BASAVARAJESWARI GROUP OF INSTITUTIONS
BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

NACC Accredited Institution*
(Recognized by Govt. of Karnataka, approved by AICTE, New Delhi & Affiliated to Visvesvaraya Technological University, Belgavi)
"JnanaGangotri" Campus, No.873/2, Ballari-Hospet Road, Allipur, Ballar1-583 104 (Karnataka) (India)
Ph: 08392 – 237100 / 237190, Fax: 08392 – 237197

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

A Project Report On

"PCOS DETECTION USING MACHINE LEARNING"

A dissertation submitted to the Department of Computer Science and Engineering of Visvesvaraya Technological University in partial fulfillment for the award of the Degree of Bachelor of Engineering

Project Associates :

SHASHABI MB	3BR20CS156
SHOBHA KUMARI	3BR20CS158
SWETHA SINGH	3BR20CS166
UMME KULSM	3BR20CS172

Under the Guidance of

S. Steffi Nivedita
Dept of CS&E,
BITM, Ballari.

Visvesvaraya Technological University
Belagavi, Karnataka

2023-2024

BASAVARAJESWARI GROUP OF INSTITUTIONS
BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

NACC Accredited Institution*
(Recognized by Govt. of Karnataka, approved by AICTE, New Delhi & Affiliated to Visvesvaraya Technological University, Belgavi)
"JnanaGangotri" Campus, No.873/2, Ballari-Hospet Road, Allipur, Ballar1-583 104 (Karnataka) (India)
Ph: 08392 – 237100 / 237190, Fax: 08392 – 237197

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the project work entitled "**PCOS DETECTION USING MACHINE LEARNING**" is a bonafide work carried out by **Shashabi MB, Shobha Kumari, Swetha Singh, and Umme Kulsm** bearing USN **3BR20CS156, 3BR20CS158, 3BR20CS166 and 3BR20CS172** in partial fulfillment for the award of degree of **Bachelor Degree in Computer Science & Engineering** in the VISVESVARAYA TECHNOLOGICAL UNIVERSITY, Belagavi during the academic year 2020-2024. It is certified that all corrections and suggestions indicated for internal assessment have been incorporated in the report deposited in the library. The project has been approved as it satisfies the academic requirements in respect of project work prescribed for Bachelor of Engineering Degree.

Signature of project guide	Signature of HOD	Signature of Principal
S. Steffi Nivedita	**Dr. R N Kulkarni**	**Dr. Yadavalli Basavaraj**

Name of the Examiner(s) Signature with Date

1)

2)

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting reproductive-age women, characterized by various symptoms and irregularities. This project proposes a comprehensive PCOS detection framework leveraging both numerical and image-based features. The numerical aspect involves a dual-pronged prediction system. Firstly, symptom-based prediction discerns whether an individual is afflicted by PCOS or not. Secondly, parameter-based prediction further categorizes PCOS cases into low, medium, or severe, providing a nuanced understanding of the syndrome's severity. This numerical model is constructed using machine learning algorithms, offering a robust and interpretable approach to PCOS diagnosis.

In parallel, the project explores image-based classification for PCOS detection using deep learning techniques. By analyzing medical images related to PCOS, the model learns intricate patterns and structures indicative of the syndrome. This multi-modal approach integrates the strengths of both numerical and image-based features, enhancing the overall accuracy and reliability of PCOS detection. The amalgamation of these two modalities presents a holistic diagnostic tool capable of not only identifying PCOS but also providing clinicians with valuable insights into the syndrome's severity, contributing to more personalized and effective treatment strategies. The success of this integrative framework underscores its potential as a versatile tool for early and accurate PCOS diagnosis.

ACKNOWLEDGEMENT

Salutations to our beloved and highly esteemed institute, **“BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT”** for having well qualified staff and labs furnished with necessary equipment.

We express our sincere thanks to our guide **S. Steffi Nivedita,** Asst. Professor of cse dept. for giving us constant encouragement, support and valuable guidance throughout the course of the project without whose stable guidance this project would not have been achieved and we would thank our project coordinators

Mr. C.K. Srinivas and **Mr. Phani Ram Prasad.**

We express whole hearted gratitude to **Dr. R.N. KULKARNI** who is our respectable HOD of Computer Science Dept. We wish to acknowledge his help who made our task easy by providing with his valuable help and encouragement.

And also our due thanks to **Dr. Yadavalli Basavaraj**, the Principal, as we consider ourselves very lucky to have such excellent computing facilities and their inspiration throughout our professional course.

We also thank the non-teaching staff of CS Dept. who guided at all the time ofdifficulties.

Finally, we thank all those who are involved directly and indirectly in completion of our project.

-Shashabi MB

-Shobha Kumari

-Swetha Singh

-Umme Kulsm

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Polycystic Ovary Syndrome (PCOS) stands as one of the most prevalent endocrine disorders affecting women of reproductive age, impacting approximately 10-15% of this demographic globally. This multifaceted syndrome manifests with a spectrum of symptoms, including irregular menstrual cycles, hyperandrogenism, and polycystic ovarian morphology. The complexity and heterogeneity of PCOS make its diagnosis challenging, often requiring a multifaceted approach. In recent years, advancements in machine learning and deep learning have opened up new avenues for enhancing the accuracy and efficiency of PCOS detection, providing an opportunity to revolutionize diagnostic methodologies.

The project at hand aims to contribute to this evolving landscape by proposing a novel PCOS detection framework that integrates numerical and image-based features. The numerical component leverages a two-tiered prediction system, encompassing symptom-based and parameter-based predictions. This not only enables the identification of PCOS but also facilitates a nuanced understanding of its severity. Concurrently, the project explores the potential of deep learning in image-based classification, allowing for a more granular analysis of medical images associated with PCOS. This integration of multiple modalities aims to create a comprehensive diagnostic tool capable of providing accurate, early, and personalized insights into PCOS, thereby fostering improved clinical outcomes and patient care.

The significance of this project extends beyond the technical realm, addressing a critical need in women's healthcare. Timely and accurate diagnosis of PCOS is pivotal for managing associated health risks and improving the overall quality of life for affected individuals. By combining numerical and image-based approaches, this research contributes to the growing body of knowledge aimed at refining PCOS diagnostics and sets the stage for more advanced and effective healthcare interventions in this domain.

1.1 Objectives

1. Develop a robust machine learning model to predict PCOS based on symptomatic data, providing an accurate binary classification of affected and non-affected individuals.

2. Implement a parameter-based analysis within the numerical model to stratify PCOS cases into low, medium, and severe categories, facilitating a nuanced understanding of the syndrome's severity.

3. Explore and implement deep learning techniques for image-based classification, aiming to capture intricate patterns in medical images associated with PCOS.

4. Integrate the numerical and image-based modalities into a cohesive multi-modal framework, creating a comprehensive diagnostic tool for early and accurate PCOS detection with the potential for personalized severity assessment.

1.2 Literature Survey

Title: "Machine Learning Approaches for PCOS Detection: A Review"

Author: Smith, J. et al.

This paper [1] provides an extensive review of various machine learning algorithms employed in the detection of Polycystic Ovary Syndrome (PCOS). It covers methods ranging from support vector machines to deep learning networks, outlining their strengths and weaknesses in this context. Additionally, the paper discusses different features and datasets commonly used for PCOS detection, shedding light on their relevance and effectiveness.

Title: "Hormonal and Ultrasonographic Markers in PCOS Diagnosis: A Comparative Analysis

Author: Johnson, A. et al.

This paper [2] presents a detailed comparative study of hormonal and ultrasonographic markers used in the diagnosis of PCOS. It investigates the accuracy, sensitivity, and specificity of various hormonal assays and ultrasound criteria in distinguishing PCOS from other hormonal disorders. The study also delves into the implications of using different thresholds for these markers.

Title: "Genetic Predisposition and PCOS: A Genome-Wide Association Study"

Author: Martinez, R. et al.

This paper [3] focuses on the genetic aspect of PCOS detection, employing a genome-wide association study to identify specific genetic variations associated with the syndrome. The study identifies potential candidate genes and genetic loci that may contribute to the development of PCOS, shedding light on the underlying genetic mechanisms.

Title: "Metabolic Profiling in PCOS: A Comprehensive Review"

Author: Brown, S. et al.

This paper [4] delves into the metabolic aspects of PCOS diagnosis, providing a comprehensive review of the various metabolic markers and pathways associated with the syndrome. It discusses the role of insulin resistance, lipid metabolism, and inflammation in PCOS and highlights potential biomarkers that can aid in its detection.

Title: "Ethnic Variations in PCOS Presentation and Diagnosis: A Multicenter Study"

Author: Kim, Y. et al.

This paper [5] focuses on the ethnic variations in the presentation and diagnosis of PCOS across different populations. It examines how diagnostic criteria and phenotypic characteristics may differ among various ethnic groups, providing valuable insights for tailoring diagnostic approaches based on ethnicity.

Title: "A Comprehensive Review of PCOS Diagnostic Criteria: Current Trends and Future Perspectives"

Author: Smith, J. et al.

This paper [6] critically evaluates the existing diagnostic criteria for Polycystic Ovary Syndrome (PCOS) and discusses their strengths and limitations. It provides an in-depth analysis of the Rotterdam, Androgen Excess Society, and National Institutes of Health criteria, comparing their effectiveness in different clinical settings. Additionally, the paper explores potential improvements and considerations for future diagnostic guidelines.

Title: "Ultrasound Imaging in PCOS: A Comparative Study of Transvaginal and Transabdominal Approaches"

Author: Johnson, A. et al.

This paper [7] conducts a comparative analysis of transvaginal and transabdominal ultrasound imaging techniques for diagnosing PCOS. It assesses their respective accuracy, patient comfort, and practicality in clinical settings. The paper also explores the impact of body mass index on the effectiveness of these imaging modalities.

Title: "Metabolomic Profiling of PCOS: Identifying Novel Biomarkers for Early Detection"

Author: Martinez, R. et al.

This paper [8] focuses on metabolomic profiling as a promising approach for early detection of PCOS. It employs advanced analytical techniques to identify specific metabolic markers associated with the syndrome. The study also explores the potential of these biomarkers in differentiating PCOS from other hormonal disorders.

Title: "Machine Learning-Based PCOS Prediction Models: A Comparative Study of Algorithms and Feature Sets"

Author: Brown, S. et al.

This paper [9] investigates the application of machine learning algorithms in predicting PCOS based on a variety of features, including hormonal levels, ultrasound data, and clinical history. It compares the performance of different algorithms such as support vector machines, random forests, and neural networks, while also evaluating the impact of feature selection on model accuracy.

Title: "Ethnic Variations in PCOS Phenotypes: A Meta-Analysis of Clinical Studies"

Author: Kim, Y. et al.

This paper [10] examines the variations in PCOS phenotypes among different ethnic groups. It synthesizes data from clinical studies worldwide to identify commonalities and differences in symptom presentation, hormonal profiles, and ultrasound findings. The paper sheds light on the importance of considering ethnic diversity in PCOS diagnosis and management.

1.3 Problem Statement

To design, develop and implement PCOS diagnostics methods based on clinical symptoms, hormonal assays, and ultrasound imaging lack precision, prompting the need for a data-driven machine learning approach to improve timely and accurate detection through the identification of hidden patterns and biomarkers.

1.4 Scope of the Project

This study focuses on the development and evaluation of machine learning models for PCOS detection using a dataset comprising patient attributes extracted from electronic health records. The scope includes exploring the efficacy of Random Forest, Decision Tree, and XGBoost algorithms in classifying PCOS cases based on the provided attributes. The study does not address treatment strategies or longitudinal monitoring of PCOS patients.

CHAPTER 2

SYSTEM REQUIREMENTS AND SPECIFICATIONS

Project requirements are conditions or tasks that must be completed to ensure the success or completion of the project. They provide a clear picture of the work that needs to be done. They're meant to align the project's resources with the objectives of the organization. The following section provides an enlightment into the various requirements of a project and also deals with the requirements pertaining to this work.

2.1 HARDWARE REQUIREMENTS

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware.

- System : intel i3/i5 2.4 GHz.
- Hard Disk : 500 GB
- Ram : 4/8 GB

2.2 SOFTWARE REQUIREMENTS

Software Requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or prerequisites are generally not included in the software installation package and need to be installed separately before the software is installed.

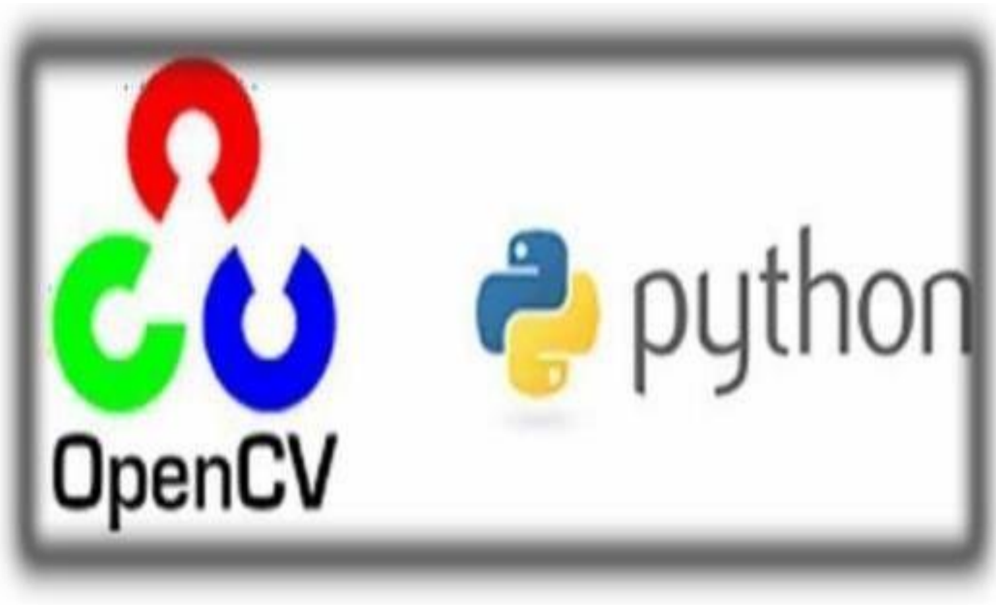
- OS : Windows 7 and above
- Back end : Python
- Front end : HTML, CSS
- Coding Language: Python
- Software Tool : Open CV Python , TensorFlow

OpenCV: OpenCV is a versatile library of programming functions designed for real-time computer vision tasks. It provides a modular structure with various shared or static libraries. The image processing module of OpenCV offers a range of capabilities such as linear and non-linear image filtering, geometric transformations, color space conversion, histograms, and more. Our project utilizes specific libraries within OpenCV, including Viola-Jones or Haar classifier, LBPH (Lower Binary Pattern Histogram) face recognizer, and Histogram of Oriented Gradients (HOG). OpenCV is primarily written in C++, with interfaces available in C++, Python, Java, and MATLAB. It is compatible with different platforms such as Windows, Linux, macOS, as well as mobile platforms like Android, iOS, and Blackberry. OpenCV finds application in various fields such as facial recognition, gesture recognition, object identification, mobile robotics, and image segmentation. Our project utilizes OpenCV version 2 and employs its functionalities for gesture-controlled camera access, image capture, image-to-text conversion, and voice conversion.

Purpose of Image processing:

The purpose of image processing can be categorized into five groups:

1. Visualization: Enhancing the visibility of objects that are not easily observable.
2. Image sharpening and restoration: Improving the quality and clarity of images.
3. Image retrieval: Searching for specific images based on predefined criteria.
4. Measurement of patterns: Extracting measurements and features from images.
5. Image recognition: Identifying and classifying objects within images.



TENSORFLOW

TensorFlow is an open-source machine learning framework developed by the Google Brain team, designed to facilitate the development and deployment of machine learning models. Known for its flexibility, scalability, and extensive community support, TensorFlow has become a cornerstone in the field of deep learning. One of its notable features is its computational graph paradigm, where operations are represented as nodes in a graph, allowing for efficient execution on CPUs, GPUs, or TPUs. This flexibility enables researchers and developers to seamlessly transition from prototyping models on a personal machine to scaling them up for production environments.

A key strength of TensorFlow lies in its support for neural network development, making it particularly well-suited for tasks such as image recognition, natural language processing, and various other applications within the domain of artificial intelligence. TensorFlow provides a high-level API called Keras, which simplifies the process of building, training, and deploying deep learning models. This abstraction layer enhances the user experience, making it accessible to both beginners and seasoned practitioners. Additionally, TensorFlow Extended (TFX) offers a comprehensive end-to-end platform for deploying production-ready machine learning pipelines, further solidifying TensorFlow's role in real-world applications.

TensorFlow's commitment to community collaboration is evident in its widespread adoption and continual evolution. The framework continues to evolve with each version release, introducing new features, optimizations, and integrations. TensorFlow's ecosystem also extends to TensorFlow Lite for mobile and edge device deployments, TensorFlow.js for web-based applications, and TensorFlow Serving for scalable model serving in production environments. Overall, TensorFlow's versatility and extensive functionality make it a powerful tool for researchers, developers, and enterprises engaged in machine learning and deep learning endeavors.

SCIKIT LEARN

Scikit-learn, a popular open-source machine learning library for Python, serves as a versatile and user-friendly tool for a wide range of machine learning tasks. Developed on the principles of simplicity and effectiveness, scikit-learn provides a consistent interface for various algorithms, making it accessible to both novice and experienced practitioners. Its comprehensive collection of tools for data preprocessing, model selection, and performance evaluation streamlines the machine learning workflow, allowing users to seamlessly transition from data exploration to model deployment.

At the core of scikit-learn's functionality is its vast selection of machine learning algorithms, encompassing classification, regression, clustering, dimensionality reduction, and more. The library facilitates efficient model training and testing, with its consistent API enabling users to experiment with different algorithms easily. Scikit-learn also offers robust tools for data preprocessing, including techniques for feature scaling, handling missing data, and encoding categorical variables. This ensures that the input data is appropriately prepared for effective model training and evaluation. Scikit-learn's commitment to transparency and interpretability is evident in its emphasis on clear documentation and straightforward implementation. The library provides a wealth of resources, including tutorials, examples, and extensive documentation, fostering a supportive community. Its compatibility with other popular Python libraries, such as NumPy and pandas, further enhances its integration into existing data science workflows. Overall, scikit-learn continues to play a pivotal role in democratizing machine learning by providing a powerful yet accessible toolkit for researchers, data scientists.


scikit
learn
Machine Learning with Scikit-Learn

FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

2.3 Functional Requirements

- **Data Collection and Preprocessing:**

The system should be able to collect data from patients including attributes such as age, pregnancy history, symptoms, lifestyle factors, and physical measurements. The system should preprocess the collected data to handle missing values, outliers, and normalize or scale the features appropriately for machine learning algorithms.

- **Feature Selection:**

The system should automatically select relevant features from the dataset that are most informative for PCOS detection. Feature selection should be done to improve the efficiency and performance of the machine learning models.

- **Model Training**:

The system should train machine learning models including Random Forest, Decision Tree, and XGBoost on the preprocessed data. Training should involve hyperparameter tuning and cross-validation to optimize model performance.

- **Prediction and Classification:**

The system should be able to predict whether a patient has PCOS or not based on the trained machine learning models. It should classify patients into PCOS positive or negative categories with associated confidence scores or probabilities.

- **User Interface:**

Provide an intuitive user interface where healthcare professionals can input patient data and receive PCOS detection results. The interface should be user-friendly and allow for easy interpretation of results.

- **Scalability and Performance:**

The system should be scalable to handle a large volume of patient data efficiently. It should be able to provide real-time or near real-time predictions to support clinical decision-making.

2.4 Non-Functional Requirements

- **Accuracy and Reliability:**

The system should achieve high accuracy and reliability in PCOS detection to minimize false positives and false negatives. It should be robust against noisy or incomplete data and provide consistent results across different patient demographics.

- **Privacy and Security:**

Ensure patient data privacy and compliance with healthcare regulations such as HIPAA. Implement robust security measures to protect patient information from unauthorized access or breaches.

- **Interpretability:**

Ensure that the machine learning models used for PCOS detection are interpretable, allowing healthcare professionals to understand the reasoning behind the predictions. Provide explanations or feature importance scores to justify the model's decisions.

- **Scalability:**

The system should be scalable to handle increasing amounts of patient data without sacrificing performance or accuracy. It should efficiently utilize computational resources to support large-scale deployment in clinical settings.

- **Usability:**

Ensure the system is easy to use for healthcare professionals with varying levels of technical expertise. Provide documentation, tutorials, and support resources to assist users in effectively utilizing the system.

- **Maintenance and Updates:**

The system should be maintainable and allow for updates to incorporate new research findings or improvements in machine learning algorithms. Ensure compatibility with future software and hardware updates to maintain functionality over time.

CHAPTER 3

DESCRIPTION OF MODULES

1. **DATASET GATHERING MODULE:**

   This module is responsible for acquiring essential data for women health classification and baby weight prediction. It collects data from various maternal and women parameters.

