*International Journal of Latest Trends in Engineering and Technology*
*Vol.(22)Issue(3), pp.001-003*
*DOI: http://dx.doi.org/10.21172/1.223.01*
*e-ISSN:2278-621X*

# EARLY DIAGNOSIS OF PCOS IN WOMEN USING MACHINE LEARNING ALGORITHM

Syed Tooba Gillani[1], Aaqib Rashid Phamda[1], Farzan Altaf Bhat[1], Syed Mohammad Talib Qadri[1], Sameen Mushtaq Kawoosa[1]

**Abstract- The most prevalent endocrine condition in women and a primary contributor to anovulatory infertility is polycystic ovary syndrome or PCOS. Sub-fertility brought on by PCOS can be treated with a variety of medical approaches, either separately or in combination. In this paper, a unique PCOS detection method is presented wherein the classification is performed by using Voting Ensemble Learning (V-EL) algorithms. The main aim of the proposed work is to increase the accuracy of the disease detection rate while lowering the complexity and processing time. To achieve this objective, a PCOS dataset is created by collecting data from various Labs across Kashmir. After this, the pre-processing technique is applied to the original dataset in order to make it more informative and balanced. Moreover, to reduce the complexity of the model, we have used Principal Component Analysis (PCA) feature selection technique that selects only informative and critical features from the processed dataset. Finally, the voting Ensemble learning (V-El) technique is employed for identifying and classifying PCOS in patients at early stages. The efficacy of the proposed V-EL model is depicted in MATLAB software for various metrics. Results revealed that the proposed V-EL model is able to detect PCOS in women with an accuracy of 0.990 respectively.**

**Keywords: PCOS, PCA, Machine learning, V-EL model.**

## I. INTRODUCTION

.In recent years, Polycystic ovarian syndrome (PCOS) in women has been one of the most prevalent but undertreated diseases. PCOS is one of the leading causes of impotence in women. It can also be defined as a medical condition that causes hormonal disorders in women during their reproductive phase [1]. This means that hormonal imbalance is the main cause of PCOS in women [2]. Small amounts of fluid known as follicles or cysts form in the ovaries as a result of this hormonal imbalance. The main issue PCOS-affected women face is that their ovaries cannot produce eggs, which leads to serious pregnancy difficulties [3]. A study found that the likelihood of PCOS is 4.8% in white Americans, 8% in African Americans, 6.8% in Spanish people, and 31.3% in Asian women 31.3%. The processes required for conception, ovulation, and the formation of a child in the womb of a mother depend entirely on unmatched hormones that must be balanced. Specifically, progesterone, luteinizing hormone (LH), estrogen, and follicle-stimulating hormone (FSH) are required. The pituitary gland produces the hormones FSH and LH, whereas the ovaries create progesterone and estrogen hormones. Both progesterone and estrogen are crucial for a healthy, balanced reproductive system in women. Polycystic ovaries can either be categorized as normal or polycystic, despite the fact that cystic ovaries are more common than polycystic ovaries. Polycystic ovary syndrome (PCOS) is the most frequent endocrine disorder during the premenopausal period; it is considered to be a multicomponent and polygenic disorder [22]. The global prevalence of PCOS constitutes 4 to 20% [23-24]. Because of high rates of prolactin and low levels of luteinizing hormone in the ovary of a woman with PCOS, follicles do not develop and

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*

mature, while in a normal ovary with proper levels of LH and FSH hormones a single follicle develops to a size of 20 mm in diameter and is ready for ovulation. PCOS's primary cause is still unknown, however, significant contributing factors include hormonal disturbance, a body mass index of >24, and irregular menstrual cycles are known [4-5]. Women with polycystic ovaries do not ovulate because the follicles do not grow and widen sufficiently to fit the numerous little antral follicles which carry eggs. Because of this reason reproductive issues are much more likely to affect women with polycystic ovaries. Such women frequently struggle to become pregnant and require treatment to increase their chances. In women with polycystic ovaries or cysts, antral follicles that are tiny cystic formations are visible on "polycystic" ultrasound scans of the ovaries that in medical terms are also referred to as PCOD.

