iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.66019

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Ovista: ML-Based Web Application for PCOS Detection

Sreedevi S[1], Abhishek J Y [2], Abhishek K[3], Chintan C Shetty[4], Dhanush H M[5]
Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering, Visvesvaraya Technological University- Belagavi, Shivamogga 577204- Karnataka

Abstract: ***Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder affecting reproductive-age women, characterized by various symptoms and irregularities. This ovista proposes a comprehensive PCOS detection framework that combines numerical and image-based features. The numerical model uses a dual-pronged system: symptom-based prediction to identify PCOS presence, and parameter-based prediction to categorize its severity into low, medium, or severe. Built using machine learning algorithms, it offers an accurate diagnosis. Additionally, deep learning techniques are applied to analyze medical images for PCOS patterns, further enhancing detection accuracy. This multi-modal approach provides a holistic diagnostic tool, helping clinicians understand PCOS severity and enabling personalized treatment strategies.Ovista is a trusted companion in detecting and managing Polycystic Ovary Syndrome (PCOS). Designed with care and precision, Ovista leverages advanced technology to empower women by providing accurate insights into their health. The name Ovista is derived from the combination of "Ova" and "Vista". "Ova," meaning egg, represents the core focus on women's reproductive health, particularly issues related to the ovaries such as PCOS. "Vista" signifies vision or perspective, highlighting the platform's mission to provide a clear and comprehensive understanding of health. Together, Ovista symbolizes a visionary approach to empowering women with insights and solutions for managing PCOS effectively.***

Keywords**: *PCOS detection, Ovista, machine learning, image analysis, feature extraction, hybrid CNN-XG boost model.***

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder affecting around 10 million women worldwide, with symptoms ranging from irregular menstrual cycles to infertility and increased risk of chronic conditions like diabetes and cardiovascular disease. Despite its prevalence, PCOS often goes undiagnosed, which delays treatment and exacerbates potential health risks. The lack of early diagnostic tools contributes to the underdiagnosis and limited awareness of this condition, leaving many women without the support needed for timely intervention. Early detection is crucial in managing symptoms, improving quality of life, and preventing further health complications. This ovista seeks to address these gaps by developing a diagnostic system using machine learning (ML) and deep learning (DL) techniques, aiming to raise awareness and provide a preliminary diagnosis for women with PCOS.

II. LITERATURE SURVEY

The reviewed studies highlight the applications and methodologies of machine learning algorithms across various domains, with a strong emphasis on healthcare. Convolutional Neural Networks (CNNs) are noted for their effectiveness in medical imaging tasks, such as disease diagnosis, but face challenges like high computational costs and the need for extensive datasets. XGBoost, known for its scalability and efficiency, is compared with Random Forest and traditional gradient boosting, showing its strengths in computational speed and accuracy, though it requires meticulous parameter tuning for optimal results.

In the context of healthcare, research on Polycystic Ovary Syndrome (PCOS) detection illustrates the utility of machine learning models such as Decision Trees, CNNs, and SMOTE-based Random Forests. These models analyze clinical, biochemical, and ultrasound data to provide early and accurate diagnosis, crucial for timely intervention. Practical applications, like mobile apps for PCOS prediction, exemplify how machine learning can enhance accessibility to diagnostic tools, bridging gaps in healthcare delivery. Additionally, techniques like Principal Component Analysis (PCA) address challenges in high-dimensional data, improving the performance of predictive models across various applications.

These studies collectively emphasize the transformative potential of machine learning in diagnostics, prediction, and decision-making.

©IJRASET: All Rights are Reserved

They highlight the need for careful model selection, parameter tuning, and addressing challenges like data imbalances, computational complexity, and the requirement for large labeled datasets. Continued research is essential to refine these models and enable their deployment in diverse, real-world settings.

III. METHODOLOGY

Ovista adopts a comprehensive methodology that integrates several essential components to ensure effective detection of Polycystic Ovary Syndrome (PCOS). The methodology includes:

