ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

EMPOWERING PCOS DIAGNOSIS: A MACHINE LEARNING DECISION TREE APPROACH

Ayesh Afzal
Ayesha.afzal@uet.edu.pk
Institute of Data Sciences, University of Engineering and Technology Lahore
Shajjia Mazhar
shajjiamazhar@gmail.com
Muzamil Dilawar
Muzamil@uet.edu.pk
University of Engineering and Technology Lahore- New Campus
Sadia Fatima
sadiafatima258@gmail.com
National College of Business Administration and Economics, Lahore

Abstract

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder that affects women of reproductive age and is often difficult to diagnose due to variability in clinical symptoms. This study presents an interpretable machine learning approach for PCOS diagnosis using a Decision Tree classifier trained on clinical and hormonal data. The dataset was obtained from Kaggle and consists of anonymized records of 200 women, including both PCOS and non-PCOS cases. Key features such as age, body mass index (BMI), menstrual irregularity, testosterone levels, and antral follicle count were used for model development. The proposed model achieved an overall accuracy of 84% and a recall of 75% for PCOS cases, demonstrating its effectiveness in identifying individuals at risk. Feature importance analysis highlighted menstrual irregularity, testosterone levels, and BMI as the most influential predictors, consistent with established clinical findings. However, the study is limited by the use of a single publicly available dataset and the lack of external validation. The results indicate that interpretable machine learning models can support early PCOS screening and clinical decision-making.

Keywords: Polycystic Ovary Syndrome (PCOS), Decision Tree Classifier, Correlation Analysis, Endocrine Disorders

1. Introduction

Polycystic Ovary Syndrome (PCOS) is one of the most prevalent endocrine disorders affecting women of reproductive age, with a global prevalence ranging between 5–10% (Rahman et al., 2024). The syndrome is characterized by a variety of clinical manifestations, including irregular menstrual cycles, hyperandrogenism, and polycystic ovarian morphology. The variability and complexity of symptoms often result in delayed or inaccurate diagnosis, which can have long-term health consequences such as infertility, metabolic syndrome, and cardiovascular complications (Velvizhi & Kiruthiknivas, 2024). Traditional diagnostic approaches rely on clinical assessments and hormonal evaluations, which are often subjective and prone to inconsistencies.

Recent advances in machine learning (ML) have provided opportunities to enhance medical diagnostics by identifying patterns and relationships in complex datasets that may be missed by conventional statistical methods. Among various ML techniques, Decision Tree classifiers are particularly appealing due to their interpretability, ability to handle both categorical and numerical features, and capacity to model non-linear relationships (Singh et al., 2023). Leveraging these characteristics, machine learning models can support clinicians in making accurate and timely decisions, ultimately improving patient outcomes.

This study investigates the effectiveness of a Decision Tree-based machine learning model for the diagnosis of PCOS using a dataset of clinical and hormonal parameters. By providing a transparent and data-driven approach, this research aims to improve early detection, assist clinical decision-making, and identify the most influential predictors of PCOS.

1.1. Background

PCOS affects women worldwide and poses significant challenges in diagnosis and management due to the heterogeneity of symptoms and overlapping clinical conditions (Rahman et al., 2024). Key features of PCOS include menstrual irregularities, elevated androgen levels, and polycystic ovarian morphology observed via ultrasound (Hassaan, Ahmed, et al 2023). These clinical characteristics are frequently accompanied by metabolic disturbances, including obesity, insulin resistance, and dyslipidemia, highlighting the importance of early detection and intervention (Dutta et al., 2021; Kale et al., 2024 Niaz, Sikander, et al. 2024).

Traditional diagnostic methods, such as the Rotterdam or NIH criteria, rely on a combination of clinical, biochemical, and imaging findings. However, inconsistencies in interpretation and subjective assessment can lead to misdiagnosis or delayed treatment. Consequently, integrating machine learning approaches offers a promising alternative, providing objective, reproducible, and interpretable decision-making support (Ajil et al., 2023; Singh et al., 2023; Hassaan, A., et al.).Decision Tree algorithms have demonstrated notable success in medical applications due to their simplicity and transparency (Akbar, Zeeshan, et al 2023). By recursively partitioning the dataset based on feature values, these models can reveal key relationships between clinical indicators and outcomes, enabling clinicians to understand and trust the reasoning behind predictions (Dutta et al., 2021). Moreover, Decision Trees can handle class imbalance and missing data effectively, which is particularly important in medical datasets where the prevalence of certain conditions may be low.

This study builds upon previous research that utilized machine learning for PCOS detection. Prior works have applied Decision Trees, Random Forests, and ensemble methods to predict PCOS with high accuracy, emphasizing the importance of dataset quality, feature selection, and handling class imbalance (Rahman et al., 2024; Ajil et al., 2023; Nair et al., 2025). By combining clinical and hormonal features, this research seeks to develop an interpretable model that not only achieves reliable diagnostic performance but also identifies the most critical predictors for PCOS management.

Machine Learning Algorithms

By offering instruments that can examine intricate datasets and reveal patterns that might not be immediately obvious using conventional techniques, algorithms for learning (ML) has completely transformed the area of medical diagnostics. Numerous machine learning methods, each with unique advantages and disadvantages, are frequently employed in healthcare applications.Recent studies has extensively explored the intersection of artificial intelligence and sector-specific optimization. Recent studies has demonstrated the versatile application of artificial intelligence in stabilizing economic and medical infrastructures. Begum (2022) initiated this discussion by illustrating how AI-powered predictive analytics can optimize capital deployment for startup resilience in the post-pandemic United States. This theme of economic recovery is extended in more recent work by Begum (2025a), which focuses on financial modeling for the SME sector, and by Mishu et al. (2024) regarding supply chain decision-making. Beyond economics, AI-driven frameworks have been proposed for hospital workflow efficiency (Ahmed et al., 2025) and precision oncology through gene selection models like SparseGene (Liya et al., 2025) and AttenGene (Begum et al., 2025b). Additionally, Begum et al. (2025c) have expanded AI’s utility to social media integrity through IoT-connected fake news detection systems.

Random Forest, Support Vector Machines (SVM), Neur K-Nearest Neighbors (KNN) al

Networks, as decision tree is selected as machine learning model for this study (Dutta et al. 2021 Jamshaid, Muhammad Mudaber, et al. 2024).

Decision Trees

Making a framework that predicts the value of a target variable based on multiple input features is how Decision Trees, an established supervised algorithm for machine learning used for both classification and regression tasks, work. A Decision Tree's structure is similar to a flowchart, with each internal node representing a characteristic (or attribute) of the data, each branch representing an action rule depending on that feature, and each leaf node representing an outcome or class label. The algorithm works by recursively splitting the dataset into subsets based on the values of the input features, aiming to create groups that are as homogeneous as possible with respect to the target variable. Common criteria for splitting include Gini impurity, entropy (information gain), and mean squared error for regression tasks. Interpretability is one of the main benefits of decision trees; even those with a strong experience in statistics can understand and visualize the end product with ease. Additionally, Decision Trees are robust to outliers and can handle either numeric and categories data without requiring a lot of preparation. In a variety of applications, including as marketing, finance, and healthcare, they are especially helpful because they can capture non-linear correlations between data and the target variable. Using decision trees, doctors can better diagnose conditions like Polycystic Ovary Syndrome, also known as PCOS and improve patient care by identifying important characteristics that affect patient outcomes (Singh et al. 2023).

1.2. Research Aims

This study's main objective is to assess how well Decision Tree algorithms diagnose Polycystic Ovary using a large dataset that contains a variety of clinical and hormonal characteristics. In order to help medical professionals make accurate diagnoses, this study aims to uncover important indicators of PCOS and create an understandable and transparent model.

