Vol. 12 No.1 Tahun 2026, pp 21-32
Jurnal Teknologi dan Manajemen Informatika
http:// http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604 E-ISSN: 2580-8044

Application of VGG16 Deep Learning Architecture and K-Means for PCOS Severity Clustering Based on Ovarian Ultrasound Images

Nurul Hidayah [1], Asfan Muqtadir [2], Andik Adi Suryanto [3]

PGRI Ronggolawe University, Indonesia

Article Info

Article History
Received : 12-03-2026
Revised : 24-04-2026
Accepted : 26-05-2026

Keywords
PCOS;
VGG16;
K-Means Clustering;
Ultrasound Image

Corresponding Author
Nurul Hidayah,
PGRI Ronggolawe University,
hid88hidayah@gmail.com

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a complex hormonal disorder that significantly impacts the reproductive and metabolic health of women. A primary challenge in PCOS analysis is the reliance on subjective manual observations of ultrasound (USG) images, which are prone to inconsistency due to the lack of objective standards. This study proposes an automated, unsupervised framework to stratify the morphological characteristics of PCOS ultrasound images without predefined clinical labels. The method integrates feature extraction using the VGG16 deep learning architecture with K-Means clustering. The structural integrity of the computational clusters was quantitatively evaluated, achieving a Silhouette Score of 0.4248 and a Davies-Bouldin Index (DBI) of 0.9175, indicating the successful formation of distinct morphological groupings. Furthermore, an evaluation using a Support Vector Machine (SVM) to assess the internal consistency and linear separability of the extracted features yielded a test accuracy of 97.54%. The results demonstrate that the integration of VGG16 and K-Means effectively partitions PCOS ultrasound images into objective, computationally stable morphological groups. This approach is designed to function as a standardized clinical decision support tool, mitigating visual subjectivity and providing a measurable baseline for women's reproductive health management.

INTRODUCTION

PCOS is a hormonal disorder commonly experienced by women of reproductive age and characterized by a combination of hyperandrogenism, ovulatory dysfunction, and polycystic ovarian morphology[1]. The World Health Organization estimates that PCOS affects 6–13% of women of reproductive age globally, with the rate of undiagnosed or unrecognized cases reaching 70%[2]. Several clinical studies indicate that the prevalence of PCOS in Indonesia ranges from 5–10% and is identified as one of the leading causes of infertility in women. This prevalence range aligns with recent literature, who estimate that PCOS may affect 2.2–20% of women of reproductive age[3]. The prevalence of PCOS among adolescents in Indonesia was found to be 6.5%, with significant risk factors including obesity, family history, irregular menstrual cycles, and hirsutism[4].

PCOS is caused by many factors, including genetic factors, hypothalamic-pituitary-ovarian axis disorders, adrenal androgens, ethnicity, insulin resistance, epigenetics, metabolism, environment, and lifestyle[5]. In the long term, PCOS can also increase the risk of type 2 diabetes, metabolic syndrome, heart disease, endometrial cancer, pregnancy complications, and mental health disorders[6]. The clinical diagnosis of PCOS is made by examining symptoms, hormonal levels, and ovarian morphology through USG to detect the number of small follicles and increased ovarian volume[7]. Ultrasound examination is

considered very useful in clinical practice, but the interpretation of USG images still depends on the visual observation of a doctor, so there is a potential for subjectivity, differences in perception, and differences in assessment standards between health facilities[8]. The subjectivity in interpreting ultrasound images not only causes delays in diagnosis and misclassification but also has the potential to lead to inaccurate treatment strategies for PCOS patients. Most PCOS assessments focus only on binary classification, i.e., PCOS or non-PCOS, without providing information related to measurable severity[9]. However, the severity of ovarian morphology greatly influences the risk of infertility, metabolic complications, and the determination of more appropriate and effective treatment strategies. Therefore, there is a need to develop a more objective method for determining the severity of PCOS.

With the development of artificial intelligence (AI) technology, various machine learning (ML) and deep learning (DL) approaches have been applied to improve accuracy and automation in PCOS diagnosis from clinical, biochemical, and ultrasound image data. Several studies have developed automated models based on algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Logistic Regression, and XGBoost, which show high accuracy of up to 95% in binary classification scenarios of PCOS and Non-PCOS[8], [10]. Conventional approaches using texture or shape feature extraction from ultrasound images are still very limited and sensitive to speckle noise, which is often found in ultrasound images. Meanwhile, deep learning approaches based on Convolutional Neural Networks (CNN) and transfer learning, such as VGG16, show significant performance improvements. In a study conducted by Suha and Islam[1], the results showed an accuracy of 99.89% with a combination of VGG16 as a feature extractor and stacking ensemble with XGBoost as a meta-learner. Meanwhile, research by Shanmugavadivel et al. [8] showed that VGG16 outperformed standard CNNs with an accuracy of 98.29% on a dataset of 3,856 ultrasound images.

