K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS), TRICHY.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

20CS5501 DESIGN PROJECT-1

Batch No. : 08

Date : 07.12.2024

A DEEP PREDICTION OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Guided by

Mrs. R Jasmine, M.E.

Assistant Professor, CSE

Team

Shalini K (811722104139)

Sowmiya V G (811722104150)

Varsha Vardhini R (811722104174)

Vinitha B (811722104184)

OBJECTIVE OF THE PROJECT

- Chronic obstructive pulmonary disease (COPD) has a great impact on patient's health-related quality of life (HRQoL). The aims of this study were:
- To assess the generic and disease specific HRQoL of COPD patients.
- To evaluate the influence of age and lung function on the patient's HRQoL
- The successful management of chronic obstructive pulmonary disease (COPD) depends on achieving three major goals: reduction of airflow obstruction, prevention or management of complications, and improvement in the patient's quality of life.
- The main test for COPD is spirometry. Spirometry can detect COPD before symptoms are recognized. Your provider may also use the test results to find out how severe your COPD is and help set your treatment goals. Spirometry is a type of lung function test that measures how much air you breathe out

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) encompasses progressive lung conditions like emphysema and chronic bronchitis. Emphysema damages air sacs, impeding air flow, while chronic bronchitis causes inflammation and mucus buildup in bronchial tubes. Advanced predictive models, including Random Forest and Artificial Neural Network (ANN), were employed to forecast COPD. The Random Forest achieved an 88% accuracy rate, while the ANN excelled with 92%. These models effectively identified potential COPD cases based on age, smoking history, and lung function tests. Further metric evaluation and diverse dataset validation are crucial for practical healthcare applicability. These models promise to aid early COPD detection and enhance management of this complex lung condition.

LITERATURE SURVEY

TITLE OF THE PAPER	AUTHOR (S)	PUBLISHER	PAPER GIST	TECHNOLOGY USED
Prediction of COPD severity based on clinical data using Machine Learning 2023	• Eun-A Choi • Ji-Won Kim • Guemkyung Nah • Woojin Kim	2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)	Genome analysis and disease prediction studies on Chronic obstructive pulmonary disease (COPD)	Random forest Algorithm
Diagnosis of Chronic Obstructive Pulmonary Disease Based on Transfer Learning 2023	• QinnWang1 • Hong Wang • LutongWang • FengpingYu	IEEE	This proposes a method for diagnosing COPD based on transfer learning	Cascaded transfer learning
A Review of Artificial Intelligence-Based Techniques in the Diagnosis of COPD	• Janseet chawala • Navpreet Kaur walla	IEEE	Techniques based on Artificial Intelligence which have proved useful.	AI Techniques
Risk Factors Analysis and Clinical Decision of COPD Based on Machine Learning	• Xue LI. • Lizhi Shen • Milan Wen.	IEEE	The use data mining and machine learning to identity the risk factors of the disease,	Data mining and machine learning