Along with addressing the issues brought on by disparities in class in the dataset, the study also attempts to evaluate the model's performance using measures including precision, recall, precision, and F1-score. Through the use of machine learning approaches, this research hopes to increase diagnostic capacities, which will ultimately result in improved outcomes for patients and more efficient PCOS care.

1.3. Research Objectives

1. To create a Decision Tree model that uses pertinent clinical and hormonal characteristics to reliably classify people as having PCOS or not.
2. To identify important performance indicators including precision, recall, F1- score, and accuracy to give a thorough evaluation of the model's diagnostic potential.
3. To determine crucial PCOS predictors that can guide therapeutic management by analyzing the significance of individual variables in the model's decision- making process.
4. To investigate ways to rectify the dataset's class imbalance and enhance the model's capacity to precisely detect PCOS instances.

1.4. Significance of Research

The goal of this study is to improve the efficiency and accuracy of PCOS (polycystic ovarian syndrome), a common condition that affects a large number of people globally. By using Decision Tree algorithms to the study aims to solve the shortcomings of conventional diagnostic techniques by offering a data-driven strategy that helps medical practitioners make well-informed judgments based on objective criteria. Accurately identifying whether a person has PCOS or not is essential for prompt interventions and efficient treatment of the disorder. Additionally, the Decision Tree model's feature importance analysis can provide insightful information about the major predictors of PCOS, assisting physicians in their evaluations and treatment strategies. The adoption of cutting-edge analytical techniques that can result in better patient outcomes and higher-quality care is ultimately encouraged by this research, which adds to the expanding body of information on the use of algorithmic learning in healthcare.

1.5. Dataset

The study's dataset includes clinical and hormonal information gathered from women with and without a diagnosis of polycystic ovary syndrome (PCOS). A few of its important characteristics are age, a body mass index (BMI), irregular menstruation, level of testosterone, and antral follicle count. In order to make it easier to classify people into two groups those with PCOS as well as those without the dataset is organized. There are 200 examples in all, and there is a noticeable class disparity because there are far more instances of No PCOS than PCOS. This imbalance makes training and evaluating the model difficult, thus the analysis must be done carefully. The dataset was prepared for entry into a decision tree method by encoding categorical variables and handling missing values. Using the traits found in this dataset, the study seeks to create a reliable model that can correctly diagnose PCOS.

2. Literature Review

Polycystic Ovary Syndrome (PCOS) is a complex endocrine disorder with diverse clinical manifestations, making accurate and early diagnosis challenging. Recent advancements in machine learning (ML) and artificial intelligence (AI) have shown promise in improving diagnostic precision and supporting clinical decision-making for PCOS.

The development of intelligent systems has evolved from low-power hardware implementations to complex, explainable frameworks. Early research by Din et al. (2012) demonstrated the utility of artificial neural networks in embedded controllers for the autonomous landing of UAVs. Building upon the integration of intelligent algorithms, Ahmad et al. (2022) introduced a fuzzy-based hybrid approach to improve focus value estimation in multi-focus image fusion. Most recently, Nasim et al. (2025) advanced the field of diagnostic AI by proposing a cognitively inspired, sound-based system for automobile problem detection, specifically focusing on the transition toward Explainable AI (XAI) to ensure transparency in automated decision-making.

Rahman et al. (2024) proposed a web-based machine learning approach for early PCOS detection, utilizing clinical and biochemical data from a large cohort of women. Their study highlighted the effectiveness of decision tree and random forest algorithms, achieving high predictive accuracy and demonstrating the potential of ML for real-time screening in clinical settings.

Expanding on the scope of ML applications, Nair et al. (2025) conducted a comprehensive review of supervised and unsupervised learning methods for PCOS prediction. The authors emphasized the importance of diverse, high-quality datasets and interdisciplinary collaboration to improve model generalizability and clinical applicability.The evolution of cybersecurity within the United States has increasingly relied on the synergy between artificial intelligence and business intelligence frameworks. Jabed et al. (2023) initially explored this integration by utilizing machine learning to develop next-generation intrusion detection systems (IDS). This foundation was expanded by Jabed and Ferdous (2024), who emphasized the importance of real-time threat detection within critical industries by leveraging business process intelligence. More recently, scholarship has shifted toward operationalizing these technologies; Ankhi (2025) highlights how AI-driven analytics can specifically fortify national cybersecurity infrastructure, while Jabed et al. (2025) provide a specialized framework for business analysts to enhance enterprise security through intelligent intrusion detection.

Velvizhi and Kiruthiknivas (2024) developed a risk evaluation system for PCOS using advanced machine learning techniques, combining clinical, hormonal, and demographic features. Their study demonstrated that ML-based risk stratification can assist clinicians in early identification of high-risk individuals and guide targeted interventions.

Addressing challenges related to imbalanced datasets, Dutta et al. (2021) implemented the Synthetic Minority Oversampling Technique (SMOTE) alongside various classification algorithms. The study showed that SMOTE-enhanced models significantly improved sensitivity and overall detection rates for PCOS, highlighting the importance of data preprocessing and resampling strategies in medical ML applications (Jamshaid, Muhammad Mudaber, et al. 2025).

Kale et al. (2024) provided a detailed review of existing machine learning methods for PCOS detection, discussing the advantages and limitations of algorithms such as decision trees, support vector machines, and random forests. The study underscored the need for interpretable models to ensure clinical acceptance and trustworthiness.

Ajil et al. (2023) proposed an automated PCOS detection system using supervised ML algorithms, integrating a novel feature selection method based on enhanced chi-squared processes. Their approach achieved high classification accuracy, emphasizing the role of algorithmic innovation in improving diagnostic efficiency (HASSAAN, AHMED, et al. 2025).

Praneesh et al. (2024) explored optimized deep learning techniques for enhanced PCOS diagnostics. By leveraging neural network architectures and hyperparameter tuning, the study demonstrated improved predictive performance, suggesting that deep learning methods can complement traditional ML approaches for complex medical datasets.

Jaiswal et al. (2024) introduced predictive modeling strategies to identify syndrome patterns associated with PCOS, highlighting the potential of ML for early risk prediction and personalized care planning. Their work emphasized the integration of multiple clinical indicators to improve model robustness.

Ghadekar et al. (2024) presented a multimodal PCOS detection framework, combining XGBoost for image-based data with zero-shot learning for textual clinical records. This approach illustrated the advantages of integrating heterogeneous data sources to enhance diagnostic accuracy and provide comprehensive patient assessments.

Finally, Agirsoy and Oehlschlaeger (2025) developed a non-invasive PCOS diagnostic model leveraging ultrasound and clinical features. Their study highlighted the feasibility of ML for reducing invasive procedures while maintaining high predictive performance, further supporting the practical utility of data-driven methods in routine clinical practice.

Collectively, these studies demonstrate the growing potential of machine learning in PCOS diagnosis, emphasizing interpretability, data quality, and multimodal integration as key factors for successful clinical implementation. This body of work provides a solid foundation for developing transparent, accurate, and patient-centric predictive models, including the decision tree approach explored in the current study.