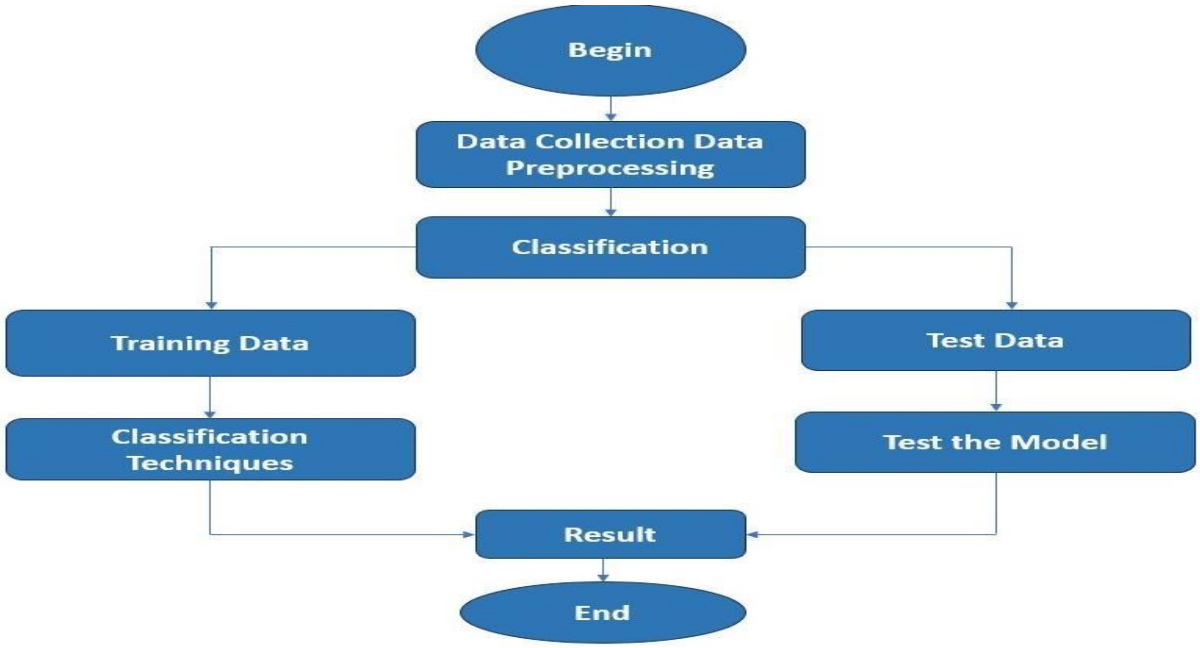


Fig 1: Flow Chart of the Ovista

1) Data Collection

The initial phase of the Ovista involves comprehensive data collection from multiple sources to ensure a robust dataset for analysis. This includes:

- Sources: Data is gathered from electronic health records, clinical databases, and patient surveys, focusing on demographic information, medical history, hormonal levels, ultrasound images, and lifestyle factors.

2) Data Preprocessing

Data preprocessing is critical for enhancing the quality and usability of the collected data. This phase includes:

- Data Cleaning: Missing values are addressed through imputation techniques (mean, median, mode), and outliers are identified and removed using statistical methods (Z-score, IQR).
- Data Transformation: Numerical features are normalized or standardized (using Min-Max scaling or Z-score normalization), and categorical variables are encoded using one-hot or label encoding.
- Image Preprocessing: Ultrasound images are resized to a consistent dimension, and augmentation techniques (e.g., rotation, flipping, zooming) are applied to increase dataset diversity.

3) Feature Selection

The feature selection phase aims to identify the most relevant features for model training. This involves:

- Feature Importance Analysis: Statistical tests are employed to evaluate the significance of features, while tree-based methods (e.g., Random Forest) rank feature importance.
- Dimensionality Reduction: Techniques such as Principal Component Analysis (PCA) are applied to reduce the feature space while retaining essential information.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

4) Model Training

In this phase, machine learning models are developed and trained on the preprocessed dataset. Key steps include:

- Algorithm Selection: Appropriate algorithms are chosen based on data characteristics, with XGBoost applied to numerical data and Convolutional Neural Networks (CNN) used for image data.
- Model Architecture: A multi-input neural network is designed to process both numerical and image data simultaneously.
- Training Process: The dataset is split into training, validation, and testing sets.

5) Model Evaluation

The evaluation phase assesses the performance of the trained models using various metrics:

- Performance Metrics: Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are calculated to evaluate model performance.
- Confusion matrices Analysis: Confusion matrices are analyzed to understand the model's classification performance and identify areas for improvement.

6) User Interface and Deployment

This phase focuses on creating a user-friendly interface and deploying the application:

- Frontend Development: A user-friendly interface is developed using HTML, CSS, and JavaScript to facilitate user interaction.
- Backend Integration: Flask is used to create a robust backend that handles data processing and model predictions.

IV. IMPLEMENTATION

The PCOS Detection System integrates hardware, software, and machine learning to deliver an efficient and reliable diagnosis of PCOS. The methodology encompasses the following key components:

