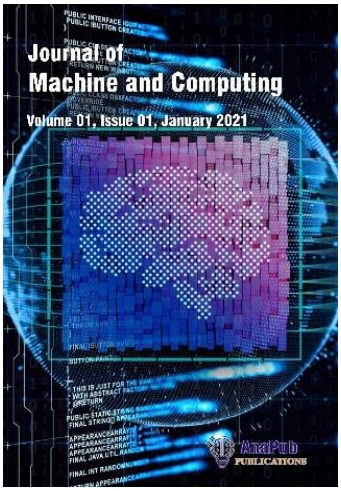
Journal Pre-proof

A Machine Learning Approach for Efficient Identification and Severity Grading of PCOS and PCOD Using Optimized Features



Suvika K V and Jyothi D G

DOI: 10.53759/7669/jmc202505156

Reference: JMC202505156

Journal: Journal of Machine and Computing.

Received 28 May 2025

Revised from 01 July 2025

Accepted 10 July 2025

Please cite this article as: Suvika K V and Jyothi D G, "A Machine Learning Approach for Efficient Identification and Severity Grading of PCOS and PCOD Using Optimized Features", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505156.

AnaPub
PUBLICATIONS

A Machine Learning Approach for Efficient Identification and Severity Grading of PCOS and PCOD Using Optimized Features

Suvika K V[1] and D G Jyothi[2]

[1]Research Scholar, Dept. of CSE, Bangalore Institute of Technology, Visvesveraya Technological University, Belagavi – 590018, kvsuvika@gmail.com

[2]Professor, Dept. of AI&ML, Bangalore Institute of Technology, Visvesveraya Technological University, Belagavi – 590018, dgjyothi@bit-bangalore.edu.in

ABSTRACT

Infertility, metabolic issues, and hormone imbalance are common symptoms of PCOS, a common endocrine illness affecting women of reproductive age. A variety machine learning technique are used in the research for PCOS severity grading and prediction. Recursive Feature Elimination (RFE) was used to choose features first, and then Random Forest and Logistic Regression, two supervised classifiers, applied. The models' efficacy in predicting PCOS was validated by their strong accuracy and AUC rating. Anti-Müllerian Hormone (AMH) was used as a crucial grading marker after patients were categorized into severity categories (Severe, Moderate, and Low) based on clinical criteria using unsupervised clustering methods, specifically K-Means and Agglomerative Clustering. Well-separated clusters were shown by the silhouette scores used to evaluate the clustering models. A comprehensive framework for early PCOS detection and phenotypic grading is provided by the combination of supervised and unsupervised techniques, which also offers insightful information for individualized treatment plans.

Index Terms – PCOS, PCOD, Machine Learning, Feature Selection, AMH (Anti-Müllerian Hormone), Severity Grading, Silhouette Score, Clinical Decision Support

1. INTRODUCTION

PCOS (Polycystic Ovary Syndrome) and PCOD (Polycystic Ovary Disease) are the most common pervasive endocrine disorders affecting women of reproductive. Globally, 8-13% of women [1] are affected approximately, and in India specifically has been affected more with rates of 22-26% [2] based on diagnostic criteria [2]. Hormonal imbalances, irregular menstrual cycles, hyperandrogenism are the few characteristics which are used for diagnosing the disorder and the presence of polycystic ovaries, often leading to obesity, infertility, insulin

resistance, and long-term cardiovascular risks [3]. The severity assessment and accurate diagnosis of PCOS and PCOD still remains clinically challenged despite their prevalence and impact because of less standardized diagnostic tools and due to heterogeneity in presentation of symptoms.

In medical diagnostics machine learning (ML) has emerged as a transformative technology in recent years, which led to analysing of datasets which are more complex for uncovering patterns which may not be evident through conventional methods [4]. ML models have demonstrated potential in lowering diagnostic subjectivity, automating clinical decision-making, and increasing the accuracy of illness diagnosis. However, many ML approaches for detecting PCOS/ PCOD rely on large features sets that also introduces overfitting, redundancy, and interpretation difficulties. These models are statistically robust but often lack clinical feasibility because of their data requirements and complexity [5].

By using optimized and minimal clinical parameters, the identification and severity grading can be done efficiently with the proposed machine learning based framework. To simplify the model and to enhance interpretability feature reduction is done which is important for real-world clinical integration. Rigorous selection techniques such as RFE (Recursive Feature Elimination) and correlation-based filtering are used to select features, ensuring that only the non-redundant and informative features are retained. The method strikes a balance between practicality and diagnostic performance, especially for application in healthcare settings with limited resources or time constraints.

In this research paper, unique component is the data-driven grading mechanism – which classifies the seriousness of the condition into clinically meaningful classes – such as severe, moderate, and mild. Grading is crucial for risk assessment, illness progression tracking, and customizing treatment regimens. The grading methods which are followed till now are either physician dependent or qualitative. To overcome this disadvantage scalable and unbiased alternative is offered, which is more algorithmic based grading system an also consistent based on clustering techniques and statistical thresholds [6].

The conceptual framework for the complete machine learning process for diagnosis of PCOS/ PCOD and severity grading. To make it effective and understandable analysis, the method merges clinical data, feature optimization, classification, and AMH-based grading. Recursive Feature Elimination (RFE) and correlation analysis are the feature selection techniques used and the main targets are to:

(i) For the sake of accurate detection of PCOS and PCOD a minimal subset of clinically significant features ate to be identified.

(ii) AMH (Anti-Mullerian Hormone) levels are considered for severity grading mechanism and implemented, where patients are structured into clinically meaningful categories. To ensure the reliability and data-driven classification, unsupervised clustering approaches are used to carry out grading, and to evaluate the quality of cluster separation silhouette scores are used.