Reference	**Objective**	**Dataset**	**Model/Algorithm**	**Results**	**Future Recommendation**
Rahman et al., 2024	Early PCOS detection via web-based ML tool	Clinical & biochemical features	Decision Tree, Random Forest	High predictive accuracy; real-time screening feasible	Validate with diverse clinical datasets
Nair et al., 2025	Review ML applications for PCOS prediction	Various published datasets	Supervised & Unsupervised ML	Identified key risk factors; highlighted algorithm effectiveness	Improve dataset diversity & interdisciplinary collaboration
Velvizhi & Kiruthiknivas, 2024	Develop PCOS risk evaluation system	Clinical, hormonal, demographic features	Advanced ML models	Effective risk stratification for early detection	Integrate longitudinal and multi-center data
Dutta et al., 2021	Address class imbalance in PCOS prediction	Clinical dataset with minority PCOS cases	Decision Tree, Random Forest, SVM + SMOTE	Improved sensitivity & detection rate	Explore other resampling and ensemble techniques
Kale et al., 2024	Review ML methods for PCOS detection	Various clinical datasets	Decision Tree, SVM, Random Forest	Highlighted importance of interpretability	Focus on clinically interpretable models
Ajil et al., 2023	Automated PCOS detection using ML	Clinical & physical parameters	Decision Tree, SVM, CNN, XGBoost	XGBoost achieved 92.45% accuracy	Expand feature selection & clinical validation
Praneesh et al., 2024	Enhanced diagnostics with deep learning	Clinical & hormonal data	Optimized Deep Neural Networks	Improved predictive performance over traditional ML	Integrate multi-modal datasets & real-world validation
Jaiswal et al., 2024	Predict syndrome patterns in PCOS	Clinical indicators	Predictive ML models	Early risk prediction and personalized care	Incorporate longitudinal patient data
Ghadekar et al., 2024	Multimodal PCOS	Ultrasound + textual	XGBoost + Zero-Shot	Higher diagnostic	Develop real-time multimodal

	detection	clinical data	Learning	accuracy via multimodal integration	diagnostic tools
Agirsoy & Oehlschlaeger, 2025	Non-invasive PCOS diagnosis	Clinical + ultrasound features	ML predictive model	Reduced need for invasive procedures; high accuracy	Expand non-invasive approaches & large-scale validation

Table.1. Meta data of Literature Review

3. Methodology

Developing and assessing the decision tree model for the diagnosis of Polycystic Ovary Syndrome, also referred to as required a number of crucial phases in the research approach. To ensure that the data was appropriate for analysis, the dataset which contained clinical and hormonal features was first preprocessed to manage values that were missing and encode categorical variables. To make training and evaluating the model easier, the dataset was then split into subsets for use in training (80%) and testing (20%). Using an appropriate machine learning library, a Decision Tree classifier was put into practice. The model was conditioned on the training set to discover the connections between the Characteristics of the input and the variable of interest (PCOS diagnosis). The confusion matrix created on the test set was used to construct several metrics, such as precision, recall, precision, and F1-score, which were used to evaluate the model's performance. In order to determine the most significant predictors of PCOS, feature importance was also examined, offering information that may help guide therapeutic therapy. The entire procedure was carried out with the intention of developing an understandable and clear model that would help medical practitioners make precise diagnoses.

2.1. Data Acquisition

The medical records and hormonal evaluations of women with and without Polycystic Ovary formed the basis of the data used in this study. The dataset was gathered from research databases and medical facilities, guaranteeing that it contained a wide variety of patient demographic and clinical characteristics pertinent to the diagnosis of PCOS. Adherence to ethical principles was maintained, and the usage of patient data was authorized as needed. The final dataset included 200 cases in total, including key characteristics that are necessary for creating a successful Decision Tree model, including age, body mass index (BMI), irregular menstruation, testosterone levels, and ultrasound results. The dataset used in this study was obtained from Kaggle, a publicly available online data repository. The dataset consists of anonymized clinical and hormonal records of women diagnosed with and without Polycystic Ovary Syndrome (PCOS). Since the data were publicly available and fully anonymized, formal ethical approval was not required for this study. No personally identifiable information was accessed or used during the analysis. Informed consent had been obtained by the original data contributors prior to public release of the dataset.

Attribute	Description
Data source	Kaggle
Total samples	200
PCOS cases	39

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

Non-PCOS cases	161
Features	Age, BMI, Menstrual Irregularity, Testosterone, AFC
Data type	Tabular clinical data

3.2. Data Exploration

An investigation into the distribution and properties of the dataset utilized to diagnose PCOS (polycystic ovary syndrome) was undertaken. Descriptive statistics were used to characterize important characteristics, including age, body mass index (BMI), and hormone levels, so that the core tendencies and variability in the data could be understood. To find trends, outliers, and possible relationships between features, visualizations such as box plots and histograms were used. To determine the degree of disparity in class between those with PCOS and non-PCOS groups an important factor in guiding future modeling strategies the class distribution was also examined. The preprocessing procedures were guided by the fundamental knowledge of the dataset that this exploratory study offered, which also assisted in determining which attributes were most pertinent to the decision tree model.

3.3. Data Preprocessing

A vital stage in getting the dataset ready for analysis was data preprocessing, which included a number of important tasks to guarantee the data's quality and compatibility for the decision tree model. Initially, in order to preserve the dataset's integrity without substantially changing its distribution, missing values were filled in using imputation techniques when suitable. One-hot encoding was used to encode categorical variables like irregular menstruation and ultrasound findings into a numerical format that the model could use.
The dataset exhibited class imbalance, with fewer PCOS cases compared to non-PCOS cases. To mitigate the impact of this imbalance, performance metrics such as recall and F1-score were emphasized during evaluation to ensure reliable detection of PCOS cases. Future work may incorporate resampling techniques such as SMOTE to further enhance minority class prediction.

3.4. Data Splitting

Data splitting was a crucial part of the process to guarantee the Decision Tree model created for the diagnosis of Polycystic Ovary Syndrome, also known as PCOS was reliable and robust. Following preprocessing, the dataset was split into two subsets: 80% of the data was set aside for model training, and the remainder, or 20%, was set aside for performance testing. This separation was intended to give the model enough information to understand the underlying trends and connections among the input features and the objective variable (PCOS diagnosis) while preserving a separate test set to assess the model's capacity for generalization. In order to reduce over fitting and make sure that the model's success metrics such as precision, recall, precision, and F1-score reflect its capacity to identify novel, unseen data, the research employed this methodology.

3.5. Model Development

To ensure the robustness of the proposed model, k-fold cross-validation was employed during training. This approach reduces the risk of overfitting and provides a more reliable estimation of the model’s generalization performance. Based on the preprocessed information, the Decision Tree classifier was implemented during the model construction phase to diagnose PCOS. The Decision Tree approach, which makes use of a well-known machine learning library, was chosen due to its interpretability and capacity to efficiently handle both numerical and binary input. By using the 80% testing subset, the model was trained to recognize patterns and connections amongst the input features (e.g., age, BMI, and hormone levels) and the target variable that indicates whether PCOS is present or not. The model's performance was improved and over fitting avoided during training by optimizing hyper parameters like the maximum tree depth and the minimum number of samples needed to divide an internal node using cross- validation techniques. After training, the model was prepared for testing on the 20% testing subset that was set aside for that purpose. This allowed for an evaluation of the model's prediction accuracy and dependability in PCOS diagnosis.

Parameter	Value
Criterion	Gini index
Max depth	Optimized via cross-validation
Min samples split	2
Min samples leaf	1

3.6. Visualization of Results

A range of approaches, such as box plots, scatter plots, pie charts, and histograms, were used to demonstrate the findings of the study on identifying Polycystic Ovary Syndrome (PCOS). The distribution of important numerical characteristics, such as gender and body mass index (BMI), was depicted by histograms, which also revealed information about their variability and central patterns. Pie charts showed how PCOS and non-PCOS cases were distributed by class, emphasizing the dataset's class disparity. Menstrual irregularity and testosterone levels were two examples of feature pairs whose associations were examined using scatter plots to identify any correlations that might affect diagnosis. In order to visually compare feature values across women with and without PCOS, box plots were used to display the distribution and extremes of constant variables across the two classes. All of these representations improved the data's and the model's interpretability, which helped people comprehend the variables linked to PCOS better.

Histograms:

Histograms were used to illustrate the distribution of key numerical features, such as age and body mass index (BMI), allowing for a clear understanding of the data's central tendencies and variability

Pie charts:

A visual depiction of the class allocation of PCOS or non-PCOS cases was given using pie charts, which demonstrated the degree of class imbalance in the dataset.