1) *User Input:* The ovista begins with the user providing symptoms or uploading ultrasound images related to Polycystic Ovary Syndrome (PCOS). This input is sent to the backend for further processing, allowing the system to analyze the provided data.
2) *Data Collection:* Based on the user input, relevant patient data such as hormonal levels, BMI, menstrual cycle characteristics, and ultrasound images are collected. The system utilizes structured forms to gather clinical data, while users can upload ultrasound images directly for analysis.
3) *Data Structuring and Cleaning:* The collected data is initially in an unstructured format and is converted into a structured tabular form using Pandas. Fields such as hormonal levels, BMI, and ultrasound image paths are organized for analysis. A data cleaning process is performed to handle missing or inconsistent values, normalize units of measurement, and remove redundant or irrelevant entries. For instance, hormonal levels are standardized to ensure consistency across different measurement units.
4) *Analysis Using Machine Learning:* The cleaned data is analyzed using machine learning algorithms. Feature selection techniques are applied to identify the most relevant features for PCOS detection, such as hormone ratios and BMI. The system then applies Extreme Gradient Boosting (XGBoost) to train a model on the numerical data and Convolutional Neural Networks (CNN) to analyze the ultrasound images. This dual approach allows for a comprehensive analysis of both numerical and image data.
5) *PCOS Diagnosis Prediction:* The trained models are evaluated using metrics such as accuracy, precision, and recall. The results are used to generate a predicted score for each patient, indicating the likelihood of PCOS diagnosis. This score helps healthcare professionals prioritize cases based on severity and provides insights into the patient's condition.
6) *Frontend Display:* The results are presented on a user-friendly front end built with HTML, CSS, and JavaScript. The front end accepts user input for symptoms or image uploads and displays the diagnosis results, severity assessment, and visualizations in a visually appealing manner.

V. RESULTS

In order to address a specific problem domain, our study used machine learning (ml) applications that Combined visual and numerical data. Preprocessing procedures were essential to guarantee the accuracy and consistency of the data; both image and numerical datasets were subjected to feature extraction and normalization techniques. Our machine learning models exhibited encouraging results, with accuracy metrics varying between 80% and 20% in various models and datasets.

Fascinatingly, feature significance analysis showed that some features extracted from picture data, including color distributions or texture patterns, were just as useful as numerical variables for forecasting the desired result. Although competitive performance was demonstrated by our models, comparative analysis indicated some minor differences in the use of numerical and picture data, pointing to possible areas for hybrid or optimization strategies. Moreover, feature analysis highlighted the significance of specific qualities in influencing model predictions, corresponding with domain expertise, and offering useful information to decision-makers. The study does have several drawbacks, though; there are still issues with model interpretability and dataset size and quality. Looking ahead, to improve the generalizability and robustness of ml models in practical applications, future work should concentrate on improving model structures, incorporating new data sources, and tackling particular difficulties.

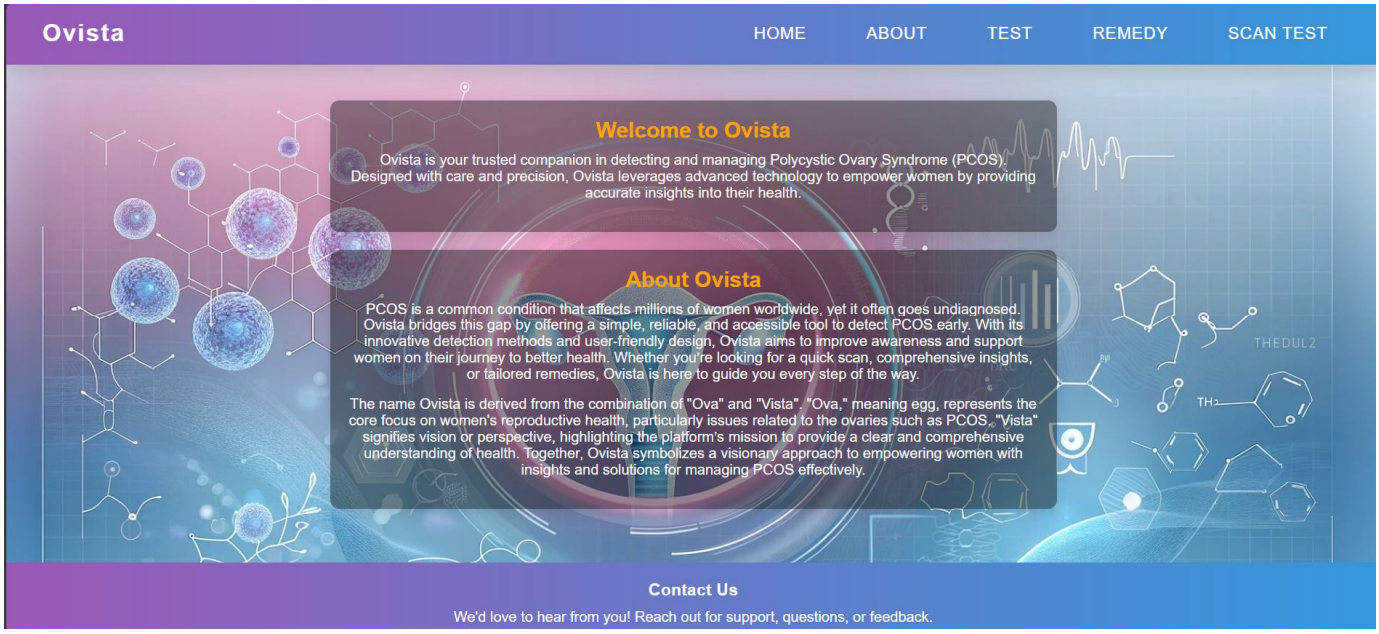


Figure 2: Landing page of the website

Figure 2 represents the front interface of our ovista's Landing page, which contains of information of our detector.

