

Chronic Obstructive Pulmonary Disease Diagnosis with Bagging Ensemble Learning and ANN Classifiers

Taskeena Siddiqui

Department of Computer Science and Information Technology, NED University of Engineering and Technology, Pakistan
siddiqui.pg3600982@cloud.neduet.edu.pk (corresponding author)

Mustafa Latif

Department of Software Engineering, NED University of Engineering and Technology, Pakistan
mustafalatif@neduet.edu.pk

Muhammad Umer Farooq

Department of Computer Science and Information Technology, NED University of Engineering and Technology, Pakistan
umer@neduet.edu.pk

Mirza Adnan Baig

Department of Computer Science, IQRA University, Pakistan
adnan.baig@iqra.edu.pk

Yusuf Sharif Hassan

Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Saudi Arabia
ysshassan@kfupm.edu.sa

Received: 21 February 2024 | Revised: 16 April 2024 | Accepted: 18 April 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7106>

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) is a persistent respiratory disease that poses a significant threat to global human health with elevated incidence and mortality rates. Timely recognition and diagnosis of COPD play a pivotal role in efficiently managing and treating the condition. The incorporation of deep learning technologies into healthcare has significant potential to enhance diagnostics and treatment outcomes. This study proposes an innovative deep-learning approach along with an ensemble technique to address the imperative need for an effective predictive model in COPD disease classification, particularly in situations with limited available data. This was achieved by leveraging the ensemble bagging technique and incorporating ANN as a classifier within this framework. Training and evaluation of the proposed ensemble ANN model were performed on a dataset comprising a variety of attributes, including demographic information, medical history, diagnostic measurements, and pollution exposures. Data were collected from people aged 18 to 60 originating from Pakistan, encompassing patients, attendants, hospital staff, faculty, and students. The effectiveness of the model in classifying COPD was measured using F1 score, recall, precision, and accuracy. The evaluation of the model produced notable results, as it achieved a 90% F1 score, 96% recall, 84% precision, and 89% accuracy in identifying the presence of COPD in individuals. Furthermore, this study carried out a comparative analysis between a standalone ANN model and the proposed ensemble ANN model which revealed that the proposed Ensemble ANN model outperforms existing methods, particularly in scenarios with limited sample size. This research provides substantial contributions to healthcare technology, as it presents an efficient tool for COPD prediction, facilitates early intervention, and significantly increases the overall standard of patient care.

Keywords-deep learning; ensemble learning; bagging; artificial neural networks; COPD; chronic diseases

I. INTRODUCTION

The five main respiratory diseases, Chronic Obstructive Pulmonary Disease (COPD), pneumonia, asthma, tuberculosis, and lung cancer, collectively contribute to approximately 20% of the global deaths [1]. Asthma and COPD stand out as the predominant and significant chronic respiratory diseases on a global scale [2-3]. COPD poses a considerable public health challenge, reducing the quality of life of those it affects [4]. Across the world, 251 million people suffer from COPD [5]. COPD is one of the primary contributors to worldwide mortality rates [6-7], accounting for approximately 90% of deaths in low- and middle-income countries [8], and it is expected that 4.5 million people worldwide will have been affected by COPD and its related conditions by 2030 [9]. People with COPD have an increased susceptibility to heart disease, lung cancer, and a variety of other health problems [10]. In Pakistan, 4.3% of the population is estimated to suffer from asthma, while 2.1% are affected by COPD [11]. These respiratory diseases are considered major health problems in the country and represent a significant proportion of patients seeking treatment in primary healthcare facilities, accounting for one-fourth of the total cases [12]. COPD is characterized by a gradual reduction in airflow resulting from persistent inflammation in the bronchial tubes, bronchioles, and lung tissue. Breathlessness, wheezing, chest discomfort, and persistent cough that produces phlegm are common signs of COPD [13]. COPD is predominantly linked to risk elements, such as tobacco use, exposure to indoor and outdoor air pollutants, and occupational hazards [14].