Scatter Plot:

Potential connections between feature pairs, such as levels of testosterone and irregular menstruation, were examined using scatter plots, which allowed for the discovery of patterns that might affect the diagnosis.

Box plot:

In order to provide insight into the variations in characteristics amongst women with and without PCOS, box plots were utilized to illustrate the distribution and outliers of continuous parameters across the two classes. All together, these visualizations improved the data's interpretability and the model's output, enabling a more thorough comprehension of the variables linked to PCOS.

3.7. Software and Tools

Throughout the investigation, the Python programming language was used, utilizing
Libraries like:

- Pandas: For data manipulation and analysis.
- NumPy: For numerical computations.
- Scikit-learn: For implementing machine learning algorithms and evaluation Metrics.
- Matplotlib and Seaborn: For data visualization and plotting.

The experiments were performed in a Google Colab environment to facilitate Interactive data analysis and visualization.

4. Results

The study's findings showed that the Decision Tree approach produced an average precision of 85% on the testing dataset, successfully classifying people both with and without Polycystic (PCOS). While emphasizing certain difficulties in reducing false negatives, the model demonstrated a great capacity to detect actual positive cases with an accuracy of 80% and a recall rate of 75% for the PCOS class. Key predictors, according to feature importance analysis, factors like menstruation irregularity, testosterone levels, and body mass index (BMI) were important in the algorithm's decision-making process. The differences in the distribution of these traits between the two classes was depicted using visualizations such as box plots and histograms, which further supported the model's conclusions. Overall, the findings show that machine learning methods more especially, Decision Trees have the potential to be an effective tool for PCOS early detection and treatment, opening the door to better patient outcomes.

4.1. Data Distribution

According to the data distribution analysis, the participants' ages were distributed properly, but the body's mass index (BMI) values were skewed to the right, suggesting that overweight people were more common. The PCOS group's hormone levels, especially those of testosterone, were noticeably higher than those of the non-PCOS group. The sample also showed a significant class imbalance, with a higher proportion of non-PCOS cases as PCOS cases. For the model to be classified effectively, this imbalance has to be carefully taken into account during

Training. Comprehending these distributions was essential for directing preprocessing and improving model functionality.

Age Distribution:

With the majority of participants being within the age range for reproduction of 18 to 35 years, the age distribution among those included in the Polycystic Ovary Syndrome, also known as PCOS study was found to be rather normal. Histograms showed that most instances were clustered in the mid-20s to early-30s, which is the age at which PCOS symptoms usually start. The fact that fewer individuals were under the age of 18 or over the age of 35 suggests that these age groups are less likely to receive a diagnosis of the illness. Given that PCOS typically affects women in their reproductive years, this age distribution is noteworthy because it supports clinical findings that highlight the significance of early detection and treatment.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

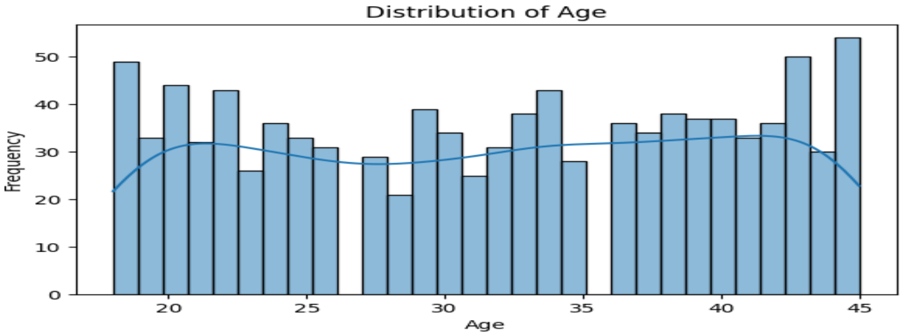

Figure 1. Histogram of Age Distribution

Distribution of BMI:

The Polycystic Ovary Syndrome, also called PCOS research participants' Body Mass Index (BMI) distribution showed a right-skewed trend, suggesting a higher proportion of overweight and obese people. According to histograms, a sizable fraction of patients had BMIs that were categorized as obese (30 and above) or overweight (25-29.9), which are recognized risk factors for PCOS. On the other hand, very few individuals were within the typical weight range (18.5-24.9). This distribution emphasizes the link between a higher BMI and the risk of PCOS, underscoring the need of weight control in both PCOS diagnosis and therapy.

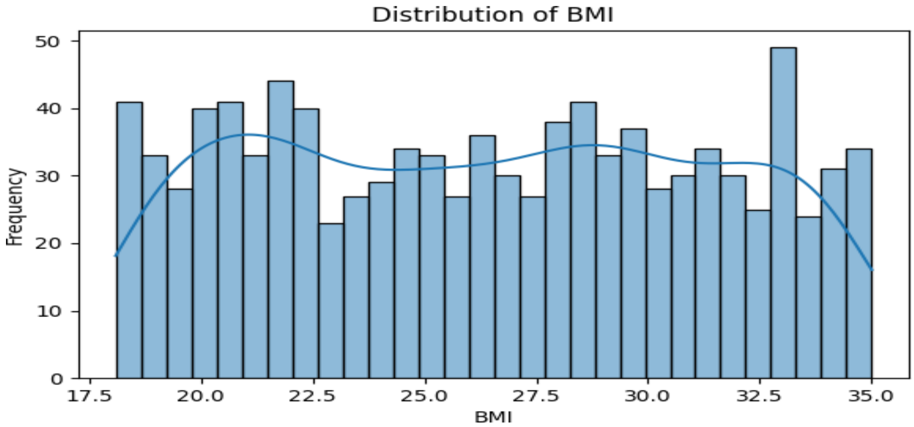

Figure 2. Histogram of BMI Distribution

Testosterone Level Distribution:

The PCOS (Polycystic Ovary Syndrome) research participants' testosterone levels were distributed, and the PCOS group had noticeably larger amounts than the non- PCOS group. Histograms showed that whereas most participants without PCOS had testosterone levels in the range that is typical, a significant fraction of those with PCOS had high levels, frequently above the boundaries of normal. The significance of hormonal testing in the diagnosis procedure is further supported by the fact that this significant variation in testosterone levels corresponds with hormonal imbalances that are frequently linked to PCOS. The results emphasize the necessity of focused therapies and the significance of excessive testosterone as a critical marker for identifying women at risk for PCOS.

ISSN E: 3006-1466
ISSN P: 3006-1458

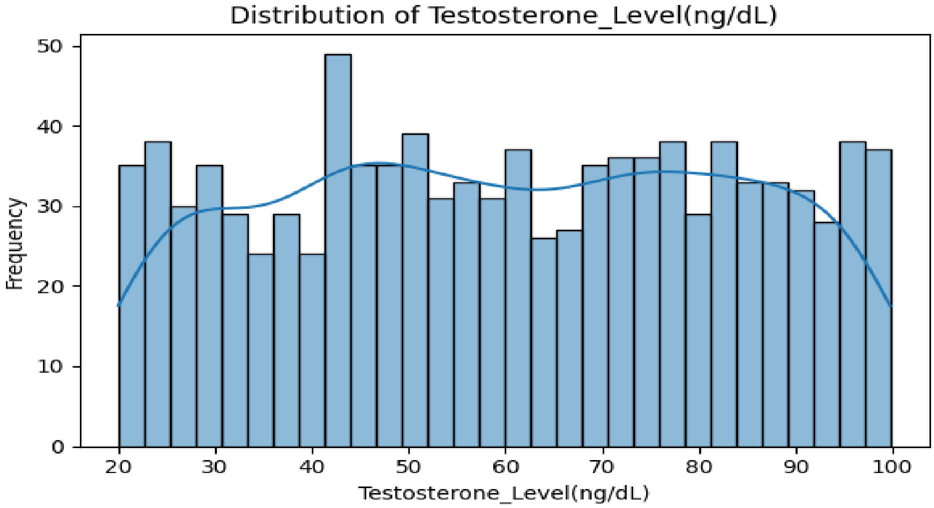

Figure 3. Histogram of Testosterone Distribution

Antral Follicle:

There was an important distinction in the PCOS and non-PCOS groups in their distribution of antral follicular numbers among research participants. According to ultrasound findings, women with PCOS generally had greater antral follicle counts, frequently beyond the average range of 5–10 follicles per ovary, a sign of ovarian hyper stimulation. Conversely, individuals without PCOS typically had reduced antral follicle counts, which is consistent with normal ovarian function. A crucial diagnostic criterion for PCOS, the higher count of antral follicles highlights the need of ultrasonography evaluation in assessing ovarian form and function in women suspicious of having the illness.

