MAJOR PROJECT-II REPORT

on

Time Series Analysis of Spirograms for Early COPD Detection and Long-Term Risk Prediction

Submitted by:

Chetan Nihith Vemula (Enrollment No.)
Abhinav V (Enrollment No.)
Sanjay Reddy (Enrollment No.)
Likhit Sagar G (Enrollment No.)
Nithish Basireddy (Enrollment No.)

under mentorship of

Dr. Mentor Name
(Assistant/Associate/Professor)

Department of Computer Science Engineering
School of Engineering and Technology
BML Munjal University, Gurugram (India)

May 2024

Candidate's Declaration

I hereby certify that the work on the project entitled, "Time Series Analysis of Spirograms for Early COPD Detection and Long-Term Risk Prediction", in partial fulfillment of requirements for the award of Degree of Bachelor of Technology in School of Engineering and Technology at BML Munjal University, having University Roll No. 1232434, is an authentic record of my own work carried out during a period from July 2023 to May 2024 under the supervision of Dr. Mentor Name.

Chetan Nihith Vemula
Abhinav V
Sanjay Reddy
Likhit Sagar G
Nithish Basireddy

Supervisor's Declaration

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Faculty Supervisor Name: Dr. Mentor Name
Signature:

Acknowledgement

I am highly grateful to Dr. Mentor Name, Assistant/Associate/Professor, BML Munjal University, Gurugram, for providing supervision to carry out this project from July 2023 to May 2024. Dr. Mentor Name has provided great help in carrying out my work and is acknowledged with reverential thanks. Without wise counsel and able guidance, it would have been impossible to complete the project in this manner.

I would like to express thanks profusely to Dr. Mentor Name for stimulating me from time to time. I would also like to thank the entire team at BML Munjal University. I would also thank my friends who devoted their valuable time and helped me in all possible ways toward successful completion.

Chetan Nihith Vemula
Abhinav V
Sanjay Reddy
Likhit Sagar G
Nithish Basireddy

Contents

List of Figures

List of Tables

List of Abbreviations

COPD Chronic Obstructive Pulmonary Disease

CNN Convolutional Neural Network

LSTM Long Short-Term Memory

AUROC Area Under the Receiver Operating Characteristic Curve

AUPRC Area Under the Precision-Recall Curve

F1-score Harmonic Mean of Precision and Recall

Chapter 1

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a serious lung condition that makes it difficult for people to breathe over time. Detecting and predicting COPD in its early stages can help prevent severe health problems and improve patient outcomes. In this study, we present a deep learning-based approach called DlSpiro to detect and predict the future risk of COPD using spirogram time series data. Since we did not have access to real patient data, we created a synthetic dataset with 30,000 entries, including cases of hospitalization and death. Our method uses a step-by-step pipeline that first smooths the breathing signal, extracts important patterns, and then uses demographic details to make predictions. The model can identify if a person currently has COPD and also predict their risk of developing the disease over the next 1 to 5 years. This system can support early screening and help doctors provide timely care to high-risk individuals.

Chapter 2

Introduction

2.1 Overview

Chronic respiratory conditions are a growing concern worldwide, especially those that cause long-term breathing difficulties and reduce the overall quality of life. One such illness is a progressive lung disease that makes it harder for individuals to breathe properly over time. It affects daily activity, increases the chances of other health problems like heart conditions, and may even lead to premature death. The sooner this disease is detected, the better the chances are to manage it and slow down its effects.

2.2 Introduction to Project

In this study, we aim to develop an intelligent system that not only detects the presence of the disease but also predicts the chances of someone developing it in the future. Our goal is three-fold: to propose a new idea using deep learning, to build this system and test it using data we generated, and to compare our results with typical methods currently used in the medical field. Instead of using real-world patient data, we created a synthetic dataset with 30,000 records.

2.3 Objectives of Project

The importance of this study lies in the fact that early and accurate predictions can lead to early medical care. With the help of advanced deep learning techniques, we can uncover hidden patterns in breathing signals that might not be visible through traditional analysis.

Chapter 3

Literature Review

3.1 Summaries of Research Work

To build a deep learning model that can detect and predict COPD using spirogram data, it is important to understand the existing research in this area. In recent years, many researchers have used deep learning to identify COPD by analyzing spirogram or lung-related data. One of the most common algorithms applied in these studies is the Convolutional Neural Network (CNN) [3]. CNNs are powerful in extracting patterns from raw data.