Machine learning and deep learning have made notable advances in the healthcare industry. In [15], Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), and Support Vector Machines (SVM) were used on a dataset containing 400 lung disease images to classify lung diseases and evaluate their performance. In [16], deep ensemble learning was employed to detect Alzheimer's disease. In [17], in-vitro classification of saliva samples was carried out on both COPD and healthy patients following a variety of machine learning techniques, including Artificial Neural Networks (ANNs), Logistic-Regression (LR), Gaussian-Naïve-Bayes (GNB), Support-Vector-Machines (SVM), and XGBoost. In [18], machine learning algorithms, specifically SVM and KNN, were applied for COPD classification. In [19], an exploratory investigation was conducted using an ANN to predict self-care behaviors in people with COPD. Symptom data were collected from patients for 21 consecutive days, covering various symptoms, entailing cough-related distress, chest tightness, mucus-related distress, activity-induced dyspnea, resting dyspnea, and fatigue. In [20], MLP networks were utilized for pattern recognition in healthcare settings in five distinct areas, including Parkinson's disease, diabetes, liver conditions, breast cancer, and heart disease. This study evaluated the efficacy, limitations, and recent advancements in employing MLP networks for pattern recognition across various healthcare datasets. In [21], data from COPD patients between 2015 and 2018 were used along with multiple locally weighted linear regression models with K-means clustering to predict disease morbidity. In [22], EEG

was deployed for dementia with Lewy bodies diagnosis, achieving a 94.4% accuracy by using a feed-forward neural network and indicating a superior classification performance compared to existing methods. In [23], medical insurance data from a large urban area were engaged to build a Bayesian network model to predict future high-cost COPD patients. In [24], a chili pepper disease diagnosis model was developed implementing MLPNN, achieving 98.91% accuracy and suggesting its potential for disease prevention and treatment, aligning with previous studies showcasing MLP's effectiveness in disease diagnosis.

The use of deep learning algorithms for disease prediction with limited tabular datasets has not yet received comprehensive attention in research investigations. This underscores the importance of exploring approaches tailored to situations with restricted data availability. This study aimed to address this gap by employing a deep learning technique within the bagging framework as a classifier. The primary focus was to overcome the challenge presented by limited data availability and develop a predictive model capable of accurately identifying the risk and probability of COPD in individuals.

II. MATERIALS AND METHODOLOGY

A. Data Collection

Data were collected with the assistance of Dr. Tahira Zubair, Senior Registrar at the Bahria University Medical and Dental College, from various individuals, including patients, attendants, hospital staff, faculty, and students aged between 18 and 60 years old. Data were sourced from individuals residing in Pakistan and its respective areas, encompassing a diverse range of regions, cities, or specific locations and ensuring a comprehensive representation of the population under investigation. Table I presents a detailed summary of the attributes considered in the dataset.

TABLE I. DATASET ATTRIBUTES

Environmental exposures	Smoking exposures	Symptoms
Pesticides exposure	Smoking exposure	Wheezing
Pets exposure	Age of starting smoking	Recurrent chest infection
Biomass fuel exposure	Type of smoke	Weight loss in last 1 year
Aerosol sprays exposure	Cigarettes per day	Muscular weakness in the last year
Gas grill exposure	Duration of smoking (years)	FEV1/FVC score
Factory smoke exposure	Passive smoking	
Wooden stove exposure		
Vehicle exhaust exposure		

B. Data Processing

This study was organized into three primary sections. Initially, the data undergoes a preprocessing stage involving cleaning and transformation to ready it for analysis. Subsequently, the investigation delves into the intricacies of a newly proposed method, offering a comprehensive explanation of its functionality. Finally, the third section centers on the research objective, representing the goal of the study. Figure 1 illustrates how these three components—preprocessing, the proposed method, and the research objective—together constitute the fundamental framework of the research.

