling to maintain proportional representation of all classes. This approach ensured balanced distribution across all splits, with the training set used for model training, the validation set for performance monitoring, and the test set for evaluating final model accuracy on unseen data.

```
mfcc_train, mfcc_val, lables_train, lables_val = train_test_split(mfccs_feature
mfcc_train, mfcc_test, lables_train, lables_test = train_test_split(mfcc_train,

print("MFCC Train Shape:", mfcc_train.shape)
print("MFCC Validation Shape:", mfcc_val.shape)
print("MFCC Test Shape:", mfcc_test.shape)

print("Labels Train Shape:", lables_train.shape)
print("Labels Validation Shape:", lables_val.shape)
print("Labels Test Shape:", lables_test.shape)
```

```
MFCC Train Shape: (697, 52)
MFCC Validation Shape: (161, 52)
MFCC Test Shape: (57, 52)
Labels Train Shape: (697, 5)
Labels Validation Shape: (161, 5)
Labels Test Shape: (57, 5)
```

Figure 9: The code snippet demonstrates the splitting of a dataset into training, validation, and testing sets.

The dataset splitting results show the following:

i. MFCC Train Shape: (697, 52) indicates 697 training samples, each with 52 features.
ii. MFCC Validation Shape: (161, 52) represents 161 validation samples with 52 features each.
iii. MFCC Test Shape: (57, 52) consists of 57 test samples, each with 52 features.
iv. Labels Train Shape: (697, 5) corresponds to 697 training labels across 5 classes.
v. Labels Validation Shape: (161, 5) and Labels Test Shape: (57, 5) represent label distributions for the validation and test sets, respectively.

These shapes confirm the dataset is properly split for training, validation, and testing.

4.8 Implementation and Results

The implemented model utilizes a hybrid architecture combining Convolutional Neural Networks (CNNs) for feature extraction and Gated Recurrent Units (GRUs) for sequence modelling, tailored for diagnosing and predicting paediatric respiratory diseases. The CNN layers employ `Conv1D` with filters of increasing sizes (16, 32, 64) and the following components; BatchNormalization and MaxPooling stabilize learning and reduce dimensionality. The final CNN output is flattened into a 1D array, which is fed into a GRU layer with 128 units to capture temporal dependencies.

The architecture also includes two fully connected dense layers with ReLU activation and dropout (0.5) for regularization. The final layer is a softmax output layer with five units representing the disease categories. The model was compiled with the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric, effectively leveraging the strengths of CNNs and GRUs for robust multi-class classification.

Table 2: A summary of a sequential neural network model with multiple layers.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 50, 16)	64
batch_normalization (BatchNo	(None, 50, 16)	64
max_pooling1d (MaxPooling1D)	(None, 25, 16)	0
...		
gru (GRU)	(None, 128)	75552
dense (Dense)	(None, 64)	8256
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2080
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 5)	165

Total params: 139821

Trainable params: 138285

Non-trainable params: 1536

4.9 Model Performance and Results

The model was trained over 50 epochs, achieving a final training accuracy of 99.14% and a loss of 0.0372, indicating a strong fit to the training data. Validation accuracy improved steadily from 55.9% to 88.20%, while validation loss fluctuated but showed an overall decline, reflecting effective learning and minimal overfitting.

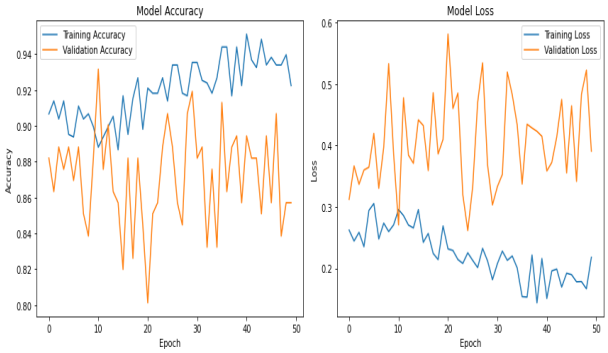

Figure 10: The training and validation accuracy and training and validation loss of a model over 50 epochs.