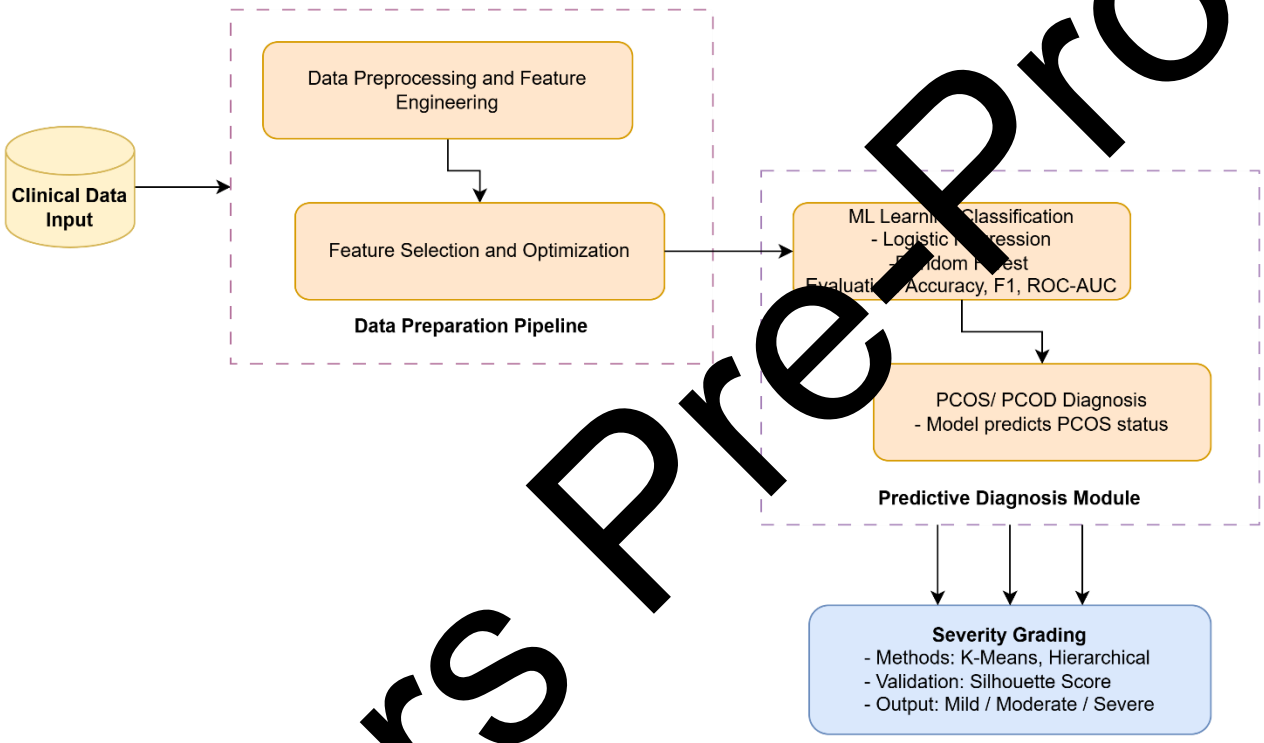


Figure 1: PCOS/ PCOD Diagnosis and Grading Framework

2. LITERATURE SURVEY

Numerous studies have been conducted on the creation of automated diagnosis tools utilizing machine learning techniques in response to the rising incidence of PCOS and PCOD. To predict PCOS based on clinical, hormonal, and ultrasound data, several studies have used classification techniques, including logistic regression, support vector machines, and ensemble methods. Feature selection has also become more popular to improve model performance and decrease dimensionality. Most of these studies, however, only consider binary or multi-class categorization for diagnosis purposes. The idea of automated severity grading, especially about objective indicators like AMH levels, has not been examined in the literature to yet, despite the clinical significance of comprehending illness severity. In addition to refining feature sets for

effective diagnosis, this study makes a distinct contribution by presenting a novel, data-driven method for grading PCOS/PCOD severity using unsupervised learning techniques. Figure 2 represents a graph where a total of 15 research papers from 2022 to 2025 were reviewed for this study.

PCOScare, created by Gandhi et al. (2023), combines classifiers such Random Forest, SVM, and XGBoost with filter, wrapper, and embedded feature-selection methods. Strong accuracy and interpretability were attained by their improved pipeline, which only addressed binary PCOS diagnosis and ignored severity grading. [7]

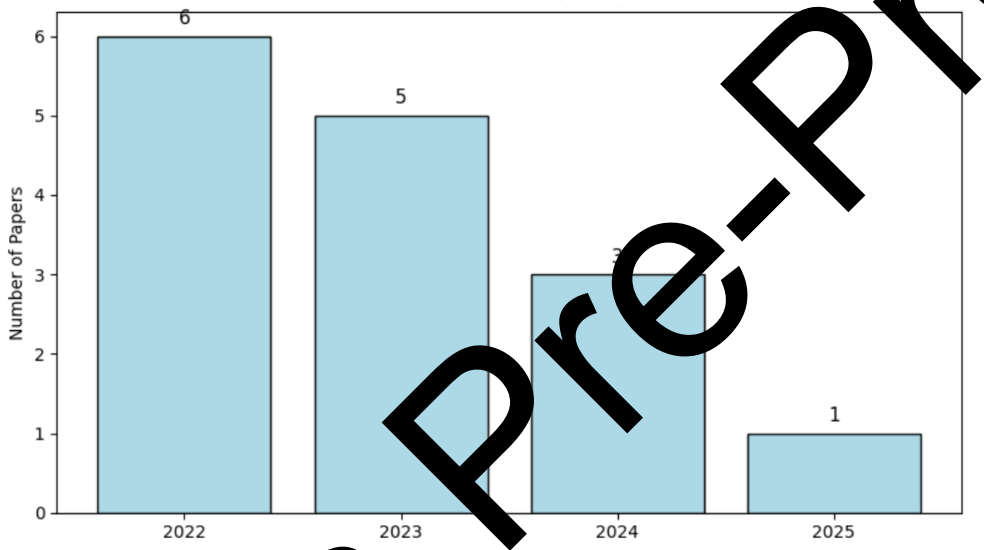


Figure 2 Number of Reviewed Papers per Year

PCONet, a CNN-based architecture for identifying PCOS from ovarian ultrasound images, was presented by Hosain et al. (2025). PCONet's accuracy of 98.12% compared to 96.56% for a fine-tuned InceptionV3 showed how effective transfer learning is for image-based diagnosis. Its application was restricted to categorization tasks, nevertheless. [8] Divekar & Sonawane (2024) used saliency maps and LIME to improve interpretability while using InceptionV3 for ultrasound picture classification. Despite achieving over 90% accuracy, their pipeline, like many others, lacked any kind of severity evaluation. [9] PCOS-WaveConvNet, introduced by Tiwari & Maheshwari (2023), preprocesses ultrasound images using wavelet transform (2D-DWT) before to CNN-based categorization. Although it demonstrated the benefits of multi-resolution analysis with an accuracy of 99.7%, it stayed solely focused on detection. [10]

A PCOS diagnosis model that combines explainable AI techniques with optimized biochemistry-based features was proposed by Elmannai et al. (2023). Their method, which used

SHAP/LIME, achieved strong classification performance without severity grading and provided insight into decision logic. [11] Alshakrani et al. (2022) used evolutionary algorithms to optimize an SVM classifier in order to increase detection accuracy and address class imbalance. While evolutionary feature selection improved model performance, the study did not stratify diseases or assess patient severity. [12] In order to improve accuracy, Faris & Miften (2022) combined SVM with evolutionary algorithms for feature selection. Although their pipeline placed a strong emphasis on optimization, it did not go beyond detection to include objective measures of PSC severity. [13]

In order to detect PCOS in its early stages, Hdaib et al. (2022) tested SVM, KNN, and Random Forest classifiers. The study only focused on diagnosis, not severity grading, even though their methodology enhanced sensitivity and specificity balances. [14] Linear Discriminant Analysis (LDA) was introduced by Joshi et al. (2022) for feature reduction before classification. Their approach demonstrated dimensionality reduction with 95.6% training and 91.7% validation accuracy, however it did not include stratified grading. [15]