Although previous studies have shown significant results in distinguishing or classifying between PCOS and non-PCOS, they still focus on binary separation without grouping the severity of PCOS based on ovarian morphology. In addition, the supervised learning approach relies on manual labeling by doctors for each ultrasound image, which in practice requires a lot of time and money, and can contain elements of subjectivity[11]. This can hinder the development of models that can be applied on a large scale with a broader dataset to train reliable and well-generalized models. Given the limitations of conventional manual feature extraction and the inherent subjectivity of human visual assessments, the novelty of this study lies in integrating the VGG16 deep learning architecture with K-Means Clustering to automatically group PCOS ultrasound image morphologies based on objective visual similarities. It is crucial to emphasize, however, that the resulting clusters represent purely computationally defined groupings within the feature space, rather than clinically validated severity classifications. To rigorously evaluate the quality, separability, and internal consistency of these clusters, a SVM is employed as a supervised validation mechanism.

This study aims to develop an automated image analysis method that partitions PCOS ultrasound images into distinct morphological groups based on visual feature patterns extracted by deep learning. Rather than claiming a clinical severity classification, this work focuses on demonstrating the feasibility of unsupervised clustering for PCOS ultrasound image stratification and validating the internal consistency of VGG16-derived features through SVM re-classification. Furthermore, it provides a preliminary computational framework that could potentially serve as a supplementary decision support tool when combined with clinical expertise. Thus, this study represents an initial methodological step toward exploring automated morphological categorization in PCOS imaging, acknowledging that clinical validation by radiologists or gynecologists remains essential prior to any real-world implementation.

METHODS

The overall research method is illustrated in Figure 1, which depicts the systematic flow from raw data processing to the evaluation of PCOS severity clustering results.

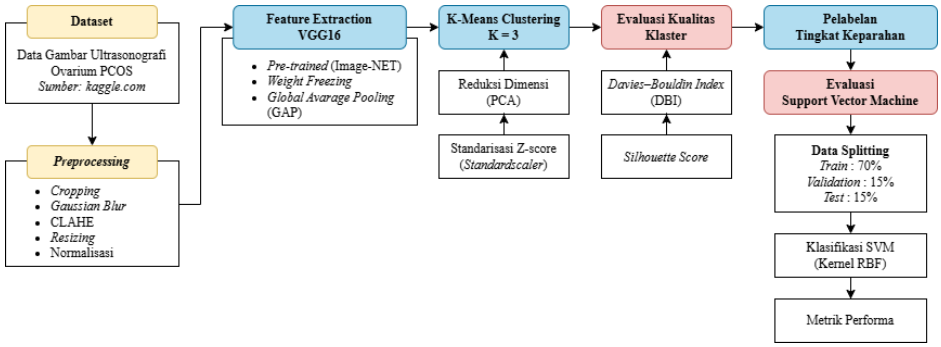

Figure 1. Research Method

Based on Figure 1, there are several stages in this study. The stages begin with the collection of raw datasets in the form of ovarian ultrasound images, which are then processed through a pre-processing stage to clarify the morphology of the follicles. After that, feature extraction is performed using a CNN model with a VGG16 architecture. At this stage, feature extraction is performed automatically to convert visual representations into high-dimensional vector matrices. The output feature vectors of VGG16 are then reduced in dimension and partitioned using the K-Means Clustering unsupervised learning algorithm to form clusters. Before labeling, the quality of the clusters formed is evaluated using the Silhouette Score and Davies-Bouldin Index (DBI). Next, the clusters are labeled and mapped to severity levels, and the stage ends with model evaluation using a SVM to prove the overall results of K-Means clustering.

Dataset

The dataset used in this study was sourced from the public Kaggle repository titled "PCOS-XAI Ultrasound: Real-World Training Dataset"[12]. The data consists of results from a retrospective ultrasound study collected from three different clinics between 2018 and 2022. The validity of the "Infected" label in this dataset has been verified through an annotation process conducted by radiology residents under the direct supervision of expert consultants, to ensure that all images truly represent PCOS cases and are not mixed with other types of ovarian cysts. A total of 6,784 images from the "Infected" class were used as the primary input, where the natural variations in follicle morphology recorded in the dataset served as the basis for computational morphological group stratification. All data underwent a de-identification process by removing patient identities and DICOM metadata to comply with research ethics standards.

