AI DRIVEN APPROACH FOR BRONCHOPNEUMONIA DETECTION IN CHILDREN

Project Reference No.: *47S_MSC_0256*

College ***:*** *St. Aloysius (Deemed To Be University), Mangaluru*
Branch ***:*** *Department of Software Technology*
Guide(S) ***:*** *Dr. Ruban S.*
Mrs. Suchetha Vijayakumar
Student(S ***:*** *Ms. Shraddha H. Shetty*
Ms. Yashaswi K.

Keywords

Artificial Intelligence, Machine Learning, Bronchopneumonia, pneumonia, children.

Introduction

Artificial Intelligence is a helpful guide in healthcare because it uses advanced algorithms to deal with health challenges. These machine learning algorithms use reliable and high-quality data to produce the required outcome. The quality of data plays an important role in effective Artificial Intelligence solutions in healthcare that ensure accurate and beneficial advancements in patient care and overall health management.

A form of pneumonia called bronchopneumonia attacks the bronchi in the lungs. Although fungal and viral infections can also cause this illness, bacterial infections are the usual cause. While symptoms can vary, fever, coughing, and breathing difficulties are frequently experienced. Breathing difficulties may arise from narrowed airways in a person with bronchopneumonia. Their lungs can receive little air due to inflammation.

Our study helps us by providing a more accurate assessment of bronchopneumonia. Through practical literature reviews and live data sets, we've validated its performance, showing superior results compared to manual methods. This innovative approach helps in improving children's respiratory healthcare and a better outcome for bronchopneumonia in children, offering clinicians a precise tool for bronchopneumonia diagnoses.

Objectives

- Our primary aim is to enhance bronchopneumonia assessment by conducting practical reviews and analyzing live data.
- We will specifically focus on detecting improvements in children suffering from bronchopneumonia, comparing outcomes against manual assessment methods.
- Through rigorous analysis, we intend to showcase superior results achieved through our approach compared to conventional manual methods.
- Ultimately, our study seeks to provide tangible evidence supporting the efficacy and efficiency of our methodology in assessing and monitoring the progression of bronchopneumonia in pediatric patients.

Methodology

Data Sources, Real patient data was collected from Father Muller Medical Hospital in Mangalore City. With this data, we establish a robust knowledge base, addressing the uncertainties in bronchopneumonia diagnosis. This data-driven approach, combined with expert opinions, enhances outcome predictions, surpassing conventional methods. To perform this study, we first tried to convert the data into a data dictionary, which gave us a brief view of the collected data. We then transformed it into a format where our model could be evaluated. A huge amount of data was collected for this process, in which most of the patients were babies ages 2 months to more than a year.

A sample of the data set which was collected from the hospital is given below:

Gender	DOA	DischargeDate	Age		Religion	CITY	DIAGNOSIS		CHIEF COMPLAINTS				HISTORY OF PRESENTING ILLNESS		PAST HISTORY
			Year	Month			1	2	1	2	3	4	1	2	
									Fever days	Cough/Cold	Breathing difficulty				
F	22-09-2021 / 15:07	24-09-2021 / 13:51	0	11	Christian	mangalore	BRONCHOPNEUMONIA		3	Cough since 1 week			Child was apparent	Not associated with	History of fever one week
M	11-08-2021 / 16:11	24-08-2021 / 08:31	0	6	muslim	mangalore	OPNEUMONIA	SEPTIC SHOCK - RESOLVE	5	Cough and cold since 5		Loose stools since 5	Child was apparent	History of cough, wet type,	History of fever for 2
M	27-07-2021 / 12:18	31-07-2021 / 06:54	0	5	Hindu	Chitradurga	OPNEUMONIA		5	Cough since 7days			Child was apparent	Fever since 5 days	No similar complai
M	26-04-2021 / 18:19	03-05-2021 / 13:29	0	1	muslim	Bhatkal	OPNEUMONIA	?PARTIAL ANOMALOUS		Cold and coryza since 3	1		One and a half month	History of hurried	significant
F	20-05-2021 / 18:37	28-05-2021 / 15:51	0	1	muslim	mangalore	BRONCHOPNEUMONIA		2	Cough x 3 days Cold x 3	1		Baby came with	History of fever since 2	No history of
M	12-07-2021 / 11:42	17-07-2021 / 16:23	0	1	muslim	mangalore	OPNEUMONIA	COVID - 19 INFECTIO	2	Cough since 2 days		bleed - 1 episode	Baby was apparent	Mother gives history	No similar complai
M	11-08-2021 / 22:38	21-08-2021 / 08:20	0	8	Hindu	Kaup	OPNEUMONIA	ONAL ANEMIA	5	Cough since 7days		Decreased activity since 2	Child was apparent	Fever since 5 days	No similar complai
F	16-08-2021 / 07:39	23-08-2021 / 16:46	0	2	hindu	bantwal	BRONCHOPNEUMONIA		Fever-1 episode	Cough since 3days	breathlessness and		Child was apparent	Child then	Nothing significant
F	05-09-2021 / 14:56	10-09-2021 / 15:34	0	4	muslim	kasaragod	OPNEUMONIA	NUTRITIONAL ANEMIA	2		Breathlessness X 1 day		Baby came with	Breathlessness in the form	History of fever two days
M	30-09-2021 / 05:25	07-10-2021 / 15:39	0	10	Hindu	ankola		RIGHT BRONCHOPNEUM	7	Cough since 7 days			Child was apparent	History of cough since 7	No similar history