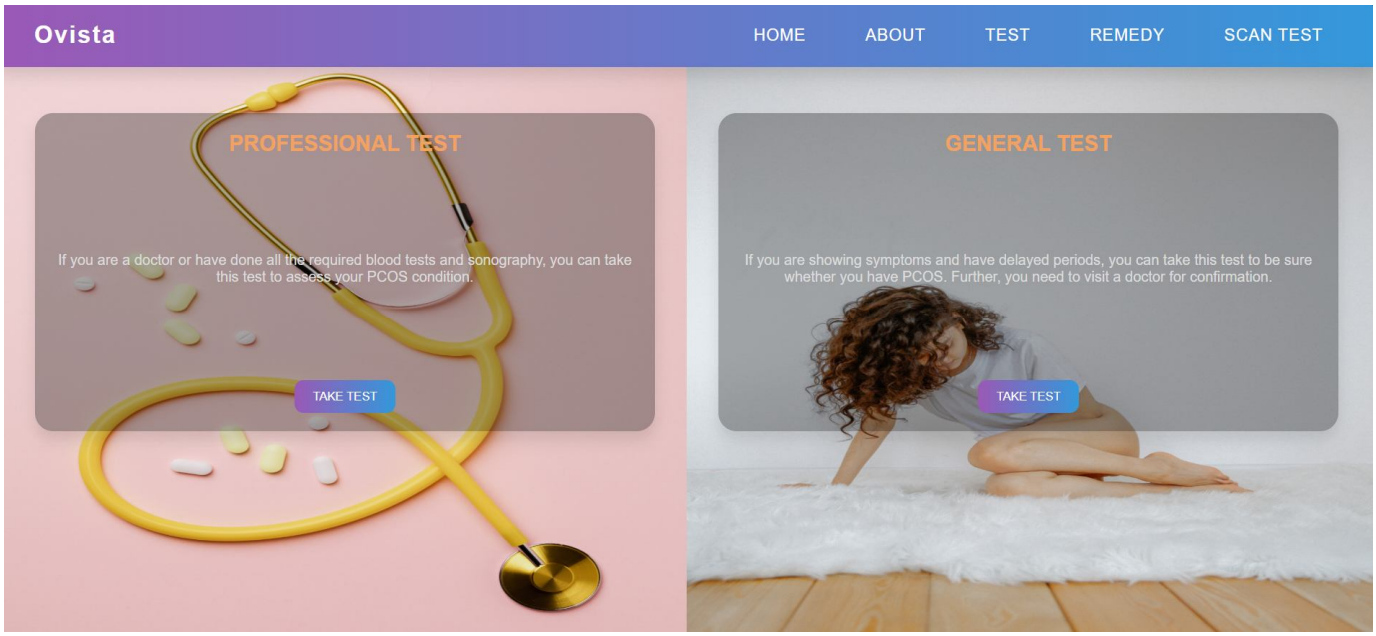


Figure 3: Choosing Test page of the website

Figure 3 represents a comprehensive and detailed representation of both the general test and the professional test, highlighting their features.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