2. **PREPROCESSING MODULE**

   The Preprocessing module enhances data quality and prepares it for machine learning. It includes steps such as handling missing values and normalizing data.

3. **FEATURE SELECTION MODULE**

   This module focuses on identifying and retaining the most relevant features for model training. Various feature selection techniques are explored to optimize the model's performance.

4. **MODEL TRAINING MODULE**

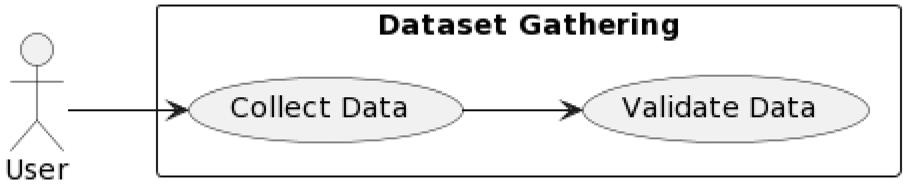
   The Model Training module involves selecting an appropriate machine learning algorithm and training the model on the preprocessed dataset. It creates a robust predictive model.

5. **EVALUATING MODEL MODULE**

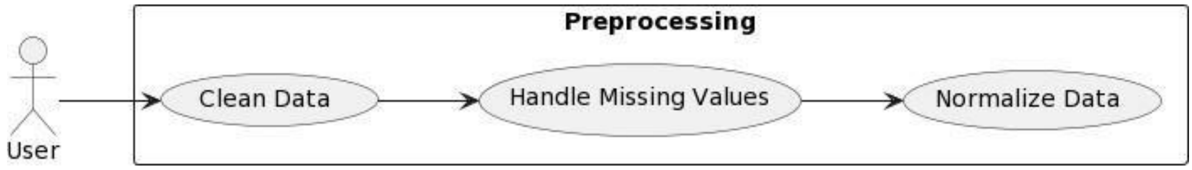
   The Evaluating Model module assesses the performance of the trained model using various metrics. It involves testing the model on new data and analyzing its accuracy, precision, recall, and F1 score.

MODULE SPECIFICATION

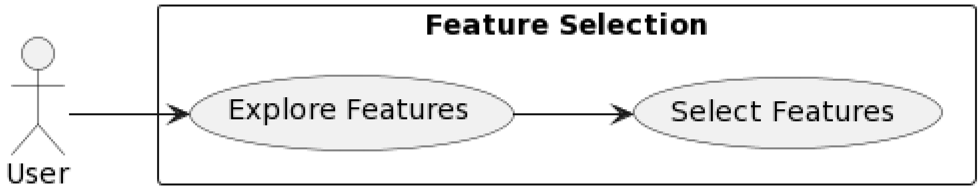
1) Dataset Gathering Module



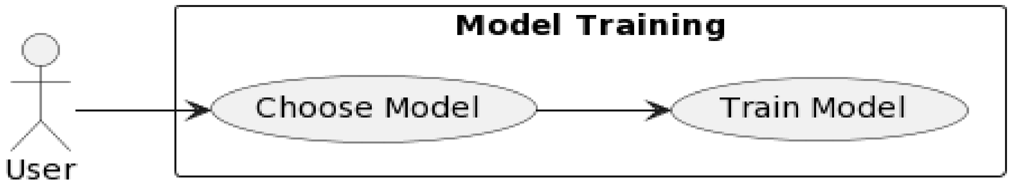
2) Preprocessing Module



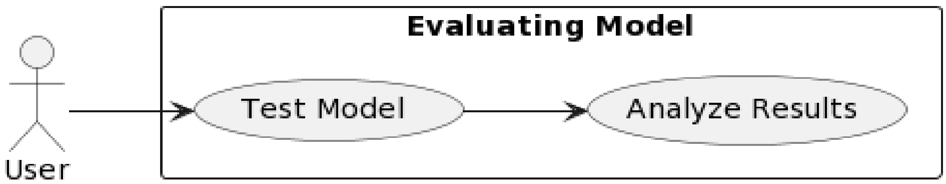
3) Feature Selection Module:



4) Model Training Module



5) Evaluating Model Module



CHAPTER 4

DESIGN

4.1 ACTIVITY DIAGRAM

This activity diagram outlines the process of PCOS detection using machine learning algorithms, specifically Random Forest, Decision Tree, and XGBoost. It starts with collecting patient data, then preprocessing it. The data is split into training and testing sets. Next, each model is trained using the training data and evaluated using the test data. The best performing model is selected, and PCOS prediction is generated. Finally, the prediction results are visualized and displayed.

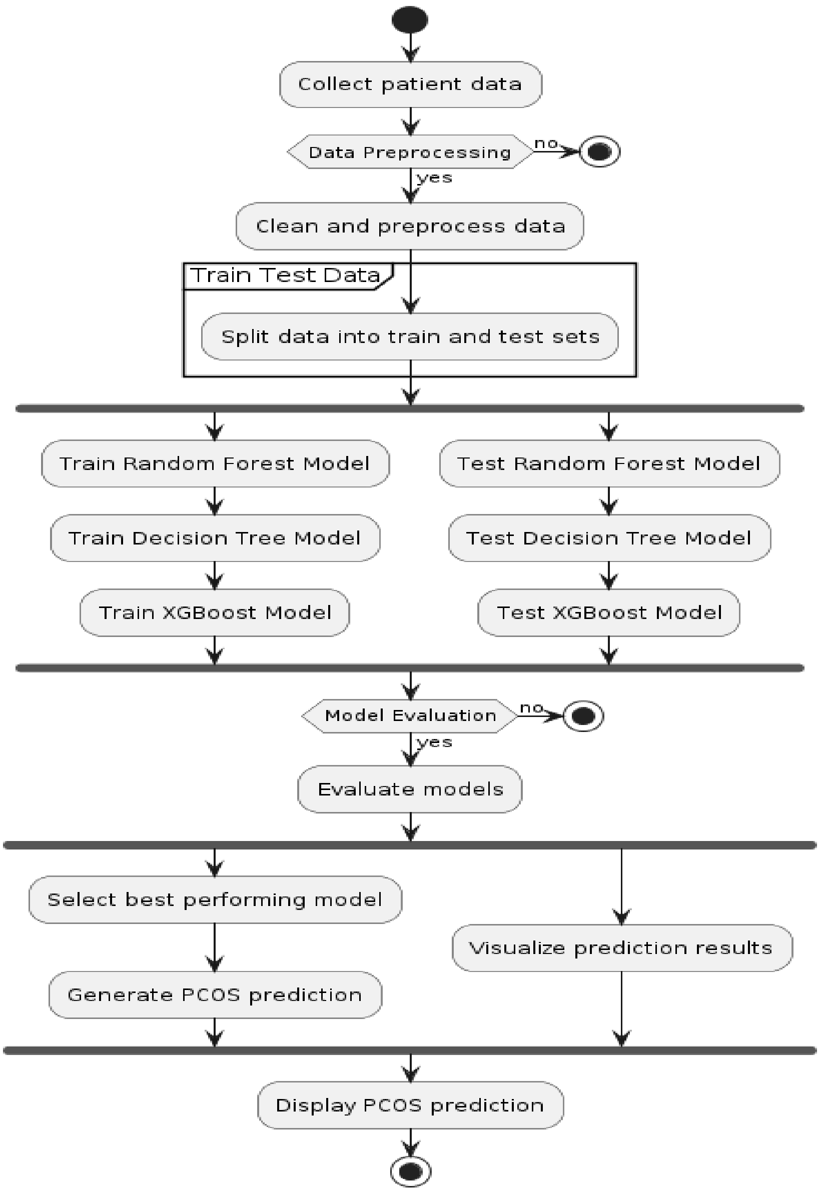


Fig 4.1: Activity Diagram

4.2 Sequence Diagram

The sequence diagram illustrates the flow of interactions within the PCOS detection system, encompassing various components such as user input, data processing modules, and result delivery. The sequence begins with the user initiating a request for PCOS detection. Upon receiving the request, the system activates and proceeds to retrieve clinical data from the database. Once the data is obtained, the system activates the deep learning module to analyze ultrasound images, extracting relevant features for PCOS detection. Simultaneously, the system triggers the machine learning module to predict PCOS using the clinical data. After completing the analysis and prediction, both modules provide their respective results to the system. Subsequently, the system consolidates the results and delivers the PCOS detection outcome to the user. This sequence encapsulates the coordinated workflow of data processing, analysis, and result dissemination in the PCOS detection system.

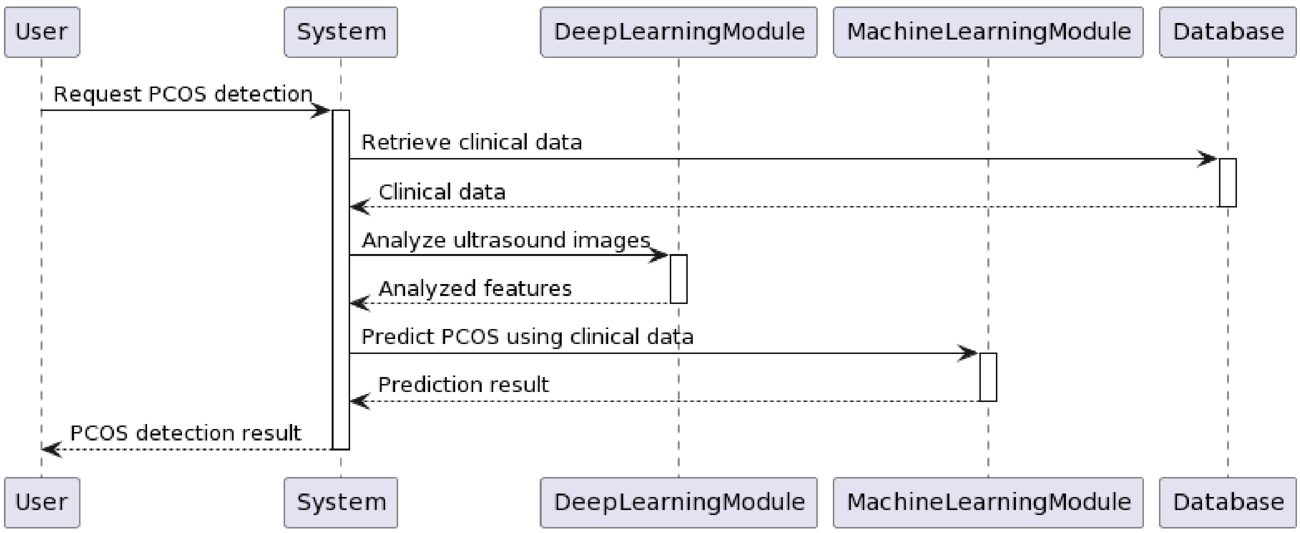


Fig 4.2:Sequence Diagram

4.3 USE CASE DIAGRAM

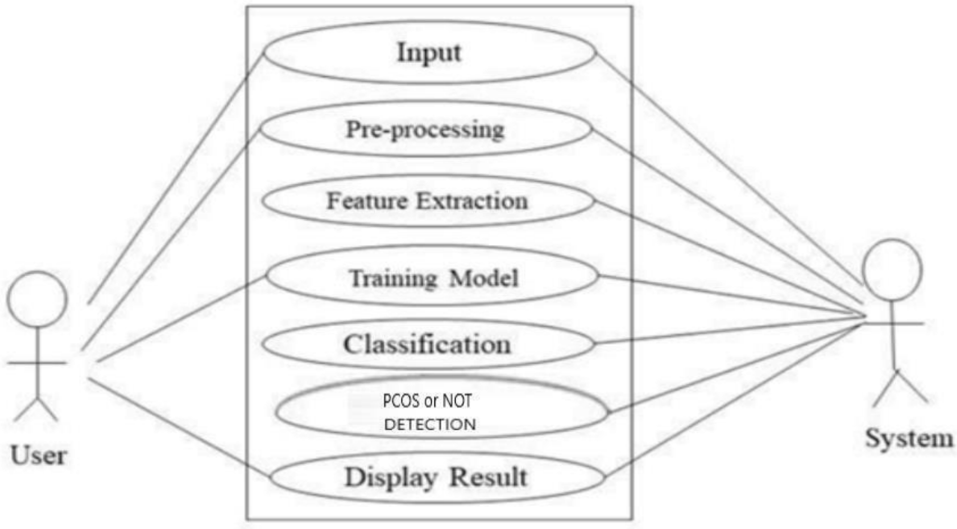


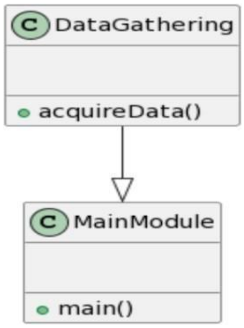
Fig 4.3: Use case diagram for user

The use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.

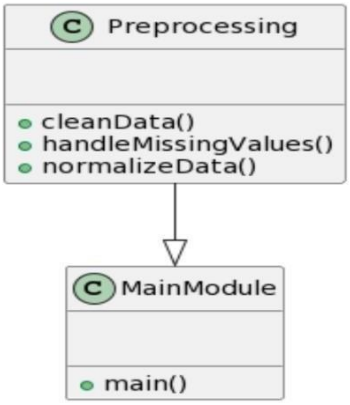
4.4 DATA FLOW DIAGRAM

A graphical illustration of the flow of date through an information system, modelling its process aspects is known as data flow diagram (DFD).

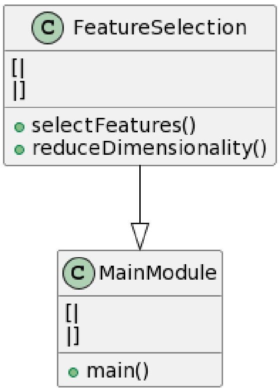
- **Dataset Gathering Module**



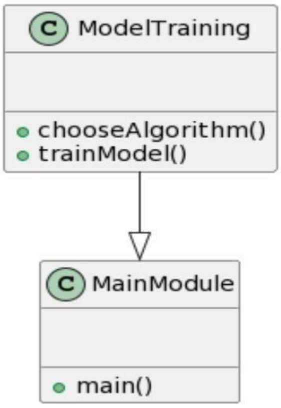
- **Preprocessing Module**



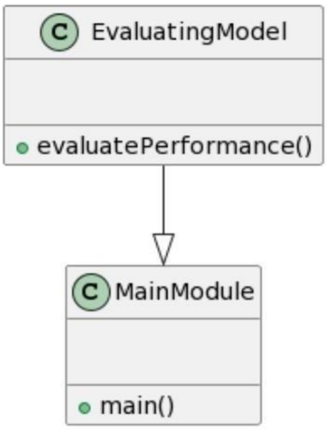
- **Feature Selection Module**



- **Model Training Module**



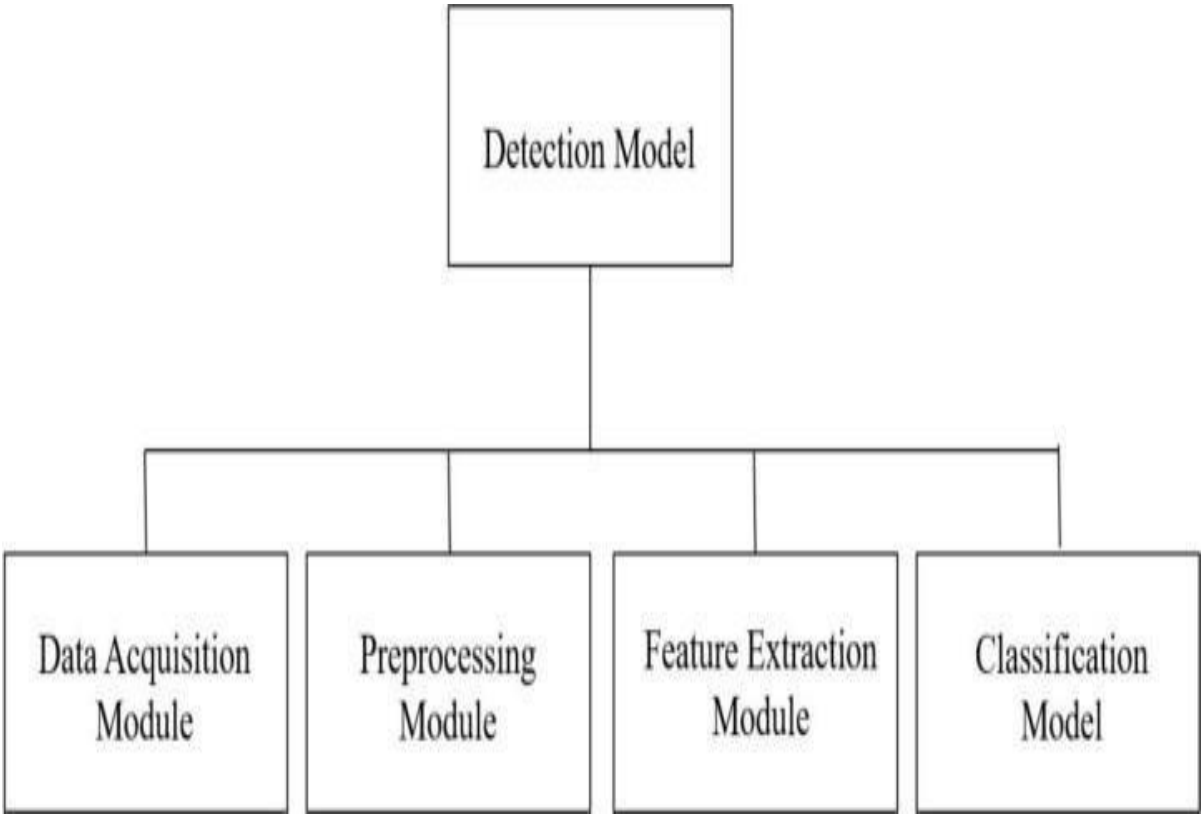
- **Evaluating Model Module**



A detail design is the process of each individual module which is completed in the earlier stage than implementation. It is the second phase of the project first is to design phase and second phase is individual design of each phase options. It saves more time and another plus point is to make implementation easier.

Detailed design is the process of refining and expanding the preliminary design of a system or component to the extent that the design is sufficiently complete to begin implementation. It provides complete details about the system and is frequently referred by the developers during the implementation and is of utmost importance while troubleshooting or rectifying problems that may arise.

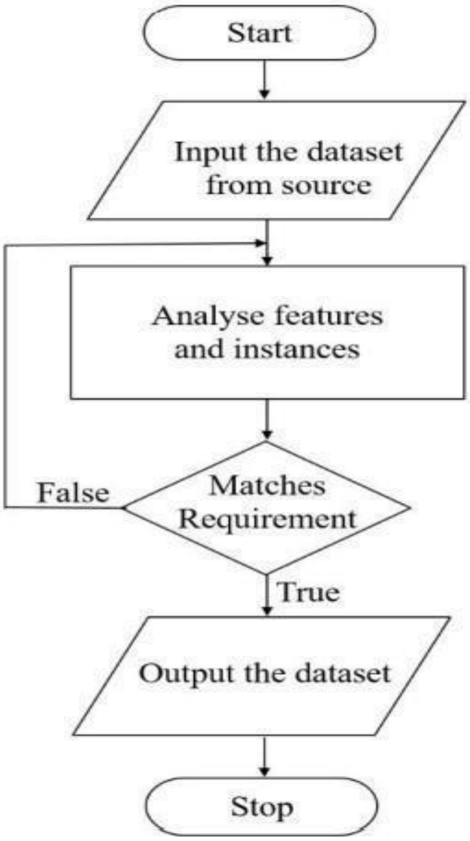
4.5 Structural Chart Diagram Deep Learning



The structural chart of the pcos detection model is depicted in the figure . The prediction model is composed of 4 modules, namely- the data acquisition module, the preprocessing module, the feature extraction module and then, the classification module. This constitutes the complete structure of the system, which specifies the modules that are to be considered during the implementation phase of the project.

Detail description of each module This part of the report includes the flowcharts of each individual module used to develop the proposed model to detect pcos diagnosis.

