International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-13 Issue-4, November 2024

A Comprehensive Approach to Predict Chronic Impairment of the Pulmonary System Through the Application of Artificial Neural Network Algorithm

Adisree. R, Mohamed Javed Khan A.

Abstract: COPD is a respiratory condition with airflow restriction and increased inflammation in the air passages. It is the main reason for sickness and death around the world, where it requires sophisticated diagnostic instruments. This research examines how Artificial Neural Networks (ANN) can be used to predict COPD. The clinical dataset has been trained and validated; ANN achieved over 93.75% accuracy. Our findings show that the ANN model is effective in aiding early COPD detection, which could enhance clinical decision-making and patient results.

Keywords: Chronic Obstructive Pulmonary Disease, Artificial Neural Networks, Forced expiratory volume, Forced vital capacity.

I. INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is a global health concern that is been characterized by ongoing inhaling symptoms and air passage restriction due to airway and/or alveolar dysfunction. These issues are usually caused by considerable disclosure of harmful substances. COPD includes chronic bronchitis and emphysema, which result in not broadening of the small airways and in return damage the lung tissue, respectively.

By the World Health Organization (WHO), it has been the third leading cause of disease worldwide. The burden of COPD is high in less-income countries. Additionally, ageing populations and the rising prevalence of smoking in developing regions contribute to the increasing incidence of COPD. In India, COPD poses a substantial public health challenge. Since 2000, there has been a marked increase in the incidence and prevalence of COPD, making it a critical concern. Several factors contribute to COPD in India:

1. Tobacco Smoking: The primary risk factor for COPD is tobacco smoking, which is highly prevalent in India. Both cigarette and bidi smoking contribute to the disease.

Manuscript received on 12 October 2024 | Revised Manuscript received on 23 October 2024 | Manuscript Accepted on 15 November 2024 | Manuscript published on 30 November 2024.
*Correspondence Author(s)
Adisree. R.*, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai (Tamil Nadu), India. Email ID: ar5751@srmist.edu.in, ORCID ID: 0009-0001-7048-4456
Mohamed Javed Khan A., Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai (Tamil Nadu), India. Email ID: mk0258@srmist.edu.in, ORCID ID: 0009-0001-0012-9598

2. Indoor Air Pollution: A proportion of the Indian population relies on biomass fuels (wood, crop residues, dung, and coal) for cooking and heating. The resultant indoor air pollution is a factor for COPD, especially among women and children.

3. Outdoor Air Pollution: Industrialization and urbanization have led to a high level of air pollution in many parts of India. Particulate matter and other pollutants intensify inhaling conditions.

4. Occupational Exposures: People working in industries such as mining, construction, and agriculture are frequently exposed to dust, fumes, and chemicals, which increase the risk of COPD.

5. Infections: Respiratory infections in childhood are found to be a greater risk of developing COPD in the future. In India, the high occurrence of such infections further intensifies the COPD.

Addressing COPD in India requires a versatile approach, including decreasing exposure to pollution, improving early diagnosis and management, and promulgating public awareness about the disease. Understanding the causes and developing effective predictive models are essential steps in fighting the rise of COPD in India and other countries.

II. METHODOLOGY

A. Data Collection and Preprocessing

i. *Data Sources***:**

The clinical dataset has been taken from the Kaggle website for this COPD Prediction. The dataset includes demographic information (such as age, and gender), detailed smoking history (including smoking status, duration, and pack-years), respiratory symptoms (like chronic cough and sputum production), and crucial pulmonary function test results (such as FEV1 and FVC measurements).

ii. *Data Cleaning:*

Data Preprocessing steps have been initiated to ensure the data is cleaned. Methods like mean and median imputation are applied to replace lost data points. Outliers and anomalies are resolved to prevent distorting the model training process.

iii. *Feature Engineering:*

This process creates new features. Important derived features include the FEV1/FVC ratio, BMI (Body

Mass Index), and categorical variables that represent different smoking statuses (e.g., non-smoker, current smoker, former smoker).