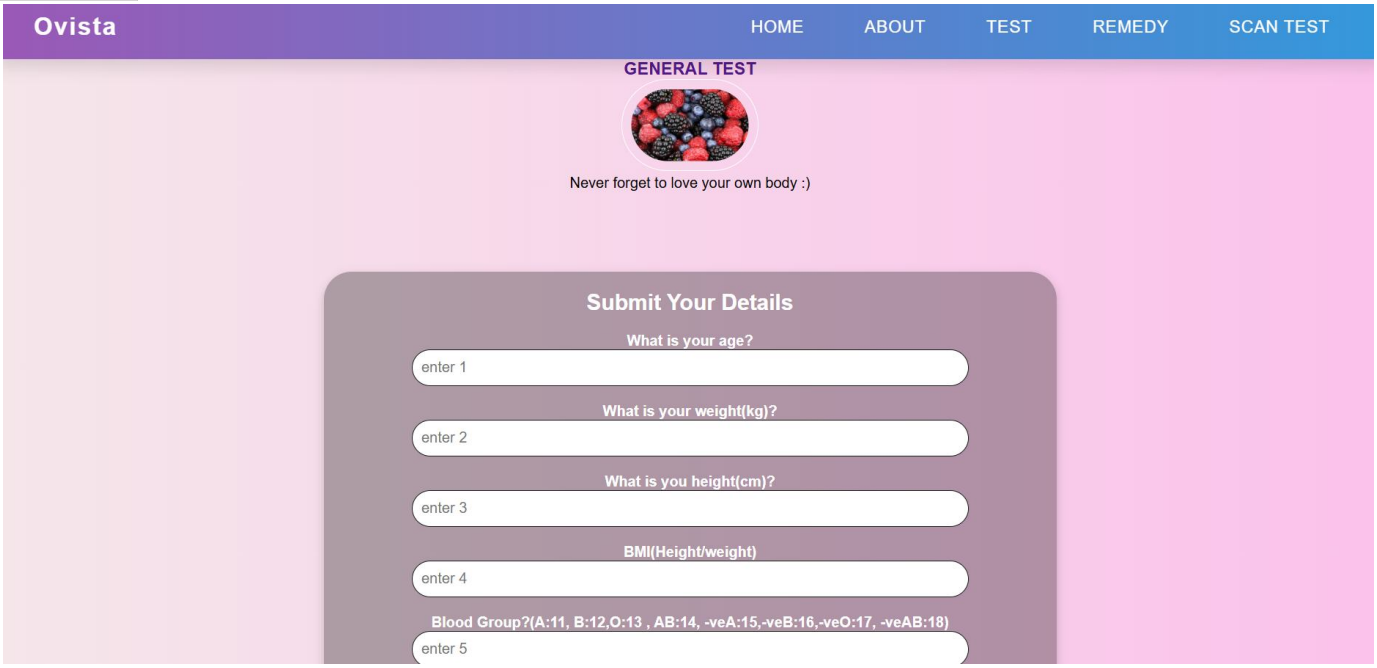


Figure 4: General test page of the website

Figure 4 represents a detailed representation of a responsive PCOS Detector interface, which includes a general test form for users to input details such as age, BMI, and symptoms.