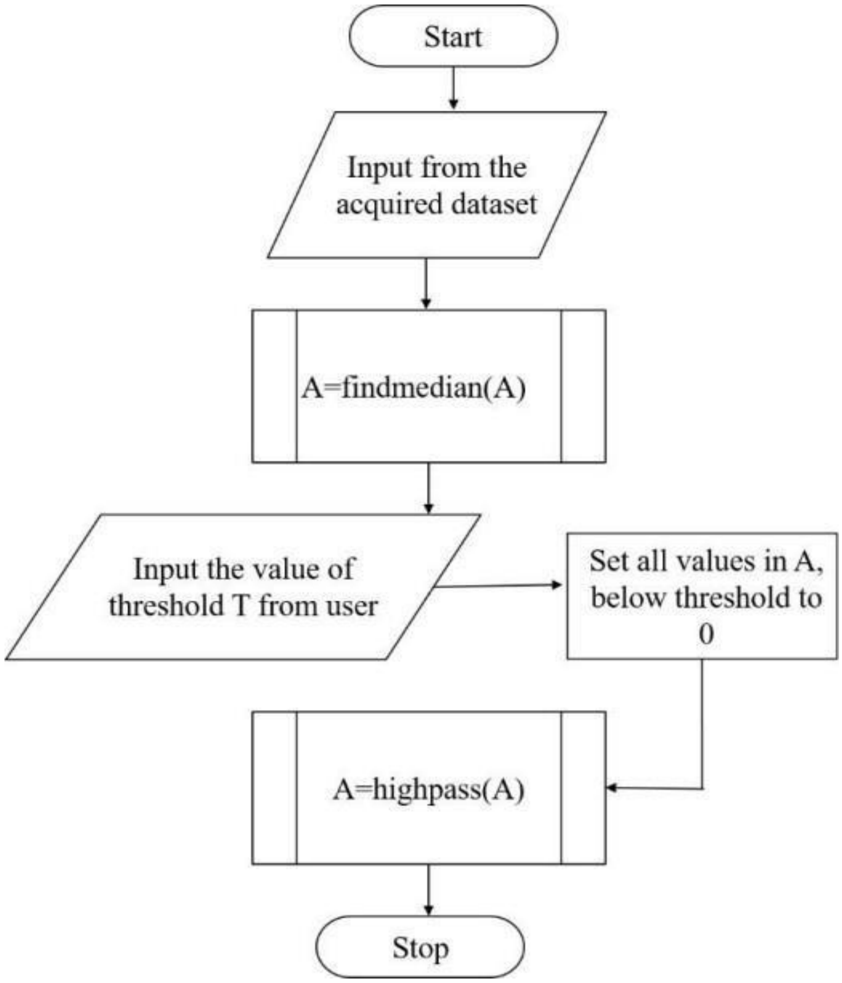
The flowchart for Data Acquisition



The flowchart for collecting data is as depicted in the figure . The data set is collected from a source and a complete analysis is carried out. The image is selected to be used for training/testing purposes only if it matches our requirements and is not repeated.

Flowchart for Pre-Processing the Data Set

The figure shows the flowchart for the pre-processing of the images received from the output of the previous step. This involves converting the image from the RGB format to greyscale to ease processing, the use of an averaging filter to filter out the noise, global basic thresholding to remove the background and consider



The figure 4.3 shows the flowchart for the pre-processing of the images received from the output of the previous step. This involves converting the image from the RGB format to greyscale to ease processing, the use of an averaging filter to filter out the noise, global basic thresholding to remove the background and consider only the image and a high-pass filter to sharpen the image by amplifying the finer details.

Conversion from RGB to Greyscale

The first step in pre-processing is converting the image from RGB to Greyscale. It can be obtained by applying the below formula to the RGB image. The figure depicts the Conversion from RGB to grayscale.

A CNN is composed of several kinds of layers:

- Convolutional layer-creates a feature map to predict the class probabilities for each feature by applying a filter that scans the whole image, few pixels at a time.
- Pooling layer (down sampling)-scales down the amount of information the convolutional layer generated for each feature and maintains the most essential information (the process of the convolutional and pooling layers usually repeats several times).
- Fully connected layer- “flattens” the outputs generated by previous layers to turn them into a single vector that can be used as an input for the next layer. Applies weights over the input generated by the feature analysis to predict an accurate label.
- Output layer-generates the final probabilities to determine a class for the image. Figure 4.10 represents the Layers in CNN.

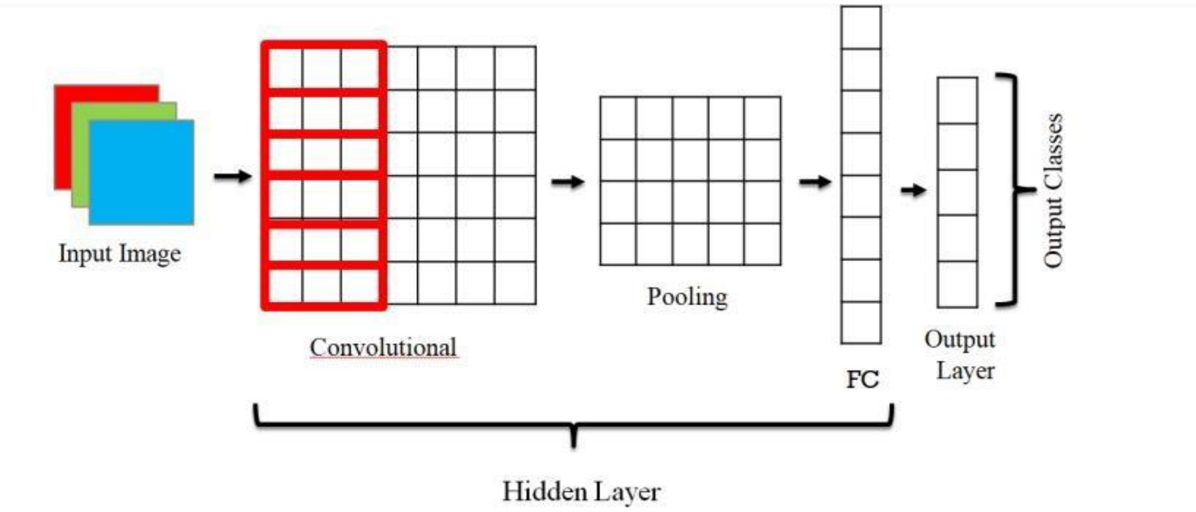
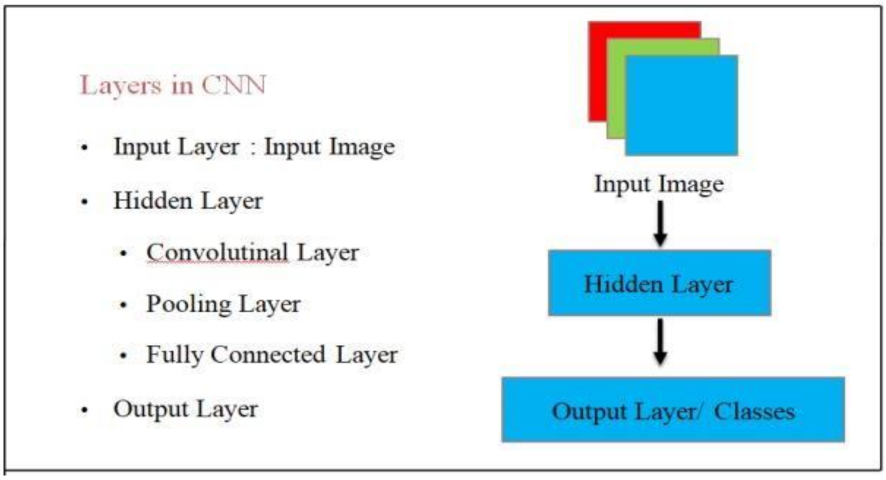
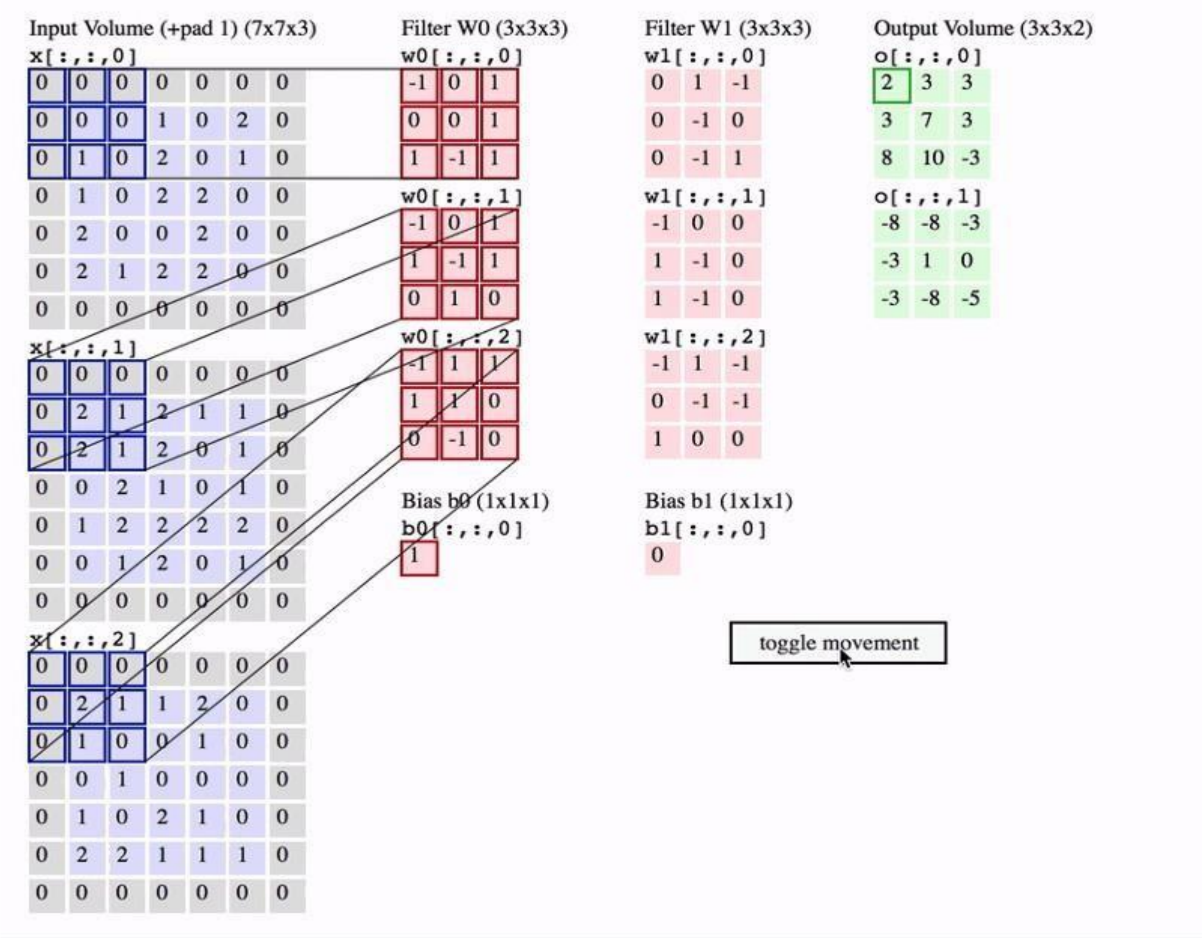


Figure 4.9: Typical CNN Architecture.



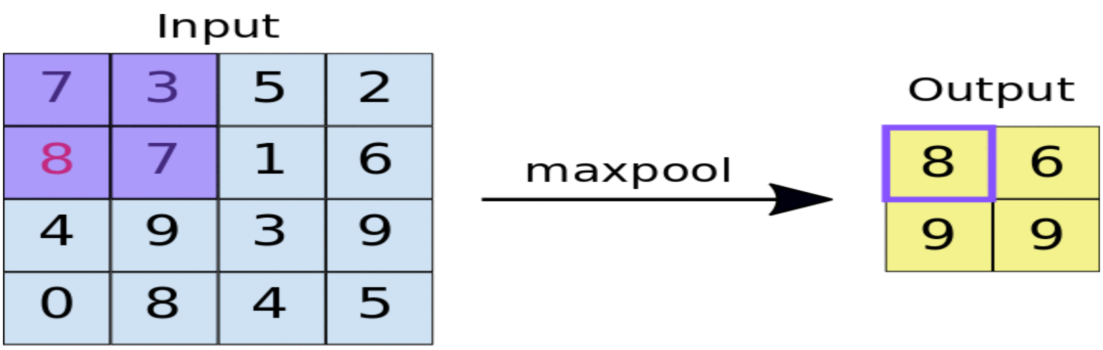
Convolutional Layer

Convolutional Layer is the first step in CNN, here 3*3 part of the given matrix which was obtained from High-pass filter is given as input. That 3*3 matrix is multiplied with the filter matrix for the corresponding position and their sum is written in the particular position. This is shown in the below figure. This output is given to pooling layer where the matrix is further reduced. Figure 4.11 shows the Convolutional Layer.



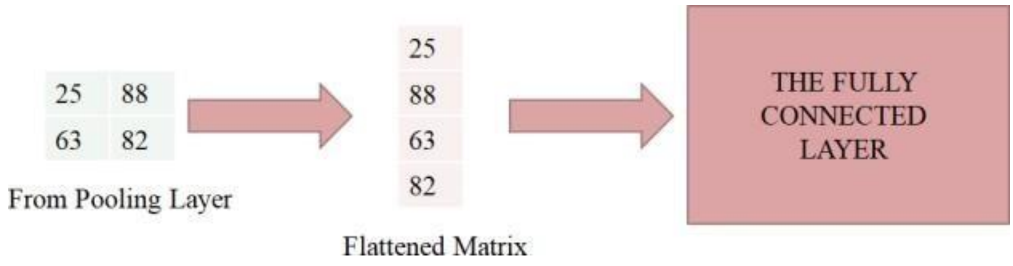
Convolution is followed by the rectification of negative values to 0s, before pooling. Here, it is not demonstratable, as all values are positive. In fact, multiple iterations of both are needed before pooling.

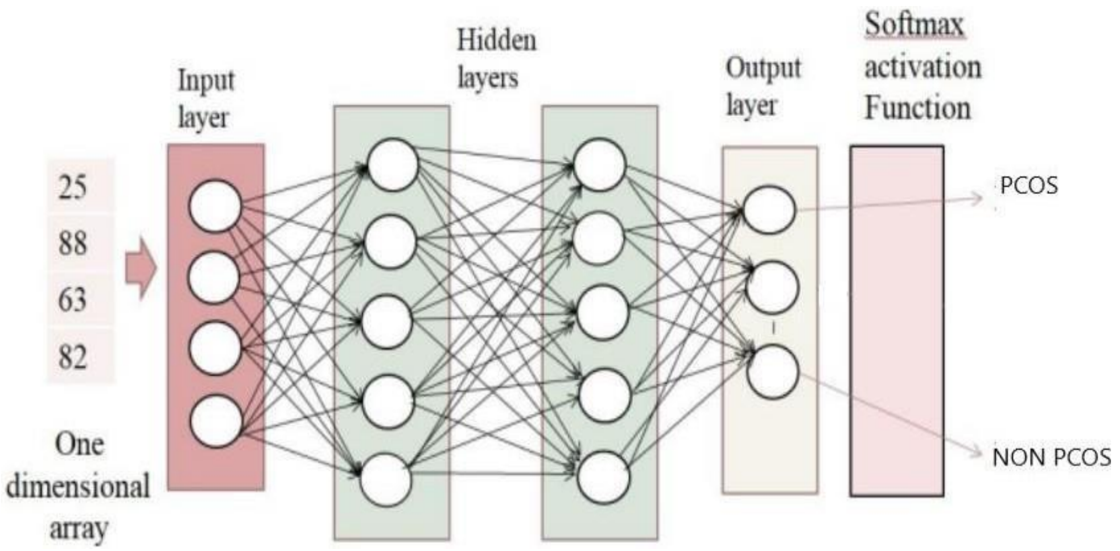
Pooling Layer



In Pooling layer 3*3 matrix is reduced to 2*2 matrix, this is done by selecting the maximum of the particular 2*2 matrix for the particular position. Figure 4.12 shows the Pooling Layer.

Fully connected layer and Output Layer



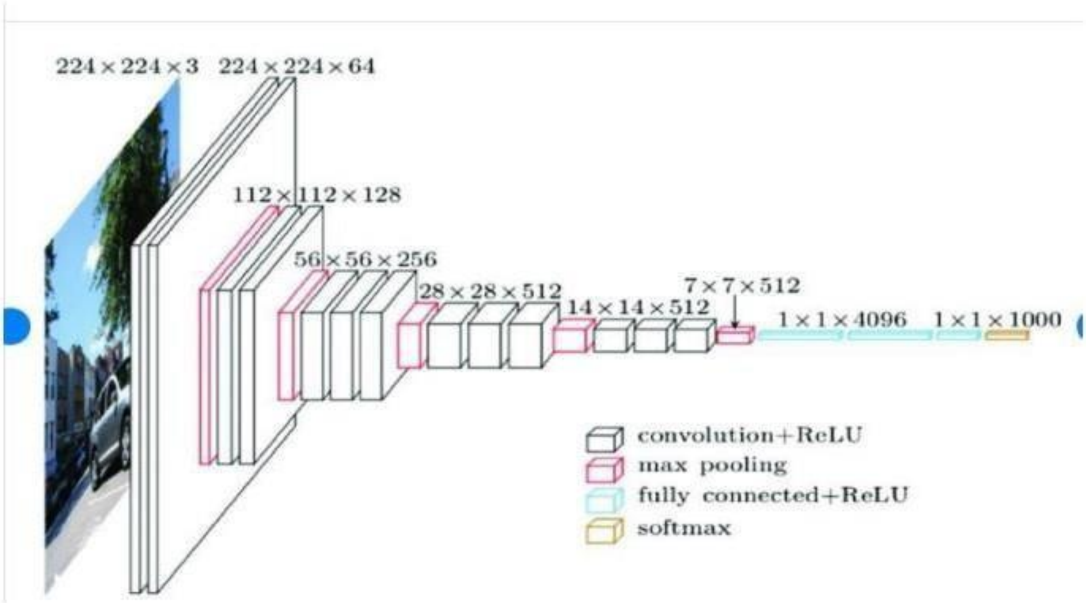


The output of the pooling layer is flattened and this flattened matrix is fed into the Fully Connected Layer. In the fully connected layer there are many layers, Input layer, Hidden layer and Output layers are parts of it. Then this output is fed into the classifier, in this case Softmax Activation Function is used to classify the image into healthy or a pcos with a particular health if present. Figure 4.13 shows the Fully connected layer and Output Layer

Various Convolutional Neural Network Models

We have made use of various CNN models to detect pcos ultrasound image and we have compared these models. We have discussed the different models that we used in our project. The architectures of these models are fixed and these can be utilized for various applications.

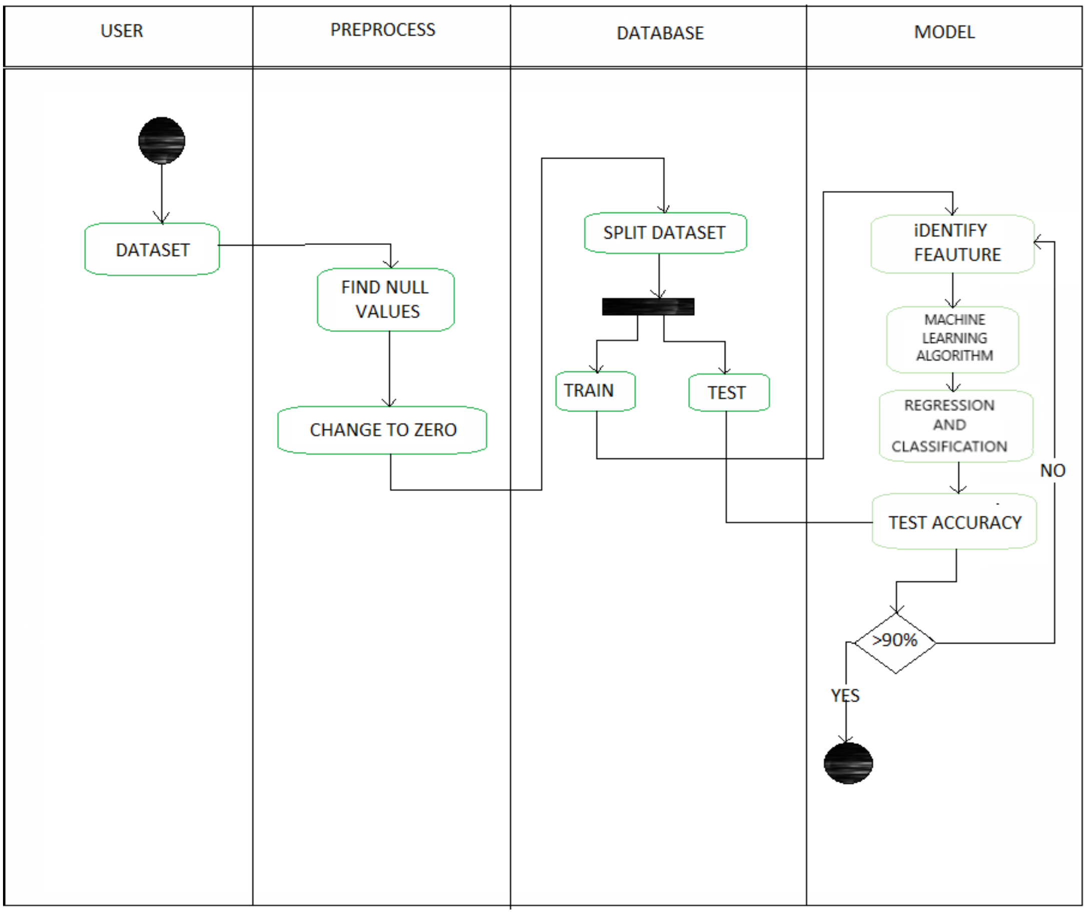
CNN Model



The first two layers are convolutional layers with 3*3 filters, and first two layers use 64 filters that results in 224*224*64 volume as same convolutions are used. The filters are always 3*3 with stride of 1

- After this, pooling layer was used with max-pool of 2*2 size and stride 2 which reduces height and width of a volume from 224*224*64 to 112*112*64.
- This is followed by 2 more convolution layers with 128 filters. This results in the new dimension of 112*112*128.
- After pooling layer is used, volume is reduced to 56*56*128.
- Two more convolution layers are added with 256 filters each followed by down sampling layer that reduces the size to 28*28*256.
- Two more stack each with 3 convolution layer is separated by a max-pool layer.
- After the final pooling layer, 7*7*512 volume is flattened into Fully Connected (FC) layer with 4096 channels and softmax output of 1000 classes

4.6 STRUCTURAL CHART MACHINE LEARNING



Dataset Gathering Module:

- **Functionality:** The primary functionality of the Dataset Gathering module is to acquire essential data required for women health classification and baby weight prediction. It involves sourcing data from various maternal and women parameters, such as age, weight gain, blood pressure, and ultrasound measurements. The module ensures a diverse and comprehensive dataset that serves as the foundation for subsequent analysis.