PCOS is very common nowadays, one in five women, or about 20% of the population, had this syndrome as per the research data. Excess weight, greasy skin, acne, high blood pressure, and metabolic dysfunction are some of the reasons that contribute to PCOS. Long-term effects include significant disorders such as endometrial hyperplasia, cardiovascular disease, and type 2 diabetes mellitus as a result of common metabolic medical conditions like hypertension, hyperinsulinemia, abdominal obesity, dyslipidemia, and an ovulatory period [6]. Researchers have also analyzed that it can result in a variety of malignancies, such as uterine or breast cancer in reproductive age. A higher ratio of androgen i.e. increased levels of male hormones [7], an abnormal menstruation period, polycystic ovaries, and metabolic issues are some of the indications of PCOS. The adoption of necessary lifestyle adjustments is supported by early diagnosis of PCOS-related indications. Women with PCOS have three times higher miscarriage chances than women without PCOS [8]. The substantial risk of miscarriage in the first trimester is demonstrated by a recent-studies [25]. Therefore, by detecting PCOS at early stages the chances of miscarriages can be avoided. Furthermore, around 12–21% of women of reproductive age are affected with PCOS, out of which 70% go undiagnosed. By taking the properly prescribed medication and making changes to lifestyle, this ailment can be cured. Birth control pills, diabetic tablets, anti-androgen medications, fertility tests, and ultrasound scans are all examples of medications that doctors prescribe. In addition to this, daily exercise reduces biochemical hyperandr ogenism, and free androgen indices aid in reducing the chances of PCOS. According to studies, as women get older and older, menopause strikes, and PCOS indicators become less severe. PCOS is divided into various phenotypes based on its clinical features. The method and criterions used to diagnose PCOS remain under dispute. The debate seems to have an impact on PCOS prevalence statistics among the criteria put out by various groups. Depending on the criteria utilized to identify this syndrome, PCOS prevalence is calculated. Ingenious traits could go unchecked in the initial stages of PCOS since symptoms appear covertly and correspond with the evolution of normal puberty. This could be a reason for the disease in a young girl going undiagnosed [9]. Patients are required to undergo an ovarian ultrasound in order to guarantee the correctness of the PCOS diagnosis. Moreover, for metabolic testing, some of the tests may even require venous sampling. It is a substantial financial burden that the average price of the beginning diagnosis and evaluation of PCOS is estimated to be $740.

Basically, PCOS manifests psychological, metabolic, and reproductive features and costs women's health significantly. The signs and test results of PCOS are frequently ignored since it is challenging to diagnose the condition because of a lack of understanding of its complex pathomechanics. Doctors are compelled to undertake a lot of clinical tests and pointless radiologic imaging procedures because of the wide range of symptoms associated with this condition. Given that PCOS increases the risk of gynecological cancer, infertility, and even miscarriage as well as mental suffering for patients due to time and financial waste, it is imperative to identify and diagnose the condition early with the least amount of testing and imaging procedures. There are few studies that apply machine learning to the prediction of PCOS [26-29, 10-11]. In this study, an Ensemble Learning based PCOS detection method is proposed which improves the accuracy of the detection rate. The major contribution of this research is:

- To implemented an effective and advanced feature selection technique for reducing dimensionality of dataset.
- Moreover, the Ensemble Learning technique is implemented on important features for detecting PCOS at early stages.
- Finally, the performance of the proposed Ensemble learning-based PCOS detection model is compared with current ML models to prove its supremacy over various metrices.

## I. LITERATURE REVIEW

The authors in [12], proposed a method that could predict the effectiveness of PCOS therapy which was based on the parameters of the ideal and minimal set. They utilized five distinct machine learning classifiers including Random Forest, SVM, Logistic Regression, Gaussian Naive Bayes, and K Neighbors to identify whether a woman had PCOS. Furthermore, they used the Chi-Squared approach for selecting important features. Results revealed that the RF Classifier had the greatest accuracy. Also, in [13], authors identified the critical features for diagnosing PCOS by using ML techniques. In order to classify huge data samples, several classification techniques were used, including SVM, LR, Gradient Boosting, RF, DT, and K-Nearest Neighbor (KNN). Results revealed that the pertinent features were not yet discovered for the many types of studies that were being done to diagnose PCOS. Moreover, in [14], authors proposed Improved Fruit Fly Optimization (IFFOA) with an artificial neural network (ANN) known as (IFFOA-ANN) for classifying follicles as abnormal or normal. In order to perform the segmentation of follicles, the adaptive k-means clustering algorithm was applied. Additionally, a feature extraction model was proposed using statistical GLCM. Furthermore in [15], to understand the Random Forest (RF) model, which was challenging to interpret for PCOS risk factors estimation, in this paper, the Local Interpretable Model-Agnostic Annotations (LIME) technique was proposed and employed the "Polycystic ovary syndrome" dataset. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value and balanced accuracy obtained from the Random Forest method were 86.03%, 86.32%, 85.37%, 93.18%, 72.92% and 85.84% respectively. Next in [16], a DCNN-based PCOS detection model was suggested. Coding for PCOS classification was developed in Python programming, and ultrasound images were used to determine if they were filled with blood or fluid. The study utilized DCNN-based image processing feature extraction to categorize PCOS in the dataset. The test dataset was therefore utilized to perform feature extraction and evaluate correctness based on performance metrics. Additionally, in [17], a novel method for diagnosing PCOS, in its initial stages was proposed wherein; an ultrasound abdominal scan image was proposed for diagnosing PCOS. Firstly, a preprocessing technique was used to discover cysts, and then an image registration technique was utilized to track their growth. The scanned picture was pre-processed for speckle reduction using the available techniques, including the wavelet, adaptive, and Gabor filters. With an accuracy of 93%, the suggested work could identify PCOS at an early stage. Also, in [18], the authors introduced an automated early detection and prediction technique in this research that could precisely predict the chance of having PCOS and related mental health problems. The Fuzzy technique for order of Preference by Similarity to the Ideal Solution (TOPSIS) method also was assessed for its effectiveness in the study. The Fuzzy TOPSIS and the well-known support vector machines (SVM) technique were contrasted utilizing local yet precise datasets gathered on a range of women. On the same dataset, both techniques were assessed. Using the Fuzzy TOPSIS technique and SVM, the precision of 98.20% and 94.01%, respectively, were attained. Moreover, an extended machine learning classification technique for PCOS prediction was proposed in [19], that was trained, and examined over 594 ovary USG images. Feature extraction from the images was done using a Convolutional Neural Network (CNN) incorporating various state-of-the-art methods and transfer learning, and then utilizing a stacking ensemble machine learning technique to differentiate between PCOS and non-PCOS ovaries utilizing conventional models as base learners and bagging or boosting ensemble models as meta-learners on that smaller feature set. The best results were attained by using by combining the "VGGNet16" pre-trained model with CNN architecture as a feature extractor and stacking ensemble method with the "XGBoost" model as the meta-learner as an image classifier with a precision of 99.89% for classifying.