An explainable machine learning pipeline that combines Random Forest and SHAP/LIME tools was created by Khanna et al. in 2025. This XAI-focused study maintained its sole focus on binary diagnosis tasks while greatly improving interpretability. [16] For PCOS identification, Venkatalakshmi & Regina (2024) examined DL (CNN) techniques in addition to conventional ML techniques (LR, SVM, RF). While summarizing the current level of early diagnoses, they pointed out that automated severity grading is conspicuously lacking in the literature. [17]

AUC values of 80–100% and detection accuracies ranging from 89–100% were reported in a comprehensive review (2022) of AI/ML studies, highlighting the fact that none provided clustering-based stratification or quantitative severity evaluation. [18] Although multimodal diagnostic approaches (2024) integrated ultrasound imaging and clinical data to increase detection accuracy, the fusion model lacked features for classifying or rating the severity of patients' conditions. [19] In their comparison of AdaBoost, GBDT, XGBoost, and CatBoost for PCOS prediction, Boosting Ensembles (2022) had good accuracy scores. Although these improved models lacked grading processes, they demonstrated diagnostic promise. [20] Though it did not concentrate on cluster-based stratification or diagnostic severity, Federated Learning for PCOS (2023) investigated privacy-preserving distributed training for treatment recommendation systems. [21]

Sl. No.	Author(s) & Year	Methodology Used	Key Contributions	Limitation (Grading Absent)
1	Gandhi et al. (2023) [7]	Filter, wrapper, embedded + RF, SVM	High accuracy and interpretability via PCOScare	No severity grading
2	Hosain et al. (2022) [8]	CNN-based PCONet	98.12% accuracy with ultrasound images	Detection only
3	Divekar & Sonawane (2024) [9]	InceptionV3 + LIME	>90% accuracy and interpretability	No grading mechanism
4	Tiwari & Maheshwari (2023) [10]	DWT preprocessing + WaveConvNet	99.7% accuracy on ultrasound data	Focused on binary classification
5	Elmannai et al. (2023) [11]	Optimized features + SHAP, LIME	Biochemical explainable PCOS detection	No clustering or grading
6	Alshakrani et al. (2022) [12]	Genetic algorithm + SVM	Feature optimization under imbalance	Severity not addressed
7	Faris & Miften (2022) [13]	SVM + GA	Robust pipeline for detection	No grading or clustering
8	Hdaib et al. (2022) [14]	RF, KNN, SVM	Improved specificity and sensitivity	Lacks stratification
9	Joshi et al. (2022) [15]	Deep LDA + ML classifiers	High validation accuracy	No severity framework
10	Khanna et al. (2023) [16]	RF + SHAP/LIME	Explainability and prediction	No grading system
11	Venkatalakshmi & Begina (2024) [17]	Literature review (ML & DL)	Highlights ML potential for early diagnosis	No implementation of grading
12	Systematic Review (2022) [18]	Meta-analysis of ML models	Reports 89–100% accuracy across methods	No severity-level insights
13	Multimodal Study (2024) [19]	Clinical + image fusion models	Increased accuracy through data fusion	Grading not implemented
14	Boosting Models (2022) [20]	AdaBoost, GBDT, XGBoost, CatBoost	High diagnostic accuracy	Focuses on prediction only
15	Federated Learning (2023) [21]	Privacy-preserving learning	Recommender system for treatment	No detection or grading

Table 1: Comparative Overview of Recent Machine Learning Approaches for PCOS/PCOD Detection Highlighting the Absence of Severity Grading Methodologies.

3. METHODOLOGY

The methodical methodology used for PCOS and PCOD detection and severity grading is described in this section. A thorough explanation of the dataset, the feature selection procedure used to determine important diagnostic parameters, and the clustering method used for severity rating are all included. The objective is to retain high accuracy and clinical relevance while ensuring an effective diagnosis with few features. The complete flow of working model is shown in Figure 3.

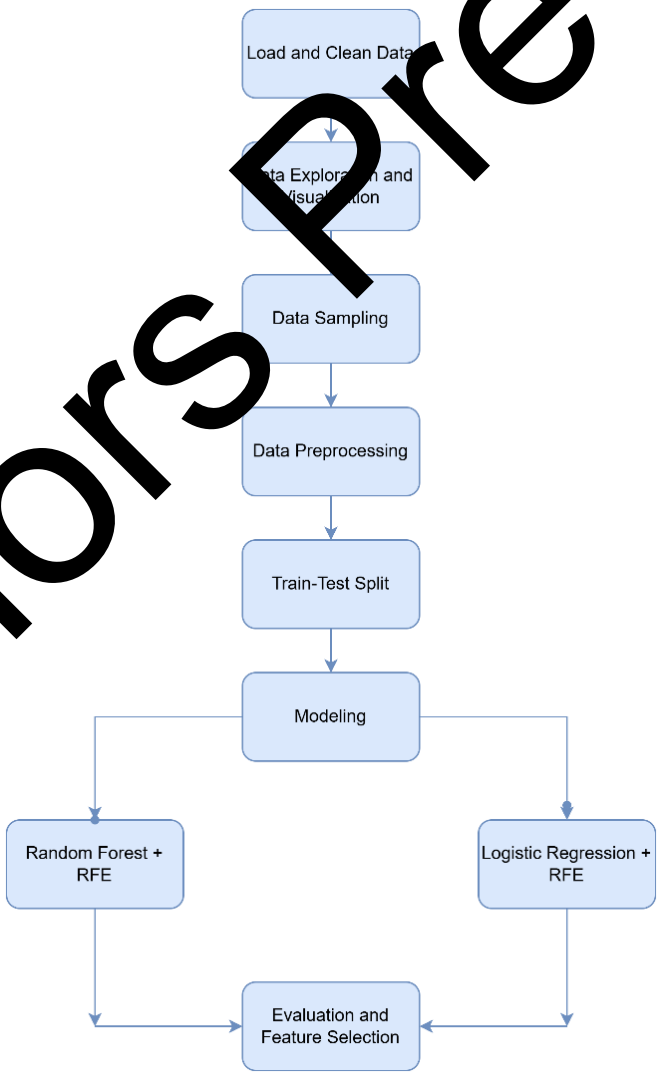


Figure 3: PCOS/PCOD Diagnosis and Modeling Workflow

3.1 DATASET DESCRIPTION

The dataset utilized in this work came from the open-access data science resource site Kaggle. It consists of two files: an Excel (XLSX) file [23] and a CSV file [24]. Clinical, demographic, and biochemical information about female patients that is pertinent to the diagnosis of PCOS and PCOD is included in both files. The CSV file includes the target variable i.e. PCOS (Yes/ No) which indicates whether patient has PCOS or not and infertility related medical values. This file has fewer columns but, includes diagnosis labels. XLSX file contains detailed patient features such as demographics, symptoms, lifestyle habits and lab results. Both files are merged based on patient file number to combine all features with actual labels. So, training of ML models can be done on complete dataset with both features and target.