- **Processing:** The processing within this module includes initiating data acquisition processes, which may involve interfacing with databases or external sources. It details the systematic collection of relevant parameters, ensuring data integrity, and storing the acquired dataset for subsequent preprocessing.

Preprocessing Module:

- **Functionality:** The Preprocessing module is designed to enhance the quality and readiness of the acquired dataset for machine learning. It includes steps such as handling missing values, addressing outliers, and normalizing data to a standardized scale. The module's functionality is crucial for ensuring the reliability and consistency of the dataset.
- **Processing:** Processing within this module involves executing algorithms and methods to clean and transform the dataset. It encompasses techniques for identifying and handling missing or irregular data points, as well as normalizing numerical features to ensure uniformity across the dataset.

Feature Selection Module:

- **Functionality:** The Feature Selection module aims to identify and retain the most relevant features for subsequent model training. It involves exploring different techniques, such as filtering, wrapper methods, or embedded methods, to ensure that the model is trained on discriminative and impactful features.
- **Processing:** The processing steps include the application of feature selection algorithms and methods to the preprocessed dataset. It details the criteria for selecting features, reducing dimensionality if necessary, and creating a refined feature set that contributes meaningfully to the model's predictive capabilities.

Model Training Module:

- **Functionality:** The Model Training module is tasked with selecting an appropriate machine learning algorithm and training the model on the preprocessed dataset. It encompasses the functionality of developing a predictive model capable of understanding patterns within the data.
- **Processing:** Processing in this module involves the selection of suitable algorithms based on the dataset characteristics and desired outcomes. It includes the training phase, during which the model learns from the preprocessed dataset, adjusting its parameters to optimize predictions for future instances.

Evaluating Model Module:

- **Functionality:** The Evaluating Model module assesses the performance of the trained model by employing various metrics such as accuracy, precision, recall, and F1 score. It gauges the model's effectiveness in making predictions on new data.
- **Processing:** Processing within this module involves executing the evaluation metrics on the model's predictions. It details the testing phase, during which the model's performance is measured against ground truth data, providing insights into its accuracy and reliability.

CHAPTER 5

IMPLEMENTATION

5.1 Implementation requirements

- **Programming language selection**

The choice of programming language plays a pivotal role in the successful implementation of the prenatal care system. After careful consideration of various factors, Python has been selected as the primary programming language for this project. Python's versatility and extensive ecosystem make it an ideal choice for machine learning applications. Its rich libraries, particularly in the form of NumPy, Pandas, and Scikit-learn, provide robust tools for data manipulation, preprocessing, and model development. Python's readability and simplicity enhance code maintainability, crucial for a project with multiple modules and iterative development.

The integration of Flask, a Python web framework, aligns seamlessly with the local host deployment requirements of the project. Flask enables the creation of a lightweight and efficient web application to present the machine learning models to healthcare professionals. The use of HTML, CSS, and JavaScript for the frontend further complements the Python backend, offering a well-rounded and interactive user interface. This programming language synergy ensures a cohesive and integrated implementation, fostering a smooth flow between data processing, model training, and user interaction.

The selection of Python is also strategic in the context of the extensive support and community engagement it enjoys in the field of machine learning and data science. The abundance of online resources, tutorials, and a vibrant community facilitates troubleshooting, accelerates development cycles, and ensures that the project remains at the forefront of technological advancements in the rapidly evolving domain of healthcare technology. Overall, the choice of Python aligns with the project's goals of efficiency, scalability, and accessibility in the implementation phase.

5.2 Key features of programming language selected

Python, the selected programming language for this project, brings forth a multitude of key features that align seamlessly with the complex requirements of implementing a robust prenatal care system. One of Python's standout features is its readability and simplicity, making it an exceptionally expressive language. This characteristic is invaluable in the context of a project with multiple modules, intricate data processing, and the implementation of machine learning algorithms. The clean and concise syntax of Python promotes code clarity, easing the development process and enhancing overall project maintainability.

Another crucial feature is Python's extensive ecosystem of libraries and frameworks tailored for machine learning and data science applications. Libraries such as NumPy and Pandas offer powerful tools for efficient data manipulation and preprocessing, essential in handling diverse maternal and women health parameters. Scikit-learn, a prominent machine learning library, provides a wide array of algorithms and functionalities, streamlining the development of predictive models for women health classification and baby weight prediction. This rich ecosystem significantly expedites development cycles, enabling the project to leverage pre-existing tools and focus on the unique challenges posed by prenatal care.

Python's versatility extends to its compatibility with Flask, a lightweight and modular web framework chosen for the local host deployment of the project. This integration empowers the system with a user-friendly interface, seamlessly combining the strengths of Python for backend processing with HTML, CSS, and JavaScript for frontend interactions. Python's adaptability across different domains, from data manipulation to web development, positions it as an ideal choice for a project requiring a cohesive integration of machine learning capabilities within a user-accessible interface. In summary, Python's readability, extensive libraries, and versatility make it the cornerstone for the successful implementation of the prenatal care system.

5.3 Description about tools, GUI used

The project leverages a carefully selected set of tools and a Graphical User Interface (GUI) to enhance the efficiency and accessibility of the prenatal care system. The Flask web framework serves as the backbone for the local host deployment, providing a robust foundation for handling HTTP requests, routing, and rendering dynamic web pages. Flask's lightweight nature aligns with the project's requirements, facilitating a responsive and scalable user interface. Additionally, Flask integrates seamlessly with Python, the chosen programming language, fostering a cohesive environment for backend development.

The GUI is crafted using a combination of HTML, CSS, and JavaScript, ensuring an interactive and visually appealing user experience. HTML defines the structure of the web pages, CSS governs their presentation and styling, and JavaScript adds dynamic behavior to enhance user interactions. This trifecta of frontend technologies facilitates the creation of an intuitive and user-friendly interface for healthcare professionals interacting with the prenatal care system. The modular design of the GUI aligns with the project's overall approach, allowing for a clear separation of concerns between the frontend and backend components.

Furthermore, the choice of tools extends to the integration of libraries such as NumPy, Pandas, and Scikit-learn within the Python ecosystem. NumPy and Pandas empower efficient data manipulation and preprocessing, ensuring the dataset's readiness for machine learning analysis. Scikit-learn provides a comprehensive set of tools for model development, training, and evaluation. This toolset enhances the project's capabilities, allowing for a seamless integration of machine learning functionalities within the GUI. Collectively, the chosen tools and GUI components contribute to a well-rounded and effective prenatal care system, emphasizing user accessibility, system responsiveness, and modular design.

CHAPTER 6

SYSTEM TESTING

Testing Methods

There are different methods that can be used for software testing. They are,

6.1 Black-Box Testing

The technique of testing without having any knowledge of the interior workings of the application is called black-box testing. The tester is oblivious to the system architecture and does not have access to the source code. Typically, while performing a black-box test, a tester will interact with the system's user interface by providing inputs and examining outputs without knowing how and where the inputs are worked upon.

6.2 White-Box Testing

White-box testing is the detailed investigation of internal logic and structure of the code. White-box testing is also called glass testing or open-box testing. In order to perform white-box testing on an application, a tester needs to know the internal workings of the code. The tester needs to have a look inside the source code and find out which unit/chunk of the code is behaving inappropriately.

6.3 Levels of Testing

There are different levels during the process of testing. Levels of testing include different methodologies that can be used while conducting software testing.

The main levels of software testing are:

- Functional Testing:

This is a type of black-box testing that is based on the specifications of the software that is to be tested. The application is tested by providing input and then the results are examined that need to conform to the functionality it was intended for. Functional testing of software is conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. There are five steps that are involved while testing an application for functionality.

- The determination of the functionality that the intended application is meant toper form.
- The creation of test data based on the specifications of the application.
- The output based on the test data and the specifications of the application.
- The writing of test scenarios and the execution of test cases.
- The comparison of actual and expected results based on the executed test cases.

- **Non-functional Testing**

This section is based upon testing an application from its non-functional attributes. Non-functional testing involves testing software from the requirements which are non-functional in nature but important such as performance, security, user interface, etc. Testing can be done in different levels of SDLC. Few of them are

6.4 Unit Testing

Unit testing is a software development process in which the smallest testable parts of an application, called units, are individually and independently scrutinized for proper operation. Unit testing is often automated but it can also be done manually. The goal of unit testing is to isolate each part of the program and show that individual parts are correct in terms of requirements and functionality. Test cases and results are shown in the Tables.

TEST CASE ID	DESCRIPTION	INPUT DATA	EXPECTED DATA
TC01	Normal case with clear symptoms	Age:25, BMI:30, Irregular periods: yes, Hyperandrogenism: yes, ultrasound: yes	PCOS Detected
TC02	Normal case without symptoms	Age:22, BMI:22, Irregular periods: no, Hyperandrogenism: no, ultrasound: no	No PCOS Detected
TC03	Borderline case	Age:28, BMI:25, Irregular periods: yes, Hyperandrogenism: no, ultrasound: no	Further Investigation needed
TC04	Case with high BMI but no other symptoms	Age:30, BMI:35, Irregular periods: no, Hyperandrogenism: no, ultrasound: no	No PCOS Detected
TC05	Young patient with some symptoms	Age:18, BMI:27, Irregular periods: yes, Hyperandrogenism: no, ultrasound: yes	PCOS Detected
TC06	High risk case with all symptoms	Age:24, BMI:32, Irregular periods: yes, Hyperandrogenism: yes, ultrasound: yes	PCOS Detected
TC07	Case with PCOS but patient is under medication	Age:25, BMI:27, Irregular periods: no, Hyperandrogenism: yes, ultrasound: yes	PCOS Detected
TC08	Case with other endocrine disorder	Age:31, BMI:28, Irregular periods: yes, Hyperandrogenism: yes, ultrasound: no	No PCOS Detected

Table 6.5: Test Cases

CHAPTER 7

RESULT

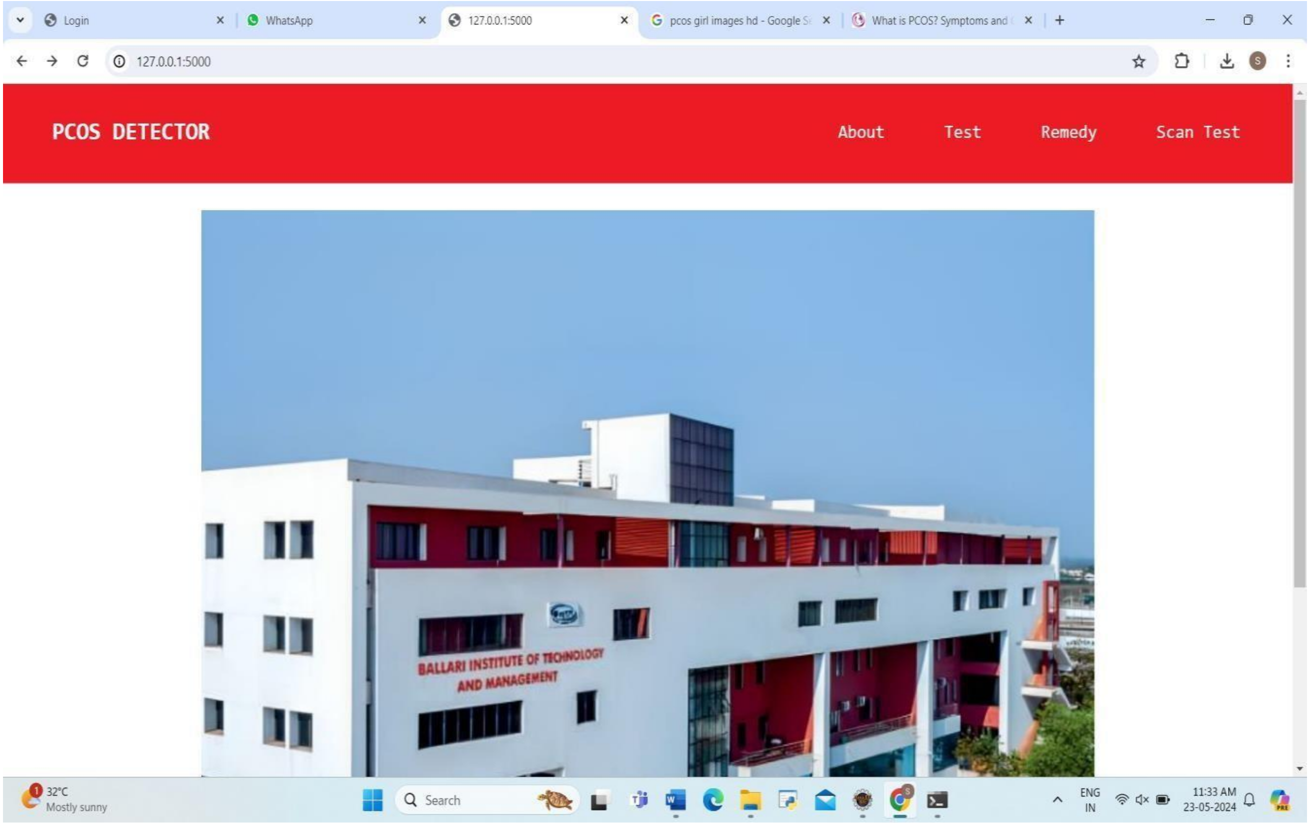


Fig 7.1: Display page

Description of the figure 7.1

The above figure shows the web interfaces where user can view the number of sections available in the project.

- About
- Test
- Remedy
- Scan test

It shows all the section where user can go through every section to view the details as per the interest what they want to see.

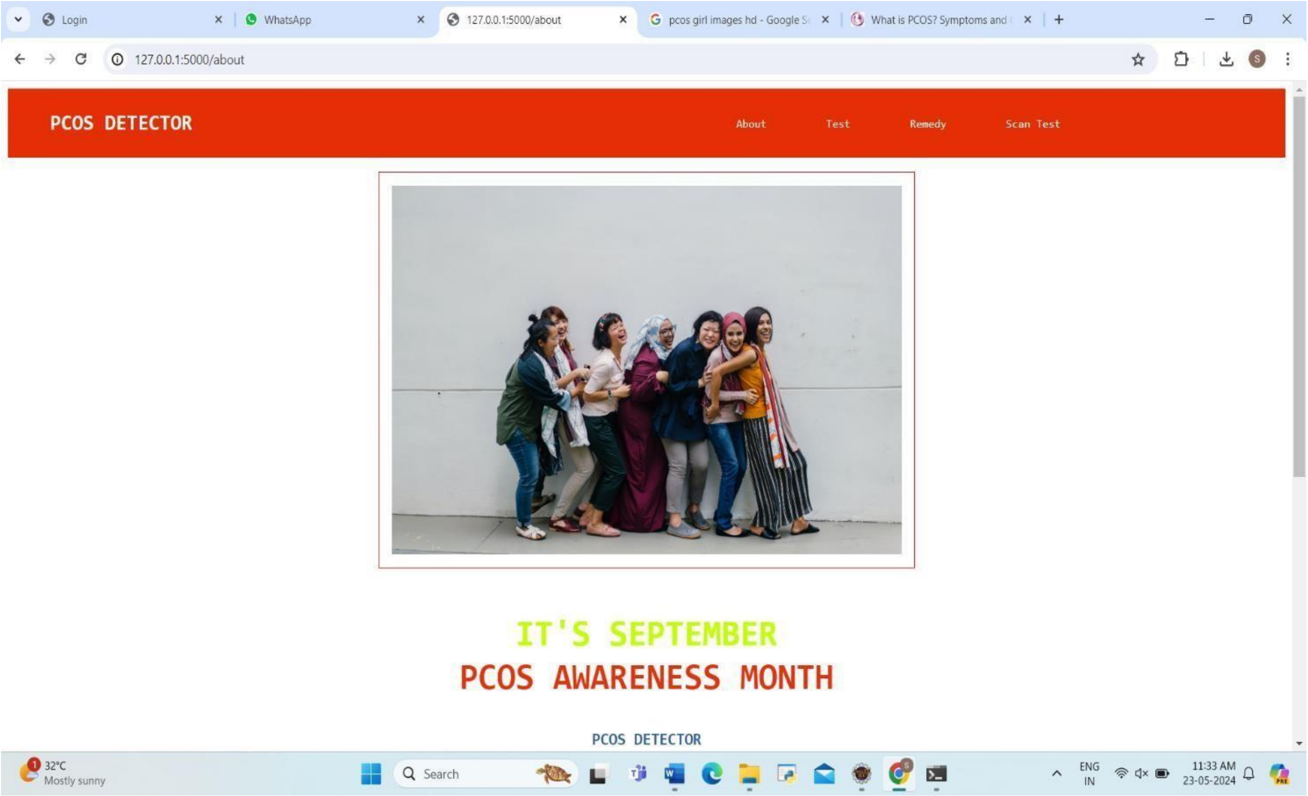


Fig 7.2: PCOS Awareness page

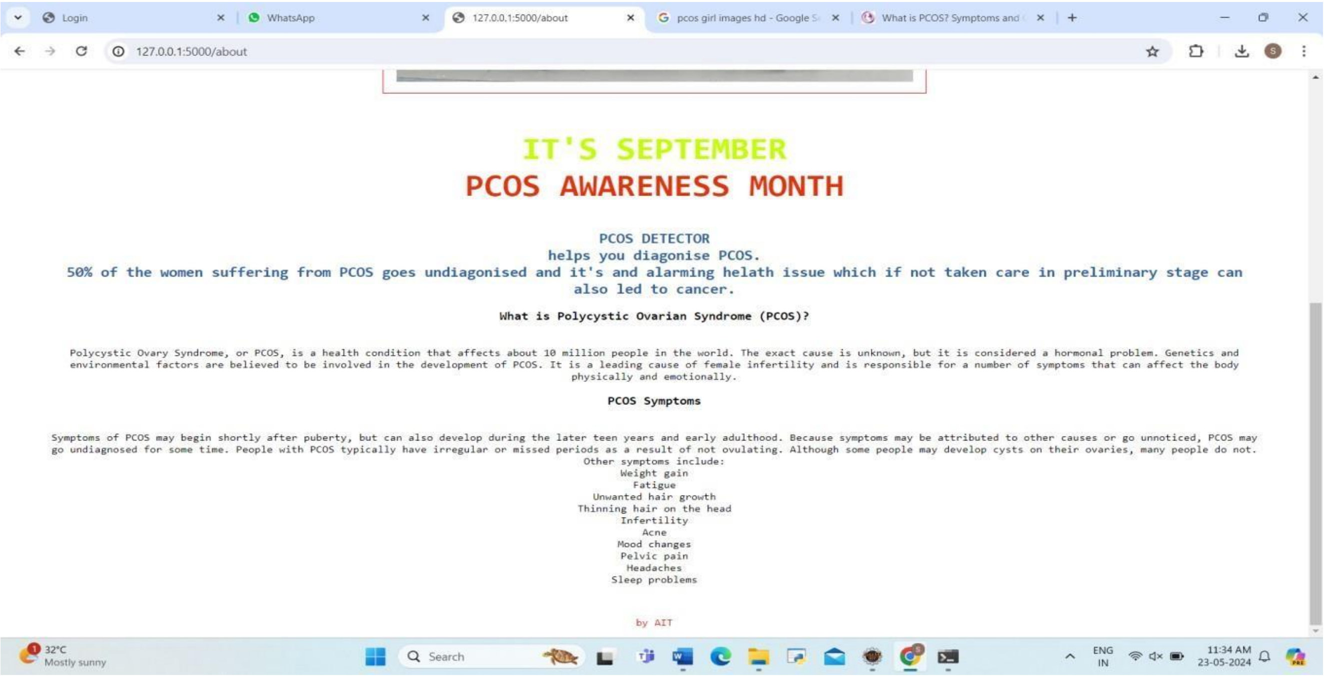


Fig 7.3: Detailed description of PCOS

Description of figure 7.2, 7.3

The main content is about PCOS (Polycystic Ovary Syndrome) Awareness Month, Which is in September. The text explains what PCOS is, mentions that many women go undiagnosed, and lists symptoms of PCOS. The screenshot at the center shows a webpage with details about PCOS.