The left plot shows training accuracy stabilizing around 94%, while validation accuracy fluctuates, stabilizing at 88–90%. The right plot indicates steady training loss reduction, but validation loss fluctuates with occasional spikes, suggesting potential overfitting.

4.10 Evaluation Metrics

The classification report, as shown in Table 4.3, provides a comprehensive evaluation of how well the model performs across the different classes.

Table 3: The precision, recall, and F1-score for a model's performance across five classes: COPD, Bronchiolitis, Pneumonia, URTI, and Healthy.

Class	Precision	Recall	F1-Score
COPD	1.00	0.70	0.82
Bronchiolitis	0.92	0.92	0.92
Pneumonia	0.69	1.00	0.82
URTI	0.73	0.67	0.70
Heathy	0.92	0.92	0.92
Overall	0.85	0.84	0.84

Table 3 summarizes the model's performance:

a. COPD: Precision 1.00, Recall 0.70, F1-Score 0.82.

b. Bronchiolitis & Healthy: High accuracy with Precision, Recall, and F1-Score of 0.92 each.

c. Pneumonia: Precision 0.69, Recall 1.00, F1-Score 0.82.

d. URTI: Precision 0.73, Recall 0.67, F1-Score 0.70.

Overall metrics show Precision (0.85), Recall (0.84), and F1-Score (0.84), indicating strong performance with improvement needed for COPD and URTI recall.

The model achieved 84% test accuracy, with strong performance in Bronchiolitis and Healthy (Precision and Recall: 92%). COPD had high precision (100%) but lower recall (70%), while Pneumonia showed 69% precision and 100% recall. The results highlight robust generalization and effective disease classification despite challenges with overlapping symptoms.

4.10.1 ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)

ROC-AUC evaluates the model's ability to distinguish between classes, with the curve plotting sensitivity vs. 1-specificity and the AUC indicating the likelihood of correctly ranking positive over negative examples.

Table 4: The confusion matrix for the model.

Actual/Predicted	COPD	Bronchiolitis	Pneumonia	URTI	Healthy
COPD	14	1	0	0	0
Bronchiolitis	1	11	2	1	0
Pneumonia	0	2	13	1	0
URTI	0	1	1	14	0
Healthy	0	0	0	0	15

The confusion matrix shows high accuracy for COPD, Pneumonia, URTI, and Healthy, with Healthy perfectly classified. Misclassifications are most common for Bronchiolitis, which overlaps with other categories like COPD, Pneumonia, and URTI.

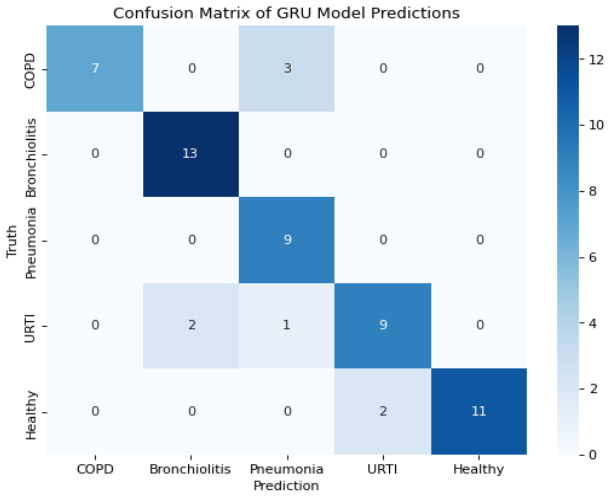

Figure 11: A confusion matrix and its visualisation to evaluate the performance of the model.