3.2 DATA PROCESSING PIPELINE

The data is loaded and cleaned by filling null values with median for categorical columns and mean for all other columns and dropped duplicate/ unnecessary columns. The data is explored and visualized by plotting bar plots for categorical features (Yes/ No) as shown in Figure 4 and histograms for numerical features as shown in Figure 5. To visualize the feature relationship correlation heatmap is plotted.

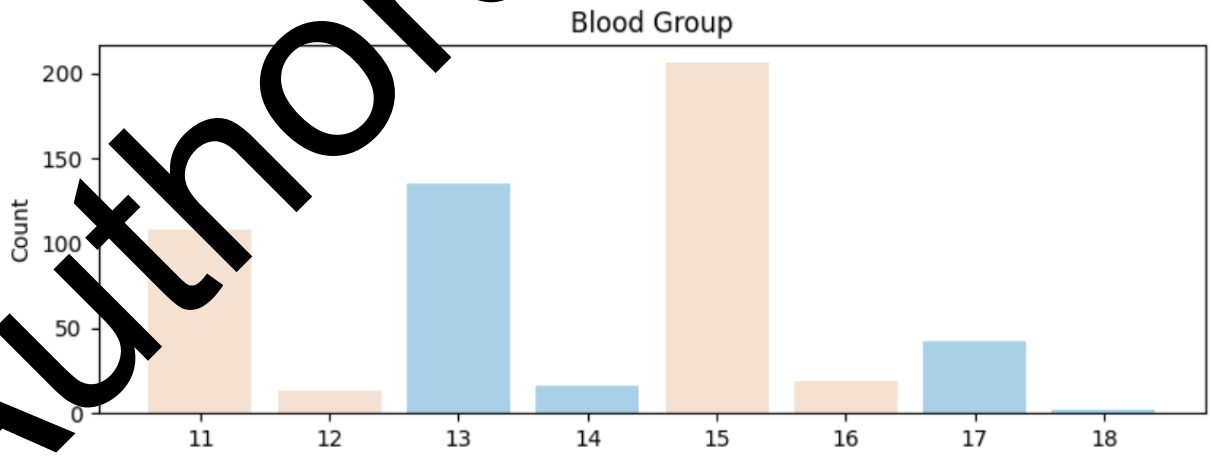


Figure 4: PCOS/PCOD Diagnosis and Modeling Workflow

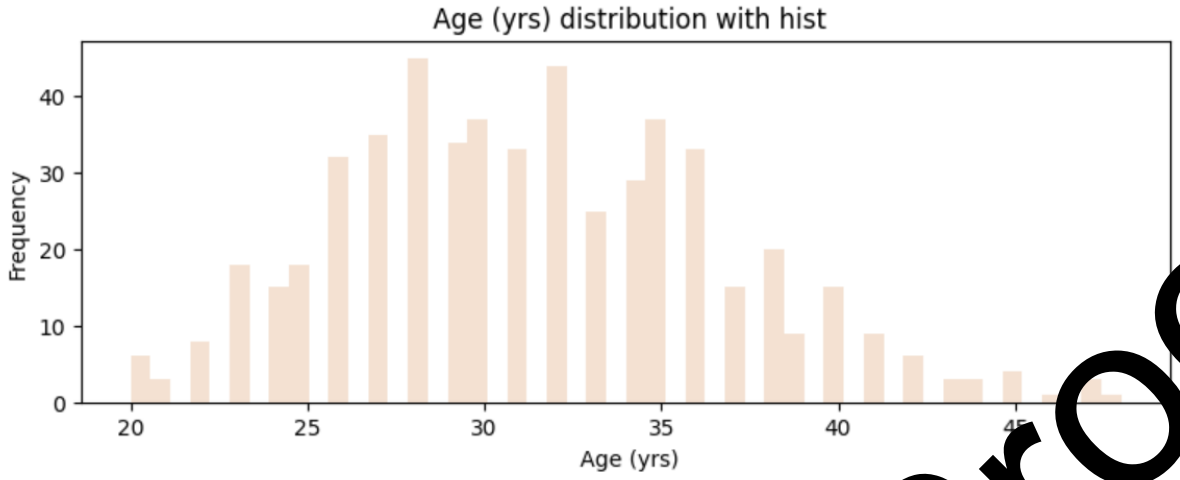


Figure 5: PCOS/PCOD Diagnosis and Modeling Workflow

To find the most pertinent characteristics affecting the diagnosis of PCOS and PCOD, a two-stage feature selection approach was used. In order to remove redundant or weakly linked features, a filter-based approach utilizing Pearson correlation analysis was used at the start of the procedure. Only features that had a moderate to strong association with the target variable were kept, using a correlation criterion of ≥ 0.25. This technique preserved crucial predictive information while drastically reducing dimensionality.

We begin with a dataset D = $\{X_1, X_2, \ldots, X_n\}$ where X_i represents each feature, and the target variable Y represents the diagnosis (e.g., PCOS or not). For each feature X_i, Pearson correlation coefficient r is calculated with the target variable Y:

$$XY_i = \frac{\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)^2 \ \sum_{j=1}^{m}(Y_j - \bar{Y})^2}} \longrightarrow \text{Equation (1)}$$

Where:

X_{ij} the value of feature i for the j-th sample, $\bar{X}_i$ is the mean of feature X_i , $\bar{Y}$ is the mean of the target variable Y, m is the number of samples.

After calculating the correlation for all features, kept only the features that have a Pearson correlation coefficient $XY_i \geq 0.25$ which is depicted in Figure 6.

After applying pearson correlation threshold dataset was found to be imbalanced. Hence, RandomOverSampler was used to balance the classes with a 0.7 sampling strategy. All features were scaled using MinMaxScaler to range [0,1] and features were converted into numpy array.

The data was split into two divisions – once with 70:30 which is used before scaling and once with 80:20 which is used after scaling for modeling.

Figure 6: Correlation Between Features with Correlation Threshold 0.25

3.3 DATA MODELLING

Recursive Feature Elimination (RFE) was combined with two supervised learning classifiers, Random Forest and Logistic Regression, in this study to create reliable prediction models for the detection of PCOS and PCOD. By training the model iteratively and classifying features based on weight or relevance, RFE systematically removes less important characteristics, producing a feature subset that is optimized and refined.

Several performance measures were calculated on the test dataset in order to assess the efficacy of the two RFE-based classification models, Random Forest and Logistic Regression. These consist of the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), F1-score, recall, accuracy, and precision. Despite using an optimum set of features chosen using RFE, both models' categorization and separability predictions differed somewhat.

A. RFE + Random Forest

The Random Forest and RFE model have shown excellent precision and recall mainly for the non-PCOS class. 0.87 and 0.91 were the precision and recall respectively for class 0, whereas for class 1 they were 0.85 and 0.80. The test accuracy was 86.29%. The overall macro-averaged F1-score was 0.85 across all class shows its strong performance. The model exhibited excellent discriminatory power, achieving a high AUC of 0.94

B. RFE + Logistic Regression

The Logistic Regression and RFE model showed accuracy of 83.06% which was marginally lower in contrast to Random Forest and RFE model. For class 0 precision was 0.86 and recall was of 0.85 whereas for class 1 it was 0.78 and 0.80 respectively. The AUC was 0.91 and the macro F1-score decreased marginally to 0.82, suggesting a solid, yet comparatively lower ability to differentiate between classes than the other model.