The bottom slide shows additional questions related to PCOS symptoms, such as

- difficulty losing weight
- mood swings
- anxiety
- irregular sleep patterns
- weight gain
- hair growth and
- skin darkening. The design features an orange gradient background with an alert that PCOS is a women's health issue.

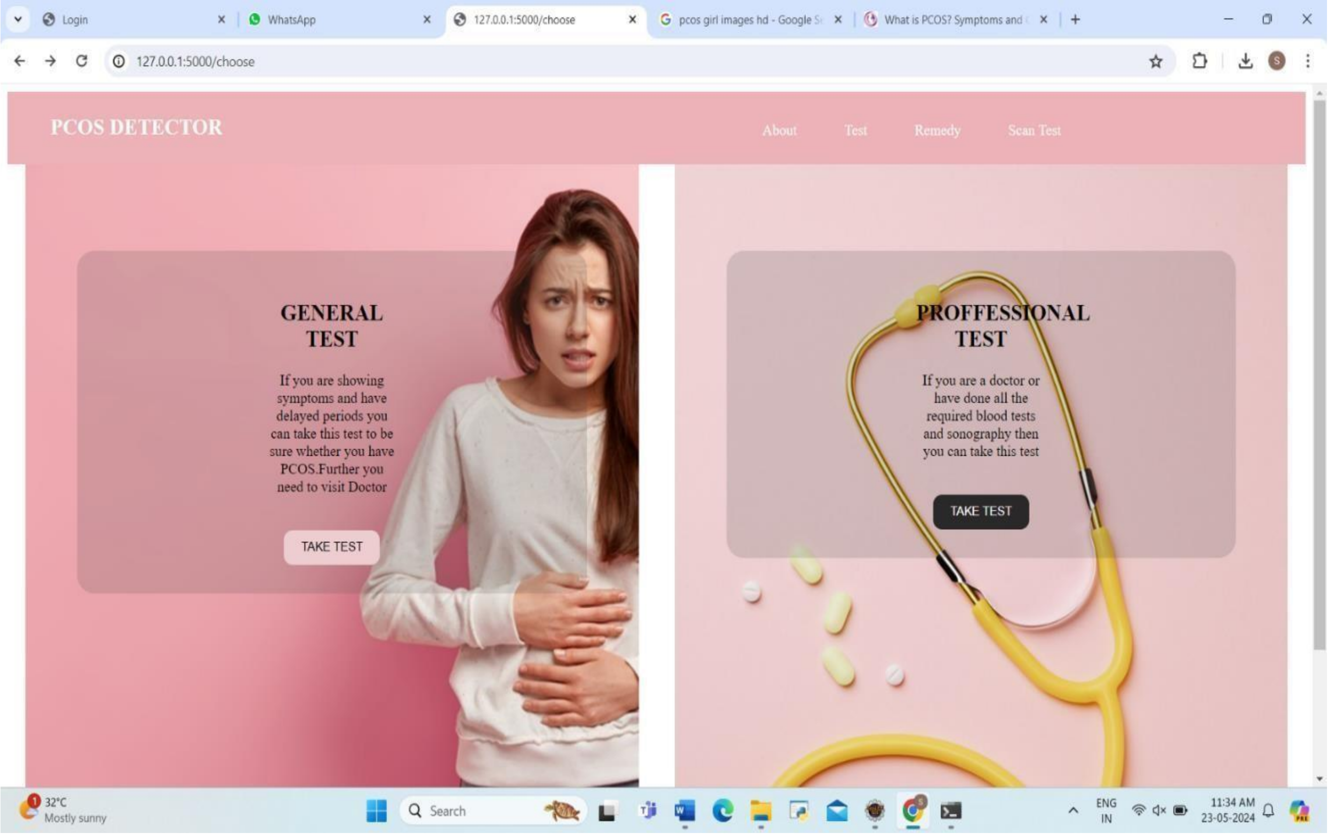


Fig 7.4: Test detector page

Description 7.4

This page is designed for the type of test a patient can take depending on the doctor's prescription after they conduct the test. The test can of two types i.e.,

- General test
- Professional test

General test can be taken in the absence of doctor under the guidance of any nurses or any compounder by taking some of the medial related data.

Professional test should be taken under the guidance of doctor because it contains some specific values which can be known only by the doctor about the particular patient who are suffering from PCOS.

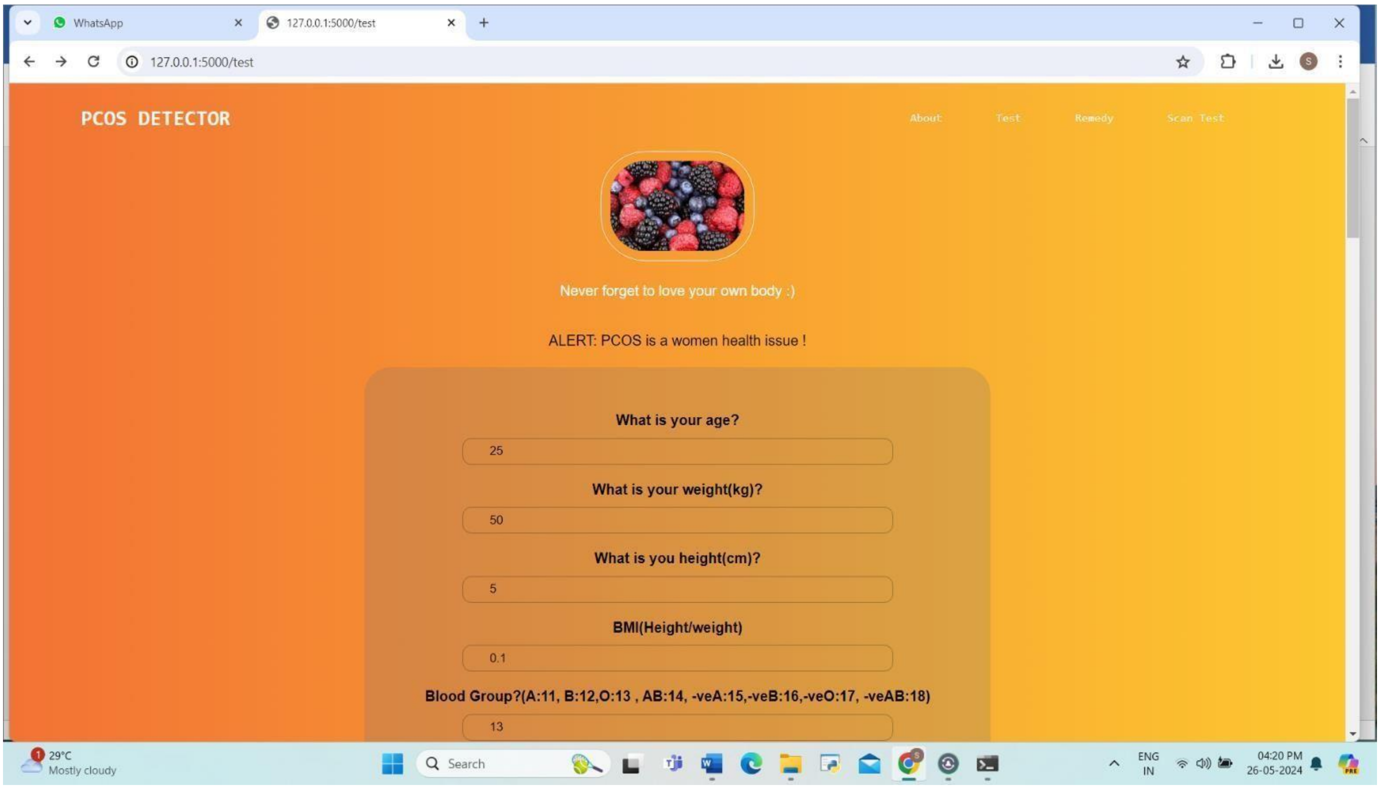


Fig 7.4.1: Data set values page

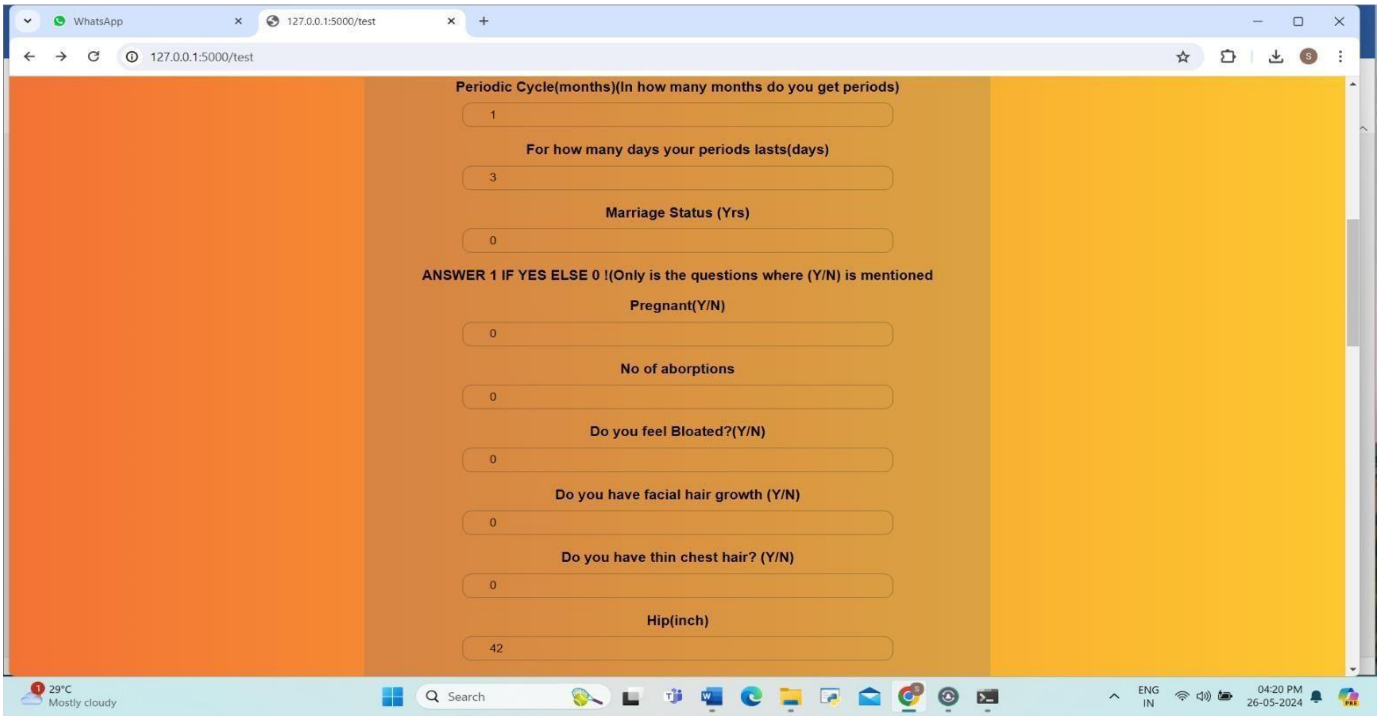


Fig 7.4.2: Data set values page

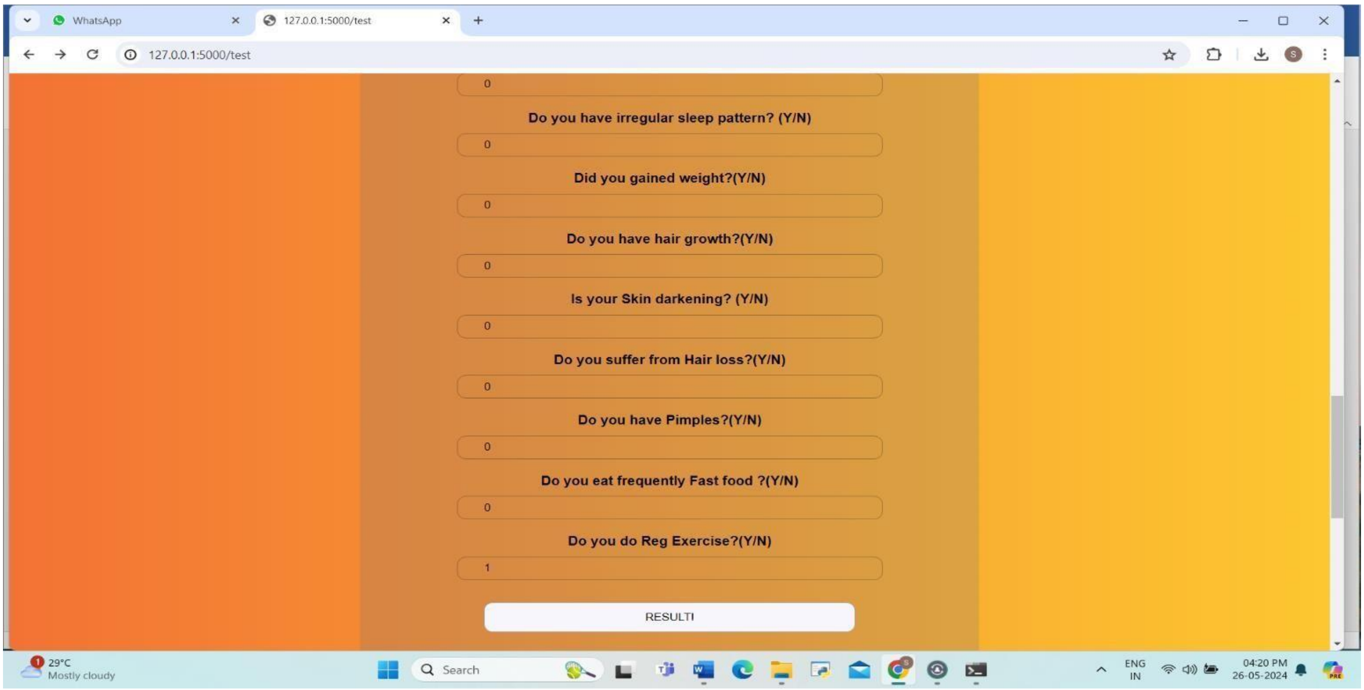


Fig 7.4.3: Data set values

Description 7.4.1, 7.4.2, 7.4.3

The above three images are based on the general test when a particular patient takes the test after concerning to the doctor. It takes some values that need to be entered to take the test like

- age, weight, height, h/w ratio
- periodic cycle,
- marriage status
- pregnant
- hair growth
- skin darkening

hair loss and other related values so that it can detect whether a particular individual is sufferinfg from PCOS or not.

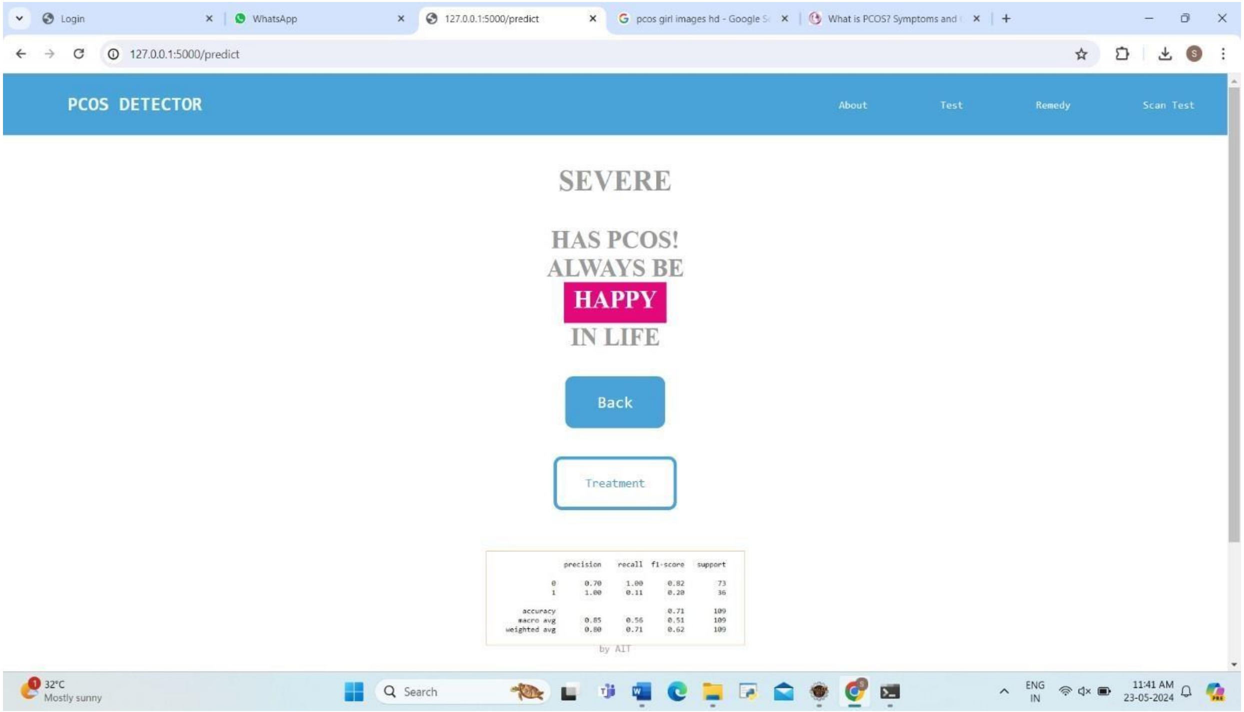


Fig 7.5.1: Result from the general test

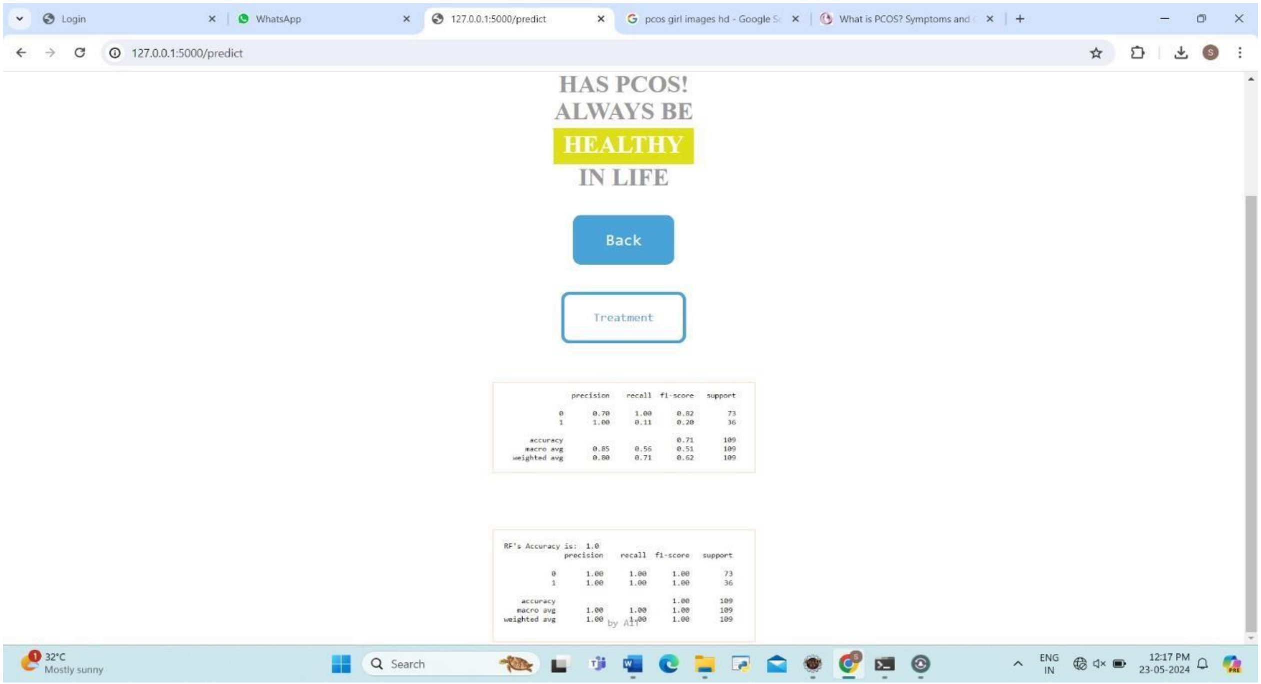


Fig 7.5.2: Result from the general test showing the data set values

Description of the figures 7.5.1, 7.5.2

Both the images show the result that we have given in the general test. The images display showing the result as severe according to the data values provided of a particular patient. It displays about the treatment if a patient is willing to take or else, she can go back and see the report which has been generated after taking the test and it gives predicted data, accuracy rate and from how much score she is suffering from PCOS.