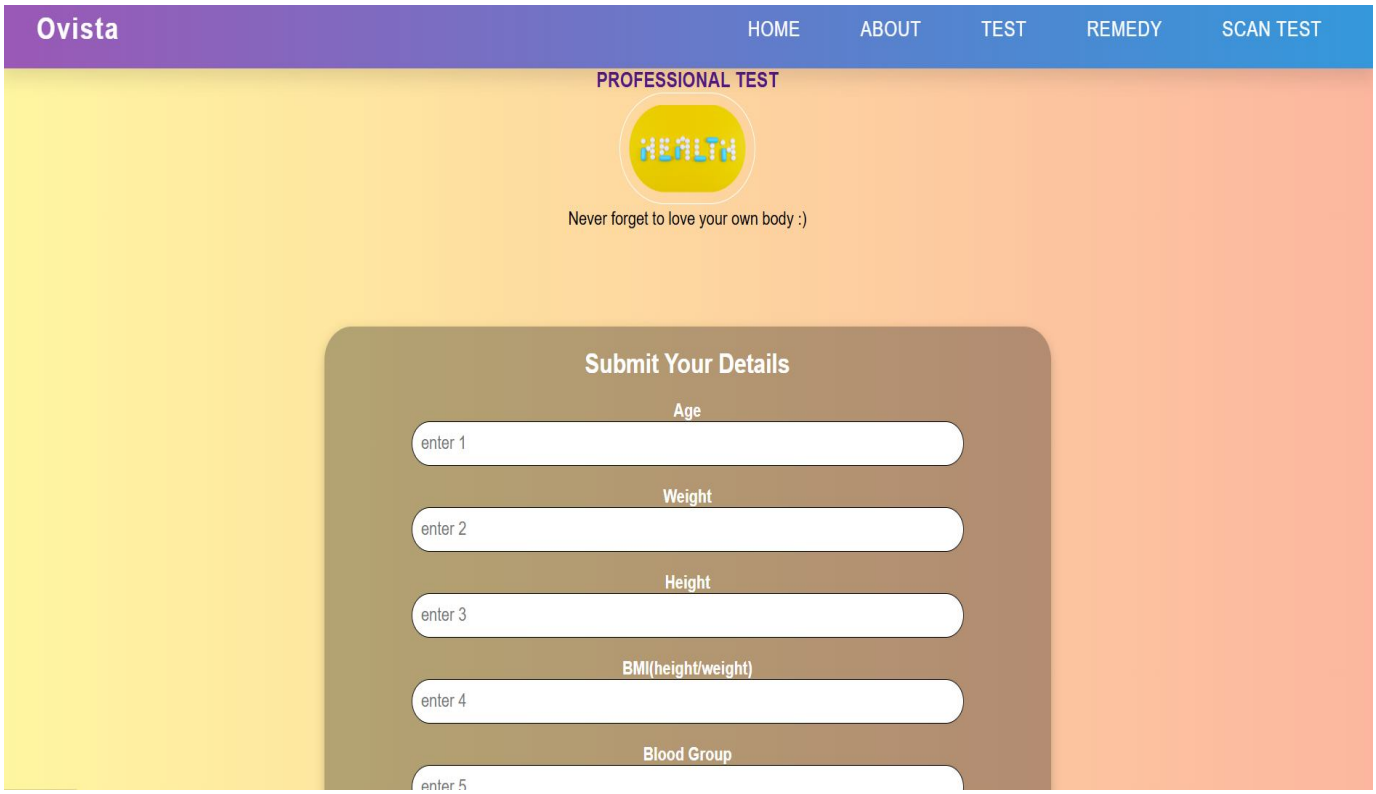


Figure 5: Professional test page of the website

Figure 5 represents a detailed representation where users input details such as age, weight, height, BMI, blood group, and pulse rate. The form also asks about symptoms like weight gain, hair growth, skin darkening, and pimples, along with lifestyle factors such as fast food consumption and exercise. Additionally, it includes blood pressure readings to assess cardiovascular health.

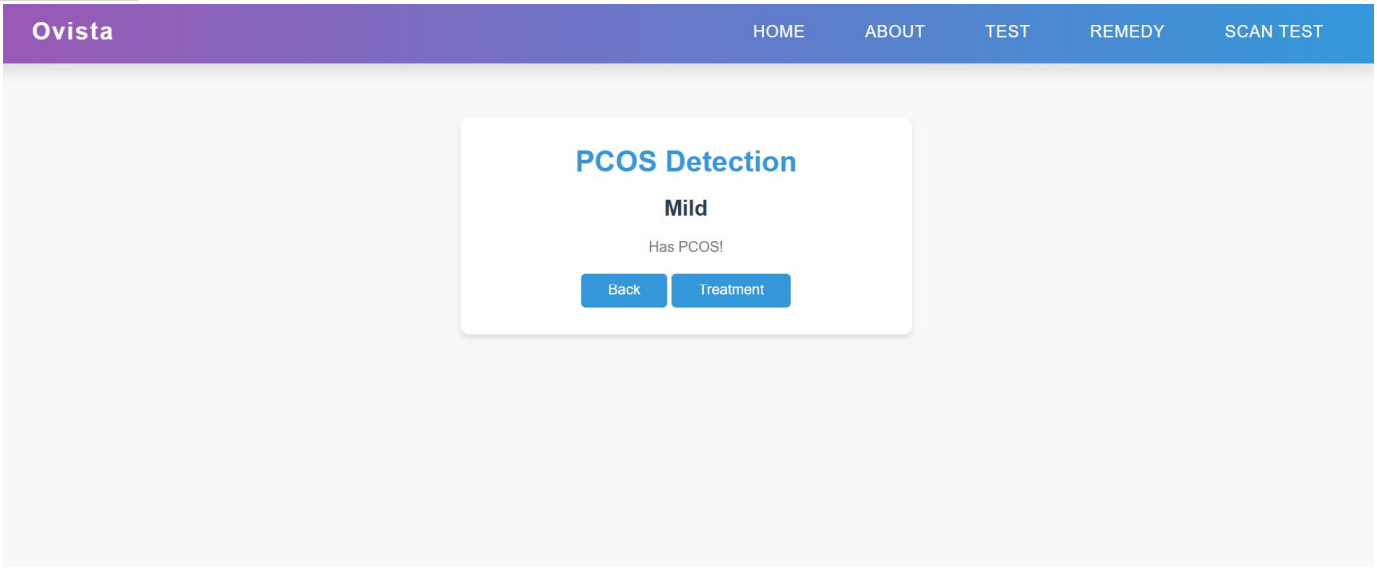


Figure 6: Result of general test

Figure 6 presents a detailed representation where,after submitting the input, the system evaluates the data and may indicate, mild you are experiencing symptoms of PCOS.

