**INTRODUCTION**Respiratory diseases are the leading causes of death and disability worldwide, with the poorest regions having the greatest disease burden. Factors such as age, smoking, environmental pollution, and body weight also contribute to the problem. Chronic respiratory diseases, including pneumonia, TB, and lung cancer, account for 7% of all deaths worldwide and are the third leading cause of death. Between 1990 and 2017, the number of deaths due to respiratory diseases increased by 18%. Over 1 billion people suffer from acute or chronic respiratory conditions, with 4 million prematurely dying from chronic respiratory disease each year. Social deprivation is the most important factor affecting death and disability rates, with the highest rates seen in the poorest regions. Early diagnosis and intervention in respiratory diseases are crucial, but 45% of WHO Member States report having less than one physician per 1000 population. Automatic and reliable tools can help doctors save time and reduce errors due to work overload.

**NEED OF THE STUDY.**

The rapid growth of respiratory diseases has led to a growing interest in integrating audio signal analysis-based techniques in medical research. Computer science has improved the ability to analyze media data automatically and process image and audio information, allowing for faster and reliable diagnosis tools. Audio signal analysis tools can help detect respiratory problems and aid in timely diagnosis in the early stages of respiratory dysfunction. Traditional stethoscopes may be exposed to external noise and cannot filter audio frequencies, leading to incorrect diagnosis. As lung and heart diseases remain the leading cause of death globally, advancements in computer vision and deep learning have contributed significantly to audio analysis. Automated algorithms for recognizing abnormalities in respiratory sounds may be of great relevance to clinical diagnosis. Researchers are also exploring combining speech and signal processing tools with image analysis-based techniques to help doctors predict or guess the presence of respiratory diseases based on verbal communication. Machine learning algorithms have been successfully used in various applications, including computerized lung sound analysis.

**RESEARCH METHODOLOGY**

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study,Data and Sources of Data, study’s variables and analytical framework. The detailsare as follows;

**3.1Population and Sample**

Lung sounds, caused by airflow during inspiration and expiration phases, are non-stationary and non-linear signals. They are difficult for physicians to recognize abnormalities. Abnormal breath sounds may indicate lung problems like inflammation or obstruction. Each lung disorder is associated with one or more lung sounds. The dominant frequency of heart sounds is typically below 150Hz, while lung sounds range between 150 and 2000Hz. The ICBHI dataset, compiled to support respiratory data analysis, is freely available for research.

**3.2 Data and Sources of Data**

The ICBHI dataset, created to support the 2017 International Conference on Biomedical Health Informatics, is now available for research. It contains both public and private datasets from the ICBHI challenge. The Respiratory Sound Database is the largest annotated, publicly available dataset. The dataset, collected by two research teams in Greece and Portugal, consists of 920 labeled audio tracks from 126 participants. The sounds were collected from six different positions and were collected in both clinical and non-clinical settings. Participants included patients with respiratory tract infections, pneumonia, bronchiolitis, COPD, asthma, bronchiectasis, and cystic fibrosis. The audios were collected from adult participants of varying ages, including those with lower and upper respiratory tract infections, pneumonia, bronchiolitis, COPD, asthma, bronchiectasis, and cystic fibrosis.

**3.3 Theoretical framework**

The ICBHI sound data, consisting of 920 annotated audio samples from 126 subjects, serves as a benchmark in the field of respiratory sound analysis. The dataset covers two broad groups: normal and problematic, with some cycles having both issues. Out of 6898 cycles totaling 5.5 hours, 3642 are healthy, while the remaining 3256 are problematic. Out of these problematic cycles, 1864 have crackles, 886 have wheezes, and 506 have both. Overall, there were 3642 healthy breath cycles and 3256 problematic breath cycles.

A single-channel respiratory sound is composed of four main components, two pauses, and two distinctive patterns. The respiratory cycle is described as starting from the inspiratory phase with a lower amplitude and regular pattern, followed by an expiratory phase with one or multiple peaks, a decreasing amplotude pattern, and higher average energy. Domain experts annotated the respiratory cycles to state the presence of crackles, wheezes, a combination of them, or no adventitious respiratory sounds. The recordings were collected using heterogeneous equipment, with duration ranging from 10 s to 90 s.