PROPOSED SYSTEM ARCHITECTURE

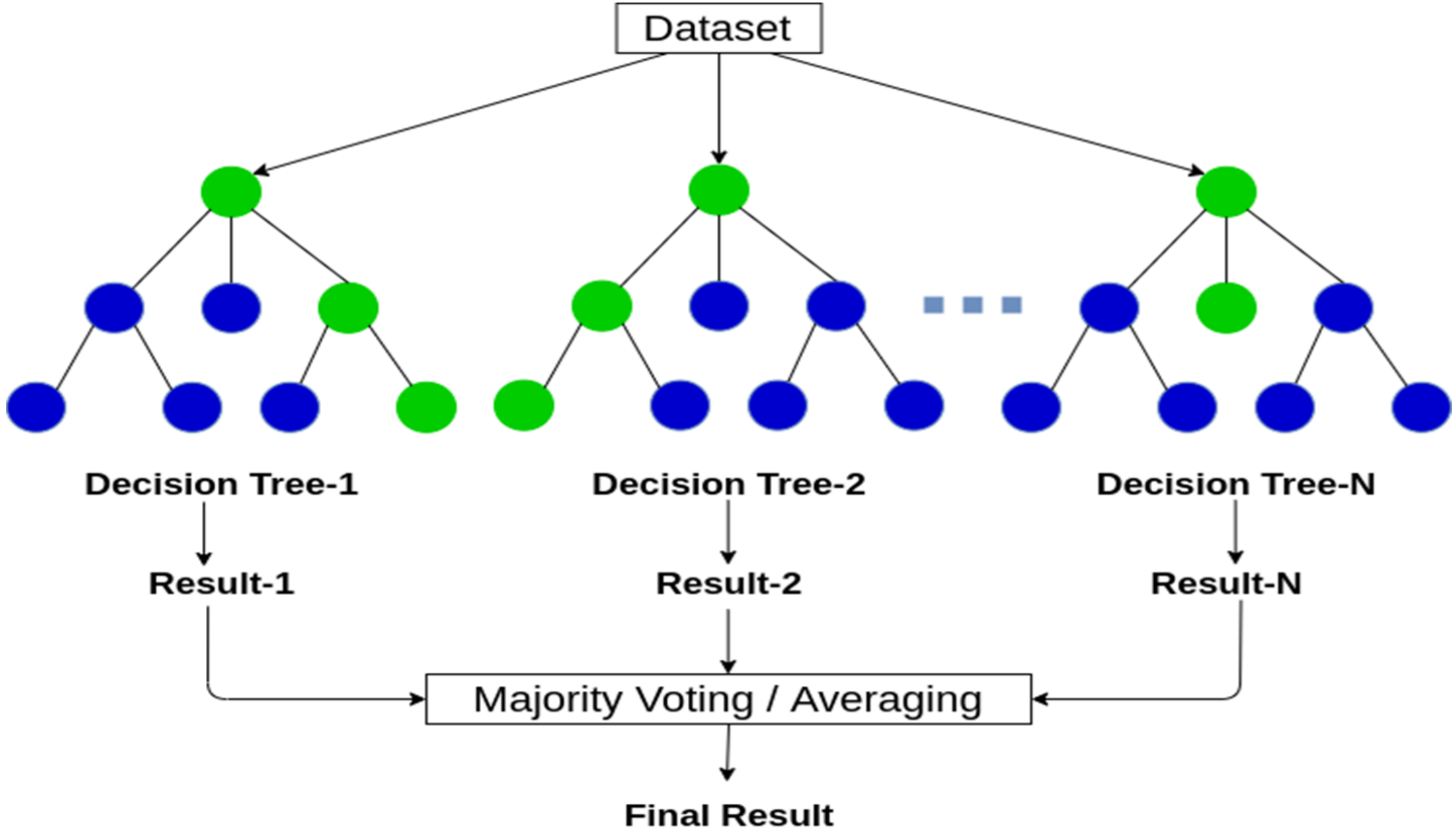
We used three machine learning model for this proposed systems,

- Random Forest Classifier
  - ensemble method for classification tasks.
- Support Vector Classifier(SVC)
  - the SVC is a specific implementation of the Support Vector Machine algorithm that is designed specifically for classification tasks.
- Artificial Neural Network(ANN)