To better understand possible effects of bisphenol A (BPA) exposure on ovarian reserve in women with polycystic ovary syndrome (PCOS), Zhou et al. measured creatinine adjusted urinary BPA (BPA_Cre) concentrations and used regression models to evaluate the association between urinary BPA level and antral follicle count (AFC), antimullerian hormone (AMH), day-3 follicle stimulating hormone levels (FSH) and inhibin B (INHB) in 268 infertile women diagnosed with PCOS. BPA was detected in all women with a median concentration of 2.35 ng/mL (the 25th and 75th percentiles of 1.47ng/mL and 3.95ng/mL). A unit increase in BPA_Cre was associated with a significant decrease of 0.34 in AFC ($\beta = -0.34$, 95% CI = −0.60, −0.08; $p = 0.01$). Likewise, BPA was negatively associated with AMH and day-3 FSH levels, but neither of them reached statistical significance. No association was observed between BPA and INHB. Their results suggest that in women with PCOS, BPA may affect ovarian follicles and, therefore, reduce ovarian reserve.

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*

Furthermore, in [21], the authors developed a unique intelligence system to categorize PCOS based on k-means paired with an LS-SVM (K-M-SVM) using fewer features. After preprocessing the original dataset, k-means were used to classify PCOS by choosing the strongest characteristics based on Euclidean distance. The k-means cluster was discovered to have a strong potential for selecting the most influential features and removing the subpar ones. Thus, from the original features, a total of six features were selected to represent PCOS data. The results demonstrated that the suggested model (K-M-SVM) performed better than the state-of-the-art and improved its accuracy to 99%.

From the above literature, it has been analyzed that a number of approaches have already been proposed for identifying and classifying PCOS in women at early stages. These methods are producing good results however, there still is a scope for improvement. It is observed that the majority of the researchers are using ML classifiers in their work for classifying PCOS, however, these classifiers are not able to handle large and complex datasets which lowered the accuracy of the system. Furthermore, few authors have selected features manually in their work which also results in reduced accuracy of PCOS detection [13-15]. While some researchers have used automated Feature selection techniques, those techniques are sensitive to small frequencies and ineffective. Keeping these limitations in mind, the need for proposing a new and improved PCOS detection method arises that can overcome the above-given limitations.

## II. PROPOSED WORK

In order to overcome the limitations of existing PCOS disease detection models, this paper presents a unique and effective PCOS detection model that is based on Ensemble Learning techniques. The main objective of the proposed model is to enhance the accuracy of classification while lowering the complexity and processing time. To achieve this objective, the proposed model undergoes through some crucial stages of data collections, data pre-processing, feature Selection and classification accordingly. Initially, all the necessary information regarding the PCOS disease is taken from dataset that is collected manually from different labs across Kashmir. Since, the dataset is collected manually, therefore, it is not balanced and contains a lot of null and repeated values. Hence, it is important to apply pre-processing technique on the original dataset. During the pre-processing phase, the empty cells are filled, repeated values are deleted and input and output data is separated. After this, principal Component Analysis (PCA) technique is employed in the proposed work for selecting only important and crucial features from the processed dataset. The main reason for using PCA in the proposed work is that it makes computation easy, solves high dimensional data issues and also speeds up the performance of classifiers. Finally, in the last phase of proposed work, we have implemented voting-based Ensemble learning technique for classifying PCOS in women. A voting classifier is a type of ML estimator which trains number of base models or estimators and makes predictions based on averaging their results. In this work, three ML estimators Logistic Regression (LR), Random Forest (RF) and Support Vector Machine (SVM) are used in the voting Ensemble Learning (V-EL) model for classifying PCOS and Non-PCOS patients. The performance of the model is evaluated in terms of parameters like accuracy, Precision etc. The step by step working of the proposed V-EL based PCOS detection model is given in the next section of this paper.

## III. METHODOLOGY

As mentioned earlier, that the proposed V-EL model undergoes through series of steps to achieve the desired objective. The brief and stepwise working of proposed V-EL model is discussed in this section of paper.