Data Pre-processing

Data pre-processing aims to clean and prepare data before further analysis. Ultrasound image pre-processing in this study involved several sequential procedures. The stage begins with center cropping the image to 85% of its original dimensions. The choice of the center cropping technique is based on technical considerations to eliminate non-clinical artifacts, ultrasound machine interface text, measurement scales, and dense speckle noise that consistently accumulate at the outer margins of the image[13]. Given that polycystic ovaries have a characteristic clinical feature of follicle distribution in the peripheral area (peripheral string of pearls), this cropping method does carry a minor risk of information loss in the outermost areas. However, the 85% ratio was selected as the optimal compromise threshold. Based on visual inspection of the dataset used, the ovarian structures were generally centered in the Region of Interest (ROI) by the sonographer during data acquisition. Therefore, this cropping step is considered crucial to prevent the VGG16 model from learning patterns from peripheral noise and text, retaining the main structural areas of the ovary as much as possible,

although the potential loss of extreme peripheral features is acknowledged as a limitation in this study.

Next, the reduction of speckle noise typical of ultrasound images is performed using a Gaussian blur, because speckle has been shown to reduce the resolution, contrast, and accuracy of diagnostic systems with ultrasound image data[14]. To further clarify the image by increasing contrast and texture, CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied, which has been shown to be effective in improving visibility in medical images[15]. Finally, after the images became cleaner and clearer, resizing and normalization were performed to ensure that the images were in accordance with the standard input requirements for the VGG16 model.

Feature Extraction using VGG16

Feature extraction was performed using the pre-trained VGG16 architecture to recognize basic visual patterns in ovarian ultrasound images and improve diagnostic accuracy[16].

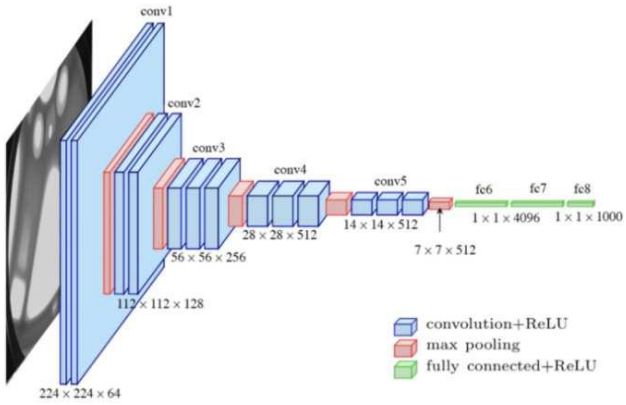

Figure 2. VGG16 architecture

Figure 2 shows the VGG16 architecture. As feature extraction, VGG16 focuses on reading image features, not on performing final classification. Therefore, the classification layer at the end of the VGG16 model or the fully connected layer is cut off. Additionally, to prevent overfitting and reduce computational load, the weights in the convolutional block are frozen, thereby transforming VGG16 into a passive sensor that only reads and extracts image patterns without parameter updates[17]. Ultrasound images that have been processed using VGG16 will produce complex feature maps, which are then summarized into a more compact form without discarding important information using the Global Average Pooling (GAP) method[18]. Mathematically, the GAP process is formulated as equation (1).

$$GAP_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} A_k\,(i,j) \tag{1}$$

Where $A_k\,(i,j)$ is the feature value on channel k at coordinate position (i,j), while H dan W represent the height and width of the feature map.

K-Means Clustering

The stage after feature extraction is the grouping of ovarian ultrasound images based on pattern similarity using the unsupervised learning algorithm K-Means Clustering. K-Means is a centroid-based clustering algorithm that divides unlabeled data into groups. K-Means works by grouping data to the nearest cluster center (centroid) and minimizing the sum of squared distances between each data point and its cluster centroid (within-cluster sum of squares/WCSS)[19]. The WCSS function is formulated mathematically in equation (2).

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{2}$$

Where C_i is the set of data points in the i -th cluster and μ_i is the average cluster center (centroid) of that cluster.

K-Means clusters data based on Euclidean distance, so that features with large values will dominate the calculation and ignore features with small values. Therefore, normalization steps using standardization, such as Z-score (StandardScaler), are needed to improve the accuracy of clustering results[20]. After that, features that are still high-dimensional are reduced while retaining important variance information using Principal Component Analysis (PCA) to eliminate noise, facilitate visualization, and speed up computation time[21].

The K-Means stage begins with initialization, which is setting the number of clusters and selecting initial centroids that are significant in the feature space. After that, the assignment is carried out by calculating the Euclidean distance of each feature vector and automatically entering it into the cluster that has the closest centroid distance. The final stage is to update the centroid by recalculating the centroid position based on the average of all new data entering the cluster[22].

Cluster Quality Evaluation (Silhouette Score and DBI)

After the clustering stage using K-Means, an internal evaluation was conducted to determine the quality of the clustering and prove that the data within the clusters was dense and clearly separated from other clusters, as well as to assess whether the clusters formed were clear and well-defined prior to labeling. This is because K-Means works without ground truth labels[23]. In this study, the evaluation was performed using two metrics, Silhouette Score and DBI.