3.2 Comparison

Another effective technique includes the use of ResNet18, a residual CNN architecture, which has shown high performance in learning from spirogram curves [6]. Table 3.1 compares key studies.

Table 3.1: Comparison of Prominent Research

Study	Method	Dataset
Avian et al. (2022)	CNN	e-nose signals
Cosentino et al. (2023)	ResNet18	UK Biobank

Chapter 4

Problem Statement

Predicting disease risk from breathing signals is not a simple task. The first challenge is that breathing data can be unstable and vary widely from person to person. The second issue is the difference in the length of breathing curves due to how long each person exhales. Third, most AI models act like black boxes, making it hard for doctors to trust the results.

Chapter 5

Methodology

5.1 Dataset

The process begins by feeding the synthetic dataset—containing 30,000 entries with spirogram and demographic data—into a preprocessing stage.

5.2 Exploratory Data Analysis

We refined noisy breathing curves using SpiroSmoother, which applies Gaussian filtering to the Time-Volume curves.

5.3 Algorithm Discussion

The architecture of our proposed model, DlSpiro, consists of four main modules:

- SpiroSmoother: Stabilizes the Time-Volume curve.
- SpiroEncoder: Extracts variable-length key patches.
- SpiroExplainer: Enhances interpretability.
- SpiroPredictor: Predicts the long-term COPD risk.

Figure 5.1: System Design Flowchart

Chapter 6

Experimental/Simulation/Analytical Analysis and Discussion of Results

6.1 Results

In this section, we evaluate the performance of DlSpiro along with three other models. DlSpiro achieved an AUROC of 0.8328, AUPRC of 0.3570, and F1-score of 0.3950 on the full dataset.

Table 6.1: Evaluation Comparison

Model	AUROC	AUPRC	F1-score
FEV1/FVC	0.7000	0.2500	0.3000
DlSpiro	0.8328	0.3570	0.3950

Figure 6.1: Evaluation Comparison

Chapter 7

Conclusions

7.1 Conclusion

In this project, we developed DlSpiro to detect and predict COPD using spirogram time series and patient demographic data. The results demonstrate that DlSpiro is a promising tool for early screening.

7.2 Future Scope

The model can be validated on real clinical data and extended to classify other respiratory conditions.

Bibliography

[1] Agustí, A. et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Am. J. Respir. Crit. Care Med.*, 207, 819–837, 2023.

[2] Ferrera, M. C., Labaki, W. W. & Han, M. K. Advances in chronic obstructive pulmonary disease. *Annu. Rev. Med.*, 72, 119–134, 2021.

[3] Avian, C. et al. Fxnet and purenet: Convolutional neural network architecture for discrimination of chronic obstructive pulmonary disease. *Comput. Biol. Med.*, 148, 105913, 2022.

[4] Dianat, B. et al. Classification of pulmonary sounds through deep learning. *Comput. Biol. Med.*, 160, 106928, 2023.

[5] Almeida, S. D. et al. Prediction of disease severity in COPD. *Eur. Radiol.*, 34, 4379–4392, 2024.

[6] Cosentino, J. et al. Inference of chronic obstructive pulmonary disease with deep learning on raw spirograms. *Nat. Genet.*, 55, 787–795, 2023.

[7] Sun, Y. et al. COVID-19 diagnosis based on swin transformer model. *Expert Syst. Appl.*, 243, 122805, 2024.

[8] Wei, X. et al. Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Syst. Appl.*, 219, 119648, 2023.

[9] Raihan, M. J. et al. Detection of the chronic kidney disease using XGBoost classifier. *Sci. Rep.*, 13, 6263, 2023.

[10] Yin, C. et al. Fractional dynamics foster deep learning of COPD stage prediction. *Adv. Sci.*, 10, 2203485, 2023.

[11] Kumar, S. et al. A novel multimodal framework for early diagnosis and classification of COPD. *Comput. Methods Prog. Biomed.*, 243, 107911, 2024.

Plagiarism Check Report

A plagiarism check has been conducted, and the similarity index is less than 10%.