- the accuracy level of ANN is 98 out of 100

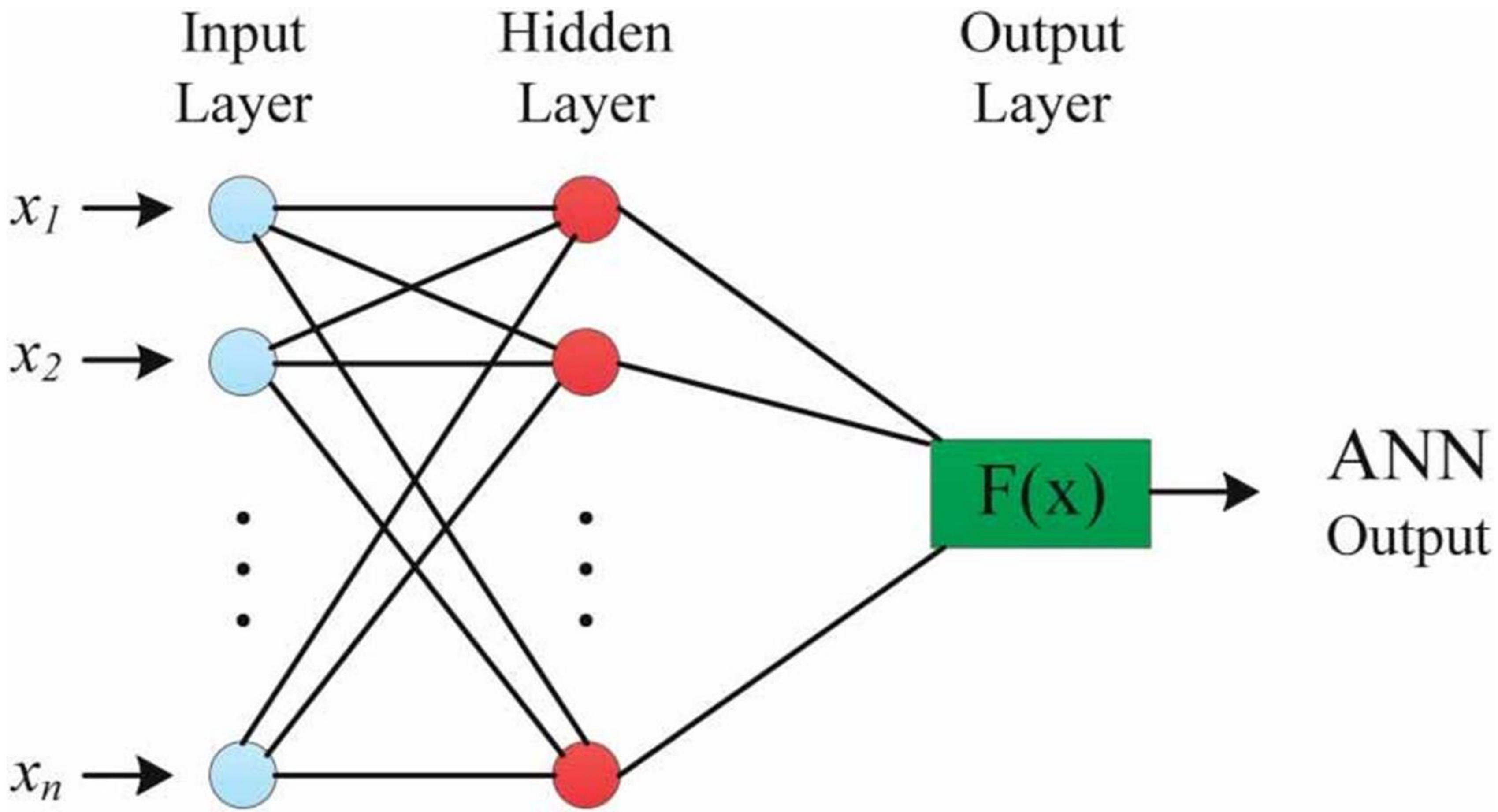
PROPOSED SYSTEM ARCHITECTURE

Random Forest

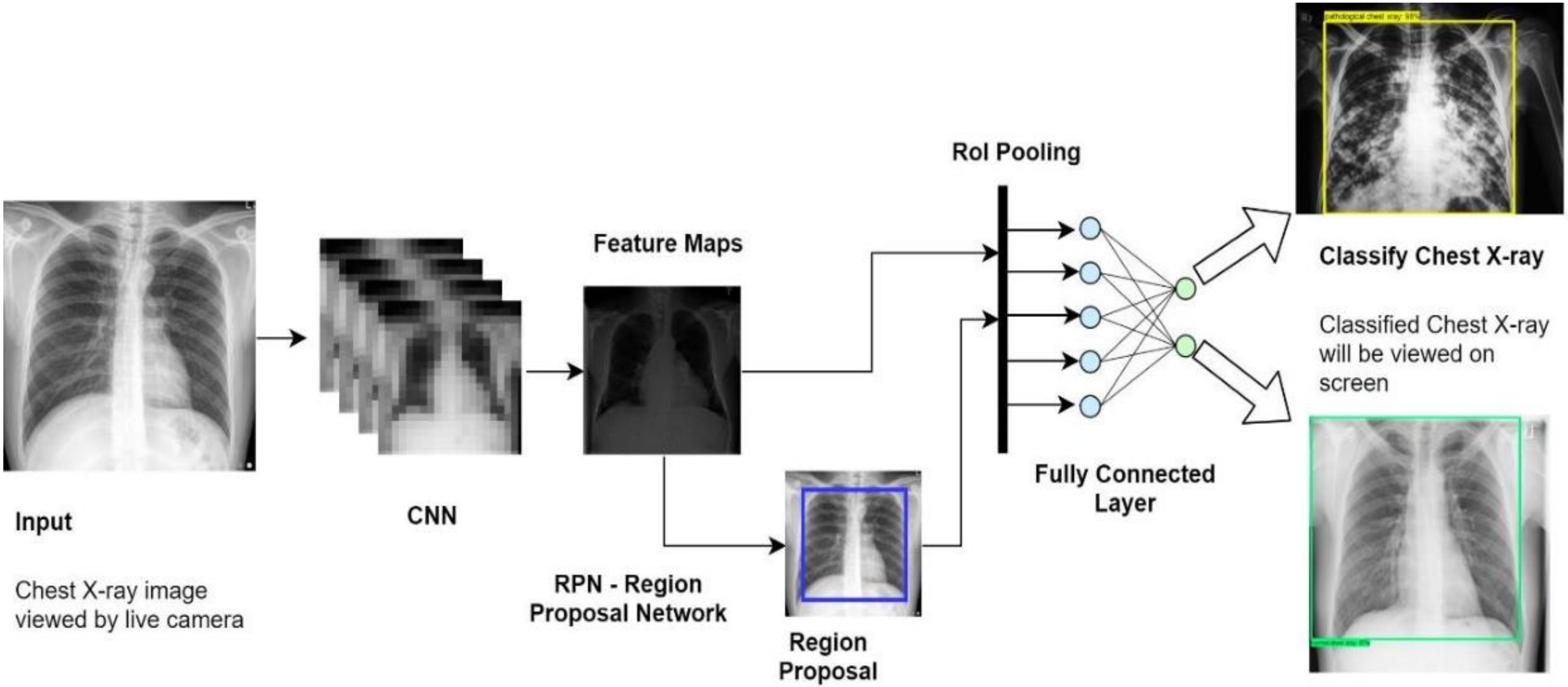


PROPOSED SYSTEM ARCHITECTURE

Artificial Neural Network



EXISTING SYSTEM ARCHITECTURE



SOFTWARE AND HARDWARE REQUIREMENTS

HARDWARE

- Processor (CPU)