Silhouette Score evaluates the quality of clustering by comparing the average distance between points to other points in the same cluster. Silhouette Score has a value range from -1 to 1. A value close to 1 indicates that the data is well separated. Conversely, a value close to 0 indicates possible data overlap, while a negative value indicates an error in clustering[24]. The DBI evaluates the trade-off between the level of compactness and separation between clusters. The lower or smaller the DBI value, the more optimal the clustering results are, indicating that the members within the cluster are very close together and the distance between clusters is far apart[25]. Next, a 2D scatter plot visualization was performed by projecting the feature matrix to determine the distribution of the three PCOS severity clusters.

Feature-Based Labeling

In the K-Means clustering stage, the groups generated from the image features are initially assigned purely numerical and non-semantic labels, such as 0, 1, and 2. To facilitate meaningful interpretation, these numerical assignments were mapped into distinct morphological categories (Cluster A, Cluster B, and Cluster C). This mapping was determined by observing the dominant visual patterns such as spatial density, textural heterogeneity, and structural complexity at each cluster's centroid, rather than relying on unverified clinical severity levels[26]. This feature-based labeling strategy effectively converts the unsupervised clustering outputs into structured, descriptive categories ready for use in the supervised learning evaluation model.

Supervised Evaluation with SVM

SVM is a supervised learning algorithm for classification and regression that works by finding the optimal hyperplane that separates data from different classes with the largest margin in feature space[27]. The use of the Radial Basis Function (RBF) kernel, which is flexible and effective for non-linear and high-dimensional data, including images and medical data[28], aims to reassure the audience of the method's capability to handle complex data structures confidently.

The stages begin with splitting the clustering results data and continue with training on the training data. After that, testing is carried out on the test data, which basically functions as a system consistency test[29]. To evaluate the SVM's performance, metrics such as accuracy, precision, and recall are used. If the SVM model classifies cluster labels with high accuracy, this indicates that the K-Means clustering results have a highly consistent and well-structured

visual pattern representation. Conversely, low accuracy suggests overlap between clusters, highlighting areas for potential improvement.

To evaluate the performance of the SVM model, the classification results from the SVM model on the test data are extracted into a Confusion Matrix. The Confusion Matrix displays the performance of the algorithm by comparing the model's prediction results with the actual labels from the test data set. Based on the Confusion Matrix, the model performance is evaluated through a Classification Report that presents the accuracy, precision, recall, and F1-score metrics for each level of PCOS severity[30].

RESULTS AND DISCUSSION

The steps for pre-processing ovarian ultrasound image data include cropping 85% of the center area of the image, Gaussian Blur to reduce speckle noise, CLAHE to increase image contrast, resizing to 224 × 224 pixels, and normalizing pixel values to the range [0,1]. The image data that has undergone the pre-processing stage is ready for use in the next stage. A comparison of the images before and after the pre-processing stage is shown in Figure 3.

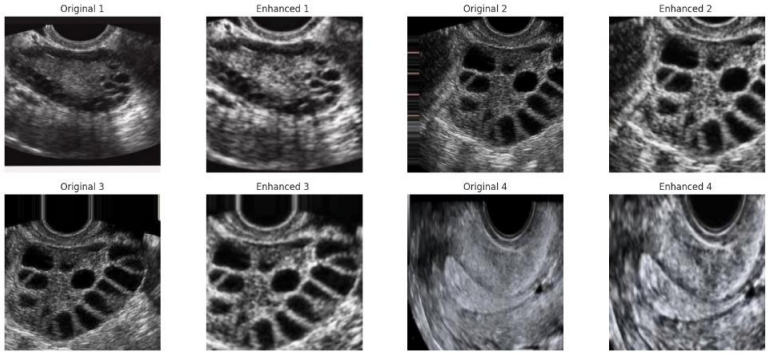

Figure 3. Visualization of Image Comparison Before and After Preprocessing

Feature Extraction using VGG16

The feature extraction stage transforms data from image or visual pixel form into numerical vectors. This stage utilizes layers from the VGG16 architecture, starting from the input layer to the last convolution block, but without the classification layer (fully connected layer). A summary of the VGG16 architecture model used for feature extraction is presented in Table 1.

Table 1. Architectural Structure of the VGG16 Feature Extractor Model

Layer Name	Output Shape	Number of Parameters
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1.792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36.928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73.856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147.584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295.168
Layer Name	**Output Shape**	**Number of Parameters**
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590.080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590.080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1.180.160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2.359.808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2.359.808

block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2.359.808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_average_pooling2d (GAP2D)	**(None, 512)**	0
Total Parameter		**14.714.688**

Based on Table 1, it can be seen that the spatial dimensions of the image undergo gradual reduction starting from the input layer, where the model receives an image measuring 224 × 224 pixels, to the last convolution layer, where the image becomes 7 × 7 pixels. This is because the model discards redundant pixel details and only retains important information that reflects the characteristics of the follicles in the image. In the GAP layer, the spatial dimensions of the last feature map (7 × 7 × 512) are reduced to a single 512-dimensional feature vector (1 × 1 × 512) and produce an output in the form of a feature matrix measuring (6,784 × 512). The output dimension (None, 512) for each ovarian ultrasound image has been standardized and is ready to be clustered by the K-Means algorithm without excessive computational load.