III. DATA SPLITTING

After preprocessing, the datasets are split into two parts. One is a training set, used for building the models, and another one is a testing set, for evaluating their performance. The split of data allocated to training is 80% and for testing, it is 20%. Stratified sampling ensures that both sets maintain the distribution of COPD-positive and COPD-negative cases proportionate to the original dataset.

A. Model Selection and Training

i. Artificial Neural Network (ANN):

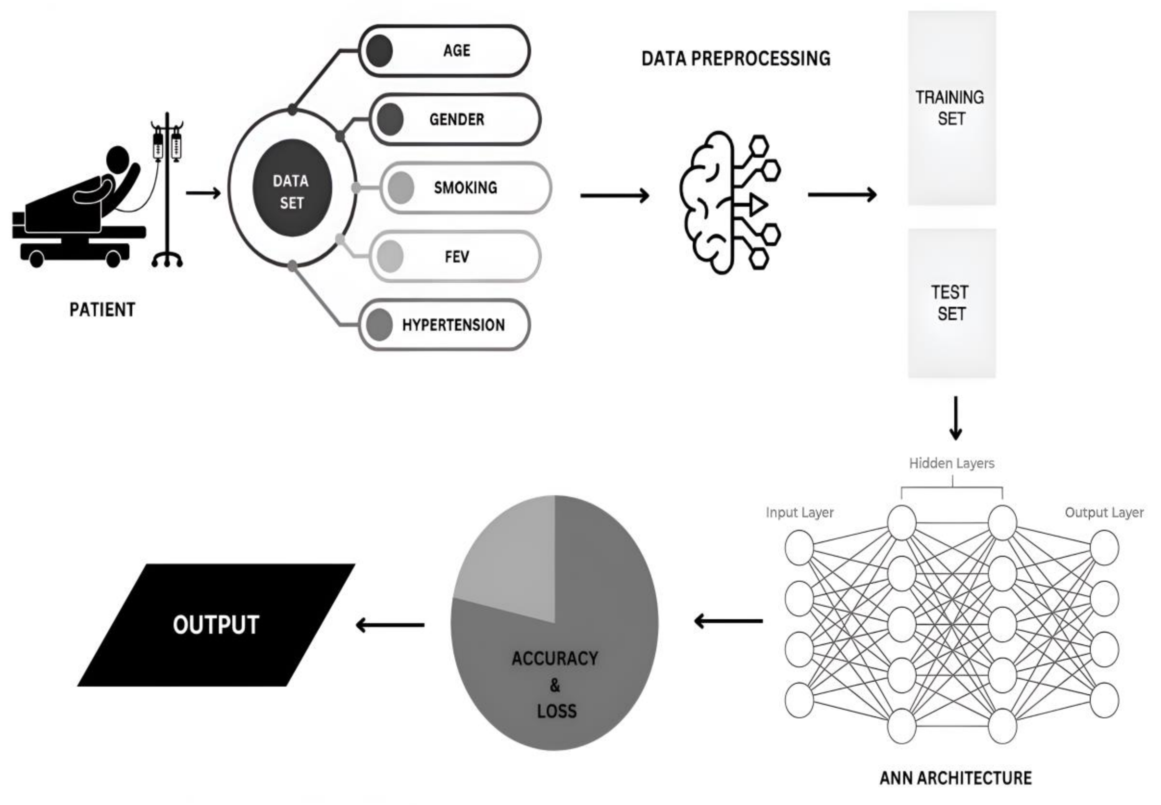
The ANN is selected for its capability to capture intricate data patterns through layers of interconnected neurons [1], resembling the human brain's neural network.

Training Process: Training of this model involved the architecture (number of layers, neurons per layer) and optimizing weights using backpropagation and optimization algorithms [2]. The model repeatedly improved the parameters to reduce prediction errors in the training data [3].

B. Model Evaluation

Performance Metrics**:** The ANN model is evaluated in testing and measures the ability of the model to generalize to unseen data and correctly classify COPD cases [4].