C. Comparison of Performance

With regard to class discrimination and overall precision and recall balance, Random Forest is the finest model in this framework according to the results as it performs better than Logistic Regression.

Metric	Random Forest + RFE	Logistic Regression + RFE
Accuracy	86.29%	83.06%
Precision (Class 0)	0.87	0.86
Recall (Class 0)	0.91	0.85
F1-Score (Class 0)	0.89	0.86
Precision (Class 1)	0.85	0.78
Recall (Class 1)	0.80	0.80
F1-Score (Class 1)	0.82	0.79
Macro Avg F1-Score	0.85	0.82
Confusion Matrix	[[68, 7], [10, 39]]	[[64, 11], [10, 39]]
AUC	**0.94**	**0.91**

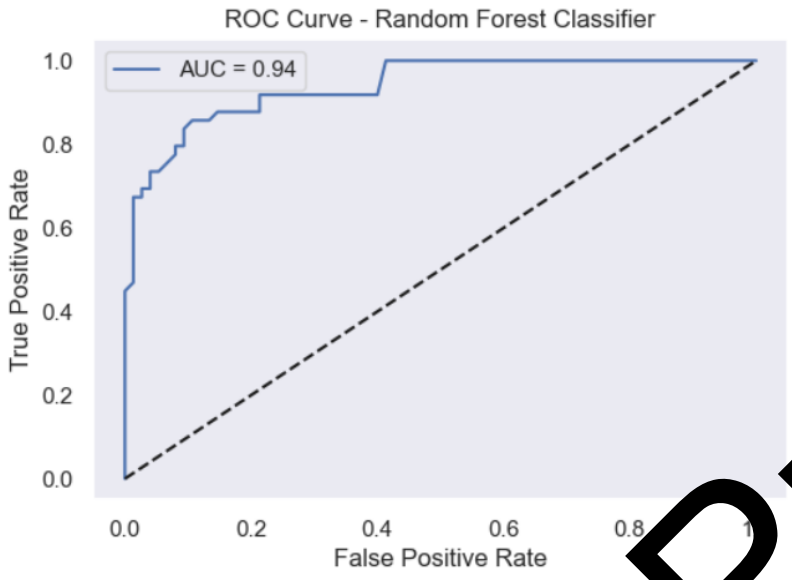


Figure 7: ROC Curve – Random Forest Classifier

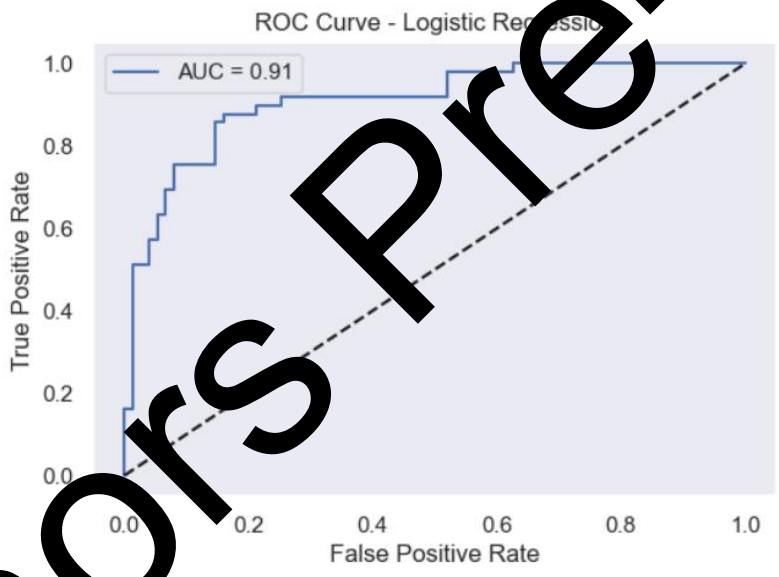


Figure 8: ROC Curve – Logistic Regression

For PCOS/PCOD disorders early and precise identification is more essential, hence AUC-ROC is a crucial statistic in medical diagnostics. At each categorization level it examines how well the model can differentiated between +ve (positive) and -ve (negative) examples. A greater AUC shows a better balance between sensitivity and specificity, which lowers false negatives and false positives. As a result, the Random Forest Model's higher AUC of 0.94 indicated that it can categorize more accurately by giving it to a more reliable tool for clinical support.

3.4 OBSERVATIONS

A. SEVERITY GRADING OF PCOS/ PCOD

Medical disease severity must be graded in order to improve patient outcomes and tailor treatment options. In this work, unsupervised clustering techniques are used to rate the severity of PCOS/PCOD cases based on clinical indicators. Unlike supervised approaches, which rely on labeled outputs, clustering finds natural groupings in the data. This makes it perfect for stratifying patient conditions when there are no explicit grading labels available. By using clustering to objectively classify patients into various severity levels, the study hopes to increase the clinical utility of the proposed approach. Anti-Müllerian Hormone (AMH) levels are specifically the main feature for clustering-based grading, as they are typically regarded as a reliable indicator of PCOS severity.

B. SIGNIFICANCE OF AMH

Clinical relevance: AMH is well known for being a robust indicator of ovarian reserve and follicle count, both of which are closely related to the pathophysiology of PCOS. Severity indicator: Higher AMH levels are frequently associated with more severe PCOS symptoms, including hormonal imbalance and an increase in follicle count. Data-driven separation: It makes sense to rank or grade severity when we cluster patients because the mean AMH levels within clusters typically vary greatly. Literature-supported: AMH is a dependable biochemical marker for categorizing PCOS traits and severity, according to numerous clinical investigations [22]. AMH stands out because it offers a quantifiable and biologically meaningful method of assigning severity ratings once clusters are formed, even though we included a lot of features in our clustering process, such as hormones, symptoms, and ultrasound data.

Because K-Means and Agglomerative Clustering are widely applicable, straightforward, and interpretable in biomedical data analysis, they were chosen for severity grading. When spherical and well-separated clusters are predicted, K-Means is an effective method for splitting huge datasets. To investigate the data structure based on layered grouping, a method of hierarchical clustering called agglomerative clustering was selected. These two approaches—Agglomerative on data hierarchy and K-Means based on centroid distances—offer complementary viewpoints, which makes them appropriate for identifying organic clusters in clinical characteristics like AMH levels. Due to their poor performance on high-

dimensional or sparse datasets like ours and sensitivity to parameter adjustment, other clustering techniques like DBSCAN were disregarded.

K-Means clustering was used to categorize patients' PCOS and PCOD severity based on 19 specific clinical and biochemical markers. StandardScaler was used to normalize the dataset and handle missing values as part of the preprocessing step. Assuming three severity levels—Severe (S), Moderate (M), and Low (L)—K-Means was used with k=3. The average Anti-Müllerian Hormone (AMH) values were used to designate the clusters; greater AMH values were linked to more severe instances.