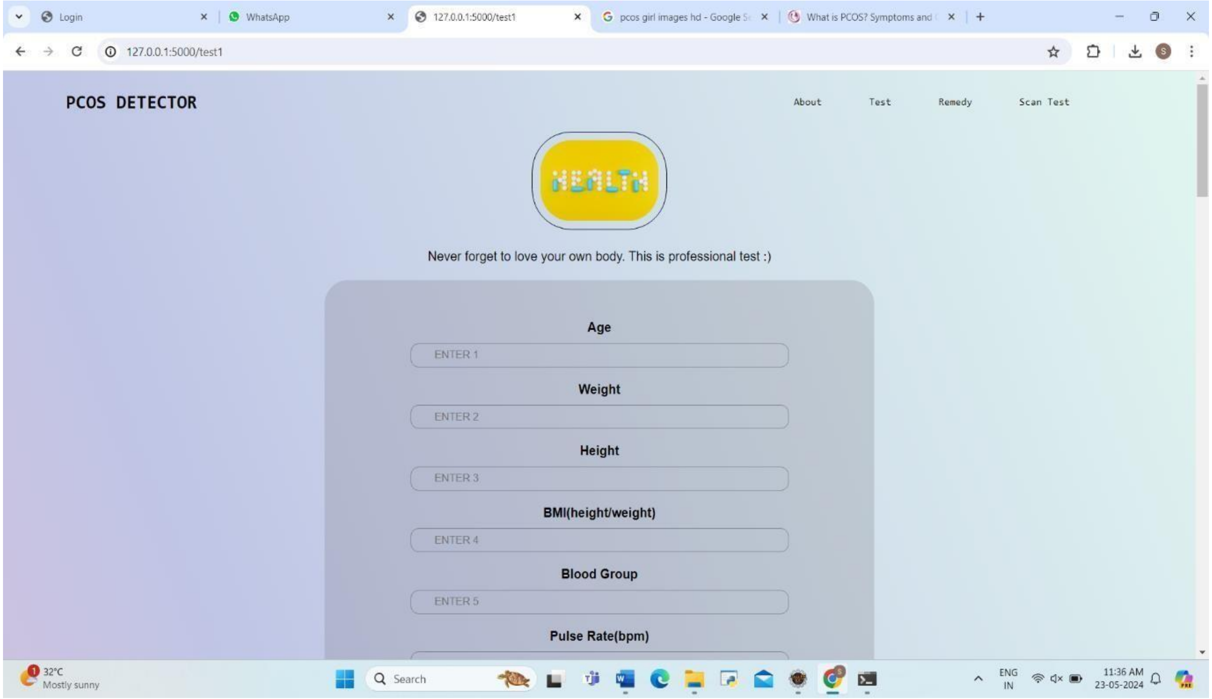


Fig 7.6.1: Professional test of data set values

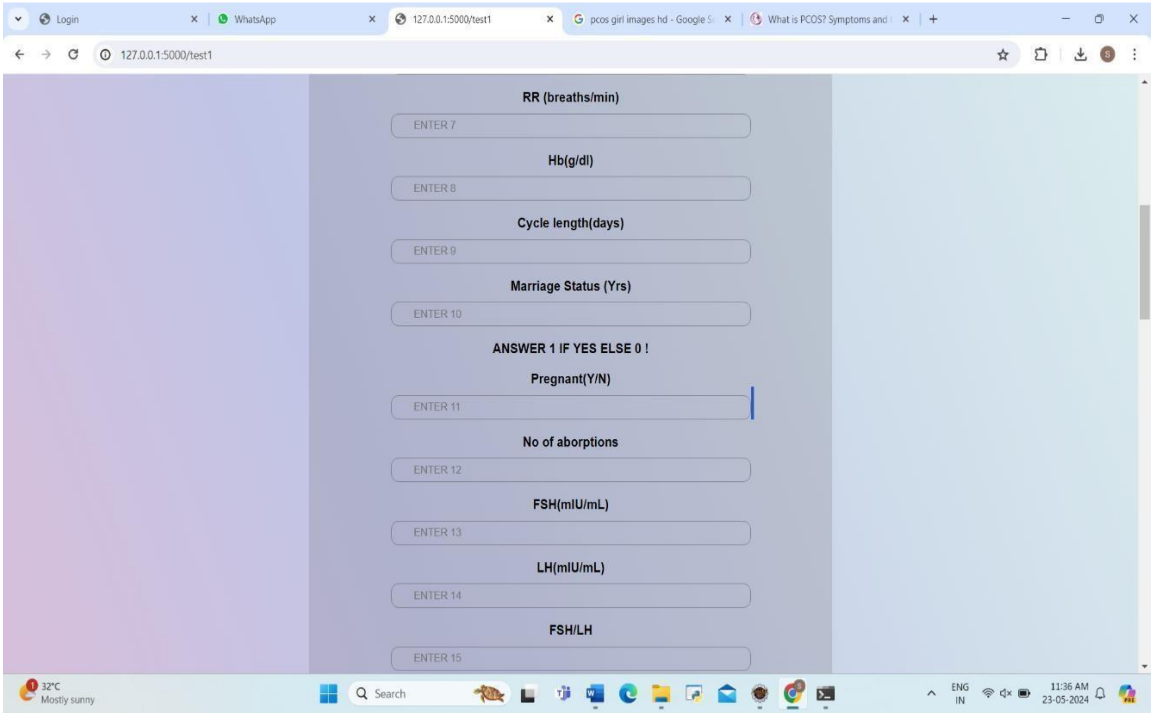


Fig 7.6.2: Professional test of data set values

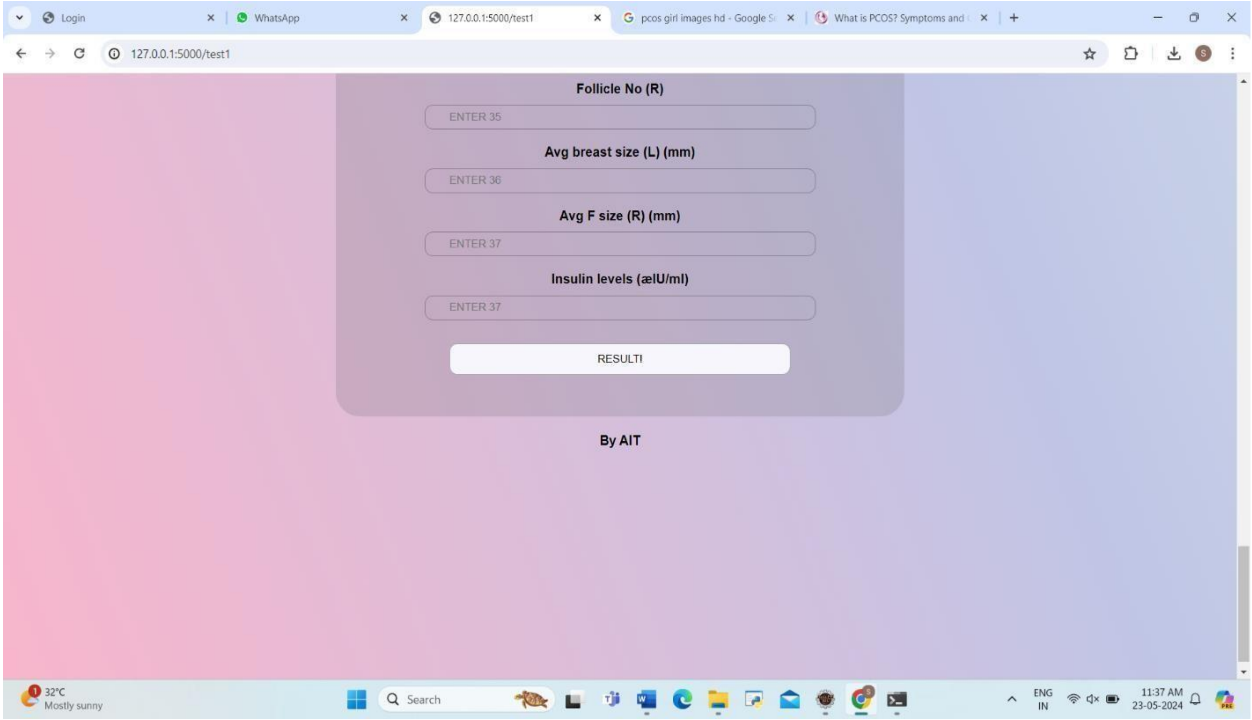


Fig 7.6.3: Professional test of data set values

Description of the images 7.6.1, 7.6.2, 7.6.3

The above three images are for the professional test that should be taken under the guidance of doctor. It takes the values so that it can be helpful to get the accurate result. The values are like

- FSH
- LH
- FSH/LH ratio,
- cycle length
- Follicle number
- Insulin levels and other related values that is required for the test.

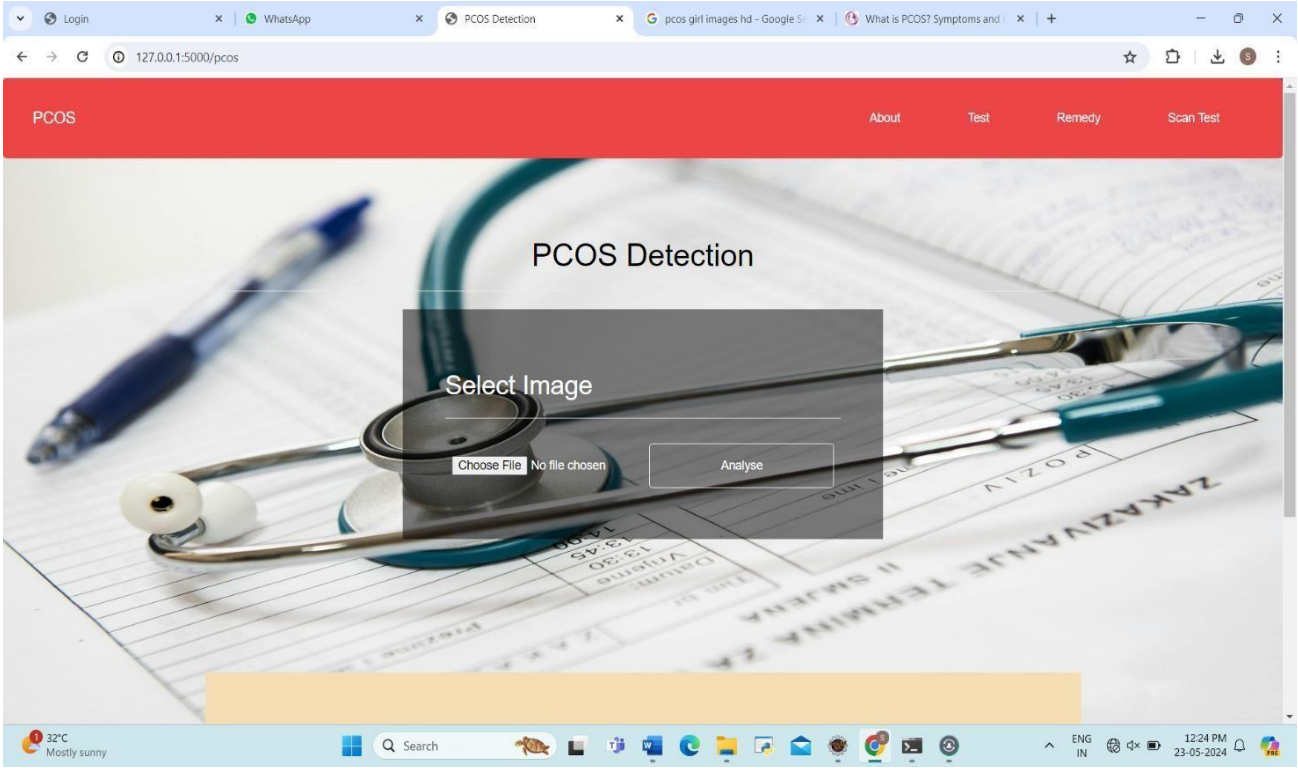


Fig 7.7: Scan test page

Description of the figure 7.7

The above image is web interface which allows the user to select the images to detect the PCOS. After clicking on the choose file it will direct the user to select an image from a specific folder where user have stored the scan image of a patient whether it can be positive or negative that can be helpful in analyzing the accuracy for the percent a particular patient is suffering from PCOS.

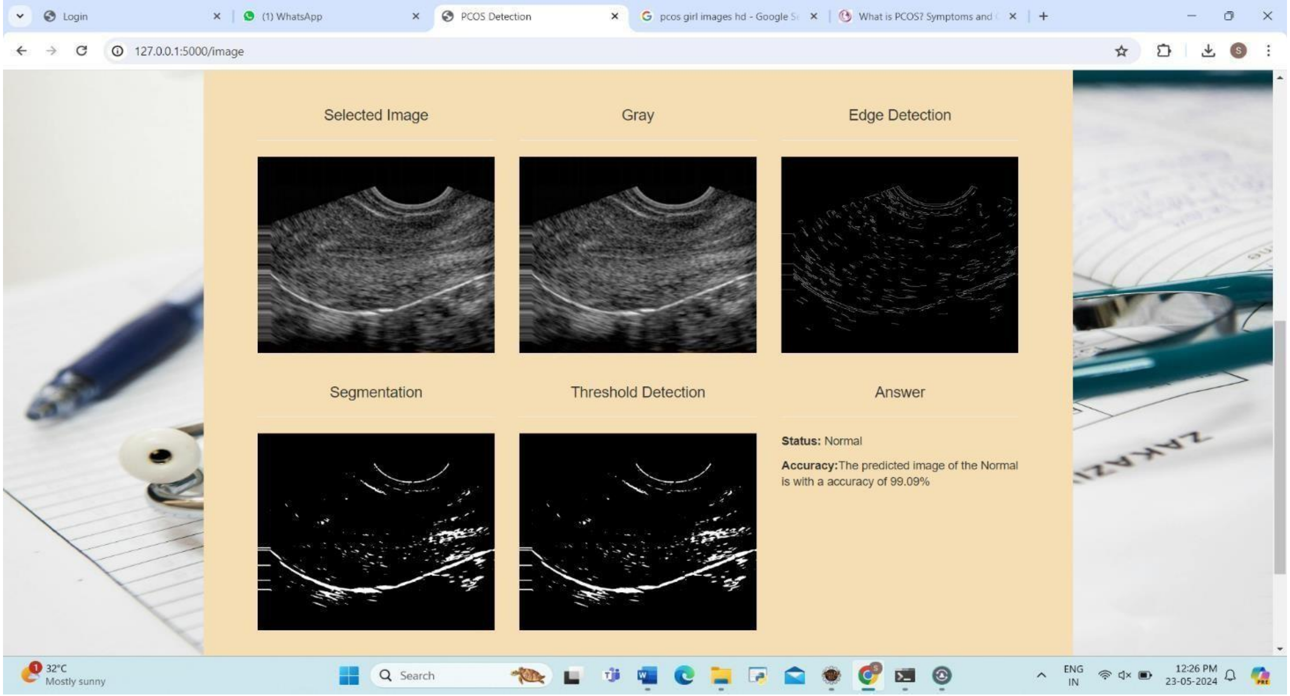


Fig 7.8: Scanned images showing result as normal

Description of the figure 7.8

The above figure shows the steps involved to detect the accuracy of a patient is having PCOS. It involved the steps like first it will select the scanned image then covert it to grey image where processing of the images can be done easily then it does edge detection edge detection is usually done for segmentation and finding the threshold value for the detection of the selected image after undergoing all the above steps it will give the status as normal if the particular patient is not suffering from PCOS. It will detect based on the CNN algorithm that uses images to predict the output.

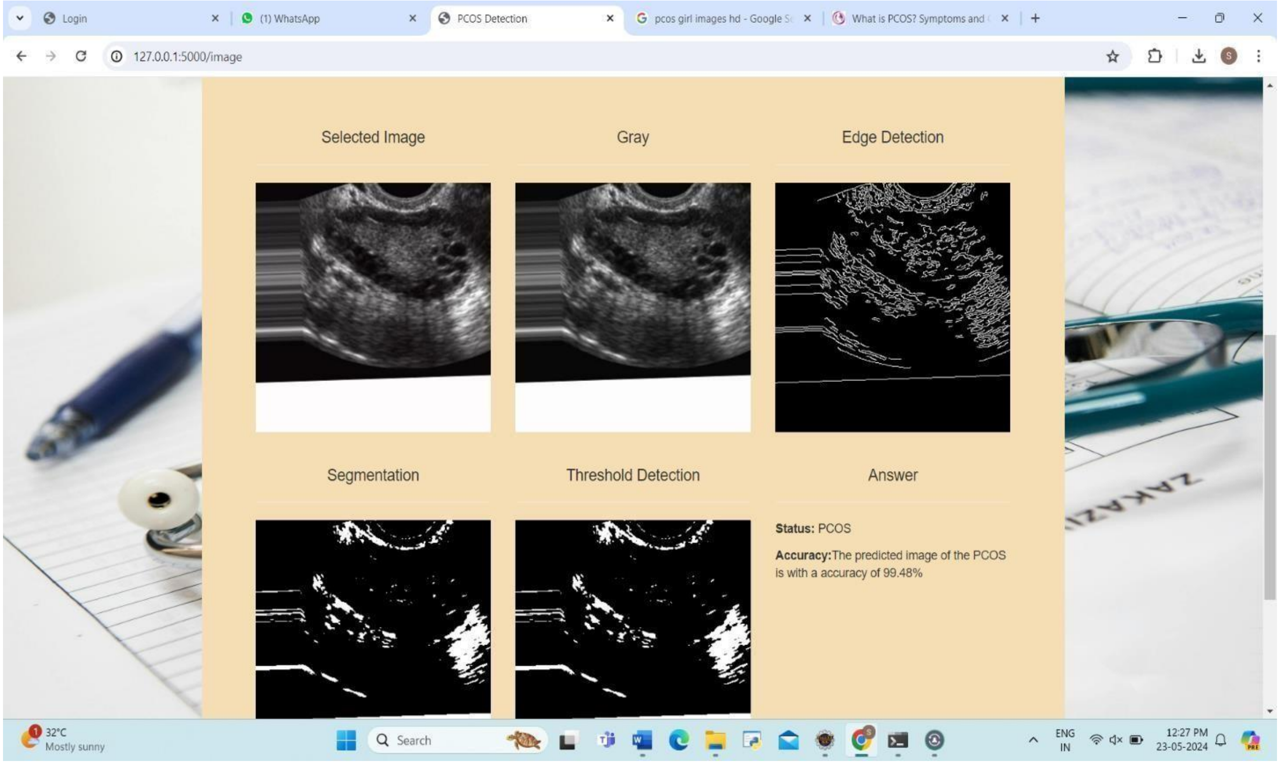


Fig 7.9: Scanned images showing patient suffering from PCOS

Description of the figure 7.9

The above figure shows the steps involved to detect the accuracy of a patient is having PCOS. It involved the steps like first it will select the scanned image then convert it to grey image where processing of images can be done easily then it does detection edge detection is usually done for segmentation and finding the threshold value for the detection of the selected image after undergoing all the above steps it will give the status as PCOS as the patient is suffering from PCOS if the particular. It will detect based on the CNN algorithm images to predict the output.