IV. ARCHITECTURE DIAGRAM



[Fig.1: Architecture Diagram]

The diagram illustrates the predictive model for COPD [5]. This utilizes the ANN model, Patient data, including age, gender, smoking status, Forced Expiratory Volume (FEV), hypertension, etc., is fed to each model. The dataset is trained and evaluated after which it is been given to the algorithm and that produces the accuracy [6].

V. IMPLEMENTATION

The data was preprocessed to remove irrelevant columns and address any inconsistencies. This data is for the training set. The remaining dataset was used for testing and validation.

Data Preprocessing: This dataset contains multiple features, including age, pack history, lung function metrics (FEV1, FVC), health scores (CAT, HAD, SGRQ), and comorbidities (diabetes, hypertension). The dataset was first cleaned by dropping off unwanted columns like 'Unnamed: 0', 'ID', 'COPDSEVERITY', 'MWT1', and 'MWT2' to ensure more accurate results [7].

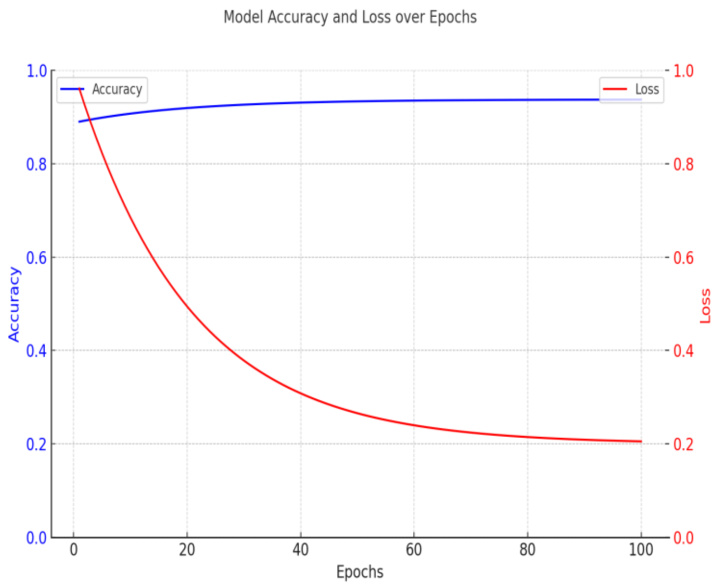
Artificial Neural Networks (ANN): The architecture of ANN has several layers such as the Input layer, a hidden layer, and an output layer to predict the four severity levels of COPD. This model has given an accuracy of 93.75%. This made ANN suitable for early detection and personalized treatment of COPD [8]. Table & Graphical Analysis of ANN Model is given below.

Retrieval Number: 100.1/ijrte.D817013041124
DOI: 10.35940/ijrte.D8170.13041124
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 1: Values of Accuracy and Loss

Epochs	Accuracy	Loss
0	0.05	1.4
10	0.45	1.0
20	0.55	0.8
30	0.65	0.7
40	0.75	0.5
50	0.80	0.4
60	0.85	0.35
70	0.88	0.3
80	0.90	0.25
90	0.92	0.22
100	0.93	0.20



[Fig.2: Graph of Accuracy and Loss over 100 Epochs]

VI. RESULTS

The study explains that the ANN model for COPD prediction, reveals nuanced performance characteristics. This model cannot be used for accurate prediction of COPD. While examining and analyzing the ANN model gives a competitive accuracy, where the potential to capture complex data patterns with the highest accuracy of 93.75%. These insights suggest complementary roles for the ANN model in enhancing COPD diagnostic accuracy.

VII. FUTURE ENHANCEMENTS

Future research is proposed to enhance the COPD prediction system's capabilities. This includes working on a larger dataset and implementing these algorithms on the dataset to find whether the accuracy predicted now matches with the larger dataset. This helps in integrating additional clinical variables. Also, in the future, we try to identify devices for more instant identification of COPD.