- **Data Collection:** This is the very first step of proposed work wherein all the necessary information regarding the disease is collected. Here, we have taken PCOS dataset manually by collecting various samples from different labs across Kashmir. The reason for collecting data manually is to check the efficiency of model on real world dataset. The count of dataset given for PCOS positive and negative patients is shown graphically in Fig 1.

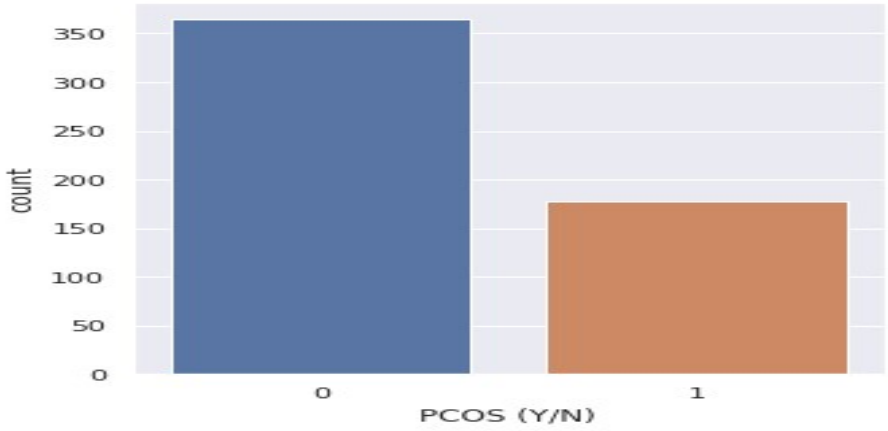


Fig 1. Count of PCOS in dataset

The blue bar present in the above graph depicts the PCOS negative samples and is represented by 0, whereas, the PCOS positive samples are shown by orange bar and represented by 1.

- **Data Pre-Processing:** Since the data is collected manually therefore, it is not balanced and contains a lot of unnecessary and redundant data, therefore it is important to apply pre-processing technique on original dataset. During the pre-processing phase, all the unnecessary, unwanted and redundant information is removed from the dataset which in turn makes it more informative.
- **Feature Selection:** Once the data is processed, it is time to select only crucial and important features from the available feature set. To accomplish this task, we have used PCA in the proposed work, that selects only 35 features from given feature vector. By doing so, the dataset dimensionality issue is resolved and complexity and processing time is also reduced.
- **Training and Testing:** Soon after selecting features, the final feature set formed is divided into categories of training data and testing data. the training data is used for training the classifiers whereas, though testing data the performance of classifiers is validated.
- **Classification:** Once the model is trained, it is time to initialize voting Ensemble learning classifiers for classification purpose. For doing so, the testing data is passed to V-EL classifier that extracts the features from testing dataset and tries to match it with the final feature vector formed. Depending on the comparison between features, the V-EL model is able to identify and classify PCOS in patients effectively.
- **Performance Evaluation:** Finally, the usefulness and productivity of the proposed V-EL model is analyzed and validated by comparing its performance with few standard PCOS detection models in terms of various metrics. The results obtained for the proposed V-EL based PCOS detection model are explained in the next section of this manuscript.

## IV. RESULTS OBTAINED

The efficiency and usefulness of the proposed Voting Ensemble learning (V-EL) algorithm is analyzed and compared with few traditional models in MATLAB software. The experimental results were demonstrated in terms of various performance dependency factors like accuracy, precision, recall, F-Score, Confusion matrix and ROC. IN this section of paper, we are going to thoroughly analyze and explain results obtained for proposed V-EL based PCOS detection model.

## V. PERFORMANCE EVALUATION

The performance of proposed V-EL based PCOS detection model is firstly analyzed in terms of its Receiver operating characteristic (ROC) values. the graph obtained for the same is shown in Fig 2., with x-axis calibrating to False Positive Rate (FPR) and Y-axis calibrating to True Positive Rate (TPR) respectively. This signifies that if the ROC curve is more deviating towards the upper left corner, the more will be the value of sensitivity and specificity respectively. The value of ROC should be 1 in an ideal case. From the given graph, it is observed that the value of ROC is around 0.998 which is very close to the value of 1. This means that proposed V-EL model has high value of sensitivity and specificity which in turn means high accuracy of disease detection.

