International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Predicting COPD Risk Using Machine Learning Algorithms

Kirupa p[1], Sri Nithiskumar N[2],Jawahar Srinath MN[3], Rethesh S[4], Saran Kumar K[5]

[1]*Assistant Professor & Department of Computer Science and Engineering*
[2]*Student & Department of Computer Science and Engineering*
[3]*Student & Department of Computer Science and Engineering*
[4]*Student & Department of Computer Science and Engineering*
[5]*Student & Department of Computer Science and Engineering*

Abstract - Chronic Obstructive Pulmonary Disease (COPD) is a progressive lung condition that affects quality of life and contributes to high global mortality. Early detection is vital but often hindered by complex, traditional diagnostic methods. This project presents a non-invasive COPD prediction system using sound analysis and machine learning. By examining respiratory sounds like coughs and wheezes from simple audio recordings, the system extracts key acoustic features to predict COPD risk. Developed as a user friendly web tool, it allows real-time predictions from uploaded recordings, making it a cost-effective and scalable solution especially in remote or low-resource areas. The model also offers interpretability, supporting informed clinical decisions and proactive care.

.

Key Words: Chronic Obstructive Pulmonary Disease (COPD), COPD Prediction System, Machine Learning, Healthcare Analytics, Data Classification, Predictive Modeling

1.INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is a long-term, progressive respiratory illness characterized by airflow limitation and breathing difficulties. It is a major public health concern, ranking among the leading causes of death worldwide. Early detection and timely intervention are critical to slowing disease progression, reducing complications, and improving patient outcomes. Traditional diagnostic methods for COPD, such as spirometry and imaging, require clinical equipment, trained personnel, and are often inaccessible in resource-limited settings. These limitations highlight the need for alternative, accessible, and non-invasive screening methods. Recent advancements in machine learning and acoustic signal processing have opened new possibilities in health diagnostics. Respiratory sounds, including coughs, wheezes, and breathing patterns, carry valuable information about pulmonary health. By analyzing these sounds using machine learning algorithms, it is possible to detect patterns associated with COPD. This project aims to develop a sound-based COPD prediction system that leverages audio recordings and machine learning to provide accurate, real-time risk assessments. The system is implemented as a web-based application, enabling users and healthcare professionals to easily upload recordings and receive immediate, interpretable predictions. This approach offers a scalable, cost-effective solution for early COPD screening, especially in underserved or remote areas.

2. RELATED SYSTEM

1.Machine Learning Applications in Chronic Disease Prediction:

Recent research explores how machine learning algorithms contribute significantly to predicting chronic diseases such as COPD. Studies demonstrate the application of classifiers like Support Vector Machines (SVM), Decision Trees, and Random Forests in identifying potential COPD cases using clinical and lifestyle data. These models improve diagnostic accuracy and assist in early intervention planning. Proper data preprocessing and feature engineering are emphasized for optimal model performance [1].

2. Electronic Health Records (EHR) and Predictive Analytics in Respiratory Diseases:

The integration of Electronic Health Records with predictive analytics enables early detection and monitoring of respiratory diseases like COPD. Research shows how structured EHR data, when analyzed through statistical and AI tools, can help in forecasting disease progression and supporting personalized treatment planning [2].

3. Clinical Decision Support Systems (CDSS) in Pulmonary Care:

Clinical Decision Support Systems play a crucial role in respiratory care. These systems, powered by rule-based logic or AI, provide healthcare professionals with recommendations based on patient data. In the case of COPD, CDSS helps assess risk factors, monitor exacerbations, and improve treatment efficiency [3].

4 Early Detection Using Supervised Deep Belief Networks (DBNs):

Supervised Deep Belief Networks (DBNs) are advanced machine learning models that combine deep learning with supervised learning techniques to enhance prediction accuracy. In COPD detection, DBNs analyze layered patterns in complex clinical data to identify early disease indicators. They enable robust feature extraction, improving classification performance for high-risk patients. This approach supports timely intervention and better management of COPD progression. [4].