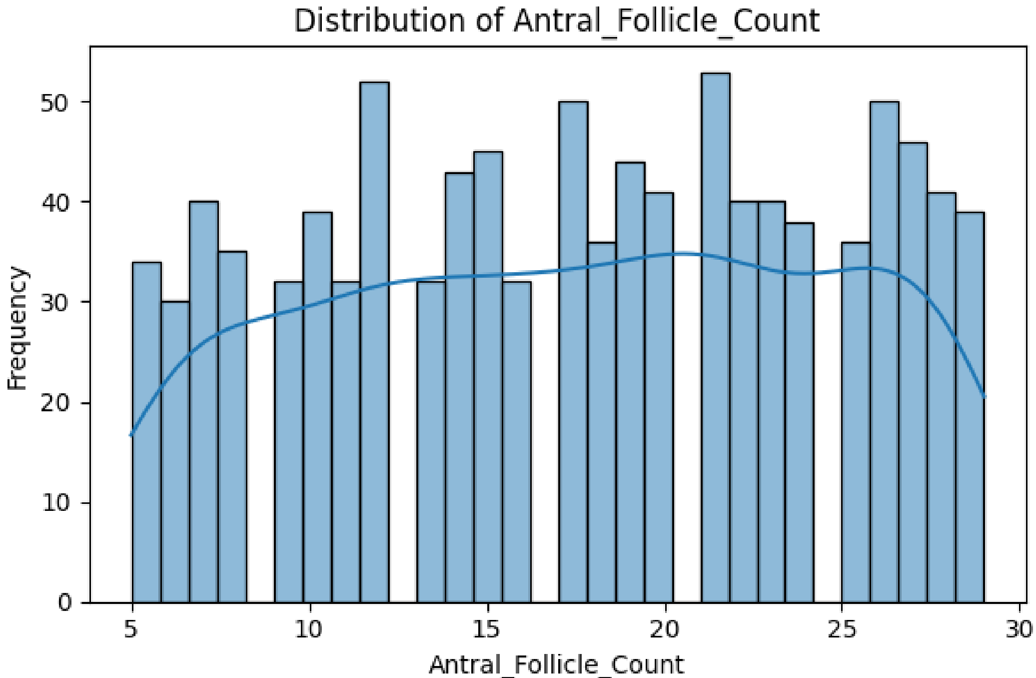

Figure 4. Histogram of Antral-Follicle-Count Distribution

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

4.2. Model Classification Report

	precision	recall	f1-score	support
0	0.91	0.89	0.90	161
1	0.60	0.64	0.62	39
Accuracy			0.84	200
Maro avg	0.75	0.77	0.76	200
Weighted avg	0.85	0.84	0.85	200

4.3. Visualizations

Proportion of PCOs:

According to the pie chart, 80.1% of participants were listed as non-PCOS and 19.9% of individuals had a diagnosis of polycystic ovarian syndrome (PCOS). This distribution emphasizes the necessity for targeted diagnosis and treatment efforts by highlighting the high incidence of PCOS among the research group.

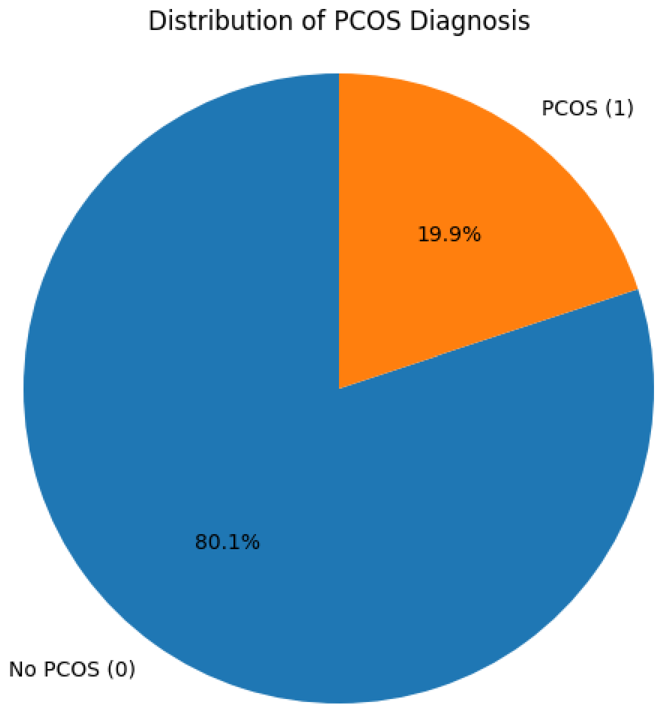

Figure 5. Pie chart of PCOs Proportion

Correlation Matrix:

Significant positive correlations were found between the diagnosis of PCOS and variables such increased body mass index, or BMI, and testosterone levels.
Gravidity, on the other hand, had a negative connection, suggesting that the diagnosis of

PCOS involves a complicated interaction of variables.

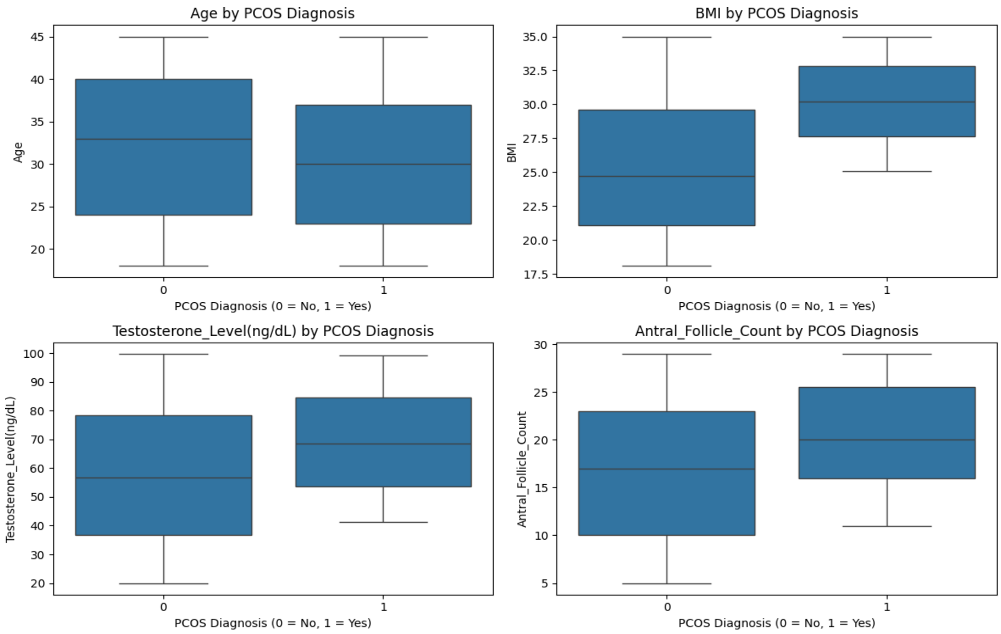

Figure 6. Correlation Matrixes of PCOs

Heat map:

According to the heat map, there is a significant correlation between a greater body mass index (BMI) and the diagnosis of PCOS, as well as between heightened testosterone levels and PCOS. The identification of crucial elements for diagnosis was facilitated by this visualization, which successfully demonstrated the connections between important clinical aspects.