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*

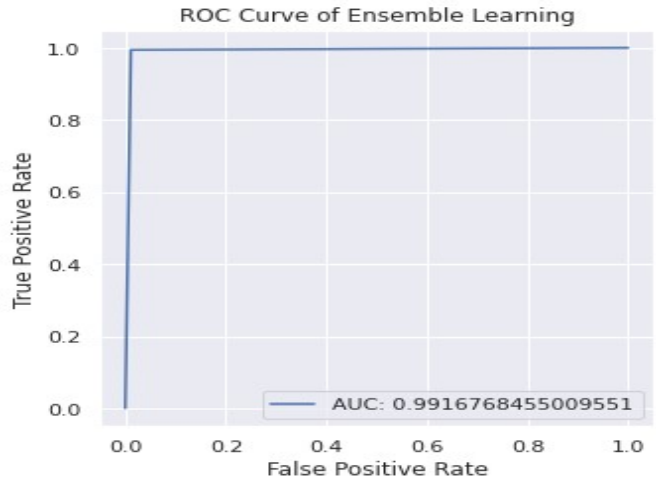


Fig 2. ROC value in proposed V-EL model

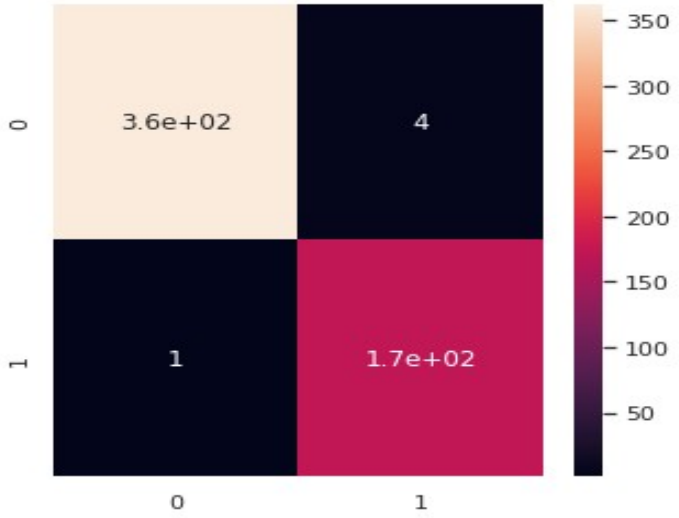


Fig 3. Confusion Matrix for proposed V-EL model

Similarly, we have also analyzed the performance of proposed V-EL based PCOS detection model in terms of confusion matrix, whose graph is shown in Fig 3. The confusion matrix represents two classes of 0 and 1 as PCOS negative and Positive respectively. From the given graph, it is observed that proposed V-EL model is able to detect PCOS in women at early sages with an overall accuracy of 99.07% which in itself is a great number.

Furthermore, to prove the supremacy of our PCOS disease detection approach, we compared its performance with few state of art PCOS detection methods in terms of accuracy. The comparative graph obtained for the same is shown in Fig 4. The x-axis and y-axis of the given graph corresponds to different classifiers and their respective accuracy values. After analyzing the given graph, it is observed that the value of accuracy was least in GNB model with 0.86, followed up by KNN, R-SVM, L-SVM, LR and RF models with 0.87, 0.88, 0.88, 0.899 and 0.90 respectively. However, the value of accuracy was 0.990% in proposed V-EL based PCOS detection model which is highest than all other approaches. The specific values of accuracy are given in table 1.

Table 1: Accuracy Comparison table

| **Algorithm** | **Accuracy** |
|---|---|
| GNB | 0.869500 |
| L-SVM | 0.876400 |
| R-SVM | 0.881100 |
| LR | 0.899800 |