Figure 11 shows the GRU model's confusion matrix for five disease categories. Most classes, including COPD, Healthy, and Bronchiolitis, were well-classified, though some misclassifications occurred, particularly between URTI and Pneumonia. High AUC scores highlight strong discriminative ability, especially for Healthy and COPD.

4.11 Diagnostic and Predictive Ability

The CNN-GRU model achieved 84% accuracy in classifying paediatric respiratory conditions using MFCCs, excelling in Bronchiolitis and Healthy cases. Despite slightly lower recall for COPD, the model's predictive ability, particularly in ambiguous cases like Pneumonia and URTI, shows practical potential. With further refinement and validation, it can serve as a non-invasive diagnostic tool for telemedicine platforms.

The CNN-GRU-based paediatric respiratory disease diagnostic system showed high accuracy, particularly in diagnosing COPD and Healthy cases, demonstrating its potential for real-world applications. This study highlights the effectiveness of combining CNNs and GRUs, addresses gaps in paediatric respiratory diagnostics, and emphasizes data augmentation's role in enhancing deep learning performance for medical applications.

5. CONCLUSION

The study concludes that the deep learning model integrating CNN and GRU demonstrates significant potential for diagnosing and predicting paediatrics respiratory diseases through audio recordings of respiratory sounds. The model performs particularly well for COPD and Healthy cases, showing high accuracy in training and validation, making it promising for real-world applications.

The research contributes to knowledge by introducing an innovative, scalable, and efficient diagnostic framework that leverages advanced feature extraction and sequential pattern analysis. This tool is particularly valuable for early disease detection in resource-constrained environments and sets a foundation for further exploration of hybrid deep learning architectures in medical audio diagnostics.

Recommendations include expanding the dataset to address class imbalances, particularly for Bronchiolitis and Pneumonia, and employing advanced audio augmentation techniques to

enhance generalization. Rigorous testing in diverse clinical settings is advised to ensure robustness before deployment. With these improvements, the system has the potential to revolutionize the early detection and management of paediatrics respiratory diseases, saving lives through accurate and timely diagnostics.

6. REFERENCES

[1] Zar, H. J., & Ferkol, T. (2014). The global burden of respiratory disease-Impact on child health. Pediatric Pulmonology, 49(5), 430–434. https://doi.org/10.1002/ppul.23030

[2] Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. Healthcare, 10(3), 541. https://doi.org/10.3390/healthcare10030541

[3] Xue, L., Liu, C., Xue, W., Xue, R., Liu, P., & Wang, F. (2022, January 1). The role of nurses in the management of respiratory disorders in children. PubMed. https://pubmed.ncbi.nlm.nih.gov/34653020/

[4] Rakel, R. E. (2023, September 13). Diagnosis | Definition, Types, Examples, & Facts. Encyclopedia Britannica. https://www.britannica.com/science/diagnosis

[5] Barakat, N., Awad, M., & Abu-Nabah, B. A. (2023). A machine learning approach on chest X-rays for pediatric pneumonia detection. Digital Health, 9, 205520762311800. https://doi.org/10.1177/20552076231180008

[6] Wang, X. (2019, July 25). Deep learning models to predict pediatric asthma emergency department visits. arXiv.org. https://arxiv.org/abs/1907.11195

[7] Chen, K. C., Yu, H., Chen, W. S., Lin, W. C., Lee, Y. C., Chen, H., Jiang, J. H., Su, T., Tsai, C. K., Tsai, T. A., Tsai, C. M., & Lu, H. H. S. (2020). Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-73831-5

[8] Chen, Y., Roberts, C., Ou, W., Petigara, T., Goldmacher, G. V., Fancourt, N., & Knoll, M. D. (2021). Deep learning for classification of pediatric chest radiographs by WHO's standardized methodology. PLOS ONE, 16(6), e0253239. https://doi.org/10.1371/journal.pone.0253239

[9] Kang, C. M., Shanmugam, S. A., & Nugroho, H. A. (2022). Respiratory Anomalies and Diseases Detection with Deep Learning. In Springer eBooks (pp. 439–448). https://doi.org/10.1007/978-981-16-8129-5_68