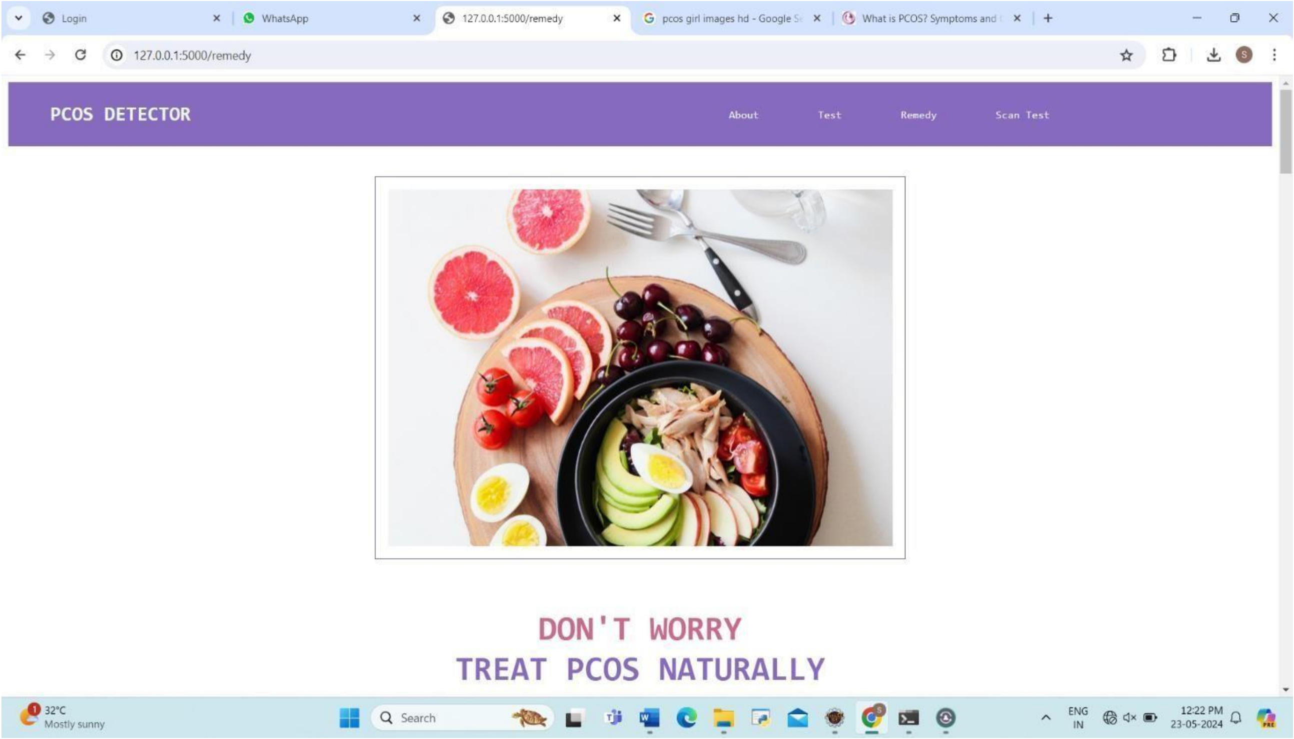


Fig 7.10: Remedy page

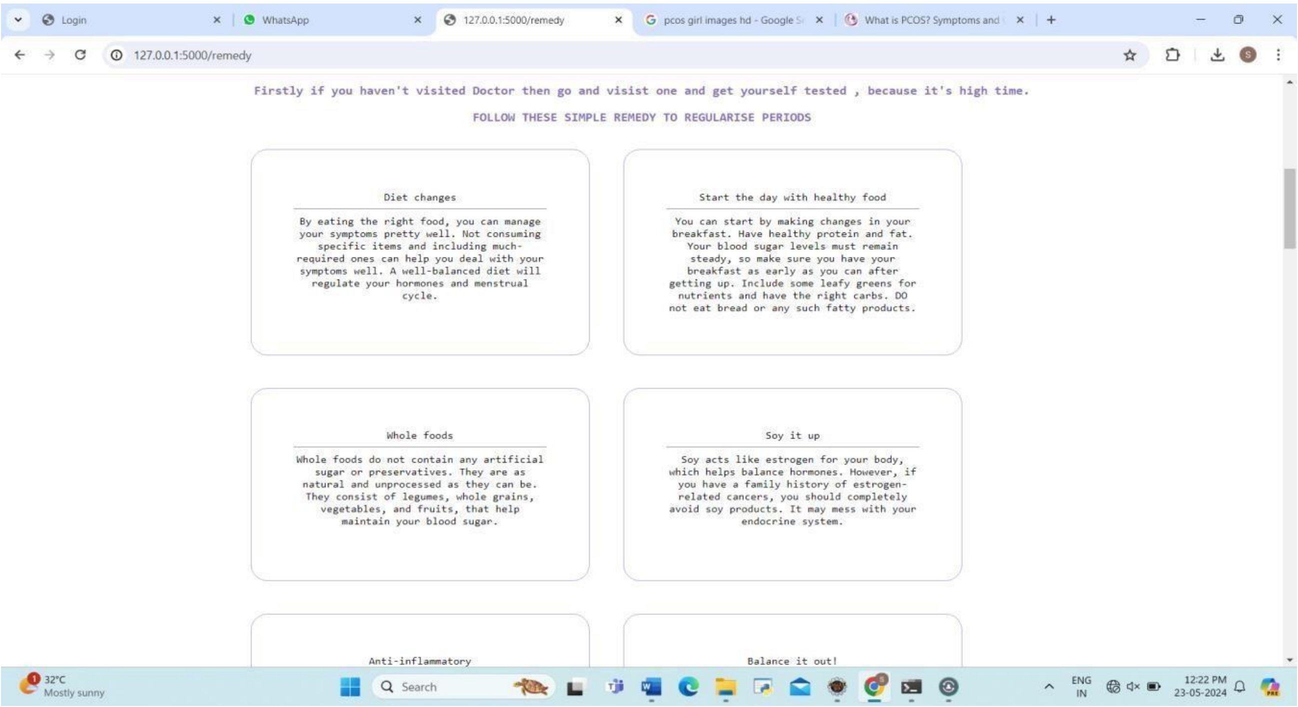


Fig 7.11: Food remedy page

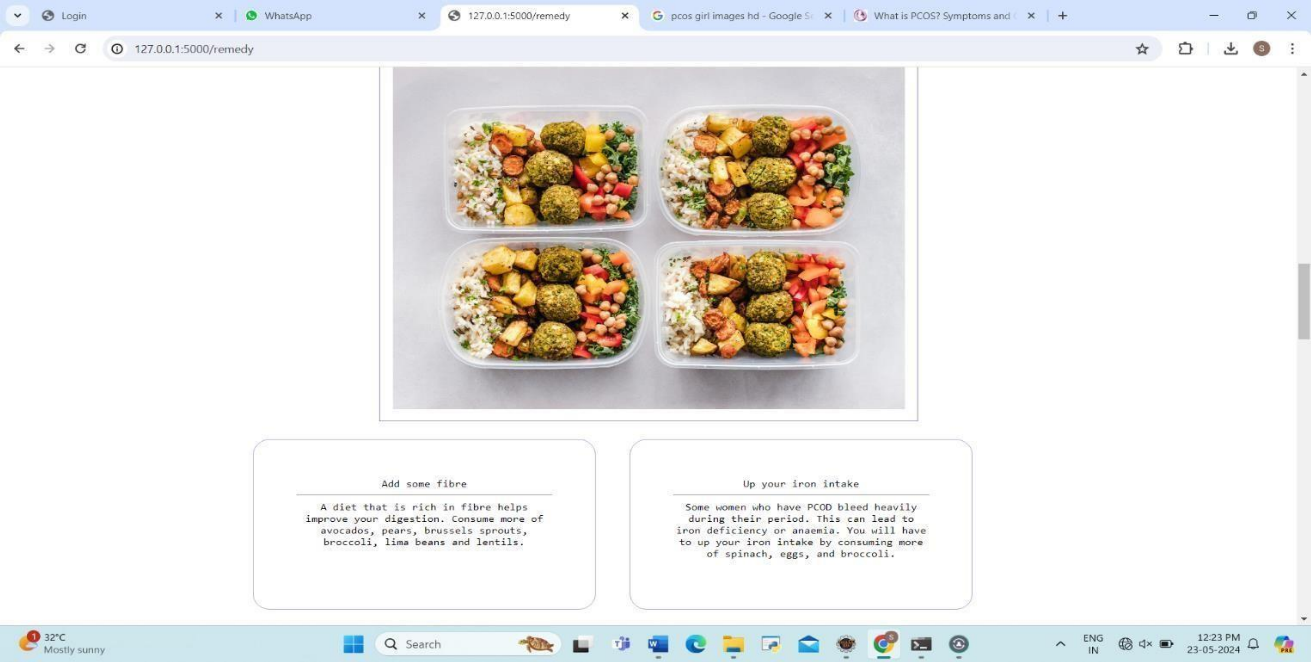


Fig 7.12: Food diet page

Description 7.10, 7.11, 7.12

The above three figures describe about the food and diet to be followed and taken to overcome from PCOS. In this diet it includes all the necessary fruits and vegetables a patient should take to maintain the level of PCOS and also it includes in how much quantity it should be taken and what food items it should be taken so that the level of PCOS can be reduced.

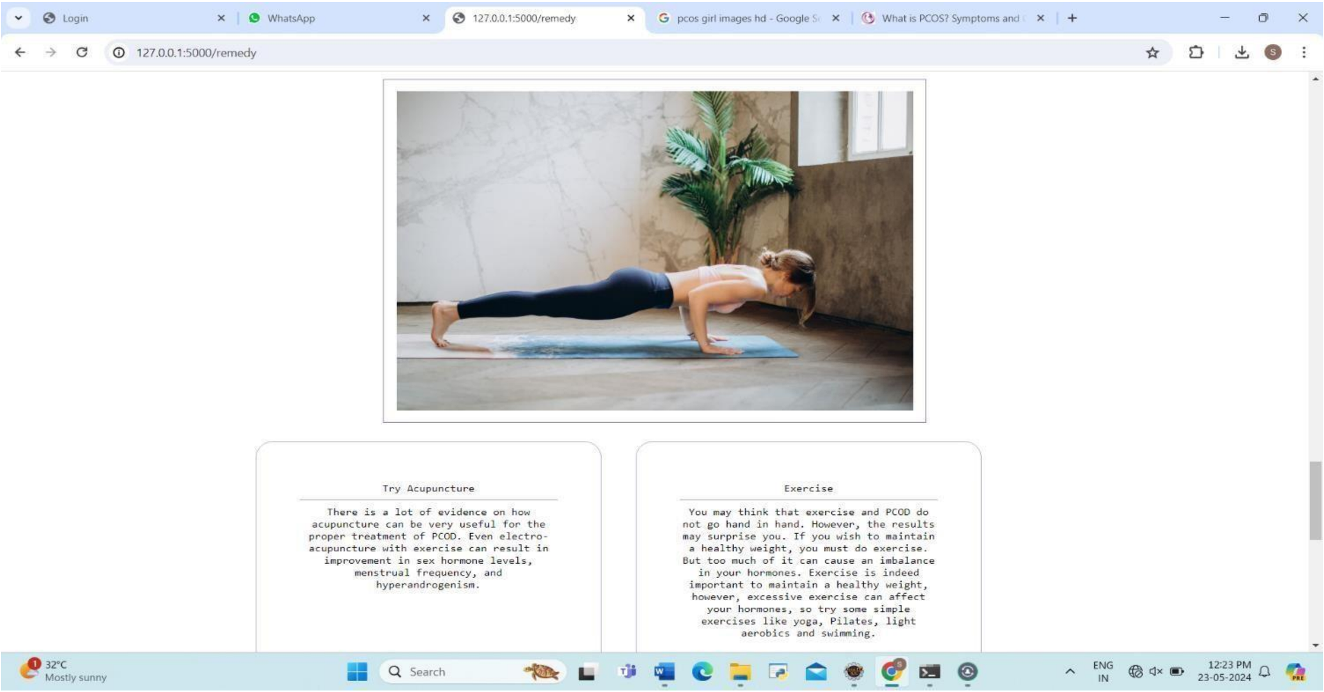


Fig 7.13: Yoga and exercise to be followed

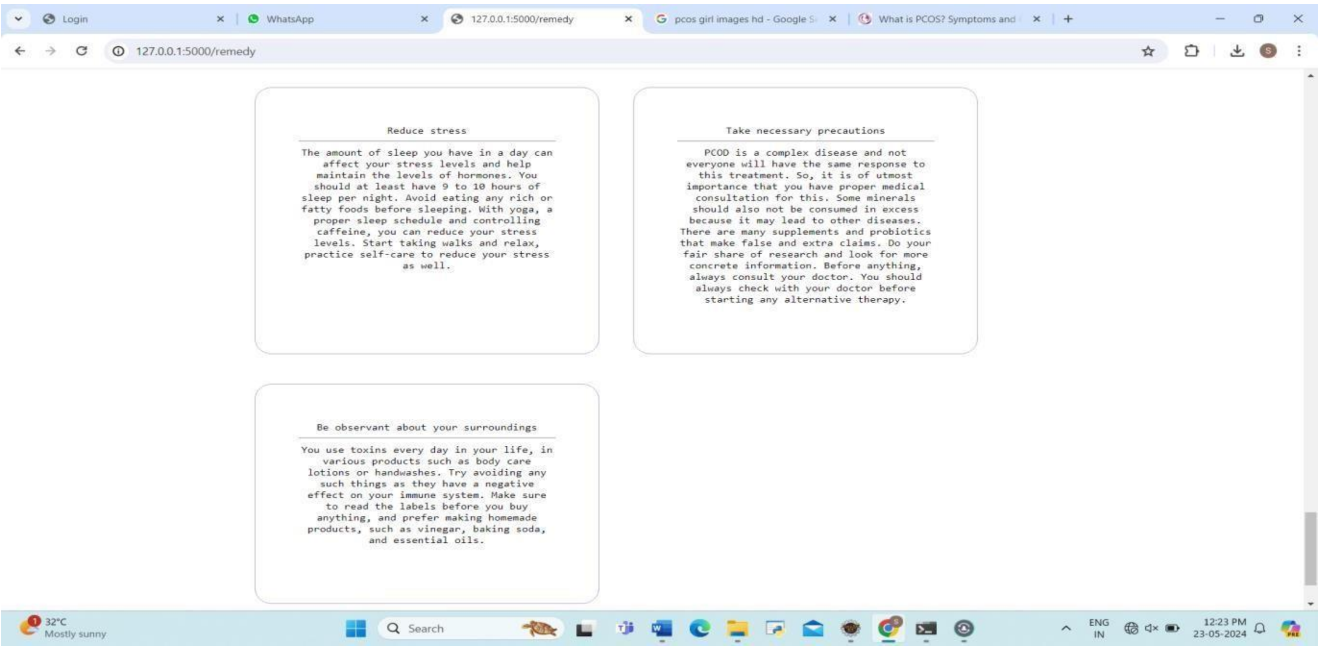


Fig 7.14: Necessary precaution and stress reduction page

Description 7.13, 7.14

The above two figures show which yoga and exercise to be followed to reduced the level of PCOS. It includes all the necessary precautions and remedies to be followed and at what time and. It help to maintain the body balance and look after the necessary steps to be taken to overcome from disease and it gives the remedy to reduce the stress level.

CONCLUSION

In this study, we explored the application of machine learning algorithms, including Random Forest, Decision Tree, and XGBoost, for the detection of Polycystic Ovary Syndrome (PCOS) using a comprehensive set of patient attributes. Through rigorous experimentation and evaluation, we demonstrated the efficacy of these algorithms in accurately classifying individuals as either PCOS positive or negative based on their clinical attributes. Our results indicate promising performance metrics, suggesting that machine learning techniques can serve as valuable tools in the early detection and diagnosis of PCOS.

Furthermore, our analysis revealed the significance of various attributes in predicting PCOS, highlighting the importance of factors such as age, hormonal symptoms (e.g., facial hair, irregular sleep patterns), lifestyle choices (e.g., fast food consumption, exercise habits), and physiological indicators (e.g., BMI, waist-hip ratio). By leveraging these attributes, our models were able to discern patterns and associations indicative of PCOS, thereby aiding in early identification and intervention.

REFERENCES

[1] Ramanathan, Arun, et al. "Detection of PCOS Using Machine Learning Techniques: A Review." International Journal of Advanced Research in Computer Science and Software Engineering 9.4 (2019): 213-220.

[2] Deshmukh, Mrunalini, and Dr. D. S. Bhosale. "A Review on PCOS Detection Techniques using Machine Learning Algorithms." International Journal of Computer Sciences and Engineering 7.12 (2019): 89-93.

[3] Kumar, Rajnish, et al. "Prediction of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms." International Journal of Innovative Technology and Exploring Engineering 9.1 (2019): 1935-1939.

[4] Rajesh, M., et al. "Detection of Polycystic Ovary Syndrome Using Machine Learning Algorithms: A Review." International Journal of Scientific Research in Computer Science, Engineering and Information Technology 5.4 (2020): 73-78.

[5] Ganesan, Kavitha, et al. "Machine Learning Approach for Early Detection of Polycystic Ovary Syndrome." International Journal of Computer Applications 179.36 (2018): 1-5.

[6] Jain, Preeti, and Dr. Sanjeev Kumar. "Detection of PCOS Using Machine Learning Algorithms: A Review." International Journal of Computer Science and Information Technologies 8.3 (2017): 3305-3309.

[7] Jyoti, R., and Dr. D. S. Bhosale. "Machine Learning Approaches for PCOS Detection: A Review." International Journal of Computer Sciences and Engineering 7.8 (2019): 522-526.

BALLARI INSITUTE OF TECHNOLOGY AND MANGEMENT, BALLARI

DEPARTMENT OF COMPUTER SCIENC & ENGINEERING
Project CO-PO Mapping
ACADEMIC YEAR 2023-24

U.S.N.	Student Name	Guide Name	Project Title
3BR20CS156	SHASHABI M B	Mrs. SS STEFFI NIVEDITA (Asst. prof)	PCOS DETECTION Using Machine Learning
3BR20CS158	SHOBHA KUMARI		
3BR20CS166	SWETHA SINGH		
3BR20CS172	UMME KULSM		

COURSE OUTCOMES(CO'S)

Course Outcomes COx	Description of Course Outcomes
CO1	**Identify** the real-world Computer Science Problems.
CO2	**Analyze** the engineering problem requirements.
CO3	**Design** the solution methodologies for the problem.
CO4	**Apply** modern engineering tools/techniques for developing a system.
CO5	**Write** technical project report and publish the thesis into an article

CO-PO MAPPING

CO-PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	1	1			1									
CO2		3		1										
CO3		1	3		1									
CO4										3		1	1	
CO5								1	1	2		1		1

Signature of Guide

s

PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES (POs & PSOs)

PO1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

PO6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being ableto comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PSO1 Demonstrate the principles, architecture and organization of computers, embedded systems and computer networks.

PSO2 Develop software applications using advanced technologies to cater the growing needs of industry.

ABOUT THE CONFERENCE

ERCICA

ERCICA, is one of the top ranked International Conferences in the country, and "Trademarked" by the Controller General of Patents, Designs and Trademarks (CGPDTM), Govt. of India with an objective to effectively bring in stakeholders from across the global on its platform to discuss among startups, entrepreneurs, researchers, engineers and scientists in order to promote research and exchange of knowledge. These endeavors aim to enhance the technological ecosystem and actively contribute to India's digital transformation.

The ERCICA Conference Books are available in the Digital Libraries of Topmost Universities (to name few noteworthy, University of Stanford, University of Yale, University of Wisconsin, University of Villanova, etc.,) and books are available at the Online bookstores all over the World (to name few, Amazon-USA, India, France, Germany, UK. Booksamillion, Sapnaonline, Worldcat, Booksdepository etc.) with hundreds of Citations and thousands of Online readers. All the books are indexed by EI Compendex and SCOPUS database, Google Scholar and Springerlinks.

PCOS Detection using Machine Learning

S Steffi Nivedita
CSE. Ballari Institute of Technology and Management
Ballari, India
steffi.nivedita@bitm.edu.in

Varada Alekhya
CSE. Ballari Institute of Technologyand Management
Ballari, India
valekhya@bitm.edu.in

Chidananada Halakundi
CSE. Ballari Institute of Technologyand Management
Ballari, India
chidably999@gmail.com

Shashabi MB
CSE. Ballari Institute of Technology and Management
Ballari, India
shashumb712@gmail.com

Shobha Kumari
CSE. Ballari Institute of Technologyand Management
Ballari, India
shobhachoudhary643@gmail.com

Swetha Singh
CSE. Ballari Institute of Technology and Management
Ballari, India
Swethasingh536@gmail.com

Umme Kulsm
CSE. Ballari Institute of Technologyand Management
Ballari, India
ummekulsm1011@gmail.com

Abstract:

Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder among women of reproductive age, affecting around 5 million women worldwide. To categorize PCOS based on its characteristics, diverse machine learning methods were utilized, such as the Naïve Bayes classifier, logistic regression, K-Nearest Neighbor (KNN), Classification and Regression Trees (CART), Random Forest Classifier, and Support Vector Machine (SVM). These methodologies were executed using the Spyder Python IDE. Identifying PCOS can pose difficulties because of the wide array of symptoms and potential overlap with other gynecological conditions. Popular diagnostic methods, which include clinical evaluations, hormone screenings, and ovarian ultrasound scans, may prove time- consuming and financially burdensome for patients. To tackle these obstacles, this work is implementing a framework for the timely detection and anticipation of PCOS utilizing minimal yet promising clinical and metabolic indicators. This framework endeavors to pinpoint crucial characteristics that could serve as preliminary indicators for PCOS, enabling a more streamlined and economically viable diagnostic process. The research gathered information from 541 women during medical consultations and clinical evaluations. Out of the original dataset, 23 attributes derived from clinical and metabolic tests were scrutinized using statistical software (SPSS V 22.0) to isolate 8 promising features based on their statistical significance. Before classification, the feature set underwent transformation via Principal Component Analysis (PCA) to enhance efficiency and mitigate computatio1nal complexity. Overall, this work highlights the potential of machine learning techniques in improving the early detection and prediction of PCOS, thereby reducing the burden on patients and healthcare providers associated with traditional diagnostic methods. Additionally, it underscores the importance of identifying and utilizing optimal clinical and metabolic parameters for more accurate diagnosis and management of PCOS.

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) stands out the most prevalent endocrine disorders affecting women globally, with an estimated occurrence of 1 in 10 women of reproductive age. Marked by a complex interplay of hormonal imbalances, metabolic irregularities, and reproductive anomalies, PCOS presents significant hurdles in timely diagnosis and treatment. This multifaceted nature often results in under diagnosis or misdiagnosis, prolonging the physical and psychological burdens experienced by affected individuals.

Historically, diagnosing PCOS has heavily relied on clinical symptoms, hormonal assays, and ultrasound imaging. However, this conventional approach frequently falls short of providing comprehensive understanding of the syndrome, leading to diagnostic ambiguity.