| | |
|---|---|
| RF | 0.909100 |
| Proposed V-EL | 0.990758 |

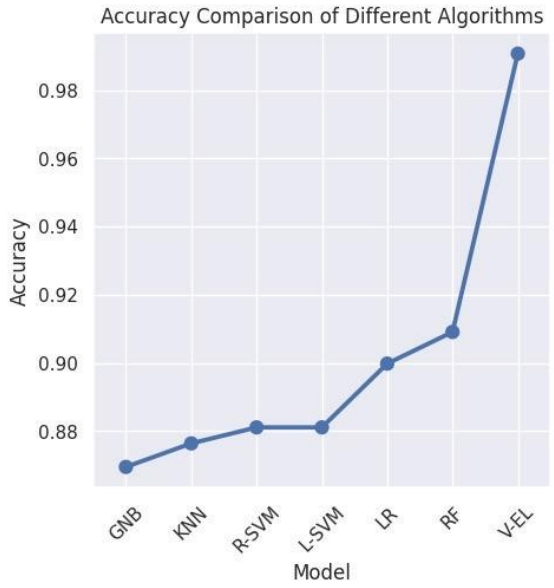


Fig 4. Comparison graph for Accuracy

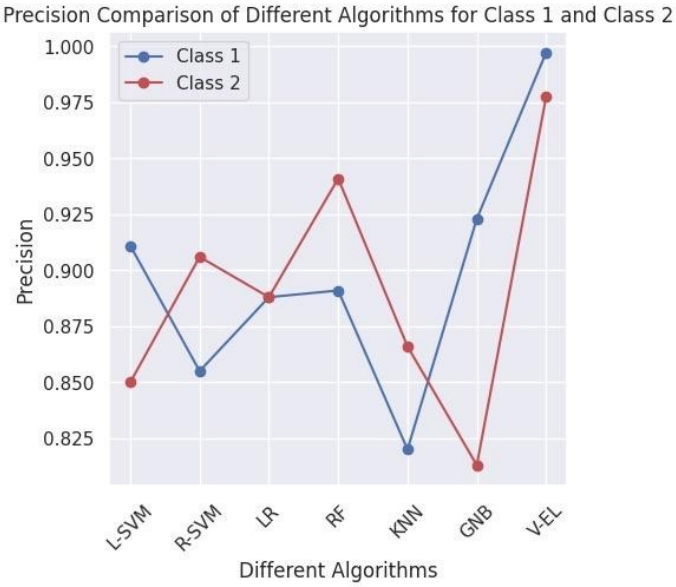


Fig 5. Comparison graph for Precision

Similarly, we have also analyzed and compared the performance of proposed V-EL model with traditional L-SVM, R-SVM, LR, RF, KNN and CNB models for two classes in terms of their precision values. The comparative graph for the same is demonstrated in Fig 5. After analyzing the graph, it is observed that value of proposed V-EL came out to be highest i.e. 0.997 and 0.977 for class 1 and class 2 respectively. On the other hand, the value of precision was only 0.91 and 0.85 in L-SVM, 0.85 and 0.90 in R-SVM, 0.88 and 0.888 in LR, 0.89 and 0.94 in RF, 0.82 and 0.866 in KNN and 0.92 and 0.81 in GNB models for class 1 and class 2 respectively. The accurate values for precision are recorded in tabular form also and is shown in Table 2.

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*

Table 2: Precision Comparison table

| Algorithm | Class 1 | Class 2 |
|---|---|---|
| L-SVM | 0.911000 | 0.85000 |
| R-SVM | 0.855000 | 0.90600 |
| LR | 0.888000 | 0.88800 |
| RF | 0.891000 | 0.94100 |
| KNN | 0.820000 | 0.86600 |
| GNB | 0.923000 | 0.81300 |
| V-EL | 0.997253 | 0.977401 |

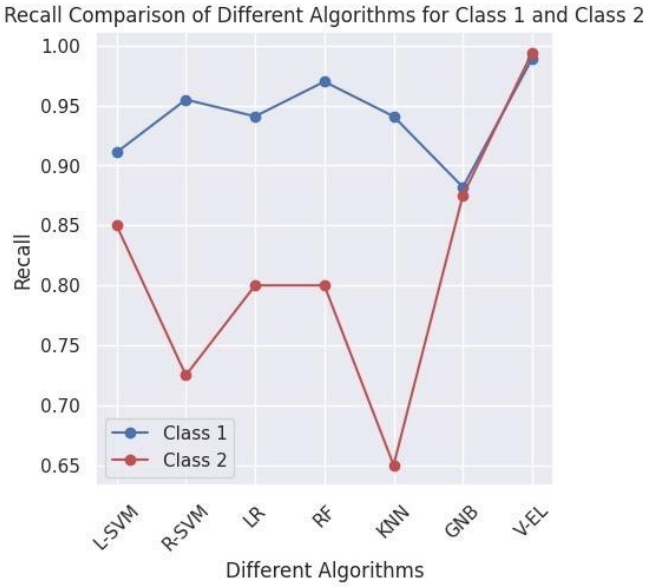


Fig 6. Comparison graph for Recall

Furthermore, the efficiency and robustness of proposed V-EL approach is also demonstrated and validated by comparing it with traditional models in terms of their recall values. Fig 5 represents the comparison graph obtained for the recall values under two classes. From the given graph, it is observed that the value of recall came out to be 0.91, 0.95, 0.94, 0.97, 0.94 and 0.88 for class 1 and 0.85, 0.72, 0.800, 0.800, 0.650 and 0.875 for class 2 in traditional L-SVM, R-SVM, LR, RF, KNN and GNB models. Whereas, the recall value was around 0.989 and 0.994 in proposed V-EL based PCOS detection model for class 1 and class 2 respectively. The exact values of recall are shown in table 3.

Table 3: Recall Comparison table

| Algorithm | Class 1 | Class 2 |
|---|---|---|
| L-SVM | 0.91100 | 0.850 |
| R-SVM | 0.95500 | 0.725 |
| LR | 0.94100 | 0.800 |
| RF | 0.97000 | 0.800 |
| KNN | 0.94100 | 0.650 |
| GNB | 0.88200 | 0.875 |
| V-EL | 0.989101 | 0.994253 |

In addition to this, the efficacy of proposed V-EL model is proved by comparing it with traditional models in terms of their Fscore values for two classes. The comparison graph for the same is shown in Fig 7. From the given graph,

it can be seen that L-SVM, R-SVM, LR, RF, KNN and GNB models yield a Fscore value of 0.91 & 0.85; 0.902 &0.805; 0.91& 0.842; 0.92 & 0.864; 0.87 & 0.742; and 0.902 & 0.843 for class 1 and class 2 respectively. However, the value of Fscore was mounted at 0.99 for class 1 and 0.985 for class 2 in proposed V-EL based PCOS detection model. The specific value for recall obtained for two classes is shown in table 4.