Several visualizations were generated to support interpretation: Severity Grade Distribution using a count plot is shown in Figure 9. Boxplot of AMH levels across severity groups, reflecting the trend between AMH and severity is shown in Figure 10. Silhouette Score Evaluation, both for k=3 and a range of k=2 to k=9, to validate the clustering structure is shown in Figure 11.

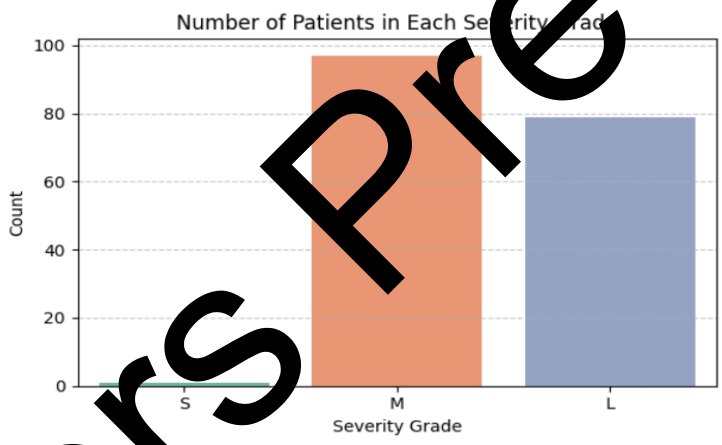


Figure 9: Severity Grade Distribution

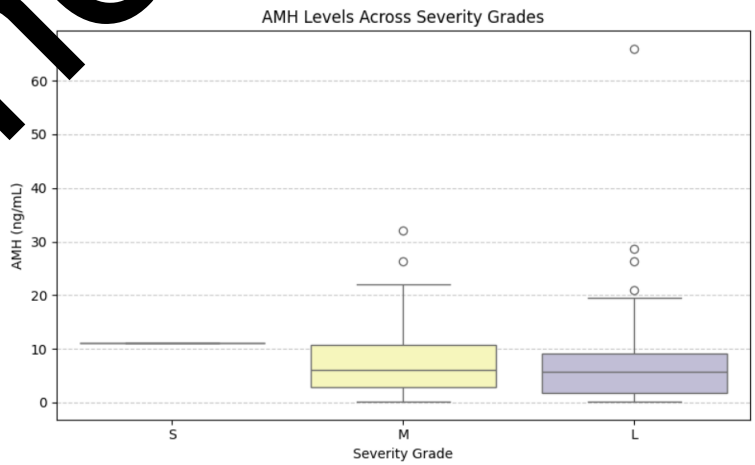


Figure 10: Boxplot of AMH levels across severity groups

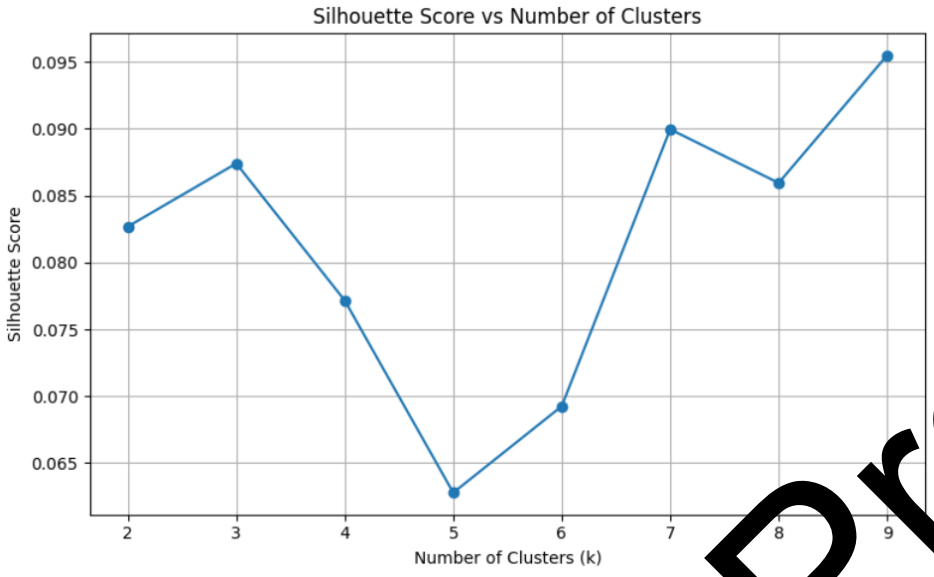


Figure 11: Silhouette Score Evaluation

By naturally classifying patient profiles according to biological markers, our unsupervised learning method provided a clinically interpretable method of grading severity.

Let X = $\{x_1, x_2, .., x_n\}$ be a set of n data points in R^d and let k be the number of clusters. Minimizing the within-cluster sum of squares (WCSS):

$argmin_C \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \longrightarrow$ Equation (2)

Where,

C_i is the set of points in cluster I, μ_i is the centroid of cluster I, $\|x - \mu_i\|^2$ is the squared Euclidian distance between a point and its cluster centroid.

ALGORITHM: K-Means Clustering
Input: Dataset X = $\{x_1, x_2, .., x_n\}$, number of clusters k **Output:** Final cluster centroids and point assignments 1. **Initialize** k centroids randomly or using the k-means++ method. 2. **Repeat** a. Assign each data point to the nearest centroid based on Euclidean distance. b. Update each centroid by computing the mean of the points assigned to it. 3. **Until** convergence (i.e., no change in assignments or minimal change in centroid positions).

Based on clinical and biochemical characteristics, PCOS severity was categorized using a bottom-up hierarchical clustering technique called agglomerative clustering. To reduce the overall within-cluster variation, the algorithm first treats each data point as a separate cluster before progressively merging the closest clusters using Ward's linkage method. Clustering was done with three clusters (n_clusters=3) after the chosen features were scaled using StandardScaler. The average AMH values were then used to map the cluster labels to severity classes (S, M, and L).

The silhouette score, which was used to evaluate the success of clustering, produced a value of 0.0746, suggesting good separation among the clusters. Figure 12 displays a boxplot of AMH levels by severity, Figure 13 uses KDE histograms to represent the distribution of AMH. To further illustrate the hierarchical merging process, a dendrogram Figure 14 was generated for a sample of the data, graphically exposing cluster cohesion and separation.