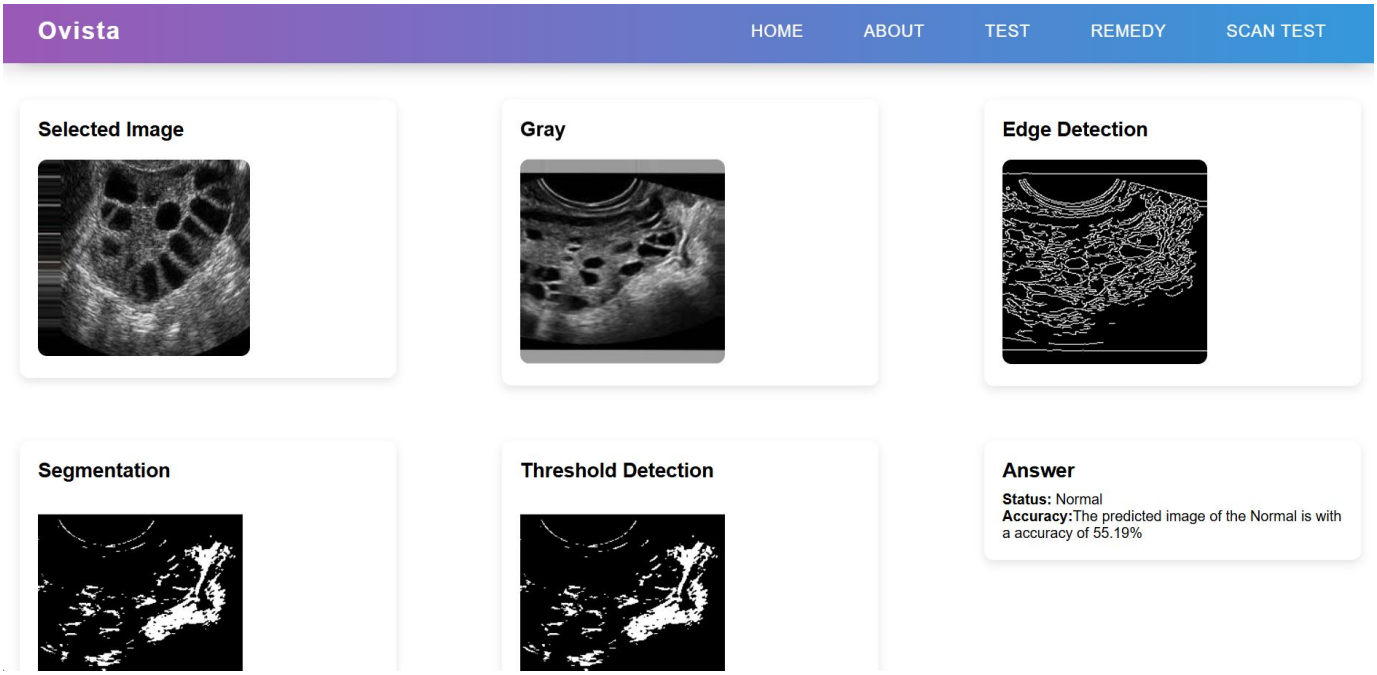


Figure 7: Scan Page

Figure 7 represents a detailed representation where a scan can be used as input for PCOS detection. The system evaluates the scan data to identify signs of PCOS, such as ovarian cysts, hormonal imbalances, or other symptoms associated with the condition. The scan results provide valuable insights into the presence and severity of PCOS, aiding in diagnosis and treatment planning.

VI. CONCLUSIONS

In this study, we propose a novel hybrid approach to enhance the diagnosis of Polycystic Ovary Syndrome (PCOS) by integrating Extreme Gradient Boosting (XGBoost) for numerical feature analysis and Convolutional Neural Networks (CNN) for image-based feature extraction from ultrasound scans. This integration aims to address the limitations of traditional diagnostic methods, providing a comprehensive and precise evaluation of PCOS.

The hybrid model combines the strengths of XGBoost in processing structured numerical data with CNN's ability to extract discriminative features from ultrasound images, focusing on key indicators such as follicle count and cyst presence. This approach enhances the model's discriminatory power and resilience in identifying complex patterns and correlations within PCOS-related datasets.

Performance evaluation using metrics such as accuracy, sensitivity, specificity, and area under the curve demonstrates the model's efficacy in distinguishing PCOS-positive cases from negative ones. Additionally, the system's flexibility and scalability allow for seamless integration into existing healthcare infrastructures, facilitating early diagnosis and enabling personalized treatment plans for PCOS patients.

Future research directions include optimizing feature selection techniques, exploring additional imaging modalities, and validating the model on larger and more diverse datasets. This study underscores the potential of machine learning models, such as the hybrid XGBoost-CNN approach, to advance PCOS diagnosis and improve patient outcomes.