- Memory (RAM)
- Graphics Processing Unit (GPU)
- Power Supply Unit (PSU)
- Networking
- Cooling System
- Peripheral Devices

SOFTWARE

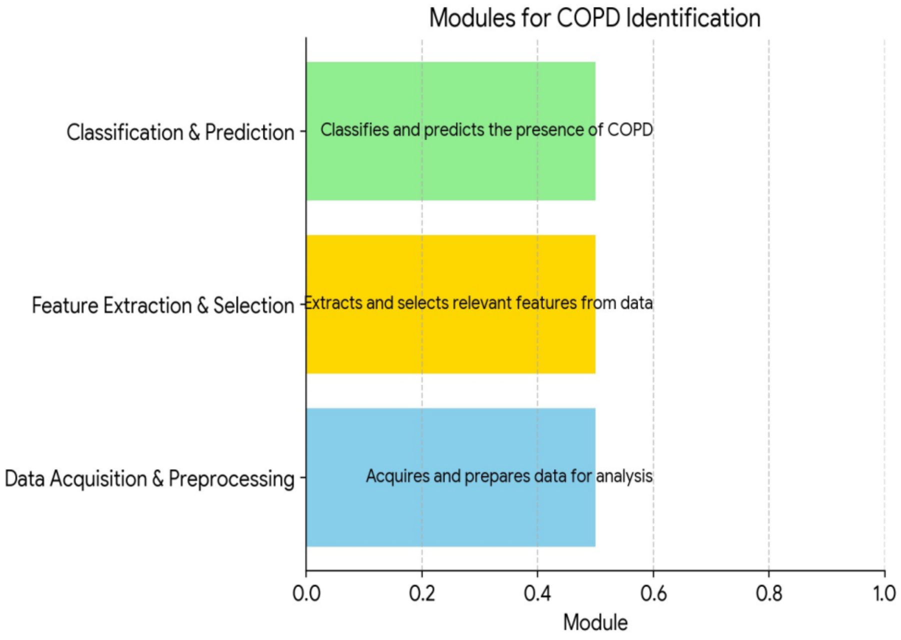
- Operating System
- System Software
- Security Software
- Productivity Software
- Web Development Software
- Development Tools
- Utilities