3. PROPOSED APPROACH

The proposed solution for addressing the inefficiencies and limitations in Chronic Obstructive Pulmonary Disease (COPD) management is a comprehensive, technology-driven system that leverages Supervised Deep Belief Networks (DBNs) and a robust technology stack including HTML, CSS, JavaScript, Python, Django, and MySQL . This system integrates advanced predictive analytics, real-time monitoring, user-friendly interfaces, and centralized data management to revolutionize how COPD is diagnosed, treated, and prevented.

At the heart of the system lies Supervised Deep Belief Networks (DBNs) , a powerful deep learning model chosen for its ability to analyze complex datasets with high accuracy and efficiency. DBNs are particularly well-suited for large-scale applications like disease prediction due to their computational scalability and robustness against challenges such as vanishing gradients. The system uses DBNs to process diverse inputs, such as sound files uploaded by users. By analyzing these inputs, the model can classify cases into "COPD" and "Healthy" categories and predict the likelihood of acute exacerbations. For instance, subtle changes in breath sounds or cough patterns captured in sound files can be identified as early indicators of respiratory abnormalities. DBNs enable the identification of COPD at an early stage, even before symptoms become severe, allowing for timely interventions. The system continuously processes real-time data (e.g., sound files, wearable device inputs) to detect worsening symptoms or predict exacerbations, sending immediate alerts to users.

To ensure accessibility and ease of use, the front-end of the system is developed using HTML, CSS, and JavaScript . These technologies create anintuitive and visually appealing interface that caters to users of all technical backgrounds. The back-end of the system is powered by Python and the Django framework, which provide a scalable and secure foundation for building the application. Django's modular architecture allows for seamless integration of machine learning models like DBNs, ensuring efficient processing of large datasets. The back-end hosts the DBN model, which processes user inputs and generates predictions.

The system uses MySQL as its database management system to store and retrieve user data, health records, and predictive insights efficiently. MySQL's reliability and scalability make it an ideal choice for handling the vast amounts of structured data generated by the system.

The proposed solution is a holistic, technology-driven system that leverages Supervised Deep Belief Networks (DBNs) and a robust technology stack to address the inefficiencies in COPD management. By integrating advanced predictive analytics, real-time monitoring, user-friendly interfaces, and centralized data management, the system democratizes healthcare, making advanced COPD prediction and management accessible to all. This innovative approach not only improves patient outcomes but also alleviates the socioeconomic burden of the disease, paving the way for transformative advancements in respiratory health.

4. ARCHITECTURAL DESIGN

Architectural diagram:

5. ANALYTICAL METHODS

Research Design:

This project follows a hybrid research design, integrating machine learning techniques with healthcare analytics to develop a predictive model for Chronic Obstructive Pulmonary Disease (COPD). The system uses a supervised learning framework to analyze clinical and lifestyle-related datasets. It aims to predict the risk of COPD in individuals based on features such as age, smoking habits, lung function metrics, and environmental exposure. The design focuses on building a user-friendly interface integrated with a robust backend for accurate and early COPD risk prediction.

Data Collection and Preprocessing:

The dataset used in this project is derived from public healthcare repositories and medical research datasets, containing relevant features such as patient demographics, smoking history, respiratory symptoms, and spirometry results. Data preprocessing includes missing value handling, normalization, feature scaling, and categorical variable encoding. Outlier detection and noise reduction techniques are applied to ensure clean input for model training, enhancing the accuracy and consistency of the predictive system.

Implementation of AI-Enhanced Features:

Advanced machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and Deep Belief Networks (DBNs) are employed to extract patterns and classify COPD risk levels. Feature selection techniques like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are used to optimize model input. AI models are trained using supervised learning techniques, with continuous tuning of hyperparameters to improve predictive performance and generalization capability.