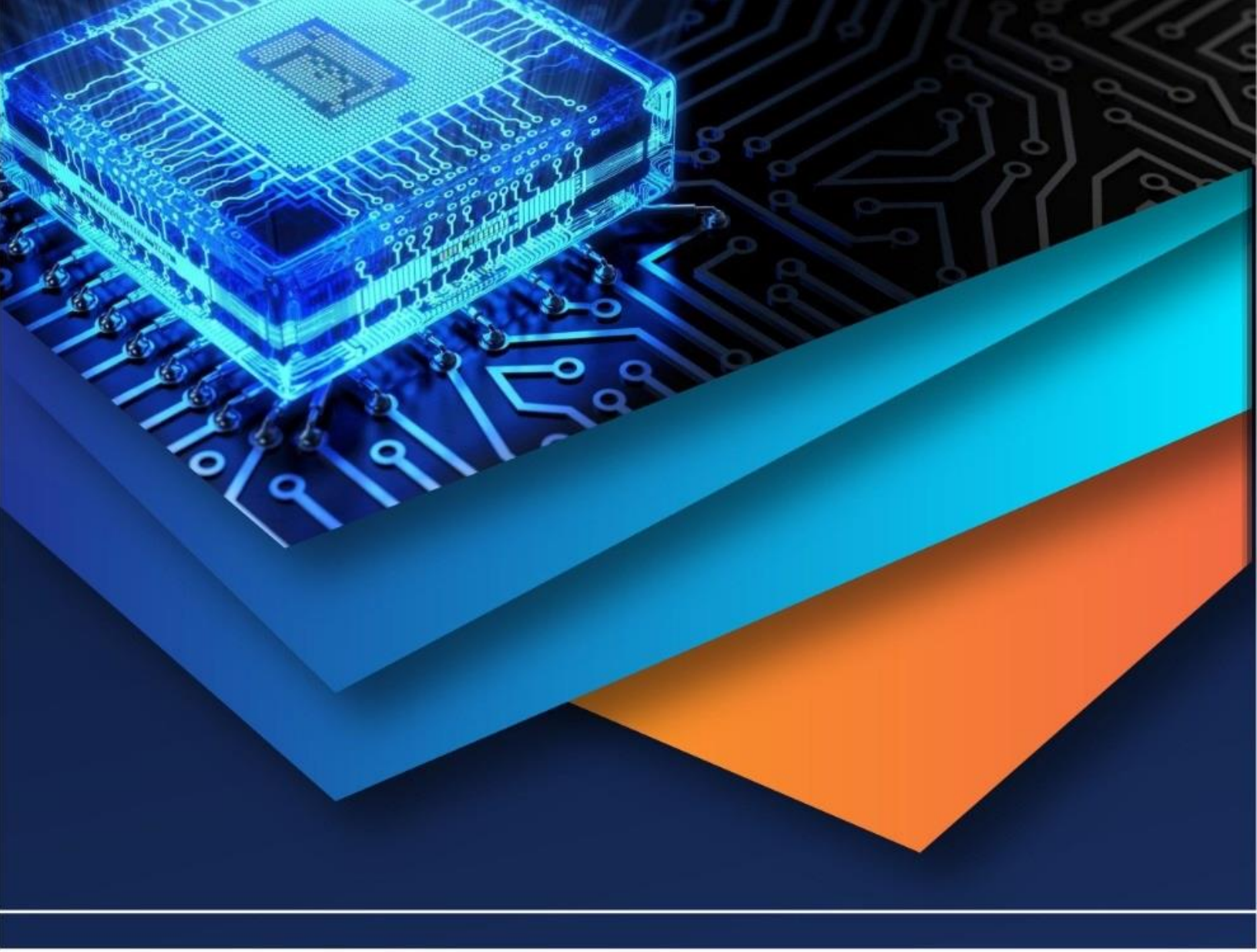
VII. ACKNOWLEDGMENT

We, the authors, express our heartfelt gratitude to Jawaharlal Nehru New College of Engineering and the Department of Computer Science and Engineering for providing the resources and support required to carry out this work.

REFERENCES

[1] T. Mitchell, Machine Learning, McGraw-Hill, 1997.
[2] S. Sridhar, M Vijayalakshmi "Machine Learning". Oxford ,2021
[3] Dewailly, D., Lujan, M.E., Carmina, E., Cedars, M.I., Laven, J., Norman, R.J. And Escobar Morreale, H.F., 2013. Definition And Significance Of Polycystic Ovarian Morphology: A Task Force Report From The Androgen Excess And Polycystic Ovary Syndrome Society. Human Reproduction Update, 20(3), Pp.334-352.
[4] Goodman, Neil F., Et Al. "American Association Of Clinical Endocrinologists, American College Of Endocrinology, And Androgen Excess And PCOS Society Disease State Clinical Review: Guide To The Best Practices In The Evaluation And Treatment Of Polycystic Ovary Syndrome-Part 1." Endocrine Practice 21.11 (2015): 1291-1300. Gourav Bathla, "Stock Price Prediction Using LSTM And SVR," 2020 Sixth International Conference On Parallel, Distributed And Grid Computing (PDGC), Waknaghat, India, Nov. 2020.
[5] Zawadzski, J. K. "Diagnostic Criteria For Polycystic Ovary Syndrome: Towards A Rational Approach." Polycystic Ovary Syndrome (1992): 39 50Chun Yuan Lai, Rung-Ching Chen, Rezzy Eko Caraka, "Prediction Stock Price Based On Different Index Factors Using LSTM", 2019 International Conference On Machine Learning And Cybernetics (ICMLC), Kobe, Japan, July 2019.
[6] Tian, Lifeng, Et Al. "Androgen Receptor Gene Mutations In 258 Han Chinese Patients With Polycystic Ovary Syndrome." Experimental And Therapeutic Medicine 21.1 (2021): 1-1.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)