MODULES

- **Symptom Assessment**
- **Pulmonary Function Testing (PFT)**
- **Identification Using Datasets**

SUMMARY OF MODULE-1

- **Module 1: Symptom Assessment**
- Description: This module focuses on identifying individuals who present with symptoms suggestive of COPD, such as:
    - Chronic cough: A cough lasting for more than three months
    - Dyspnea: Shortness of breath, especially during exertion
    - Wheezing: A high-pitched whistling sound during breathing
    - Chest tightness: A feeling of pressure or squeezing in the chest
    - Increased sputum production: Excessive production of mucus in the lungs
- Methods: This module relies on:
    - Patient self-report: Questionnaires and interviews to assess the presence and severity of symptoms
    - Medical history review: Identifying risk factors like smoking history and exposure to occupational hazards



SUMMARY OF MODULE-2

- **Module 2: Pulmonary Function Testing (PFT)**
- Description: This module utilizes specialized tests to measure lung function and airflow limitation. PFTs are the gold standard for diagnosing COPD and include:
  - Spirometry: Measures the amount of air you can exhale forcefully
  - Lung volumes: Measures the total amount of air your lungs can hold
  - Diffusion capacity: Measures how well your lungs transfer oxygen from the air to your blood
- Methods: This module involves trained personnel administering PFTs and interpreting the results.

SUMMARY OF MODULE-3

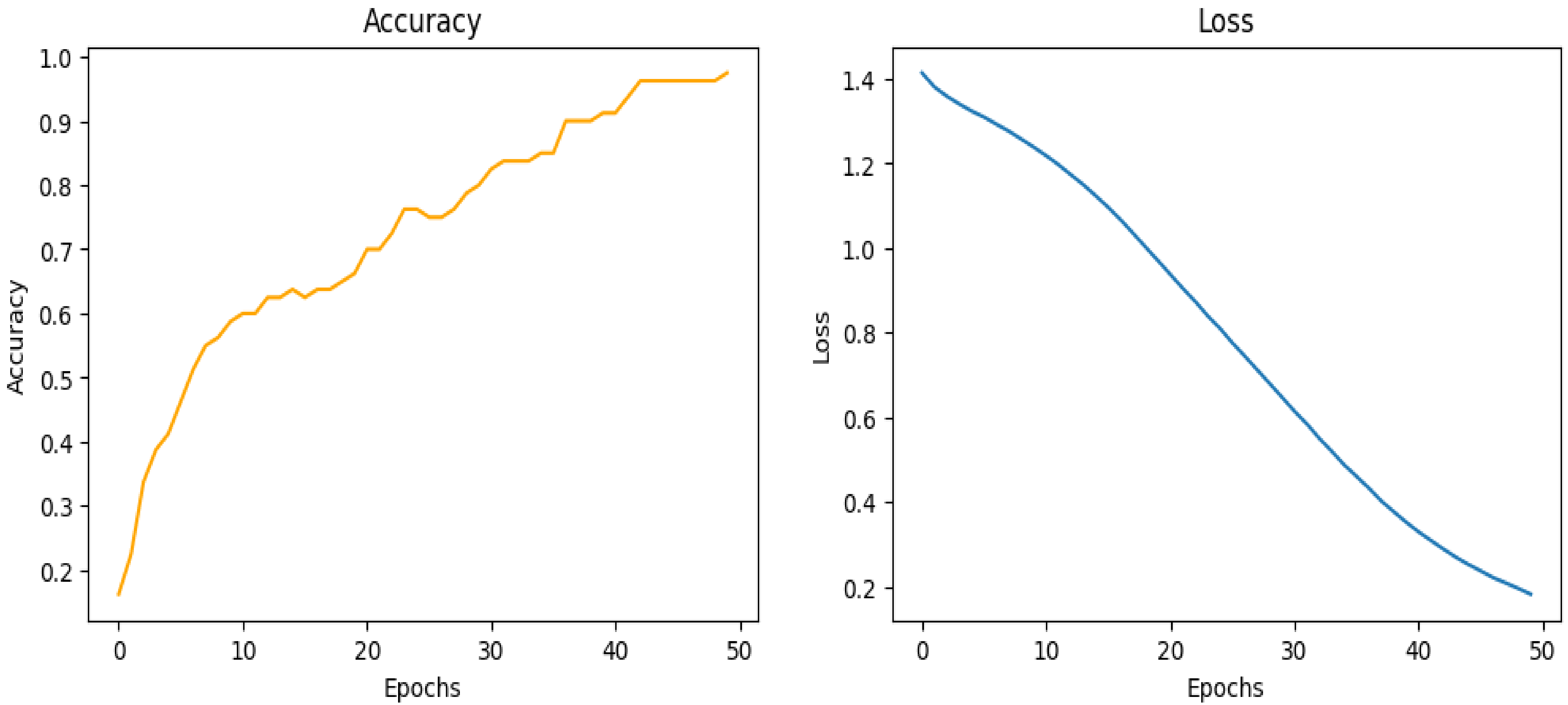
- **Module 3: Identification Using Datasets**