The initial dataset consisted of 517 observations. Including numerous variables allowed for a comprehensive examination of factors relevant to disease diagnosis. However, it should be noted that the dataset featured a significant number of missing cells with incomplete or unrecorded data. This challenge had the potential to affect the reliability of the findings. To mitigate this, appropriate techniques for handling missing data, such as imputation, were employed. Despite the presence of missing cells, it was essential to ensure the uniqueness of the data by confirming the absence of duplicate rows. Before the model was developed, categorical features were transformed implementing label encoding. Subsequently, normalization was applied to ensure consistency across all features.

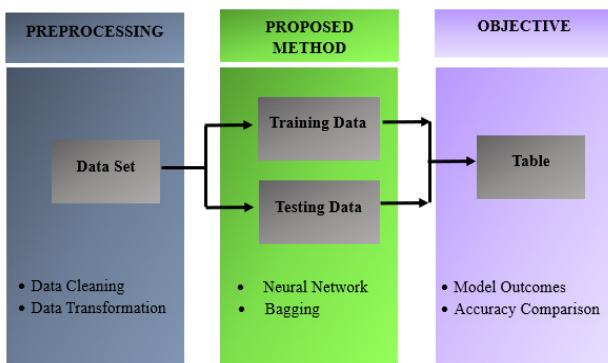


Fig. 1. The proposed framework.

III. MODEL DEVELOPMENT

This study employed a robust ensemble learning approach alongside an Artificial Neural Network (ANN) to develop predictive models for COPD given the constraints of a limited dataset. Figure 2 illustrates the block diagram of the proposed ensemble ANN model.

The dataset was partitioned into two main subsets for testing and training. The training set, comprising 80% of the data, was used to train the model, while the remaining 20% was reserved to evaluate its performance on unseen data. The training set was further divided into three portions to construct individual ANN models within the ensemble, all based on identical specifications. Figure 3 presents a visual representation of the structure of the ANN models. The partitioned training data set was further divided into training and testing sets, with 80% of the data assigned to training and 20% to testing the ANN model. Standardization of input features was performed deploying the standard scaler to ensure stable model training. The ANN architecture consisted of five layers, including an input layer with 68 neurons, followed by three hidden layers with 52, 36, and 28 neurons, respectively, each employing the Rectified Linear Unit (ReLU) activation function. The output layer consisted of a single neuron that utilized the sigmoid activation function for the binary classification of COPD. Model training utilized the Adam optimizer with a regularization parameter of 0.01 for L2 regularization. Training was carried out over 200 epochs with a batch size of 32, and a 10% validation split was applied to monitor model performance during training.

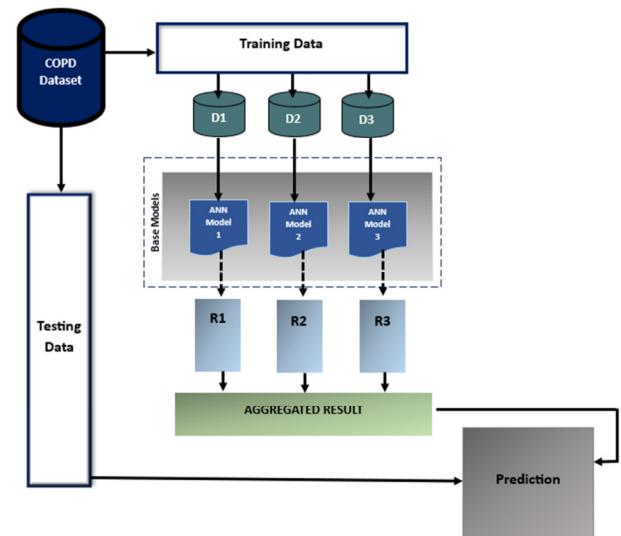


Fig. 2. The proposed ensemble ANN model.