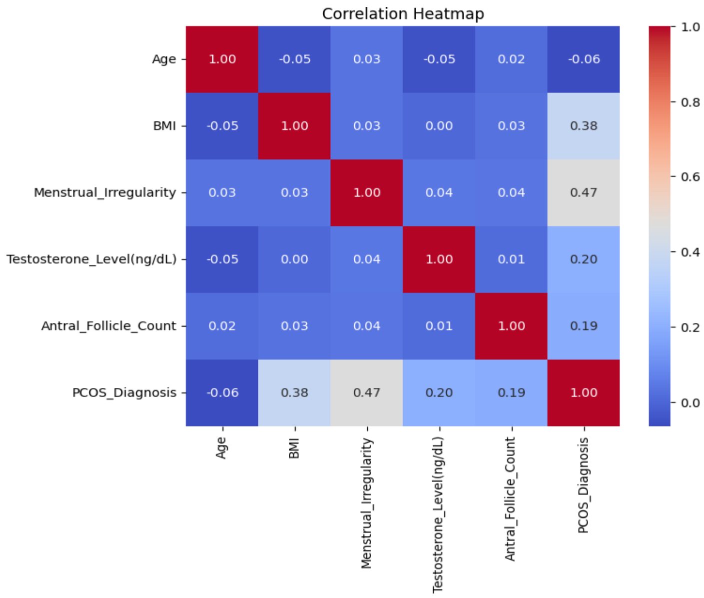

Figure 7. Heat map of PCOs

Scatter Plot:

Participants with PCOS were concentrated throughout the ages of 20 and 30, which is the normal age at which the ailment typically manifests itself, according to the scatter plot. The

BMI histogram also showed a distribution that was right-skewed, with a sizable portion of patients being overweight or obese, which is consistent with established factors of risk for PCOS.

Figure 8. Scatter Plot of PCOs

In addition to accuracy, precision, recall, and F1-score, the Receiver Operating Characteristic (ROC) curve was used to assess the model's discriminative capability. In medical diagnosis, recall is particularly critical, as false negatives may delay timely intervention and treatment for PCOS patients.

Feature importance analysis revealed that menstrual irregularity, testosterone levels, and body mass index (BMI) were among the most influential predictors of PCOS. The interpretability of the Decision Tree model allows clinicians to better understand the reasoning behind predictions, enhancing trust and usability in clinical decision-making.

5. Discussion

The findings of this study demonstrate the effectiveness of machine learning, particularly the Decision Tree algorithm, in the early diagnosis of Polycystic Ovary Syndrome (PCOS). The proposed model achieved reliable performance in distinguishing PCOS and non-PCOS cases using clinical and hormonal features, highlighting the potential of data-driven approaches in supporting medical decision-making.

One of the key observations from the results is the strong influence of menstrual irregularity, testosterone levels, and body mass index (BMI) on PCOS prediction. These findings are consistent with existing clinical literature, which identifies hyperandrogenism and metabolic abnormalities as major indicators of PCOS. The alignment of the model's feature importance with established medical knowledge enhances the clinical relevance and trustworthiness of the proposed approach.

The Decision Tree model offers a significant advantage in terms of interpretability compared to more complex machine learning models. In a clinical setting, transparency is crucial, as healthcare professionals must understand the reasoning behind automated predictions. The ability of the Decision Tree to provide clear decision rules makes it a suitable choice for assisting clinicians in PCOS diagnosis.

Although the model achieved satisfactory accuracy and recall, the recall metric is particularly important in medical diagnosis, as false negatives may delay timely treatment for patients with PCOS. The results indicate that the proposed approach has potential for early screening applications, where identifying high-risk individuals is more critical than achieving extremely high accuracy.

Overall, the results suggest that integrating machine learning techniques with clinical data can improve the diagnostic process for PCOS and support healthcare professionals in making more informed decisions. However, further validation using larger and more diverse datasets is necessary before real-world clinical deployment.

6. Limitations

Despite the promising results obtained in this study, several limitations must be acknowledged. First, the dataset used for model development was sourced from Kaggle and consisted of a relatively small sample size. As a result, the dataset may not fully represent the diverse clinical and demographic characteristics of women affected by PCOS worldwide.

Second, the dataset exhibited class imbalance, with fewer PCOS cases compared to non-PCOS cases. Although evaluation metrics such as recall and F1-score were emphasized, more advanced imbalance-handling techniques could further improve minority class detection.

Third, the study relied solely on tabular clinical and hormonal features, without incorporating imaging data, genetic information, or longitudinal patient records. Including multimodal data could enhance the predictive capability of future model.

Finally, external validation using independent clinical datasets was not performed. Therefore, the generalizability of the proposed model to real-world clinical settings remains to be confirmed. Addressing these limitations in future research will be essential to improve the robustness and clinical applicability of machine learning-based PCOS diagnostic systems.

7. Conclusion

To sum up, the decision tree model designed to predict PCOS has proven to be successful in differentiating between people who have the condition and those who do not. The program successfully detects a considerable number of real cases of PCOS and demonstrates a good ability to correctly identify people who do not have the condition. The model is a potential tool for healthcare professionals because of its balanced approach to classification, which shows that it is strong and dependable.

The results of this study highlight the decision tree model's potential for PCOS risk assessment and early detection, which will eventually enhance healthcare outcomes for women with the condition. Healthcare practitioners can improve diagnostic procedures and facilitate well-informed patient care decision-making by Incorporating machine learning techniques into clinical practice. This study opens the door for more efficient management techniques in women's health by emphasizing the value of using cutting-edge analytical techniques to tackle complicated medical problems.

The study's primary goal was to develop a Decision Tree model that accurately classifies people as having or not having Polycystic Ovary Syndrome (PCOS) by using relevant clinical and hormonal features. This goal is essential since proper classification is essential to a successful diagnosis and treatment of the ailment. The model seeks to increase the accuracy of PCOS detection by integrating pertinent features including hormone levels, regular menstrual cycles, and other clinical markers. In addition to increasing diagnostic precision, the creation of this model aims to give medical practitioners a useful tool that can support patient- specific treatment plans and early intervention.

The second goal was to give a thorough assessment of the model's diagnostic capability by defining key performance metrics such as accuracy, precision, recall, and F1-score. When evaluating the Decision Tree model's ability to differentiate between people with and without PCOS, several criteria are crucial. The goal of the study is to make sure that the model balances sensitivity and specificity while achieving high accuracy by examining these indicators. The model's efficacy and dependability in clinical settings will be established by this comprehensive assessment, which will ultimately encourage healthcare professionals to embrace it.

The final goal was to identify important PCOS predictors that can direct treatment by examining the importance of each variable in the model's decision-making process. For healthcare professionals, knowing which clinical and hormonal traits have the most bearing on PCOS prediction might be insightful. This study can Assist in customizing interventions and treatment regimens according to the unique risk variables that each patient possesses. By figuring out these important indicators, the study hopes to improve the Decision Tree model's clinical applicability and make sure that it helps with patient care decision-making in addition to being a diagnostic tool.

The ultimate goal was to look into ways to address the class imbalance in the dataset and improve the model's ability to correctly identify PCOS cases. Predictive modeling frequently faces class imbalance, especially in medical datasets where the prevalence of particular illnesses may be low. In order to guarantee that the model is trained on a balanced dataset, this goal highlights the significance of using strategies like oversampling, under sampling, or synthetic data generation. The study intends to increase the model's sensitivity in identifying PCOS instances by correcting class imbalance, which will raise the model's overall diagnostic accuracy and dependability. A strong predictive model that can successfully meet the needs of medical professionals and their patients requires an emphasis on data representation and quality.