K-Means Clustering and Data Distribution Analysis

In the feature extraction stage using VGG16, a 512-dimensional feature vector was obtained. K-Means clustering operates by partitioning data based on Euclidean distance metrics, which can cause features with larger magnitudes to disproportionately influence clustering assignments. To address this issue, Z-score standardization (StandardScaler) was applied to normalize all 512-dimensional feature vectors to unit variance, ensuring equitable feature contribution to distance calculations. The K-Means algorithm was implemented to cluster the induced data with the parameter K = 3.

After standardization, PCA was applied to reduce the 512-dimensional feature space to 2 principal components for visualization purposes. This 2D projection is employed solely for visual exploration and interpretation, not for actual clustering computations. The K-Means clustering algorithm itself operated in the original 512-dimensional space with standardized features. The PCA transformation, while useful for understanding cluster separation in lower-dimensional space, inherently discards significant variance information:

a. PC1 (First Principal Component) explained 10.69% of the total variance.
b. PC2 (Second Principal Component) explained 7.01% of the total variance.
c. Combined explained variance (PC1 + PC2): 17.70%.
d. Unexplained variance retained in dimensions 3-512: 82.30%.

The 2D scatter plot visualization (Figure 4) provides only a partial projection of the underlying high-dimensional cluster structure. Consequently, it serves primarily as a descriptive visual aid rather than definitive evidence of cluster separability. To ensure methodological rigor, the actual cluster quality metrics namely the Silhouette Score and Davies-Bouldin Index, were computed directly within the original 512-dimensional feature space, thereby reflecting the true clustering performance independent of any visualization artifacts.

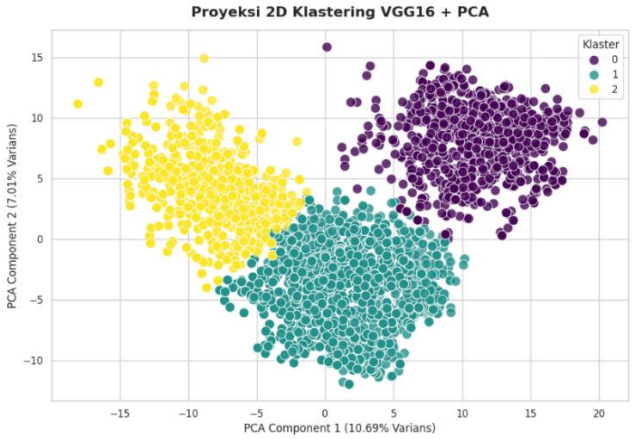

Figure 4. Scatter Plot Projection of K-Means Clustering Results

Based on Figure 4, the K-Means algorithm has formed three clusters with fairly clear spatial separation, namely cluster 0 (purple), cluster 1 (turquoise green), and cluster 2 (yellow). Although there are boundary areas that are close to or overlap between groups, the main components, PC1 and PC2, attempt to map the differences in the visual characteristics of the ultrasound images. Specifically, cluster 0 is separated at the top right, cluster 1 dominates the lower middle area, and cluster 2 is scattered in the upper left area. The results of the clustering process show the distribution of the number of images in each cluster, with details of the data distribution presented in Table 2.

Table 2. Distribution of K-Means Clustering Results (K=3)

Cluster Label	Number of Images	Percentage
0	1.068	15,74 %
1	3.691	54,41 %
2	2.025	29,85 %

Based on Table 2, the clustering results show an imbalance in the distribution of the number of members in each cluster. The dominance of the number in cluster 1 indicates that most images have homogeneous morphological characteristics.

Cluster Quality Evaluation

The quality of clusters generated by the K-Means algorithm was evaluated using two metrics, namely Silhouette Score and DBI. Based on the evaluation metric calculations, the average Silhouette Score value obtained was 0.4248. This value is close to 0.5 and is in the moderate positive range, indicating that the clustering structure has been formed reasonably well, where images have a higher degree of similarity to their own clusters than to neighboring clusters.

While the primary clusters are distinct, the feature space exhibits a degree of data overlap at the boundary areas. This overlapping characteristic reflects a natural intersection at the morphological thresholds. Because polycystic ovary syndrome manifests as a continuous biological spectrum rather than strictly discrete categories, the morphological transition from sparse follicular structures to dense configurations inherently produces intersecting visual features. Consequently, the moderate Silhouette Score accurately captures this natural continuum of structural variations present in the ultrasound images.