Application of Secure Data Handling:

To ensure patient data privacy, the system implements secure data encryption techniques for all inputs and outputs. Access control protocols and user authentication mechanisms are applied to safeguard sensitive information. The platform is developed with compliance to data privacy standards such as HIPAA, ensuring ethical handling of healthcare data and maintaining trust in the system.

6. LEARNING IN PRIVACY PRESERVATION

Collaborative Model Training:

The proposed COPD Prediction System incorporates collaborative model training using federated learning to improve prediction accuracy while safeguarding patient data privacy. Each local model is trained on patient datasets within secure environments (e.g., hospitals or clinics), and only encrypted model updates are sent to a central server. This approach ensures that sensitive patient data, such as medical history and lifestyle factors, never leave the local system, promoting secure and privacy-aware AI in healthcare diagnostics.

Decentralized Data Management:

The system adopts a decentralized data management architecture where patient records remain stored within the healthcare provider's local infrastructure. By avoiding centralized data repositories, the system minimizes the risk of mass data breaches and ensures compliance with data protection regulations like HIPAA. This setup enables patients and healthcare institutions to maintain control over medical data while

still participating in the global improvement of AI-driven COPD detection algorithms.

Benefits and Challenges:

This privacy-preserving learning model ensures that individual health information is never directly exposed, enhancing trust in AI adoption for clinical diagnostics. It allows for model performance enhancement across institutions without compromising data confidentiality. However, challenges such as synchronization of model updates, managing heterogeneous datasets, and increased computational demands exist. Techniques like differential privacy, secure multiparty computation, and homomorphic encryption are integrated to address these concerns while maintaining model efficiency and accuracy.

7. METHODOLOGY

Requirement Gathering and Analysis:

The initial phase of the project involves comprehensive research and analysis to understand the requirements of both end-users and healthcare providers. The system is designed to collect patient data, including symptoms, smoking history, age, gender, environmental exposure, and spirometry test results. The goal is to develop a predictive model that can detect COPD risk levels accurately using machine learning techniques. This includes identifying the right datasets, selecting key features for prediction, and determining the necessary functionalities for user interaction and result visualization.

Data Collection and Preprocessing:

Relevant COPD-related datasets are collected from medical sources or clinical records. Data preprocessing involves handling missing values, normalizing data, and encoding categorical variables to prepare it for model training. Feature selection techniques are used to identify the most influential factors contributing to COPD diagnosis, ensuring the model focuses on meaningful inputs.

Model Development:

Supervised machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, or Deep Belief Networks (DBNs) are employed to build the predictive model. The model is trained and validated using a split dataset to ensure generalizability and accuracy. Performance metrics like accuracy, precision, recall, and F1-score are used to evaluate the effectiveness of the model in predicting COPD.

8. RESULT AND DISCUSSION

Analysis of System Efficiency:

The implementation of the COPD Prediction System has demonstrated considerable efficiency in early identification of Chronic Obstructive Pulmonary Disease through machine learning-based prediction models. The system successfully integrates data preprocessing, supervised learning algorithms, and a user-friendly interface to analyze patient inputs and predict COPD risk levels with high accuracy. By leveraging structured clinical datasets and advanced classification algorithms, the system enables healthcare practitioners and patients to assess the likelihood of COPD quickly and effectively, facilitating early diagnosis and timely intervention.

Impact on Accuracy and Performance:

The model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. Results indicated that algorithms like Random Forest and Deep Belief Networks (DBNs) achieved high classification accuracy in detecting COPD risk. The use of feature selection techniques significantly improved the model's prediction reliability by reducing noise and focusing on clinically relevant parameters such as age, smoking history, and respiratory symptoms. Preprocessing techniques like normalization and outlier removal also contributed to enhanced model performance.