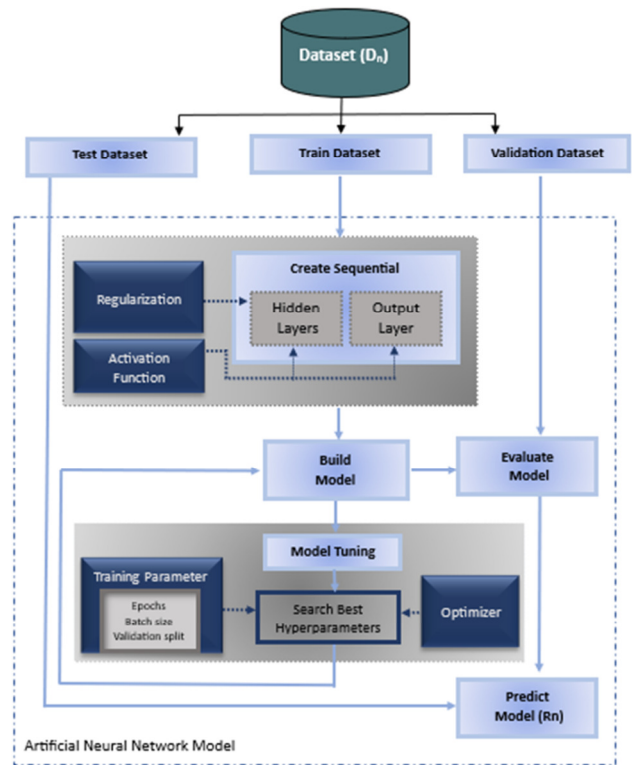


Fig. 3. The architecture of the ANN model.

Once the ANN models were trained and validated on their respective subsets, they were integrated into the ensemble framework. The bagging ensemble model aggregated the predictions of the base models to enhance the classification performance. This ensemble model, renowned for its adaptability to various architectures, was trained on the aggregated predictions of the base models. The evaluation of the ensemble's effectiveness included generating a confusion matrix and a detailed classification report, offering insights into

the model's performance on unseen test data. This comprehensive method aimed to leverage the collective strength of multiple ANNs, resulting in enhanced classification outcomes and robustness across diverse scenarios, even with a limited dataset. The methodology displayed in Figures 2 and 3 encompasses three primary steps outlined as follows.

A. Bootstrap Sampling

The split training dataset was further divided into multiple training samples.

$$D_b = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

where D represents the training dataset, x_n denotes input features, y_n denotes corresponding outputs, and $b = 1, 2, \dots, n$.

B. Training Base Learner

The ANN was used as a base classifier, denoted by L . Each base ANN learner h_b was trained on a specific bootstrap sample D_b , resulting in n base ANN learners.

$$h_b = L(D_b) \quad (2)$$

where $b=1, 2, 3, \dots, n$. An artificial neural network consists of multiple layers. Its structure typically encompasses an input layer where data are received, hidden layers where data are processed, and an output layer where the network produces results. The neurons in each layer process information through weighted connections and activation functions, allowing the network to learn complex patterns and make predictions. The general expression for the output Y_i of a neuron in an ANN is denoted as:

$$Y_i = f \sum_{j=1}^n w_{ij} \cdot x_j + b_i \quad (3)$$

where Y_i denotes the neuron output, f is the activation function applied to the weighted sum, $\sum_{j=1}^n w_{ij} \cdot x_j + b_i$ represents the weighted sum of inputs (x_j being the inputs and w_{ij} the weights), and b_i is the bias term. Designing an ANN for specific tasks involves critical considerations such as selecting an activation function (f) and determining the network structure, which encompasses configuring the number of neurons and hidden layers. The activation functions employed in this study were ReLU and sigmoid.

C. Aggregation

The final prediction was obtained by aggregating the predictions made by the individual base ANN learners.

$$Y_e = \operatorname{argmax} (\sum_{b=1}^B Z (h_b(x) = c)) \quad (4)$$

where c represents the class labels, Z is the indicator function and Y_e is the final classification prediction. Algorithm I elaborates on the logical sequence of actions involved in the model.