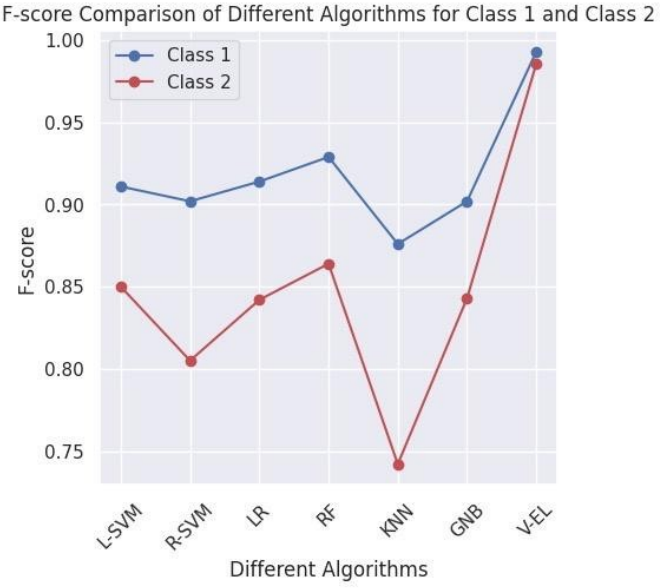


Fig 7. Comparison graph for Fscore

Table 4: Fscore comparison table

| Algorithm | Class 1 | Class 2 |
|---|---|---|
| L-SVM | 0.91100 | 0.850000 |
| R-SVM | 0.902000 | 0.805000 |
| LR | 0.914000 | 0.842000 |
| RF | 0.929000 | 0.864000 |
| KNN | 0.876000 | 0.742000 |
| GNB | 0.902000 | 0.843000 |
| V-EL | 0.99316 | 0.985755 |

From the graphs given in figure 5, figure 6, figure 7 and values given in table 1, table 2, table 3 and table 4, it is clear that proposed V-EL based PCOS detection model is outperforming all other PCOS detection models in terms of accuracy, precision, recall and Fscore to prove its superiority over them.

All the data required to understand the PCOS condition is painstakingly gathered from various labs throughout Kashmir. The dataset total contains 41 features, out of which we have used 35 features in our proposed model. The dataset is unbalanced and has many null and duplicate values because it was manually compiled. Therefore, it is crucial to do pre-processing on the given dataset. During the pre-processing phase, the empty cells are filled, duplicate values are removed, and input and output data are segregated. Furthermore, as we studied earlier that it is important to reduce the dimensionality of dataset to improve the performance of entire model, therefore we have implemented PCA feature selection technique in proposed work. The PCA is responsible for selecting only important and informative features from processed dataset and remove all the unnecessary details and redundant values which in turn reduced dataset dimensionality. One of the major reasons for using PCA In proposed work is that it speeds up the working of Machine Learning algorithms and is easy to compute. Furthermore, it is also generating effective results when dealing with high dimensional data.

## VI. CONCLUSION

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*

In this paper, an effective and highly accurate PCOS detection model is presented that is based on Voting Ensemble Learning (V-EL) Algorithms. The performance of the proposed model is examined and validated in MATLAB software under various metrics. The simulating results revealed that proposed V-EL model achieved a highest accuracy of 0.9907 whereas, it was only 0.86 in traditional GNB, 0.87 in KNN, 0.88 in R-SVM and L-SVM, 0.899 in LR and 0.909 in RF respectively. Moreover, we have also observed that proposed V-EL model is outperforming traditional models in terms of precision, recall and Fscore for class 1 and class 2 respectively. In addition to this, results were achieved in terms of ROC curve that was mounted at 0.998 i.e. close to 1, signifying high sensitivity and specificity values. These values prove the supremacy of proposed V-EL model over other similar models.

## REFERENCES


[1] Dipamoni Morang, Pankaj Chasta and K. Kaushal Chandrul A Review on “Polycystic Overy Syndrome (PCOS), Int. J. Trend Sci. Res. Dev. (IJTSRD) (2019)

[2] I. Kyrou, E. Karteris, T. Robbins, K. Chatha, F. Drenos, and H. S. Randeva, 617 ‘‘Polycystic ovary syndrome (PCOS) and COVID-19: An overlooked 618 female patient population at potentially higher risk during the COVID-19 619 pandemic,’’ BMC Med., vol. 18, no. 1, pp. 1–10, Jul. 2020. 620

[3] B. J. Sherman, N. L. Baker, K. T. Brady, J. E. Joseph, L. M. Nunn, and 621 A. McRae-Clark, ‘‘The effect of oxytocin, gender, and ovarian hormones 622 on stress reactivity in individuals with cocaine use disorder,’’ Psychophar- 623 macology, vol. 237, no. 7, pp. 2031–2042, May 2020.

[4] Gibson-Helm, M., Teede, H., Dunaif, A., Dokras, A.: Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J. Clin. Endocrinol. Metab. 102(2), 604–612 (2017)

[5] C. Gopalakrishnan and M. Iyapparaja, “A Detailed Research on Detection of Polycystic Ovary Syndrome from Ultrasound Images of Ovaries”, Int. J. Recent Technol. Eng. (IJRTE) (2019)

[6] M Sumathi and P Chitra, “R Sakthi Prabha and Srilatha K Study and detection of PCOS related diseases using CNN ICRIET (2020)”,

[7] E. Khashchenko, E. Uvarova, M. Vysokikh, T. Ivanets, L. Krechetova, 628 N. Tarasova, I. Sukhanova, F. Mamedova, P. Borovikov, I. Balashov, and 629 G. Sukhikh, ‘‘The relevant hormonal levels and diagnostic features of 630 polycystic ovary syndrome in adolescents,’’ J. Clin. Med., vol. 9, no. 6, 631 p. 1831, Jun. 2020. 632

[8] M. Woźniak, R. Krajewski, S. Makuch, and S. Agrawal, ‘‘Phytochemicals 633 in gynecological cancer prevention,’’ Int. J. Mol. Sci., vol. 22, no. 3, 634 p. 1219, Jan. 2021.