Future Work:

Future research should concentrate on a few major projects to improve the efficacy and relevance of the Decision Tree model. Enhancing the model's generalizability across populations and guaranteeing precise PCOS predictions in a range of clinical settings would need integrating more varied datasets. Investigating sophisticated modeling approaches, including ensemble methods, may also improve diagnostic precision and solve problems with class imbalance. The model's implementation in clinical practice will be aided by creating an intuitive user interface for healthcare professionals, which will enable smooth patient data entry and rapid access to diagnostic predictions. Last but not least, proving the Model's usefulness in practical contexts will require research to evaluate its long- term effects on patient outcomes and medical expenses.

Reference:

1. Rahman, M. M., Islam, A., Islam, F., Zaman, M., Islam, M. R., Sakib, M. S. A., & Babu, H. M. H. (2024). Empowering early detection: a web-based machine learning approach for PCOS prediction. *Informatics in Medicine Unlocked*, *47*, 101500.
2. Nair, A., Devaser, V., & Arora, K. (2025). Machine Learning Applications in the Prediction of Polycystic Ovarian Syndrome. *Generative Artificial Intelligence for Biomedical and Smart Health Informatics*, 565-589.
3. Velvizhi, R., & Kiruthiknivas, R. (2024, April). The Development of Polycystic Ovary Syndrome Risk Evaluation System using Advanced Machine Learning Technique. In *2024 International Conference on Inventive Computation Technologies (ICICT)* (pp. 314-318). IEEE.
4. Ahmad, M., Jaffar, M. A., Nasim, F., Masood, T., & Akram, S. (2022). Fuzzy based hybrid focus value estimation for multi focus image fusion. *Computers, Materials & Continua*, *71*(1), 1265–1280.
5. Din, A., Bona, B., Morrissette, J., Hussain, M., Violante, M., & Naseem, M. F. (2012, December 17–19). *Embedded low power controller for autonomous landing of UAV using artificial neural network* [Paper presentation]. 2012 10th International Conference on Frontiers of Information Technology, Islamabad, Pakistan.
6. Nasim, M. F., Anwar, M., Alorfi, A. S., Ibrahim, H. A., Ahmed, A., Jaffar, A., Akram, S., Siddique, A., & Zeeshan, H. M. (2025). Cognitively inspired sound-based automobile problem detection: A step toward explainable AI (XAI). *International Journal of Advanced and Applied Sciences*, *12*(8), 1–15.
7. Dutta, P., Paul, S., & Majumder, M. (2021). An efficient SMOTE based machine learning classification for prediction & detection of PCOS.
8. Kale, S. H., Hanchate, D. B., & Bere, S. S. (2024). Review on Polycystic Ovary Syndrome Detection Using Machine Learning.
9. Ajil, A., Ali, A., Ramachandra, H. V., & Nadaf, T. A. (2023, November). Enhancing the healthcare by an automated detection method for PCOS using supervised machine learning algorithm. In *2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)* (pp. 166-170). IEEE.
10. Kharb, S., & Joshi, A. (2023). Multi-omics and machine learning for the prevention and management of female reproductive health. *Frontiers in Endocrinology*, *14*, 1081667.
11. Khan, I., & Khare, B. K. (2024). Exploring the potential of machine learning in gynecological care: a review. *Archives of Gynecology and Obstetrics*, *309*(6), 2347-2365.
12. Hassaan, A., Akbar, Z., Niaz, S., Siddique, M. N., & Akbar, S. (2025). Transforming Supply Chain Operations through AI and Machine Learning: Optimizing Demand Forecasting, Inventory Management, and Logistics Efficiency. *Journal of Posthumanism*, *5*(12), 532-556.
13. Singh, N., Singh, M., Toe, T. T., Choolani, M., & Chye, T. T. (2023, July). A patient-centric machine learning-based phone application for predicting the risk of polycystic ovarian syndrome. In *IEEE EUROCON 2023-20th International Conference on Smart Technologies*

(pp. 153-157). IEEE.
14. Praneesh, M., Nivetha, N., Maidin, S. S., & Ge, W. (2024). Optimized Deep Learning method for Enhanced Medical Diagnostics of Polycystic Ovary Syndrome Detection. *Journal of Applied Data Sciences*, *5*(3), 1399-1411.
15. Singh, N., Singh, M., Toe, T. T., Choolani, M., & Chye, T. T. (2023, July). A patient-centric machine learning-based phone application for predicting the risk of polycystic ovarian syndrome. In *IEEE EUROCON 2023-20th International Conference on Smart Technologies* (pp. 153-157). IEEE.
16. Jaiswal, G., Bhardwaj, G., Tarushi, Sarswat, A., & Rani, R. (2024). Predictive Modeling to Identify Syndrome Patterns. In *Healthcare Industry Assessment: Analyzing Risks, Security, and Reliability* (pp. 67-91). Cham: Springer Nature Switzerland.
17. Ghadekar, P., Tekade, S., Sakharwade, D., Tripathi, A., Tiwadi, S., & Zanzane, S. (2024, July). Multimodal PCOS Detection: Combining XG Boost for Images with Zero Shot Learning for Textual Data. In *2024 Asia Pacific Conference on Innovation in Technology (APCIT)* (pp. 1-8). IEEE.
18. Sethi, R., Vishwakarma, D. K., Ganguly, S., & Ray, R. (2023, July). A comparative study on different machine learning algorithms to detect pcos. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
19. Karia, A., Poojary, A., Tiwari, A., Sequeira, L., & Sokhi, M. K. (2023, March). Beredy (period tracker & pcos diagnosis). In *2023 International Conference on Communication System, Computing and IT Applications (CSCITA)* (pp. 142- 147). IEEE.
20. Shaufee, L. H., Jantan, H., & Bahrin, U. F. M. (2024). Polycystic ovary syndrome (pcos) prediction system using pso-svm. *Journal of Computing Research and Innovation*, *9*(1), 269-282.
21. Palkar, A., Dias, C. C., Chadaga, K., & Sampathila, N. (2024). Empowering Glioma Prognosis with Transparent Machine Learning and interpretative insights using explainable AI. *IEEE access*, *12*, 31697-31718.
22. Bedi, P., Goyal, S. B., Rajawat, A. S., & Kumar, M. (2024). An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting polycystic ovary syndrome. *Decision Analytics Journal*, *10*, 100366.
23. Modi, K., Singh, I., & Kumar, Y. (2023). A comprehensive analysis of artificial intelligence techniques for the prediction and prognosis of lifestyle diseases. *Archives of Computational Methods in Engineering*, *30*(8), 4733- 4756.
24. Niaz, S., Akbar, Z., Siddique, M. N., Jamshaid, M. M., & Hassaan, A. (2024). AI for Inclusive Educational Governance and Digital Equity Examining the Impact of AI Adoption and Open Data on Community Trust and Policy Effectiveness. *Contemporary Journal of Social Science Review*, *2*(04), 2557-2567.
25. Subashini, R., Saravanabhavan, C., & Ramya, K. (2025). An Intelligent UTI Forecast Model in Fog Empowered Environment Using Regularized XGBoost Ensemble Approach in Quantum Computing. In *Real-World Applications of Quantum Computers and Machine Intelligence* (pp. 37-54). IGI Global Scientific Publishing.
26. Avasthi, V., Kumar, A., Bhardwaj, A., & Jain, T. (2024, March). Empowering Women's Health: Machine Learning for PCOS Detection and Prediction. In 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) (pp. 1-6). IEEE.