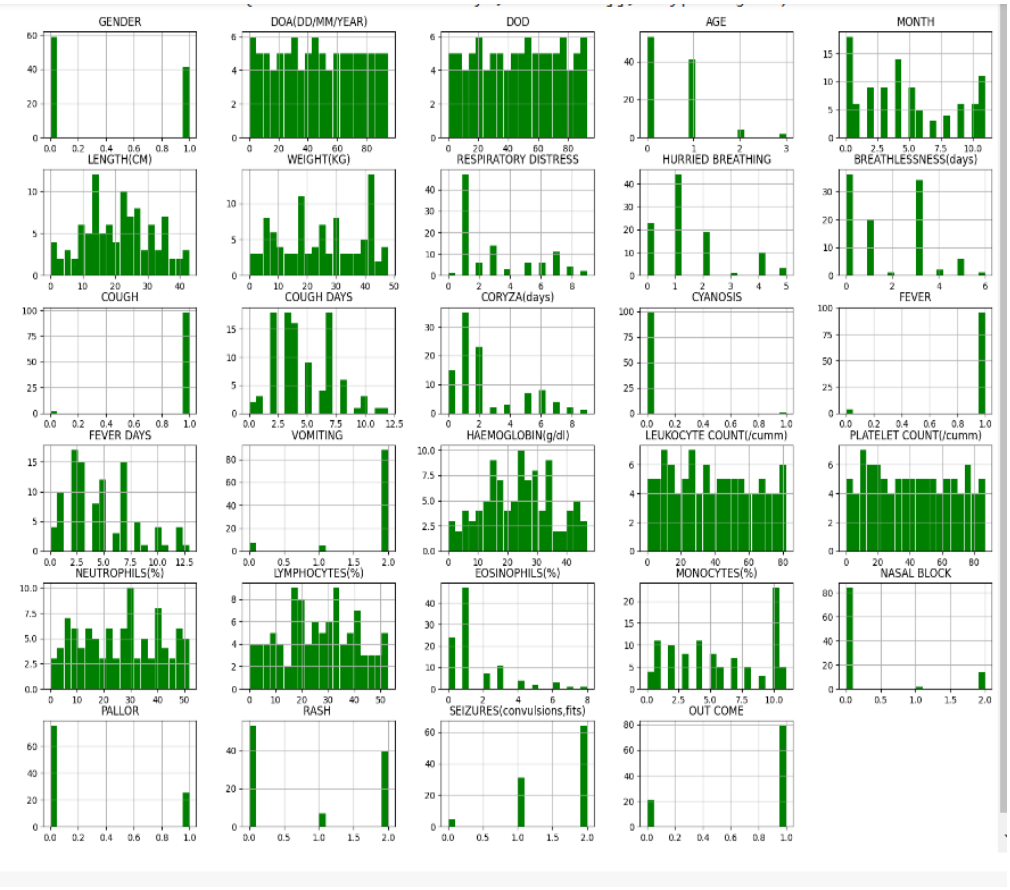


Fig: Visualization of the data set

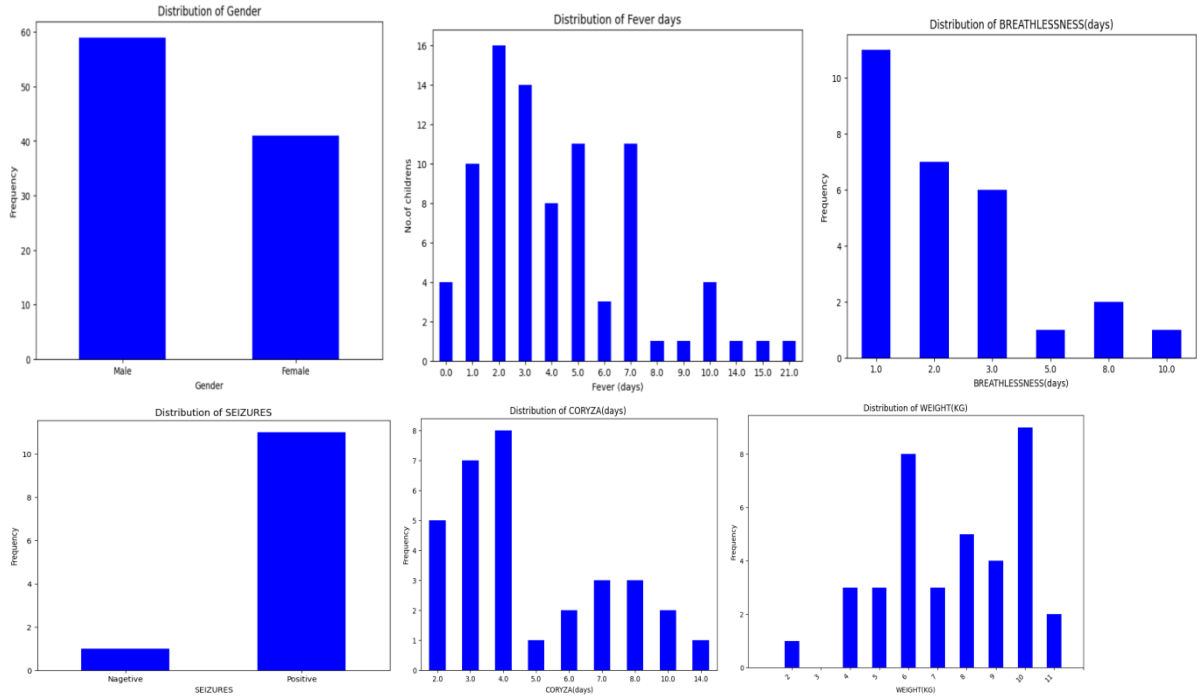
Data Preprocessing:

- Data Cleaning: Data cleaning was undertaken to address issues associated with missing values, illegible handwriting, and redundant data present in the clinical notes.
- Data integration: Data integration was performed to integrate the data from the medical records to create a uniform dataset for analysis.
- Data transformation: The data is then converted into a format that is process-friendly, which includes the date of admission and the date of discharge.
- Data reduction: In this, we removed redundant and unnecessary data. Only the required efficient data were kept.

Data Processing: In this project, we meticulously amassed a comprehensive dataset, focusing on identifying essential features critical for our subsequent analyses. This dataset was methodically organized into an Excel file, with numeric values carefully allocated to each feature. The dataset was developed iteratively through a methodical approach that included replacing missing values with means in order to handle missing values strategically and removing duplicate columns in order to maintain data integrity. The dataset's analytical compatibility was improved by transforming categorical data into numerical values using label encoding. In order to provide a consistent and adaptable format that facilitates easy integration into various analytical models, the last refinement involved transforming the entire dataset to float values. To facilitate the extraction of insights from the initially extensive information, we judiciously employed both Excel and the Google Collaboratory tool.

The development of a prediction model is the next stage. The whole collection of data is divided into "train" and "test" sets for this purpose. We developed a

supervised machine learning model, as this model's task is categorization. Classification algorithms such as Support Vector Machines, Random Forests, Naive Bayes, and Linear Discriminant Analysis were used in the creation of the model. After that, the model was fed input data, also known as training data, which was compiled from the clinical notes of the patients treated for bronchopneumonia at the Medical College Hospital. We utilized the 'train' set to teach the model, or fit it, so that it could discover patterns and relationships in the data. Using data that hasn't been seen before, the model's working accuracy is determined using the 'test' set. Using NumPy, Pandas, Matplotlib, Scikit-learn, and Seaborn, among other libraries and packages, the program was written in Python 3.7.