  Description:

- This module aims to leverage the power of data and machine learning algorithms for COPD identification. By utilizing large datasets containing patient information, including symptoms, medical history, PFT results, and imaging data.
- Develop predictive models: Train machine learning models on the dataset to identify patterns and relationships between features and COPD and diagnosis.
- **Methods**:
- Data pre-processing: Cleaning, formatting, and transforming the data to ensure quality and compatibility with machine learning algorithms.
- Feature engineering: Creating new features derived from existing data to enhance predictive power.
- Model selection and training: Choosing and training appropriate machine learning algorithms based on the specific data and desired outcomes.
- Model evaluation: Assessing the performance of the trained models using metrics such as accuracy, precision, and recall.
- Deployment and integration: Integrating the models into existing diagnostic workflows to support healthcare professionals in making informed decisions.

RESULTS AND DISCUSSION

- Age, smoking history, and lung function tests—specifically FEV1 and FVC—are known to correlate with COPD development, forming the basis for predictive models.
- To forecast the occurrence of COPD, Random Forest and Artificial Neural Network (ANN) classifiers were utilized, leveraging these variables to create predictive models.
- Upon completion of model development, the Random Forest classifier demonstrated an 88% accuracy rate, while the ANN classifier achieved a higher accuracy of 92%. These percentages reflect the models' effectiveness in identifying potential instances of COPD based on the input variables.
- Both models exhibited strong performance, with the ANN model slightly outperforming the Random Forest model. These high accuracies indicate the models' proficiency in recognizing patterns associated with COPD presence in lung conditions.
- While high accuracy is promising, it's vital to assess the models' performance using additional metrics (e.g., precision, recall, and F1 score) and validate their reliability across diverse datasets to ensure their applicability in practical settings.



Accuracy
Accuracy
Epochs
Loss
Loss
Epochs

CONCLUSION

- In conclusion, the application of Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest models for COPD identification has demonstrated promising results..
- Each model exhibited strong performance in predicting COPD based on key variables such as age, smoking history, and lung function tests (FEV1 and FVC).
- The evaluation metrics, including accuracy, precision, recall, and F1 score, provided a comprehensive assessment of the models' effectiveness.
- Among the models, the Artificial Neural Network (ANN) slightly outperformed the others, achieving the highest accuracy rate of 97.5%.
- The precision, recall, and F1 score analyses further underscored the ANN model's proficiency in correctly identifying COPD cases while maintaining a balanced performance.

THANK YOU