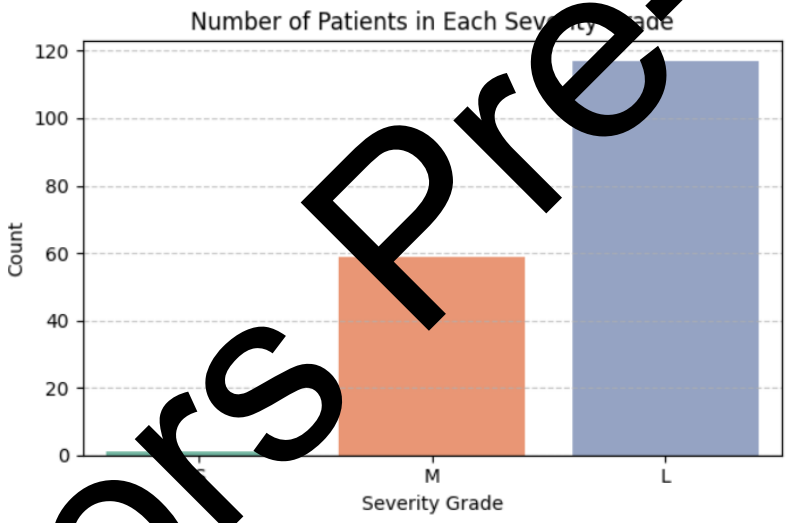


Figure 12: Severity Grade Distribution

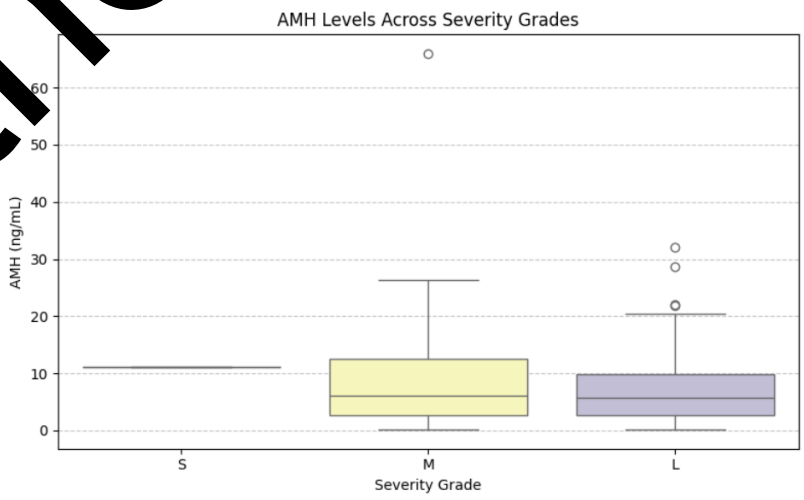


Figure 13: Boxplot of AMH levels across severity groups

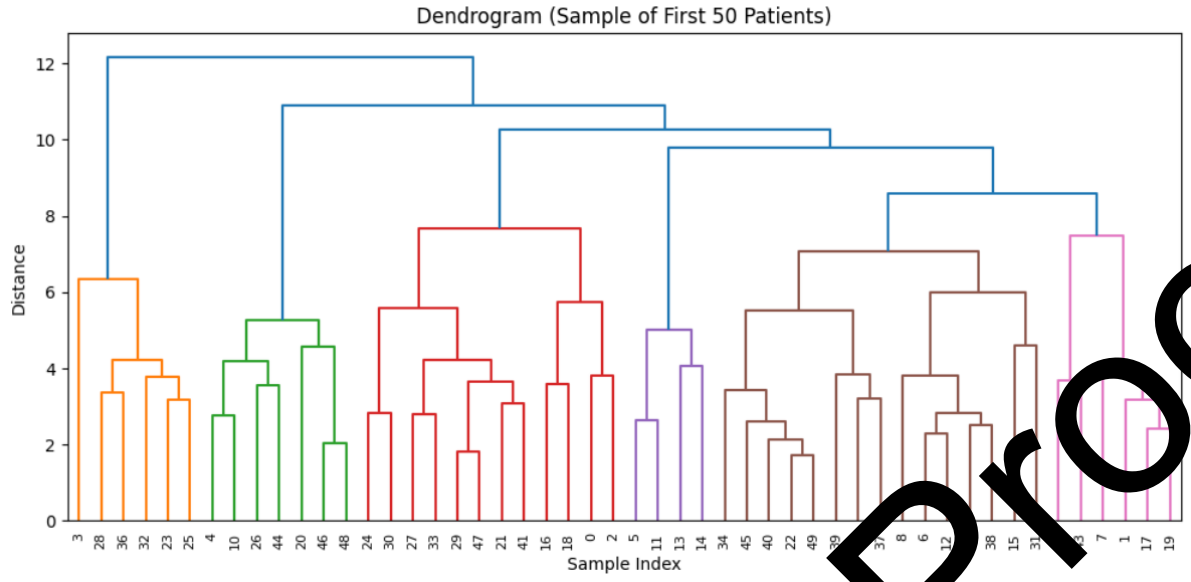


Figure 14: Dendrogram

Let X = $\{x_1, x_2, \ldots, x_n\}$ be the set of scaled feature vectors. At each iteration, the pair of clusters (C_i, C_j) that minimizes the increase in total within-cluster variance is merged:

Ward's Distance: $D(C_i, C_j) = \frac{n_i n_j}{n_i + n_j} \left\| \bar{x}_i - \bar{x}_j \right\|^2$ → Equation (3)

Where, $\bar{x}_i$ *and* $\bar{x}_j$ are the centroids of clusters C_i *and* C_j, *and* $n_i n_j$ are their sizes.

4. RESULTS AND DISCUSSION

There are significant discrepancies between the way patients are categorized by PCOS severity based on AMH levels when comparing the clustering results from K-Means and Agglomerative Clustering. Both algorithms showed concordance in detecting extreme hormonal variation, consistently classifying one patient into the severe (S) group. Nonetheless, there is a notable difference in the distribution between the moderate (M) and low (L) severity classes. Agglomerative Clustering produced 59 moderate and a larger group of 117 patients in the low severity category, while K-Means classified the majority of patients (97) as moderate and 79 as low.

The underlying distinction in clustering mechanisms is highlighted by this contrast: Agglomerative Clustering employs a hierarchical method that takes into account the connection of instances, potentially producing tighter or broader groupings than K-Means, which creates spherical clusters centered around means that can be sensitive to initialization. The findings imply that although both models offer a structured classification of severity, K-Means tends to

classify patients according to centroid-defined hormonal profiles, while Agglomerative Clustering may enable better segmentation of less severe instances.

As shown in Figure: Silhouette Score Comparison, the Silhouette Score metric was calculated for both K-Means and Agglomerative Clustering in order to assess the clustering quality. Agglomerative Clustering produced a silhouette score of 0.0746, whereas the K-Means technique obtained a little better score of 0.0874 which is shown in Figure 15.