ALGORITHM I: THE PROPOSED MODEL

Input: training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Output: Ensemble of ANN models $\{h_1, h_2, \dots, h_n\}$

for $b = 1$ to n do
 # Generate a bootstrap sample D_b

```

D_b = randomly_sample_with_replacement(D)
# Train an ANN model on the bootstrap sample
h_b = TrainANN(D_b)
# Store the trained ANN model
ensemble.append(h_b)

# Training ANN Function - Trains an ANN model on a
# provided bootstrap sample
function TrainANN(bootstrap_sample):

  # Define ANN architecture and hyperparameters
  # Hidden layer sizes are determined based on
  # requirements
  ANN_model=ANNClassifier(hidden_layer_sizes =
  (As per the requirement), max_iter, activation =
  'ReLU', solver = Adam)
  # Extract features and labels from the bootstrap
  # sample
  X_train_bootstrap = bootstrap_sample.features
  Y_train_bootstrap = bootstrap_sample.labels

  # Train ANN on the bootstrap sample
  ANN_model.fit(X_train_bootstrap,
  Y_train_bootstrap)

  # Return the trained ANN model
  return ANN_model

# Aggregation
Input: Ensemble of ANN models {h_1, h_2, ..., h_n}, Test
data X_test

Output: Final classification prediction for the test
data Y_e

for each data point x in X_test do
  # Aggregate predictions of individual ANN models
  predictions = [h_b.predict(x) for h_b in ensemble]

  # Choose the most frequently predicted class
  # label
  Y_e(x) = majority_vote(predictions)

# Final classification predictions
Output: Y_e

```

IV. MODEL EVALUATION

Confusion matrix is a fundamental tool to evaluate the performance of classification models. This matrix provides a detailed analysis of a model's predictions, enabling a nuanced evaluation of its strengths and limitations. A confusion matrix has four essential parts when it comes to binary classification.

- TP (True Positive): Cases where the model accurately predicts the positive class.
- TN (True Negative): Cases where the model accurately predicts the negative class.
- FP (False Positive): Cases where the model incorrectly predicts the positive class.
- FN (False Negative): Cases where the model incorrectly predicts the negative class.

These parameters form the basis for calculating crucial performance metrics, such as Precision, F1-Score, Recall, and Accuracy. These metrics can be determined using the formulas outlined below.

$$\text{Precision (P)} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall (R)} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

$$\text{Accuracy} = \frac{\text{Sum of True Values (TP+TN)}}{\text{Sum of All Values (TP+TN+FP+FN)}} \tag{8}$$

V. RESULTS AND DISCUSSION

A. Results of the Ensemble ANN Model

This study employed an ensemble approach incorporating an ANN to develop a predictive model for COPD. An ANN was incorporated as the foundational classifier, and bagging served as the ensemble technique. A comprehensive array of metrics was deployed to evaluate the model's performance, designed specifically to address challenges arising from a limited dataset. The inherent constraints associated with a small dataset were mitigated through the application of bagging, while the efficacy of the model benefited from the integration of ANN. The model was trained and tested on the collected dataset, and Figure 4 demonstrates the results.

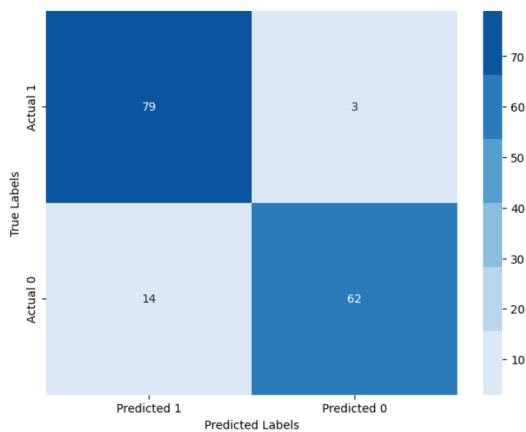


Fig. 4. Confusion matrix: results of the proposed ensemble ANN model.