[9] Azziz R, Marin C, Hoq L, Badamgarav E, Song P. Health Care-Related Economic Burden of the Polycystic Ovary Syndrome During the Reproductive Life Span. J Clin Endocrinol Metab (2005) 90(8):4650–8.

[10] Mehrotra, P., Chatterjee, J., Chakraborty, C., Ghoshdastidar, B., Ghoshdastidar, S.: Automated screening of polycystic ovary syndrome using machine learning techniques. In: Paper Presented at the 2011 Annual IEEE India Conference, 2011 (2011)

[11] Bharati, S., Podder, P., Mondal, M.R.H.: Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms. In: 2020 IEEE, pp 1486–1489 (2020)

[12] Thakre, V., and A. S. Sonawani. "PCOcare: PCOS detection and prediction using machine learning algorithms." Biosci. Biotechnol. Res. Commun. 13.14 (2020): 240-244.

[13] Aggarwal, Shivani, and Kavita Pandey. "An Analysis of PCOS Disease Prediction Model Using Machine Learning Classification Algorithms." Recent Patents on Engineering 15.6 (2021): 53-63.

[14] Nilofer, N. S. "Follicles Classification to Detect Polycystic Ovary Syndrome Using Glcm And Novel Hybrid Machine Learning." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.7 (2021): 1062-1073.

[15] Cicek, Ipek Balikci, Zeynep Kucukakcali, and Fatma Hilal Yagin. "Detection of risk factors of PCOS patients with Local Interpretable Model-agnostic Explanations (LIME) Method that an explainable artificial intelligence model." The Journal of Cognitive Systems 6.2: 59-63.

[16] Shubham Bhosale, Lalit Joshi, Arun Shivsharan, “PCOS (Polycystic Ovarian Syndrome) Detection Using Deep Learning”, International Research Journal of Modernization in Engineering Technology and Science, Vol. 04, Issue 01, January-2022

[17] Ramamoorthy, Suganya, and Rajaram Sivasubramaniam. "Monitoring the growth of polycystic ovary syndrome using mono-modal image registration technique: Application of medical big data in healthcare." Proceedings of the ACM India Joint International Conference on Data Science and Management of Data. 2019.

[18] Ashwini Kodipalli and Susheela Devi, “Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM”, Frontiers in Public Health, November 2021, Vol. 9.

[19] Suha, S.A., Islam, M.N. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Sci Rep 12, 17123 (2022).

[20] Wei Zhou, Fang Fang, Wenting Zhu, Zi-Jiang Chen, Yanzhi Du and Jun Zhang, “Bisphenol A and ovarian reserve among infertile women with polycystic ovarian syndrome." International journal of environmental research and public health 14.1 (2017): 18.

[21] Faris, NajlaaNsrulaah, and Firsas Saber Miften. "An intelligence model for detection of PCOS based on k-means coupled with LS-SVM." Concurrency and Computation: Practice and Experience 34.21 (2022): e7139.

[22] Barthelmess, E.K.; Naz, R.K. Polycystic ovary syndrome: Current status and future perspective. Front. Biosci. (Elite Ed.) 2014, 6, 104–119.

[23] Azziz, R. Polycystic Ovary Syndrome. Obstet. Gynecol. 2018, 132, 321–336.

[24] Naz, M.S.G.; Tehrani, F.R.; Majd, H.A.; Ahmadi, F.; Ozgoli, G.; Fakari, F.R.; Ghasemi, V. The prevalence of polycystic ovary syndrome in adolescents: A systematic review and meta-analysis. Int. J. Reprod. Biomed. 2019, 17, 533–542.

[25] Mei-Lien Pan, Li-Ru Chen and Kuo-Hu Chen (2021), “The Risk of Subsequent Miscarriage in Pregnant Women with Prior Polycystic Ovarian Syndrome: A Nationwide Population-Based Study”, Int J Environ Res Public Health. 18(16): 8253.

[26] Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril. 2009 Feb;91(2):456–88.

[27] Riestenberg C, Jagasia A, Markovic D, Buyalos RP, Azziz R. Health Care-Related Economic Burden of Polycystic Ovary Syndrome in the United States: Pregnancy-Related and Long-Term Health Consequences. J Clin Endocrinol Metab. 2022 Jan 18;107(2):575–85.

[28] Sirmans SM, Pate KA. Epidemiology, diagnosis, and management of polycystic ovary syndrome. Clin Epidemiol. 2013 Dec 18; 6:1–13.
[29] Barry JA, Azizia MM, Hardiman PJ. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. Hum Reprod Update. 2014;20(5):748–58.

[1] *Department of Electronics and Communication Engineering, Institute of Technology, University of Kashmir*