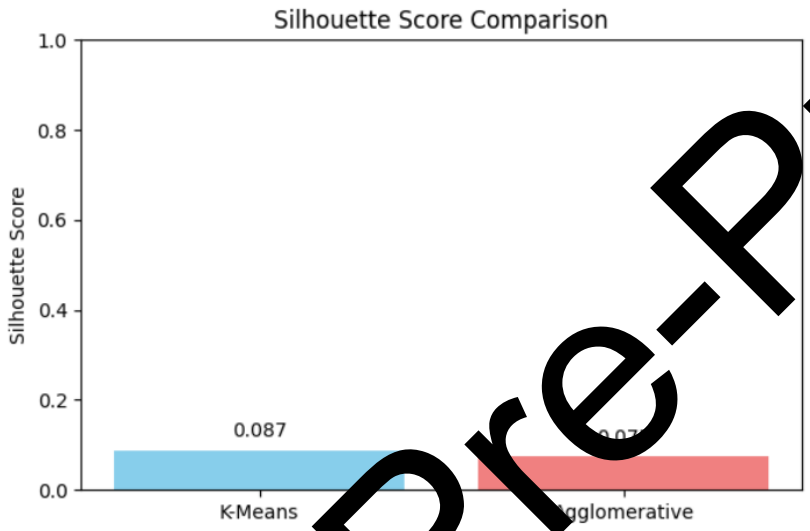


Figure 15: Silhouette Score Comparison

The somewhat better performance of K-Means indicates that it is suitable for capturing the spherical structure of hormone and symptom-based PCOS feature distributions in the dataset, even though both scores are rather low—indicative of overlapping clusters or substantial intra-cluster variance. This differentiation is further supported by the visual comparison, which provides a quantitative view of the separation and cluster compactness attained by each method.

A. CASE STUDY

A patient record which was collected during survey was analysed using both classification and clustering approaches to demonstrate the practical applicability of the developed models. The patient's clinical features, symptoms were collected where patient was having AMH (ng/mL) of 9.5. This patient exhibited classic signs of PCOS, high AMH and symptoms of weight gain, hair growth and skin darkening. The models detected that patient has PCOS but, also appropriately graded the severity as S (Severe) on biomedical markers, validating the approach.

CONCLUSION

A thorough machine learning pipeline was used in this work to examine and understand patient data related to PCOS. Initially, Recursive Feature Elimination (RFE) was used to choose features for supervised classification models - Random Forest, and Logistic Regression, which produced high accuracy and AUC scores in predicting the existence of PCOS. The severity of PCOS was then graded using unsupervised learning techniques, namely K-Means and Agglomerative Clustering, based on important clinical parameters. The interpretation of cluster severity was guided by Anti-Müllerian Hormone (AMH) values. To enhance clinical decision-making and individualized treatment planning, the clustering-based grading provided a data-driven method for classifying patients into mild, moderate, and severe phenotypes. Based on silhouette scores and distribution analysis, K-Means demonstrated superior performance and more balanced clustering, making it more suitable for PCOS severity grading in this context.

REFERENCES

[1] G. Bozdag *et al.*, "The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis," *Hum. Reprod.*, vol. 31, no. 12, pp. 2841–2855, Dec. 2016.

[2] R. Nidhi *et al.*, "Prevalence of PCOS in Indian adolescents," *J. Pediatr. Adolesc. Gynecol.*, vol. 24, no. 4, pp. 223–227, Aug. 2011.

[3] R. Azziz *et al.*, "The Androgen Excess and PCOS Society criteria for the PCOS diagnosis: a review," *Fertil. Steril.*, vol. 91, no. 2, pp. 456–488, Feb. 2009.

[4] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015.

[5] Y. Kumar *et al.*, "PCOS diagnosis using machine learning techniques: A review," *Biomed. Signal Process. Control*, vol. 66, 102648, May 2021.

[6] A. Rawat and V. Varma, "A machine learning model for severity classification in PCOS using clustering and correlation techniques," *Int. J. Healthc. Inf. Syst. Informatics*, vol. 15, no. 4, pp. 1–15, Oct.–Dec. 2020.

[7] V. C. Gandhi and S. Majumdar, "A comparative study for detecting polycystic ovary syndrome using a machine learning framework," in *Proc. IEEE Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Dec. 2023.

[8] H. K. Hosain, M. K. Mehedi, and I. E. Kabir, "PCONet: A convolutional neural network architecture to detect polycystic ovary syndrome (PCOS) from ovarian ultrasound images," in *Proc. 8th Int. Conf. Eng. Emerg. Technol. (ICEET)*, Oct. 2022, pp. 296–302.

[9] A. Divekar and A. Sonawane, "Leveraging AI for automatic classification of PCOS using ultrasound imaging," *arXiv preprint arXiv:2501.01984*, Dec. 2024.

[10] M. Uddin *et al.*, "Intelligent detection for polycystic ovary syndrome (PCOS)," in *Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, 2022.

[11] H. Elmannai *et al.*, "Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence," *Diagnostics*, vol. 13, no. 8, 2023.

[12] A. Alshakrani *et al.*, "Optimized machine learning for early detection of PCOS: an ensemble-based approach with feature selection," *Sensors*, vol. 25, no. 4, 2025.

[13] S. Faris and S. Miften, "Automated high precision PCOS detection through a segment-based machine learning pipeline," *Int. J. Med. Inform.*, 2023.

[14] D. Hdaib *et al.*, "Comparative performance of ML classifiers for early PCOS detection," *IEEE Access*, 2022.

[15] R. Joshi *et al.*, "Succinct differentiation of disparate boosting ensemble learning methods for prognostication of polycystic ovary syndrome diagnosis," *arXiv preprint arXiv:2201.00418*, Jan. 2022.

[16] N. Khanna *et al.*, "Explainable AI based pipeline using SHAP and LIME for PCOS detection," in *Proc. IEEE Region 10 Conf. (TENCON)*, 2023.

[17] S. Venkatalakshmi and P. Regina, "A comprehensive overview of ML and DL approaches for PCOS diagnosis," *IEEE J. Biomed. Health Inform.*, vol. –, 2024.

[18] A. Modi *et al.*, "International evidence-based guideline for the assessment and management of polycystic ovary syndrome," *Semin. Reprod. Med.*, vol. 41, no. 1–02, pp. 3–4, Mar. 2023.

[19] Follicles Net *et al.*, "Follicles Net: An efficient tool for the diagnosis of polycystic ovarian syndrome," *Diagnostics*, vol. –, 2024.

[20] "Systematic review: Detection of polycystic ovary syndrome using AI/ML," *Front. Endocrinol.*, 2023.

[21] A. Biswas *et al.*, "Federated learning for PCOS treatment recommendations," *arXiv preprint*, 2023.

[22] R. Dewailly, J. Laven, D. H. C. De Groot, S. E. M. Dolleman, B. C. J. M. Fauser, and E. S. Teede, "Anti-Müllerian hormone as a diagnostic and prognostic marker in polycystic ovary syndrome," *Hum. Reprod. Update*, vol. 22, no. 5, pp. 726–740, Sep. 2016, doi: 10.1093/humupd/dmw020.

[23] "PCOS Dataset," *Kaggle*, [Online]. Available: https://www.kaggle.com/code/swanieeeeeeee/pcos-dataset?select=PCOS_data_without_infertility.xlsx.

[24] "PCOS Dataset," *Kaggle*, [Online] Available: https://www.kaggle.com/code/swanieeeeeeee/pcos-dataset?select=PCOS_infertility.csv.