TABLE II. EVALUATION METRICS OF ENSEMBLE ANN MODEL

Evaluation Metrics	Results
Precision	84%
Recall	96%
F1-Score	90%
Accuracy	89%

The confusion matrix provided a comprehensive breakdown of how well the model predicted COPD. For the cases where COPD is present, the matrix revealed that 79 instances were correctly identified as positive (TP), 14 instances were incorrectly classified as positive (FP), 62 instances were correctly identified as negative (TN), and 3 instances were incorrectly classified as negative (FN). These values offer a nuanced evaluation of the model's ability to correctly classify instances, providing valuable insights into its performance in the identification of COPD presence. Table II presents recall, precision, accuracy, and F1-Score to provide a

comprehensive insight into the model's effectiveness in predicting the presence of COPD.

B. Comparative Analysis between the ANN and the Proposed Model

A comparative analysis was carried out on a standalone ANN and the proposed model, with identical specifications, to evaluate their efficacy. However, the number of neurons was varied (64, 48, 36, 24, 10, 1) to investigate the model's performance across diverse scenarios.

TABLE III. COMPARISON OF THE PROPOSED ENSEMBLE ANN AND SIMPLE ANN MODELS

Metrics	Standalone ANN	Ensemble ANN model
Precision	72%	82%
Recall	98%	98%
F1-Score	83%	89%
Accuracy	79%	88%

The proposed ensemble ANN model demonstrates greater efficacy in disease diagnosis and accurate prediction compared to the standalone ANN. The results suggest that a viable approach to mitigate the challenges posed by limited clinical datasets involves integrating bagging techniques with neural networks. Figure 5 offers a visual representation of the comparative performance of the standalone ANN and the proposed ensemble ANN model on the COPD dataset.

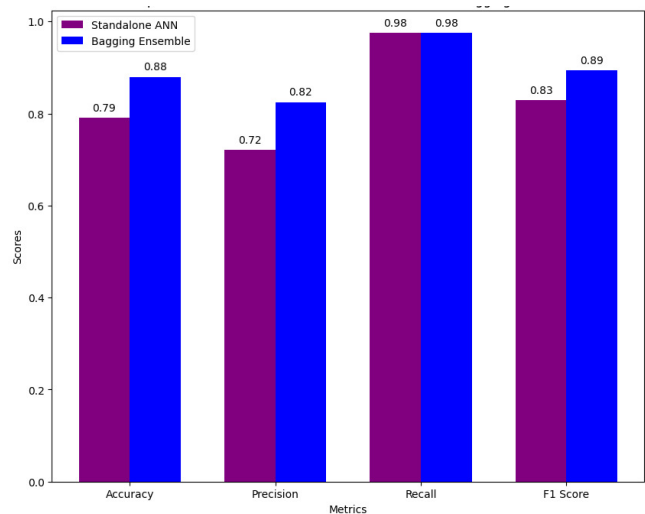


Fig. 5. Comparison of results between the standalone ANN and the proposed ensemble ANN model.

VI. CONCLUSION

This study aimed to address the issue of disease prediction with limited datasets, focusing particularly on COPD through the utilization of deep learning methods. This study combined the ensemble bagging method with an ANN, departing from the traditional practice of employing machine learning algorithms such as decision trees and random forests in bagging. The findings demonstrate the effectiveness of this approach, highlighting notably enhanced classification results with an accuracy rate of 89%. This study contributes to the

healthcare technology domain by showcasing the potential of ensemble techniques to enhance disease prediction accuracy, particularly when dealing with constrained datasets. By demonstrating the effectiveness of ensemble bagging with ANN, this study expands the current understanding of ensemble methods in clinical diagnosis and highlights their applicability in scenarios with limited data availability. Moreover, this study paves the way for future investigations into alternative ensemble techniques and their synergies with deep learning algorithms. By exploring these avenues, researchers can further refine disease detection methodologies and achieve even more advanced classification results.

ACKNOWLEDGMENT

The authors express their sincere gratitude to Dr. Tahira Zubair for sharing the dataset used in this study, as its availability allowed us to perform our analysis and come to insightful conclusions.