[10] Kim, B. J., Kim, B. S., Mun, J. H., Lim, C., & Kim, K. H. (2022). An accurate deep learning model for wheezing in children using real world data. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-25953-1

[11] Mahmoudi, R., Benameur, N., Mabrouk, R., Mohammed, M. A., Garcia-Zapirain, B., & Bedoui, M. H. (2022). A Deep Learning-Based Diagnosis system for COVID-19 detection and pneumonia screening using CT imaging. Applied Sciences, 12(10), 4825. https://doi.org/10.3390/app12104825

[12] Podder, P., Das, S. R., Mondal, M. R. H., Bharati, S., Maliha, A., Hasan, J., & Piltan, F. (2023). LDDNET: A deep learning framework for the diagnosis of infectious lung diseases. Sensors, 23(1), 480. https://doi.org/10.3390/s23010480

[13] Priyadarsini, M. J. P., Kotecha, K., Rajini, G. K., Hariharan, K., Raj, K. U., Ram, K. B., Indragandhi, V., Subramaniyaswamy, V., & Pandya, S. (2023). Lung diseases detection using various deep learning algorithms. Journal of Healthcare Engineering, 2023, 1–13. https://doi.org/10.1155/2023/3563696

[14] Kwak, G. H. (2019, September 1). DeepHealth: Review and challenges of artificial intelligence in health informatics. arXiv.org. https://arxiv.org/abs/1909.00384 6

[15] Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

[16] Kundu, R., Das, R., Geem, Z. W., Han, G., & Sarkar, R. (2021). Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLOS ONE, 16(9), e0256630. https://doi.org/10.1371/journal.pone.0256630

[17] Tran, T. T., Pham, H. H., Nguyễn, T. V., Le, T. T., Nguyen, H., & Nguyen, H. Q. (2021). Learning to automatically diagnose multiple diseases in pediatric chest radiographs using deep convolutional neural networks. medRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2021.08.12.21261954

[18] Li, L., Alimu, A., Luan, Q., Yang, B., Yilamujiang, S., Gong, H., Zulipikaer, A., Xu, J., Zhong, X., Ren, J., & Zou, X. (2022). Prediction and diagnosis of respiratory disease by combining convolutional neural network and bi-directional Long Short-Term memory methods. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.881234

[19] Newman-Toker, D. E., Schaffer, A. C., Yu-Moe, C. W., Nassery, N., Tehrani, A. S. S., Clemens, G., Wang, Z., Zhu, Y., Fanai, M., & Siegal, D. (2019). Serious misdiagnosis-related harms in malpractice claims: The "Big Three" – vascular events, infections, and cancers. Diagnosis, 6(3), 227–240. https://doi.org/10.1515/dx-2019-0019

[20] Pham, L. (2020, January 21). Robust Deep Learning Framework For Predicting Respiratory Anomalies and Diseases. arXiv.org. https://arxiv.org/abs/2002.03894

[21] Sharan, R. V., & Rahimi-Ardabili, H. (2023). Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review. International Journal of Medical Informatics, 176, 105093. https://doi.org/10.1016/j.ijmedinf.2023.105093

[22] Srinath, S., Jacob, P., Sharma, E., & Gautam, A. (2019). Clinical practice guidelines for assessment of children and adolescents. Indian Journal of Psychiatry, 61(8), 158. https://doi.org/10.4103/psychiatry.indianjpsychiatry_580_18

[23] Yu, G., Yu, Z., Shi, Y., Wang, Y., Liu, X., Li, Z., Zhao, Y., Sun, F., Yu, Y., & Shu, Q. (2021). Identification of pediatric respiratory diseases using a fine-grained diagnosis system. Journal of Biomedical Informatics, 117, 103754. https://doi.org/10.1016/j.jbi.2021.103754