This finding is reinforced by the DBI value of 0.9175. A DBI value below the threshold of 1.0 indicates that the cluster separation quality is relatively good, where the ratio between the internal distribution of data (intra-cluster) is relatively smaller than the separation distance between clusters (inter-cluster). This shows that the K-Means algorithm can form a cluster structure that is quite representative in distinguishing the characteristics of PCOS ultrasound image data. Overall, the combination of a positive Silhouette Score and a low DBI confirms

that the division of data into 3 clusters is not a random result but rather represents the natural structure of the ultrasound image morphology successfully captured by the VGG16 feature.

Feature-Based Labeling

In the K-Means clustering stage, labeling was performed to map each numerical cluster into distinct morphological categories (Cluster A, Cluster B, Cluster C). This mapping was achieved by analyzing the proportion of the data distribution and the average visual feature characteristics concentrated at each centroid. To maintain strict methodological rigor and avoid unverified medical claims, these groups are defined exclusively by their visual feature similarities such as textural heterogeneity, edge patterns, and spatial densities extracted by the VGG16 network. Referring to the K-Means data distribution (Table 2), the structural mapping was established as follows:

1. Cluster A (from Computational Cluster 0)
   This cluster contains 1,068 images (15.74%). Based on the visual features mapped near this centroid, the VGG16 network identifies patterns consistent with lower morphological complexity. Images in this group typically exhibit a more homogenous ovarian stroma with a relatively minimal number of visible anechoic structures (follicles). The lower variance in pixel intensity suggests a simpler structural composition.
2. Cluster B (from Computational Cluster 2)
   This cluster contains 2,025 images (29.85%) and occupies an intermediate position within the feature space. The visual features in this group show a moderate increase in textural heterogeneity and follicle density compared to Cluster A. This represents a transitional morphological pattern, capturing ovaries with developing multi-follicular states that bridge the gap between sparse and highly dense configurations.
3. Cluster C (from Computational Cluster 1)
   Covering the majority of the dataset with 3,691 images (54.41%), this dominant cluster represents the highest level of visual complexity. The VGG16 feature maps for this group strongly activate on patterns representing high contrast, significant structural enlargement, and dense cystic congregations (visually correlating to >12 follicles). The dense spatial distribution of these features distinctively separates this cluster from the other two.

Evaluation with SVM

The final stage of the methodology employs a SVM classification algorithm to evaluate the internal consistency and linear separability of the computational clusters. In this framework, the SVM serves as a supervised validation mechanism to test whether the high-dimensional visual features extracted by VGG16 are sufficiently robust and distinctive to reliably reproduce the decision boundaries established by the unsupervised K-Means algorithm. To conduct this evaluation, the clustered dataset was partitioned into 70% training data, 15% validation data, and 15% testing data. Based on the evaluation of the test set, the SVM model achieved a highly significant classification performance with an overall accuracy of 97.54%. The detailed performance metrics for each morphological group are presented in Table 3.

Table 3. Classification Report

Category	Precision	Recall	F1-Score	Support
Cluster A	0.97	0.97	0.97	160
Cluster B	0.97	0.98	0.97	554
Cluster C	0.98	0.98	0.98	304
Accuracy			0.98	1018
Macro avg	0.97	0.97	0.97	1018
Weighted avg	0.98	0.98	0.98	1018

Referring to Table 3, the model demonstrates balanced performance across all computational categories, with Precision, Recall, and F1-Score values consistently ranging between 0.97 and 0.98. These high evaluation scores quantitatively indicate that the visual features extracted by VGG16 possess highly distinctive mathematical characteristics. This distinctiveness enables the SVM hyperplane to effectively separate the three feature-based

clusters with a minimal margin of error. A more in-depth analysis of the classification distribution is illustrated in the confusion matrix in Figure 5.

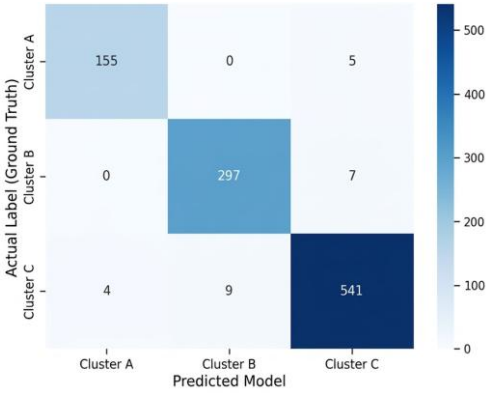

Figure 5. Confusion Matrix of SVM prediction results on test data

As shown in Figure 5, the model exhibits a highly effective prediction pattern. While the overall performance is stable, a minor number of misclassifications occurred, primarily concentrated in the feature-space transition areas between adjacent categories. Specifically, 4 images from Cluster C were predicted as Cluster A, 9 images from Cluster C as Cluster B, 7 images from Cluster B as Cluster C, and 5 images from Cluster A as Cluster C. Computationally, these misclassifications occur due to the inherent similarities in texture patterns and pixel densities among ultrasound images located strictly at the mathematical thresholds between clusters. Nevertheless, the strong dominance of values along the main diagonal of the confusion matrix statistically confirms that the integration of VGG16 feature extraction and K-Means clustering yields a highly consistent and computationally stable morphological stratification.