REFERENCES

- [1] G. J. Gibson, R. Lodenkemper, B. Lundbäck, and Y. Sibille, "Respiratory health and disease in Europe: the new European Lung White Book," *European Respiratory Journal*, vol. 42, no. 3, pp. 559–563, Sep. 2013, <https://doi.org/10.1183/09031936.00105513>.
- [2] A. N. Aggarwal, K. T. Prasad, and V. Muthu, "Obstructive lung diseases burden and COVID-19 in developing countries: a perspective," *Current Opinion in Pulmonary Medicine*, vol. 28, no. 2, pp. 84–92, Mar. 2022, <https://doi.org/10.1097/MCP.0000000000000836>.
- [3] S. A. A. Biabani and N. A. Tayyib, "A Review on the Use of Machine Learning Against the Covid-19 Pandemic," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8039–8044, Feb. 2022, <https://doi.org/10.48084/etasr.4628>.
- [4] S. Sharma, D. Salibi, and N. Tzenios, "Modern approaches of rehabilitation in COPD patients," *Special journal of the Medical Academy and other Life Sciences*, vol. 1, Jul. 2023, <https://doi.org/10.58676/sjmas.v1i16.39>.
- [5] Y. K. Wu, W. L. Su, M. C. Yang, S. Y. Chen, C. W. Wu, and C. C. Lan, "Associations Between Physical Activity, Smoking Status, and Airflow Obstruction and Self-Reported COPD: A Population-Based Study," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. Volume 17, pp. 1195–1204, May 2022, <https://doi.org/10.2147/COPD.S337683>.
- [6] D. Adeloje, P. Song, Y. Zhu, H. Campbell, A. Sheikh, and I. Rudan, "Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis," *The Lancet Respiratory Medicine*, vol. 10, no. 5, pp. 447–458, May 2022, [https://doi.org/10.1016/S2213-2600\(21\)00511-7](https://doi.org/10.1016/S2213-2600(21)00511-7).
- [7] GBD Chronic Respiratory Disease Collaborators *et al.*, "Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: A systematic analysis for the global burden of disease study 2017," *The Lancet Respiratory Medicine*, vol. 8, no. 6, pp. 585–596, Jan. 2020, [https://doi.org/10.1016/S2213-2600\(20\)30105-3](https://doi.org/10.1016/S2213-2600(20)30105-3).
- [8] J. O. Obira and R. Sinde, "Development of a Sensor-Based Heartbeat and Body Temperature Monitoring System for Remote Chronic Patients," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7375–7380, Aug. 2021, <https://doi.org/10.48084/etasr.4216>.
- [9] A. Rehman, H. Shafiq, S. Jawed, and F. Behram, "Chronic Obstructive Pulmonary Disease (COPD) Screening is Still a Challenge in Pakistan: COPD in Pakistan," *Journal of Aziz Fatimah Medical & Dental College*, vol. 1, no. 1, pp. 18–23, Apr. 2019, <https://doi.org/10.55279/jafimdc.v1i1.27>.
- [10] K. Kousalya, K. Dinesh, B. Krishnakumar, J. G. Kavyapriya, C. Kowsika, and K. Ponmathi, "Prediction of Optimal Algorithm For Diagnosis of Chronic Obstructive Pulmonary Disease," in *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichirappalli, India, Apr. 2023, pp. 1–8, <https://doi.org/10.1109/ICEEICT56924.2023.10156915>.
- [11] M. A. Khan *et al.*, "Feasibility of delivering integrated COPD-asthma care at primary and secondary level public healthcare facilities in Pakistan: a process evaluation," *BJGP Open*, vol. 3, no. 1, Apr. 2019, <https://doi.org/10.3399/bjgpopen18X101632>.
- [12] M. A. Khan, M. Ahmed, S. Anil, and J. Walley, "Strengthening the delivery of asthma and chronic obstructive pulmonary disease care at primary health-care facilities: study design of a cluster randomized controlled trial in Pakistan," *Global Health Action*, vol. 8, no. 1, Dec. 2015, Art. no. 28225, <https://doi.org/10.3402/gha.v8.28225>.