VIII. CONCLUSION

In conclusion, this paper has shown the efficacy of employing the ANN model as a predictive modeling technique for Chronic Obstructive Pulmonary Disease (COPD). By using demographic, clinical, and spiro metric data, the developed COPD prediction system exhibits good performance in early detection and risk assessment. The application of the ANN model enables accurate identification of developing COPD, thereby facilitating personalized treatment strategies. In summary, the adoption of the ANN model for COPD prediction exemplifies a paradigm shift toward healthcare solutions.

Retrieval Number: 100.1/ijrte.D817013041124
DOI: 10.35940/ijrte.D8170.13041124
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it has been conducted without any external sway.
- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. X. S. Roy, T. Issac, and J. S. Terrance, "Optimizing Travel Itineraries: A Multi-Objective Machine Learning Model Approach," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 421–425, 2024, DOI: https://doi.org/10.1109/ICAAIC60222.2024.10575378
2. R. Zhao, "The Water Potability Prediction Based on Active Support Vector Machine and Artificial Neural Network," 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR), pp. 110–114, 2021, DOI: https://doi.org/10.1109/ICBAR55169.2021.00032
3. M. Soykan and P. S. Bölük, "Tor Network Detection by Using Machine Learning and Artificial Neural Network," 2021 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6, 2021, DOI: https://doi.org/10.1109/ISNCC52172.2021.9615730
4. Chari, K. K., Babu, M. C., & Kodati, S. (2019). Classification of Diabetes using Random Forest with Feature Selection Algorithm. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 1, pp. 1295–1300). doi. https://doi.org/10.35940/ijitee.l3595.119119
5. T., G., M., V. Y., M., U., D., R., & K., R. B. (2020). Prediction of Lung Cancer Risk using Random Forest Algorithm Based on Kaggle Data Set. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 6, pp. 1623–1630). doi. https://doi.org/10.35940/ijrte.f7879.038620
6. Sistla, S. (2022). Predicting Diabetes u sing SVM Implemented by Machine Learning. In International Journal of Soft Computing and Engineering (Vol. 12, Issue 2, pp. 16–18). doi. https://doi.org/10.35940/ijsce.b3557.0512222
7. Nagar, K., & Chawla, M. P. S. (2023). A Survey on Various Approaches for Support Vector Machine Based Engineering Applications. In International Journal of Emerging Science and Engineering (Vol. 11, Issue 11, pp. 6–11). doi. https://doi.org/10.35940/ijese.k2555.10111123
8. Kumar, C. S., & Thangaraju, P. (2019). Improving Classifier Accuracy for diagnosing Chronic Kidney Disease Using Support Vector Machines. In International Journal of Engineering and Advanced Technology (Vol. 8, Issue 6, pp. 3697–3706). doi. https://doi.org/10.35940/ijeat.f9377.088619

AUTHORS PROFILE

Adisree R., is a final year student pursuing her Bachelor of Technology in Computer Science Engineering, specializing in Artificial Intelligence and Machine Learning at SRM Institute of Science and Technology. She boasts a solid educational background and strong programming skills, along with a wide-ranging knowledge of technology, which allows her to quickly grasp new concepts and excel in her academics. In addition to that, she possesses excellent communication skills, enabling her to convey complex ideas clearly. Her interests are in the field of Machine Learning, Deep Learning, and Artificial Intelligence. Her strengths are being friendly and adaptable, by thriving in new environments.

Mohamed Javed Khan A., is a final year student, pursuing his education in the field of Computer Science Engineering specializing in Artificial Intelligence and Machine Learning at SRM Institute of Science and Technology, Chennai. His reputation is based on his affable personality and willingness to embrace different viewpoints, with expertise ranging over multiple areas. He engages himself in student clubs and tech meetups, that encourage the growth of ideas. He is highly interested in the fields of User Interface/ User Experience (UI/UX) design, Deep Learning, Machine learning, and Artificial Intelligence. His strengths include being punctual, friendly, and empathetic in nature.