Machine Learning Model Development:

- Supervised machine learning models were developed for classification tasks.
- Algorithms such as Support Vector Machines, Random Forests, Naive Bayes, and Linear Discriminant Analysis were evaluated.
- The dataset was divided into training and testing sets.
- Models were trained on the training set and evaluated on the test set to assess their performance.

Results and Conclusions

Exploratory Data Analysis was performed on data from Father Muller Medical College, Mangalore, enabling machine learning tasks. For classification, algorithms like Support Vector Machine, Naïve Bayes, and Random Forest were evaluated, with metrics such as Accuracy, Precision, Recall, and F-Measure considered for performance assessment.

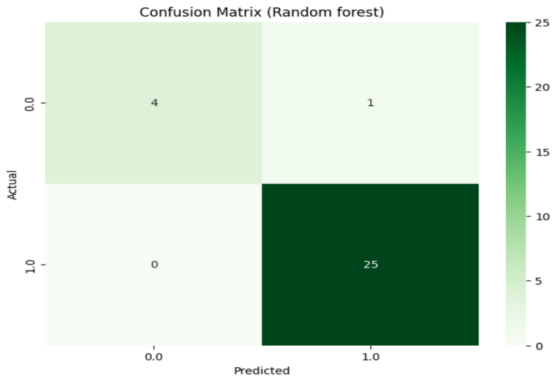


Figure 1: Confusion Matrix for Random Forest

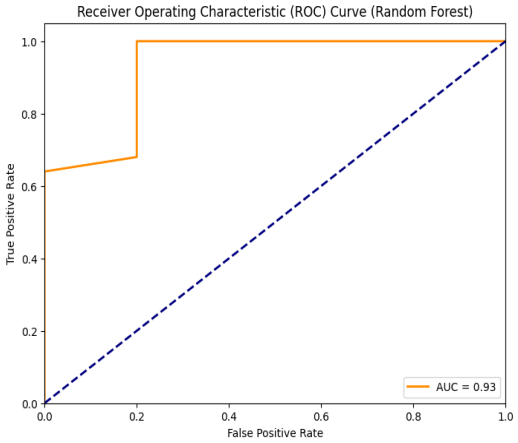


Figure 2 : ROC Cure for Random Forest

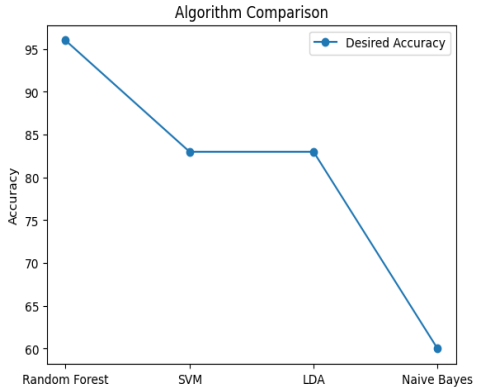


Figure 3 : Comparison of Algorithms

Table 1. Performance comparison of Different Machine Learning Algorithms

Algorithm	Precision	Sensitivity	F1-Score	Accuracy
Random Forest Classifier	0.9615	1.0	0.9804	96
SVM	0.8333	1.0	0.9090	83
LDA	0.9545	0.84	0.8936	83
Naive bays	1.0	0.52	0.6842	60

Innovation in the Project

The Bronchopneumonia detection project lies in its integration of artificial intelligence and machine learning, enabling advanced and data-driven diagnostic approaches. By leveraging real patient data and employing comprehensive data processing techniques, the project automates the diagnostic process and enhances accuracy. Through the development and evaluation of machine learning models, clinicians gain access to quicker, more consistent, and reliable diagnostic tools, ultimately improving patient care outcomes.

Scope for Future Work

Collaborating with diverse healthcare institutions enhances the model's adaptability, capturing regional variations in bronchopneumonia cases. Acquiring more data is crucial for higher accuracy, facilitating better pattern recognition and

diagnostic predictions. Integration of X-ray data significantly enhances diagnostic accuracy, improving the overall effectiveness of the diagnostic system and offering a more nuanced evaluation of bronchopneumonia cases.