- [13] Q. Wang, H. Wang, L. Wang, and F. Yu, "Diagnosis of Chronic Obstructive Pulmonary Disease Based on Transfer Learning," *IEEE Access*, vol. 8, pp. 47370–47383, Jan. 2020, <https://doi.org/10.1109/ACCESS.2020.2979218>.
- [14] D. Adeloje, P. Song, Y. Zhu, H. Campbell, A. Sheikh, and I. Rudan, "Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis," *The Lancet Respiratory Medicine*, vol. 10, no. 5, pp. 447–458, May 2022, [https://doi.org/10.1016/S2213-2600\(21\)00511-7](https://doi.org/10.1016/S2213-2600(21)00511-7).
- [15] B. M. Boban and R. K. Megalingam, "Lung Diseases Classification based on Machine Learning Algorithms and Performance Evaluation," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, Jul. 2020, pp. 0315–0320, <https://doi.org/10.1109/ICCSP48568.2020.9182324>.
- [16] N. An, H. Ding, J. Yang, R. Au, and T. F. A. Ang, "Deep ensemble learning for Alzheimer's disease classification," *Journal of Biomedical Informatics*, vol. 105, May 2020, Art. no. 103411, <https://doi.org/10.1016/j.jbi.2020.103411>.
- [17] P. S. Zarrin, N. Roekendorf, and C. Wenger, "In-Vitro Classification of Saliva Samples of COPD Patients and Healthy Controls Using Machine Learning Tools," *IEEE Access*, vol. 8, pp. 168053–168060, 2020, <https://doi.org/10.1109/ACCESS.2020.3023971>.
- [18] S. Vora and C. Shah, "COPD Classification using Machine Learning Algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 06, no. 04, 2019.
- [19] A. Bugajski, A. Lengerich, R. Koerner, and L. Szalacha, "Utilizing an Artificial Neural Network to Predict Self-Management in Patients With Chronic Obstructive Pulmonary Disease: An Exploratory Analysis," *Journal of Nursing Scholarship*, vol. 53, no. 1, pp. 16–24, 2021, <https://doi.org/10.1111/jnu.12618>.
- [20] A. A. Safar, D. M. Salih, and A. M. Murshid, "Pattern recognition using the multi-layer perceptron (MLP) for medical disease: A survey," *International Journal of Nonlinear Analysis and Applications*, vol. 14, no. 1, pp. 1989–1998, Jan. 2023, <https://doi.org/10.22075/ijnaa.2022.7114>.
- [21] Z. Huang *et al.*, "Predicting the morbidity of chronic obstructive pulmonary disease based on multiple locally weighted linear regression model with K-means clustering," *International Journal of Medical Informatics*, vol. 139, Jul. 2020, Art. no. 104141, <https://doi.org/10.1016/j.ijmedinf.2020.104141>.
- [22] G. Anuradha and D. N. Jamal, "Classification of Dementia in EEG with a Two-Layered Feed Forward Artificial Neural Network," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7135–7139, Jun. 2021, <https://doi.org/10.48084/etasr.4112>.
- [23] S. Lin, Q. Zhang, F. Chen, L. Luo, L. Chen, and W. Zhang, "Smooth Bayesian network model for the prediction of future high-cost patients with COPD," *International Journal of Medical Informatics*, vol. 126, pp. 147–155, Jun. 2019, <https://doi.org/10.1016/j.ijmedinf.2019.03.017>.
- [24] S. Nuanmeesri and W. Sriurai, "Multi-Layer Perceptron Neural Network Model Development for Chili Pepper Disease Diagnosis Using Filter and Wrapper Feature Selection Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7714–7719, Oct. 2021, <https://doi.org/10.48084/etasr.4383>.