CONCLUSIONS

Based on the results of this study, the integration of the VGG16 deep learning architecture for feature extraction with the K-Means clustering algorithm successfully partitioned the PCOS ultrasound image dataset into three distinct morphological groups (Cluster A, Cluster B, and Cluster C). The quality of this computational stratification was confirmed by a Silhouette Score of 0.4282 and a DBI of 0.9175, indicating a robust clustering structure with clear inter-cluster separation, notwithstanding the natural feature intersections. Furthermore, the supervised evaluation using a SVM demonstrated high internal consistency of the extracted features, successfully reproducing the morphological boundaries with a test accuracy of 97.54%.

Within a practical context, the generated clusters represent mathematically optimized feature-space groupings based on visual similarities. Therefore, the proposed framework is designed to function as a clinical Decision Support Tool. By providing an objective computational baseline, this system effectively mitigates the potential for subjectivity inherent in human visual assessments. It serves as a supplementary analytical instrument to assist clinicians, rather than a replacement for professional medical diagnosis.

For future developments, further research is recommended to explore data balancing techniques to enhance the model's sensitivity across all morphological variations. Additionally, integrating patient clinical variables, such as hormonal profiles and BMI, would be highly beneficial in developing a more comprehensive, multimodal classification system with deeper clinical relevance.

REFERENCES

[1] M. Bashir, F. Amin, H. Amin, T. A. Teli, and T. A. Tali, "Polycystic ovary syndrome: a comprehensive review of pathophysiology, diagnosis, and emerging management strategies," *Int. J. Reprod. Contracept. Obstet. Gynecol.*, vol. 14, no. 10, pp. 3618–3623, Sep. 2025, doi: 10.18203/2320-1770.ijrcog20253125.

[2] World Health Organization, "Polycystic ovary syndrome." Accessed: May 01, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome

[3] D. A. Sari, E. Y. Kurniawati, and M. A. Ashari, "SKRINING DAN DETERMINAN KEJADIAN SINDROM OVARIUM POLIKISTIK (SOPK) PADA REMAJA," *Jurnal Ilmu Kebidanan*, vol. 9, no. 2, 2023, doi: 10.48092/jik.v9i2.211.

[4] Habiburrahman Said, Rheina Weisch Fedre, Saurie Hernandez, Sophia Lucille Rodriguez, Fatimah Mursyid, and Irna Nettles, "The Prevalence and Risk Factors for Polycystic Ovary Syndrome (PCOS) among Adolescents in Indonesia: Implications for Early Intervention," *Sriwijaya Journal of Obstetrics and Gynecology*, vol. 1, no. 2, 2023, doi: 10.59345/sjog.v1i2.83.

[5] H. Bai, H. Ding, and M. Wang, "Polycystic Ovary Syndrome (PCOS): Symptoms, Causes, and Treatment," 2024, *IMR Press Limited*. doi: 10.31083/j.ceog5105126.

[6] K. M. Hoeger, A. Dokras, and T. Piltonen, "Update on PCOS: Consequences, Challenges, and Guiding Treatment," Mar. 01, 2021, *Endocrine Society*. doi: 10.1210/clinem/dgaa839.

[7] S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21724-0.

[8] K. Shanmugavadivel, M. D. M S, M. T R, T. Al-Shehari, N. A. Alsadhan, and T. E. Yimer, "Optimized polycystic ovarian disease prognosis and classification using AI based computational approaches on multi-modality data," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02688-9.

[9] J. Kermanshahchi, A. J. Reddy, J. Xu, G. K. Mehrok, and F. Nausheen, "Development of a Machine Learning-Based Model for Accurate Detection and Classification of Polycystic Ovary Syndrome on Pelvic Ultrasound," *Cureus*, Jul. 2024, doi: 10.7759/cureus.65134.

[10] K. V. Suvika and D. G. Jyothi, "A Machine Learning Approach for Efficient Identification and Severity Grading of PCOS and PCOD Using Optimized Features," *Journal of Machine and Computing*, vol. 5, no. 4, pp. 1994–2005, Oct. 2025, doi: 10.53759/7669/jmc202505156.