Comparative Performance with Traditional Diagnostic Methods:

Compared to conventional COPD diagnostic methods like spirometry or manual clinical evaluation, the proposed system offers faster and more accessible risk prediction. While traditional methods require physical testing and expert interpretation, the COPD Prediction System provides preliminary risk analysis based on input data, reducing diagnostic delay and aiding proactive healthcare decisions. Additionally, the system's portability and digital nature make it suitable for remote health monitoring and telemedicine applications.Framework Evaluation

Performance Metrics:

The effectiveness of the Real-Time Language Translator Bot is evaluated using key performance indicators, including translation accuracy, speech recognition efficiency, response time, and user experience. Translation accuracy measures how well the system converts spoken language into the target language while maintaining contextual meaning. Speech recognition efficiency assesses the ability to accurately transcribe spoken words with minimal errors. Response time evaluates the system's capability to process and deliver translations in real time, ensuring minimal latency. Additionally, usability is assessed based on the intuitiveness of the user interface, ensuring smooth interaction for both technical and non-technical users.

Comparative Analysis with Existing Approaches:

Compared to traditional offline translation tools and manual interpretation, the proposed system offers real-time speech translation with improved accessibility and convenience. Unlike conventional translation applications that rely solely on preloaded language data, this system leverages cloud-based AI models, ensuring up-to-date translations with evolving linguistic patterns. While standard translation apps may struggle with speech-to-text conversion in noisy environments, the integrated SpeechRecognition module enhances accuracy through noise filtering and adaptive learning. Furthermore, traditional translators often lack real-time text-to-speech conversion, whereas the proposed solution seamlessly converts translations into natural-sounding speech using GTTS, improving accessibility for visually impaired users and language learners.

9. CONCLUSION

The COPD Prediction System successfully demonstrates the potential of machine learning in the early detection and prevention of Chronic Obstructive Pulmonary Disease. By integrating advanced algorithms and user-friendly interfaces, the system provides an efficient and accessible solution for identifying at-risk individuals based on clinical and behavioral data. The use of predictive analytics allows healthcare providers to make timely interventions, improving patient outcomes and reducing the burden on healthcare systems. Overall, the project highlights the value of AI in enhancing preventive healthcare, and it lays a strong foundation for further advancements in medical diagnostics and personalized treatment planning.

REFERENCES

1.G. T. Omlor et al., "Predicting COPD using machine learning techniques: A comprehensive review," Journal of Translational Medicine, vol. 18, no. 1, pp. 1–10, 2020.

2.A. K. Gupta and R. A. Raj, "A Machine Learning Approach to Predict Chronic Obstructive Pulmonary Disease," International Journal of Advanced Computer Science and Applications, vol. 11, no. 9, pp. 124–130, 2020.

3. A. A. Alshamrani et al., "Early prediction of chronic obstructive pulmonary disease using artificial neural networks," IEEE Access, vol. 8, pp. 74110–74120, 2020.

4. S. W. Park, J. S. Lee, and H. K. Lee, "Risk factor identification for COPD using machine learning algorithms," Healthcare Informatics Research, vol. 27, no. 2, pp. 125–132, 2021.

5. T. L. Davis, "Use of supervised learning algorithms in COPD diagnosis and severity prediction," Journal of Medical Systems, vol. 44, no. 3, pp. 1–9, 2020.

6. A. H. Bashir et al., "Application of deep learning models in respiratory disease prediction," BioMed Research International, vol. 2021, Article ID 9218742, 2021.

7. L. D. Ríos and A. L. Ríos, "Data preprocessing in medical diagnosis using machine learning," Procedia Computer Science, vol. 177, pp. 161–167, 2020.

8. N. Srivastava et al., "Chronic Disease Prediction using Supervised Learning: A COPD Case Study," International Journal of Engineering and Technology, vol. 9, no. 3, pp. 456–463, 2020.

9. M. A. Khan and S. Saba, "Data Mining in Healthcare for COPD Prediction using Decision Tree," International Journal of Computer Applications, vol. 175, no. 3, pp. 20–25, 2020.

10. D. R. Wilson and T. R. Martinez, "Improved data reduction techniques for k-nearest neighbor classification," Intelligent Data Analysis, vol. 4, no. 1, pp. 15–24, 2000.