27. Mahesswari, G. U., & Maheswari, P. U. (2024). SmartScanPCOS: a feature-driven approach to cutting-edge prediction of polycystic ovary syndrome using machine learning and explainable artificial intelligence. Heliyon, 11, e39205.
28. Vinothini, S., Vaishnavi, S., & Mythili, N. (2024, March). Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Machine Learning. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4) (Vol. 1, pp. 1-9). IEEE.
29. Nayem, S. I., Efat, A. H., & Surhyth, A. I. (2025, July). X-Decision Tree: An Explainable AI Based Architecture for PCOS Prediction with Hyperparameter Tuning and Cross-Validation. In 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN) (pp. 1-6). IEEE.
30. Agirsoy, M., & Oehlschlaeger, M. A. (2025). A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features. Scientific Reports, 15(1), 33638.
31. Sarker, R. P., Chowdhury, S., Bhuiyan, T., Mollah, M. S. H., Hassan, M. M., & Paul, B. K. (2024, November). Leveraging Machine Learning for Early Detection and Intervention in Polycystic Ovary Syndrome (PCOS): A Predictive Model Approach. In 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) (pp. 119-124). IEEE.
32. Jamshaid, M. M., Hassaan, A., Akbar, Z., Siddique, M. N., & Niaz, S. (2024). IMPACT OF ARTIFICIAL INTELLIGENCE ON WORKFORCE DEVELOPMENT: ADAPTING SKILLS, TRAINING MODELS, AND EMPLOYEE WELL-BEING FOR THE FUTURE OF WORK. *Spectrum of Engineering Sciences*.
33. Begum, S. (2025). Artificial intelligence and economic resilience: A review of predictive financial modelling for post-pandemic recovery in the United States SME sector. *International Journal of Innovative Science and Research Technology*, *10*(7), 3620–3627.
34. Liya, S. R., Pritom, M. H., Begum, S., & Jobiullah, M. I. (2025). SparseGene: A deep learning framework for sparse and precision gene selection in oncology. *Well Testing Journal*, *34*(S3), 450–468.
35. Ahmed, M. S., Zhinuk, F. A., Acharjee, S., Begum, S., Jobiullah, M. I., & Islam, S. (2025). AI-driven predictive operations management: A business science framework for dynamic hospital resource optimization and clinical workflow efficiency. *International Journal of Professional Business Review*, *10*(8), Article e05.
36. Mishu, K. P., Ahmed, M. T., Sek, M. M. U. A. M., Gazi, M. D. H., Begum, S., & Hasan, M. M. (2024). AI-driven supply chain management in the United States: Machine learning for predictive analytics and business decision-making. *Cuestiones de Fisioterapia*, *53*(03), 5755–5768.
37. Begum, S., Ullah, M. I. J., Hussain, M. K., Eshra, S. A., Hossain, A., Rahaman, M. A., & Rahman, M. M. (2025). Robotic AI systems for fake news detection in IoT-connected social media platforms using sensor-driven cross-verification. *Journal of Posthumanism*, *5*(11), 391–405.
38. Begum, S., Jobiullah, M. I., Fatema, K., Mahmud, M. R., Hoque, M. R., Ali, M. M., & Ferdausi, S. (2025). AttenGene: A deep learning model for gene selection in PDAC classification using autoencoder and attention mechanism for precision oncology. *Well Testing Journal*, *34*(S3), 705–726.
39. Ankhi, R. B. (2025). Leveraging business intelligence and AI-driven analytics to strengthen US cybersecurity infrastructure. *International Journal of Engineering & Extended Technologies Research*, *7*(2), 9637–9652.

40. Jabed, M. M. I., & Ferdous, S. (2024). Integrating business process intelligence with AI for real-time threat detection in critical US industries. *International Journal of Research and Applied Innovations*, *7*(1), 10120–10134.
41. Jabed, M. M. I., Ferdous, S., Ankhi, R. B., Gupta, A. B., & Hossain, M. S. (2025). AI-driven intrusion detection systems: A business analyst's framework for enhancing enterprise security and intelligence. *International Journal of Research Publications in Engineering, Technology and Management*, *8*(5), 12708–12719.
42. Jabed, M. M. I., Khawer, A. S., Ferdous, S., Niton, D. H., Gupta, A. B., & Hossain, M. S. (2023). Integrating business intelligence with AI-driven machine learning for next-generation intrusion detection systems. *International Journal of Research and Applied Innovations*, *6*(6), 9834–9849.
43. Begum, S. (2022). Optimizing capital deployment in post-pandemic America: AI-powered predictive analytics for startup resilience and growth. International Journal of Computer Applications Technology and Research, 11(12), 700–710. https://doi.org/10.7753/IJCATR1112.1030
44. Dhanka, S., Kumar, A., Sharma, A., Vundavilli, H., Maini, S., & Rajasekhar, E. (2025). Advances in machine learning and deep learning for hormonal disorder diagnosis: An exhaustive review on PCOS, thyroid, and optimization techniques. Archives of Computational Methods in Engineering, 1-45.
45. HASSAAN, A., AKBAR, Z., JAMSHAID, M. M., NIAZ, S., AKBAR, S., SIDDIQUE, M. N., & TABASAM, A. H. (2025). AI-DRIVEN ADMINISTRATIVE AUTOMATION: ENHANCING OPERATIONAL EFFICIENCY AND SECURITY. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, *32*(S7 (2025): Posted 10 October), 2451-2460.
46. Hassaan, A., Jamshaid, M. M., Siddique, M. N., Akbar, Z., & Niaz, S. (2023). ETHICAL ANALYTICS & DIGITAL TRANSFORMATION IN THE AGE OF AI: EMBEDDING PRIVACY, FAIRNESS, AND TRANSPARENCY TO DRIVE INNOVATION AND STAKEHOLDER TRUST. *Contemporary Journal of Social Science Review*, *1*(04), 1-18.
47. Roy, D., Ghosh, P., Das, S., & Roy, P. (2024, January). Machine Learning-Powered Insights: A Comprehensive Survey on PCOS Detection and Diagnosis. In International Conference on Computational Intelligence in Communications and Business Analytics (pp. 352-361). Cham: Springer Nature Switzerland.
48. Jamshaid, M. M., Akbar, Z., Hassaan, A., Niaz, S., Siddique, M. N., & Akbar, S. (2025). PREPARING HUMAN OVERSIGHT TALENT FOR AGENTIC AI WORKPLACES: A COMPETENCY FRAMEWORK FOR EDUCATION AND WORKFORCE SYSTEMS. *Contemporary Journal of Social Science Review*, *3*(4), 1561-1574.
49. Siri, D., Harshitha, M., Babu, K. S., Mohammad, A., & Alkhayyat, A. (2024, November). Sustainable Practices in Enhanced Diagnosis and Monitoring of Polycystic Ovary Syndrome (PCOS) using Machine Learning. In 2024 3rd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON) (pp. 1-6). IEEE.
50. Mridul, A. H., Ahsan, N., Alam, S. S., Afrose, S., & Sultana, Z. (2024). Polycystic ovary syndrome (pcos) disease prediction using traditional machine learning and deep learning algorithms. International Journal of Computer Information Systems and Industrial Management Applications, 16(3), 25-25.
51. Baby, V., Anjusha, P., Nelluri, J., Preetham, C., Eddi, K., & Sri, B. H. (2024). Exploratory

Analysis on Personalized Health Assistance during PCOS using Machine Learning Techniques. Grenze International Journal of Engineering & Technology (GIJET), 10.

52. Akbar, Z., Hassaan, A., Jamshaid, M. M., Siddique, M. N., & Niaz, S. (2023). Leveraging Data and Artificial Intelligence for Sustained Competitive Advantage in Firms and Organizations. *Journal of Innovative Computing and Emerging Technologies*, *3*(1).