[11] D. N. Varshitha *et al.*, "Automated Detection of Polycystic Ovary Syndrome Using Convolutional Neural Networks on Ultrasound Images," *Salud, Ciencia y Tecnologia*, vol. 5, Jan. 2025, doi: 10.56294/saludcyt20252272.

[12] Ibadeus, "PCOS-XAI Ultrasound: Real-World Training Dataset." Accessed: Nov. 12, 2025. [Online]. Available: https://www.kaggle.com/datasets/ibadeus/pcos-xai-ultrasound-dataset

[13] A. A. Al-Jaburi and A. H. Al-Sudani, "Medical Ultrasound Image Quality Enhancement and Regions Segmentation," *Iraqi Journal of Science*, vol. 63, no. 10, pp. 4518–4533, 2022, doi: 10.24996/ijs.2022.63.10.35.

[14] K. Sikhakhane, S. Rimer, M. Gololo, K. Ouahada, and A. M. Abu-Mahfouz, "Evaluation of Speckle Noise Reduction Filters and Machine Learning Algorithms for Ultrasound Images," *IEEE Access*, vol. 12, pp. 81293–81312, 2024, doi: 10.1109/ACCESS.2024.3411709.

[15] T. B. Nguyen-Tat, T. Q. Hung, P. T. Nam, and V. M. Ngo, "Evaluating pre-processing and deep learning methods in medical imaging: Combined effectiveness across multiple

modalities," *Alexandria Engineering Journal*, vol. 119, pp. 558–586, Apr. 2025, doi: 10.1016/j.aej.2025.01.090.
[16] A. Chauhan and I. Kumar, "Deep feature extraction and optimized VGG16-SVM architecture for breast cancer characterization," *Discover Computing*, vol. 28, no. 1, Dec. 2025, doi: 10.1007/s10791-025-09736-6.
[17] M. S. K. Inan, F. I. Alam, and R. Hasan, "Deep Integrated Pipeline of Segmentation Guided Classification of Breast Cancer from Ultrasound Images," Feb. 2022, doi: 10.1016/j.bspc.2022.103553.
[18] M. J. Ramesh, D. R. Manavalan, and D. K. M. Prabusankarlal, "Feature Extraction Of Ultrasound Prostate Image Using Modified Vgg-19 Transfer Learning," 2021.
[19] H. Liu, J. Chen, J. Dy, and Y. Fu, "Transforming Complex Problems Into K-Means Solutions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 9149–9168, Jul. 2023, doi: 10.1109/TPAMI.2023.3237667.
[20] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS One*, vol. 19, no. 12, Dec. 2024, doi: 10.1371/journal.pone.0310839.
[21] B. M. S. Hasan and A. M. Abdulazeez, "A Review of Principal Component Analysis Algorithm for Dimensionality Reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, Apr. 2021, doi: 10.30880/jscdm.2021.02.01.003.
[22] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Annals of Data Science*, vol. 11, no. 5, 2024, doi: 10.1007/s40745-022-00428-2.
[23] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, Dec. 2023, doi: 10.59653/ijmars.v2i01.504.
[24] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," 2020. doi: 10.1109/DSAA49011.2020.00096.
[25] A. Sujjada, G. Purnama Insany, and S. Noer, "Analisis Clustering Data Penyandang Disabilitas Menggunakan Metode Agglomerative Hierarchical Clustering dan K-means," *Jurnal Teknologi dan Manajemen Informatika*, vol. 10, pp. 1–12, 2024, doi: https://doi.org/10.26905/jtmi.v10i1.10654.
[26] C. H. Yeh, S. X. Yu, and M. C. Lin, "Meibography phenotyping and classification from unsupervised discriminative feature learning," *Transl. Vis. Sci. Technol.*, vol. 10, no. 2, pp. 1–11, Feb. 2021, doi: 10.1167/tvst.10.2.4.
[27] K. L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions," Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/math12243935.
[28] S. Pal, L. H. Trang, V. T. Hieu, D. D. Nguyen, D. Q. Vu, and I. Prakash, "Investigation of Support Vector Machines with Different Kernel Functions for Prediction of Compressive Strength of Concrete," *Journal of Science and Transport Technology*, vol. 4, no. 2, 2024, doi: 10.58845/jstt.utt.2024.en.4.2.55-68.
[29] S. Avcı Azkeskin and Z. Aladag, "Evaluating regional sustainable energy potential through hierarchical clustering and machine learning," *Environ. Res. Commun.*, vol. 7, no. 1, 2025, doi: 10.1088/2515-7620/ada2e5.
[30] S. Simhadri, M. Ponnam, R. Rajitha, and R. Balamurugan, "Enhanced Multi-Class Model Evaluation: Analyzing BERT, GPT-2, and LLaMA with Precision, Recall, and F1-Score Metrics," in *Proceedings of the 4th International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2025*, 2025. doi: 10.1109/ICIMIA67127.2025.11200914.