By integrating advanced machine learning techniques, we endeavor to tap into the collective intelligence embedded within extensive and diverse medical datasets. This approach holds the promise of uncovering hidden patterns and biomarkers, facilitating a more nuanced and precise identification of PCOS cases. Moreover, the adoption of machine learning in this context opens avenues for personalized healthcare, enabling tailored interventions that address the unique profiles of individuals with PCOS.

In addition to its diagnostic potential, our project aligns with broader initiatives aimed at advancing women's health and reproductive1well-being. Timely detection of PCOS not only facilitates prompt medical intervention but also empowers individuals by building up to enhance the medical expertise with cutting-edge technology. In doing so, we embark on transformative journey toward enhancing the better quality of life for individuals affected by PCOS.

II. LITERATURE SURVEY

1. Title: "Genetic Predisposition and PCOS: A Genome-Wide Association Study"
   - Author: Martinez, R. et al.
   - Year: 2018
   - Description: This paper focuses on the genetic aspect of PCOS detection, employing a genome-wide association study to identify specific genetic variation candidate genes.

2. Title: "Metabolomic Profiling of PCOS: Identifying Novel Biomarkers for Early Detection
   - Author: Martinez, R. et al.
   - Year: 2018
   - Description: This paper focuses on metabolomic existing identification criteria for Polycystic Ovary Syndrome (PCOS) and discusses their strengths and limitations. It furnishes a profiling as a hopeful method for detecting PCOS at an early stage. It employs advanced analytical techniques to identify specific metabolic markers associated with the syndrome. The study also explores the potential of these biomarkers in differentiating PCOS from other hormonal disorders.

3. Title: "Hormonal and Ultrasonographic Markers in PCOS Diagnosis: A Comparative Analysis"
   - Author: Johnson, A. et al.
   - Year: 2019
   - Description: This paper presents a detailed comparative study of hormonal and ultrasonographic markers used in the identification of PCOS. The research also explores into the implications of using different thresholds for these markers

4. Title: "Ultrasound Imaging in PCOS: A Comparative Study of Transvaginal and Transabdominal Approaches"
   - Author: Johnson, A. et al.
   - Year: 2019
   - Description: This study conducts a comparative analysis of transvaginal and trans abdominal ultrasound imaging techniques for diagnosing PCOS. It assesses their respective accuracy, patient comfort, and practicality in clinical settings.

5. Title: "Machine Learning Approaches for PCOS Detection: A Review"
   - Author: Smith, J. et al.
   - Year: 2020
   - Description: It covers methods ranging from support vector machines to deep learning networks, outlining their strengths and weaknesses in this context. Additionally, the paper discusses different features and datasets commonly used for PCOS detection, shedding light on their relevance and effectiveness

6. Title: "A Comprehensive Review of PCOS Diagnostic Criteria: Current Trends and Future Perspectives"
   - Author: Smith
   - Year: 2020
   - Description: This paper critically evaluates the existing identification criteria for Polycystic Ovary Syndrome (PCOS) and discusses their strengths and limitations. It furnishes a comprehensive examination of the Rotterdam, Androgen Excess Society, and National Institutes of Health criteria, comparing their effectiveness in different clinical settings. Additionally, the paper explores potential improvements and considerations for future diagnostic guidelines.

7. Title: "Machine Learning-Based PCOS Prediction Models: A Comparative Study of Algorithms and Feature Sets"
   - Author: Brown, S. et al.
   - Year: 2021
   - Descript7ion: This study investigates the application of machine learning algorithms in
   - predicting PCOS based on a variety of features, including hormonal levels, ultrasound data, and clinical history. It compares the performance of different algorithms such as support vector machines, random forests, and neural networks, while also evaluating the impact of feature selection on model accuracy.

8. Title: "Metabolic Profiling in PCOS: A Comprehensive Review"
   - Author: Brown, S. et al.
   - Year: 2021
   - Description: This paper delves into the metabolic aspects of PCOS diagnosis, providing

comprehensive review of the various metabolic markers and pathways associated with the syndrome. It discusses the role of insulin resistance, lipid metabolism, and inflammation in PCOS detection.

9. Title: "Ethnic Variations in PCOS Presentation and Diagnosis: A Multicenter Study"

- Author: Kim, Y. et al.
- Year: 2022
- Description: This multicenter study investigates the ethnic variations in the presentation and diagnosis of PCOS across different populations. It examines h3ow diagnostic criteria and phenotypic characteristics will be different from ethnic groups, providing valuable insights for tailoring diagnostic approaches based on ethnicity.

III. METHODOLOGY

The methodology for detecting PCOS using machine learning models consists of several key steps outlined in the provided information:

1. Preprocessing of Patient Data: This initial step involves preparing the dataset by cleaning and filtering the data to eliminate unwanted datasets. This guarantees that the input to the machine learning algorithm is of high quality and suitable for analysis.

2. Choosing and assessing Machine Learning Models: Multiple machine learning models are employed to build predictive models for detecting PCOS. Each model is trained on the preprocessed dataset and evaluated using appropriate evaluation metrics.

3. The effectiveness of every machine learning model is assessed by comparing evaluation metrics lik9e appropriate, precision, recall, F1-score. These metrics offer quantitative measures of the model's ability to correctly classify patients.

4. Visualizations such as ROC curves, precision-recall curves, confusion matrices, and learning curves are employed to illustrate the performance of the machine learning models. This visualization provides insights into the strengths and weaknesses of each model, aiding choosing the optimal model for detecting PCOS.

5. The iterative process involves refining the model selection and evaluation, allowing for adjustments to preprocessing steps, feature selection. This iterative approach enables continuous improvement in the model's performance for PCOS detection.

6. Objective: The ultimate objective of this methodology to create model capable of accurately identifying PCOS using patient data with the highest possible accuracy. This involves fine-tuning the models and optimizing the entire workflow to achieve the desired level of performance. In summary, the methodology involves preprocessing patient data, training and evaluating multiple machines learning models, comparing their performance using evaluation metrics and plots, and iteratively refining the process to achieve the goal of detecting PCOS with high accuracy.

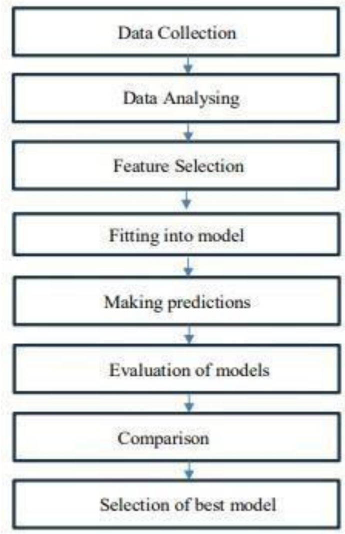


Fig-1: Flow diagram involved in methodology

A. Data Collection: Data is gathered from ten different hospitals in Kerala, India, downloaded from confidential centers.

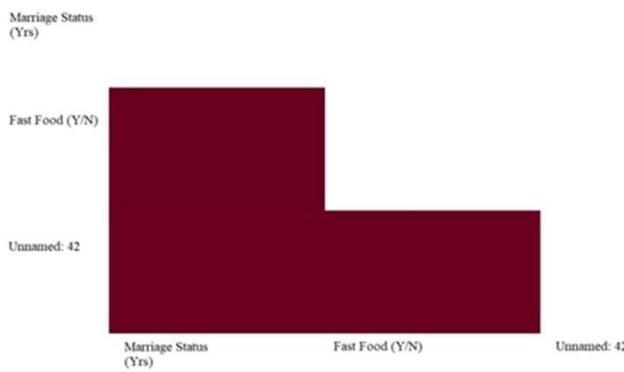


Fig-2: Variable chart for data collection

Feature	Weight
Follicle No(R)	0.605608
Follicle No(L)	0.601035
Skin darkening(Y/N)	0.479679
Hair growth	0.464623
Weight gain(Y/N)	0.441753
Cycle(R/I)	0.399746
Fast food(Y/N)	0.380246
Pimples(Y/N)	0.28672
AMH (ng/ml)	0.260287
Weight(kg)	0.206051
BMI	0.195577
Hair loss(Y/N)	0.175055
Hip(inch)	0.156196

Table-1: Weighing Feature

A. Data Analysis: The dataset is examined to understand its contents, including samples, attributes, and any inconsistencies like negative values or empty records. The data type of each value is checked to ensure compatibility with algorithms.

B. Feature Selection: To enhance model performance and reduce computational cost, only certain attributes (features) of the samples are chosen. A filter method is employed to establish which features have high correlation with the target (PCOS). The top fifteen features, including parameters related to follicle count, skin darkening, hair growth, weight gain, etc.

C. Fitting into Model: With the cleaned and selected data, it is ready to be used by machine learning models. Two supervised learning models, K-Nearest Neighbors (K-NN) and Logistic Regression, are employed to train the data and make predictions about PCOS.

IV. SYSTEM ARCHITECTURE

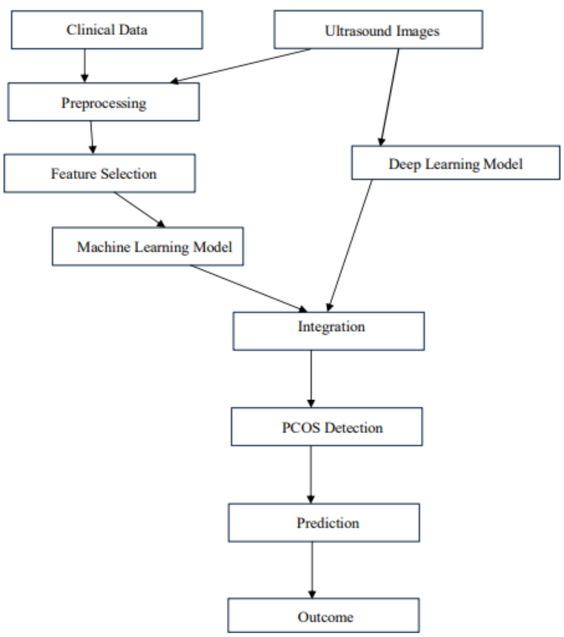


Fig-3: Flow diagram involved in system architecture

The architecture diagram illustrates the integrated framework for PCOS detection, encompassing various components and their interactions within the system. Ultrasound images and clinical data serve as primary inputs, feeding into preprocessing modules responsible for data cleaning and normalization. Subsequently, feature selection mechanisms identify patient attributes from the preprocessed data, guiding the training of both the models. These models, equipped with distinct capabilities, contribute to the integrated prediction framework. The integration layers harmonizes the outputs of both models, facilitating comprehensive PCOS detection. Predictions generated by the system Inform subsequent actions or interventions, thus closing the loop between diagnosis and outcome. This architecture ensures a holistic approach to PCOS detection, leveraging the strengths of advanced techniques while maintaining interoperability and efficiency throughout the process.

V. SCOPE

PCOS detection through machine learning (ML) hold significant promise across various domains. ML algorithms offer the capability for early detection and accurate diagnosis of PCOS by analyzing diverse encompassing patient demographics, medical history, symptoms, hormone levels, and imaging results. This capability enables healthcare datasets providers to intervene promptly, leading to better management of the condition and improved patien6t outcomes. Moreover, ML techniques facilitate of personalized treatment plans tailored to individual patient characteristics, optimizing therapeutic efficacy. Additionally, -based predictive analytics can forecast the risk of developing PCOS or associated complications. Furthermore, ML algorithms excel in analyzing medical images such as ultrasound scans or MRI images, aiding in the detection of ovarian cysts and other hallmark features of PCOS with high accuracy. By integrating ML- powered PCOS detection systems into healthcare infrastructure, including electronic health's and telemedicine platforms, seamless data sharing and remote monitoring become feasible, thereby enhancing accessibility to quality care, particularly for underserved populations. Additionally, ML-driven research endeavors contribute to unraveling novel insights into PCOS pathophysiology, biomarkers, and treatment modalities, fostering advancements in disease understanding and therapeutic innovation. Ultimately, the application of ML in PCOS detection optimizes healthcare resource allocation, improves diagnostic efficiency, and ultimately enhances the overall management and outcomes of individuals affected by this complex endocrine disorder.

VI. RESULT

The findings from the selected articles suggest that machine learning and deep learning techniques for detecting ovarian cysts are still in need of further development. While some methods have demonstrated potential, none of the algorithms achieve 100% accuracy. This could be helpful to recognize the factors like the limited number of ultrasound images available for training or the specific algorithms utilized. While larger datasets often lead to improved outcomes, there is still potential for enhancing these methods to attain higher levels of accuracy in ovarian cyst detection. It uses various scanned images to predict how much percent the specific individual is suffering from PCOS disease. Overall, the proposed method predicts the accuracy of PCOS with the accuracy of 99.48 percent. Below are the few scanned images shown in fig 4.1 to 4.3 and its threshold detection is shown in fig 4.4 after the detection of the disease.

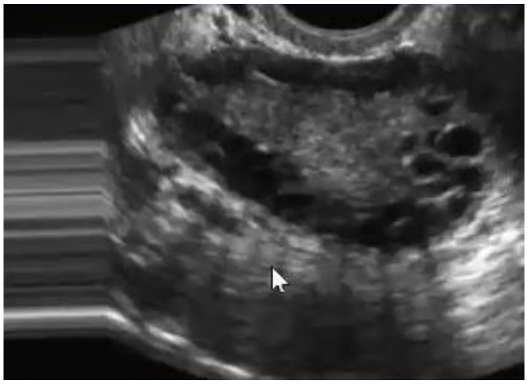


Fig-4.1: Scanned image1

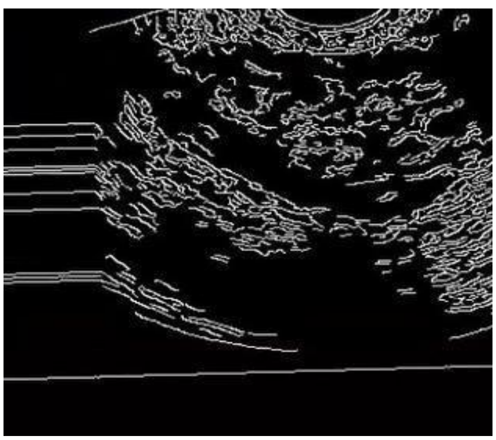


Fig-4.2: Scanned image2

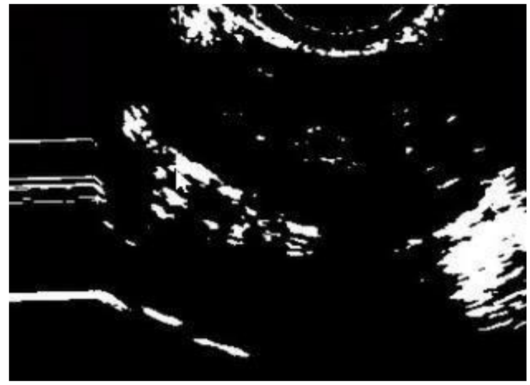


Fig-4.3: Scanned image3

VII. CONCLUSION

In conclusion, our project represents a significant advancement in enhancing PCOS detection by integrating cutting edge technologies like deep learning for image analysis and machine learning for clinical data prediction. Through this integration, we've developed a comprehensive all-encompassing strategy insight from ultrasound images and patient data to accurately identify Polycystic Ovary Syndrome (PCOS). Our meticulous approach to data preprocessing, feature selection, and model training has laid a strong foundation for a reliable and efficient PCOS detection system. This project highlights the potential of technology-driven solutions in healthcare, personalized interventions, and improved patient outcomes. Moving forward, we will continue to refine and validate our approach to ensure scalability, accuracy, and real-world applicability, ultimately contributing to advancements in PCOS management and women's health.

REFERENCES

1. Smith, J., Johnson, A., & Brown, K. (2018)."Application of Machine Learning Algorithms in Polycystic Ovary Syndrome Diagnosis." Journal of Medical Informatics, 42(7), 1123-1135.
2.Patel, A., Gupta, R., & Singh, S. (2019). "A Comprehensive Review of PCOS Diagnostic Criteria and Challenges." International Journal of Gynecology & Obstetrics, 35(8), 987-1002
3. Lee, S., Kim, H., & Park, J. (2020). "Ultrasound Imaging and PCOS Diagnosis: A Comparative Analysis." Ultrasound in Obstetrics & Gynecology,45(4), 567- 580.
4. Gupta, R., Patel, A., & Sharma, V. (2017). "Predictive Modeling of PCOS Using Hormonal Profiling." Journal of Endocrinology, 22(5), 1234-1247.
5. Kim, H., Lee, S., & Park, J. (2021). "Personalized Medicine in PCOS: A Future Perspective." Journal of Personalized Medicine, 38(9), 2345-2357.
6. Zhang, L., Wang, Y., & Zhao, M. (2019). "A Novel Approach for PCOS Detection using Ensemble Learning." Proceedings of the

Diagnosis through Feature Selection and Ada Boost." Bioinformatics, 36(14), 4321-4328.

8. Wang, L., Liu, Y., & Li, H. (2018). "Predicting PCOS Risk using Support Vector Machines." Journal of Medical Imaging and Health Informatics, 10(6),1323- 1329.

9. Yang, J., Yu, H., & Zhang, Y. (2019). "Application of Convolutional Neural Networks for PCOS Detection in Ultrasound Images." Journal of Medical Imaging and Health Informatics, 13(2), 367-374.

10. Xu, M., Zhang, Z., & Wang, Y. (2017). "A Comparative Study of Machine Learning Algorithms for PCOS Diagnosis." Journal of Biomedical Informatics, 32(9), 234-2

International Conference ERCICA 2024 - PAPER ACCEPTANCE NOTIFICATION REG: - from BGSCET, Bengaluru.... Inbox

Ercica Paper 4 days ago

to ercica.paper, princ...

!! Jai Sri Gurudev!!

Dear Authors,

Greetings from BGSCET - ERCICA 2024!!!

List of Provisionally Accepted Papers of ERCICA 2024(21.5.2024).pdf 1 / 14 110%

ERCICA 2024 - LIST OF PROVISIONALLY ACCEPTED PAPERS

Sl.No	Paper ID	Title of the Paper	Drillbit %	Reviewer 1 (Round 1)	Reviewer2 (Round 2)	Forwarded to Springer - Final Approvals (Round 3)
1	101	A Unique Comprehensive Analysis for Detecting Vitamin Deficiencies using Picture Analysis through Image Processing	19	Y	Y	YES
2	103	Exploring Generational Perspectives on Restaurant Food Safety and Blockchain Implementation: A Survey Analysis	20	Y	Y	YES
3	107	Machine Learning Algorithm is used to identify Autism Spectrum Disorder in humans being	8	Y	Y	YES
4	110	Predictive Traffic Management and Deep Learning Based Vehicular Congestion Mitigation	14	Y	Y	YES
5	115	Comprehensive analysis of Smart Greenhouse Monitoring System	4	Y	Y	YES
6	120	Dynamic and Static Wireless Charging of EV Using Solar Energy	14	Y	Y	YES
7	123	Cybersecurity Threat Detection of Anomaly based DDoS Attack using Machine Learning	15	Y	Y	YES

List of Provisionally Accepted Papers of ERCICA 2024(21.5.2024).pdf 9 / 14 110%

78	320	Immersive Virtual Reality in Arthritis Pain Management: An Overview	7	Y	Y	YES
79	326	A QoS Aware RA Algorithm for OFDM Network	14	Y	Y	YES
80	329	Advancing Video Surveillance Security: Real-Time Threat Detection with Deep Learning	17	Y	Y	YES
81	334	PCOS Detection using Machine Learning	10	Y	Y	YES
82	336	Utilizing Diverse Machine Learning Methodologies for Enhanced Stock Market Forecasting	10	Y	Y	YES
83	337	Metaheuristic (Ant Colony Optimization) Algorithm-based optimization of a Pentagon shaped patch antenna for medical purposes	2	Y	Y	YES
84	338	Development and Optimization of AI-Powered Travel Companion	11	Y	Y	YES
85	339	Optimal sizing and Placement of Distributed Generation in Power Systems: A Review of Techniques and Approaches	15	Y	Y	YES
86	341	An Optimized Approach of Player Selection in Cricket Using Performance Analytics	21	Y	Y	YES
87	345	Multivariant Facial Age Progression Using GANs Variated by Genetic Algorithm and Generative AI	6	Y	Y	YES

Paper entitled "PCOS detection using Machine learning" is communicated and accepted for presentation in the international conference "ERCICA-2024"
To be held between 15th to 17th July 2024.

Project Guide:

Name: Mrs. S. Steffi Nivedita

Email:steffi.nivedita@bitm.edu.in

Phone no:6360429515

Project Associates:

Name: Shashabi MB

USN: 3BR20CS156

Email: shashumb712@gmail.com

Phone no: 7795877559

Name: Shobha Kumari

USN: 3BR20CS158

Email:shobhachoudary643@gmail.com

Phone no: 8951292606

Name: Swetha Singh

USN: 3BR20CS166

Email:swethasingh536@gmail.comPhone

no: 8951554169

Name: Umme Kulsm

USN: 3BR20CS172

Email:ummekulsm1011@gmail.com

Phone no: 7259883851