Polycystic Ovary Syndrome (PCOS) Detection Using Machine Learning Models

**Abstract**

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder that affects women of reproductive age. Early diagnosis is crucial for effective management of the condition. This report focuses on the application of machine learning (ML) techniques to detect PCOS using MRI images. Four distinct models—Neural Networks (NN), Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM)—were implemented and evaluated. Each model was trained and tested using an MRI dataset specific to PCOS cases. The performance of these models was compared in terms of accuracy, and computational efficiency. The results highlight the strengths and weaknesses of each method and provide insights into the feasibility of using ML for PCOS detection.

**Introduction**

Polycystic Ovary Syndrome (PCOS) is a complex and often misunderstood condition characterized by hormonal imbalances, irregular menstrual cycles, and ovarian cysts. Accurate and early detection is paramount to mitigate long-term complications such as infertility, diabetes, and cardiovascular diseases. Recent advancements in medical imaging and machine learning have opened new avenues for the automated detection of PCOS. Among the imaging modalities, Magnetic Resonance Imaging (MRI) offers high-resolution insights into ovarian morphology, making it a suitable candidate for automated diagnosis using ML techniques.

This study explores the integration of ML algorithms for PCOS detection. Specifically, four algorithms were employed: a traditional Neural Network (NN), Convolutional Neural Network (CNN) for image-specific tasks, and classical ML models K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). By comparing these models, this study aims to identify the most effective approach for PCOS detection using MRI data.

**Related Works**

Machine learning has been extensively used in medical diagnostics, ranging from cancer detection to neurological disorder classification. In the domain of gynecology, ML has shown promise in diagnosing PCOS through clinical and biochemical data. However, studies leveraging MRI data for PCOS detection remain limited.

1. **Image-Based Diagnostics:** CNNs have been widely used for analyzing MRI images in various fields, such as brain tumor segmentation and cardiac analysis. This motivated the choice of CNN for this study.
2. **Classical ML Approaches:** Models like KNN and SVM have been effective in low-dimensional diagnostic tasks, particularly when dealing with structured datasets.
3. **Hybrid Models:** Recent studies emphasize the combination of image features and clinical parameters for improved accuracy in PCOS detection. While this study focuses solely on MRI data, future work could integrate hybrid methods for enhanced performance.

**Methods**

**Data Collection and Preprocessing**

The dataset comprised MRI images labeled as either PCOS-positive or PCOS-negative. Each image was preprocessed to enhance contrast, remove noise, and normalize pixel intensity.

* **Data Augmentation:** Techniques such as rotation, flipping, and scaling were applied to expand the dataset and improve model generalization.
* **Feature Extraction:** For KNN and SVM, features were extracted using Histogram of Oriented Gradients (HOG) and Principal Component Analysis (PCA).

**Model Architectures**

1. **Neural Networks (NN):** A basic feedforward NN was implemented with fully connected layers, ReLU activations, and dropout regularization. The input was a flattened vector of image pixels.
2. **Convolutional Neural Networks (CNN):** A CNN with convolutional, pooling, and fully connected layers was designed to leverage spatial hierarchies in MRI images.
3. **K-Nearest Neighbors (KNN):** KNN was applied using Euclidean distance for classification, with the number of neighbors optimized through cross-validation.
4. **Support Vector Machines (SVM):** SVM with a linear kernel was used to classify MRI features into PCOS-positive or PCOS-negative categories.

**Training and Evaluation**

The dataset was split into training (80%) and testing (20%) subsets. Models were trained using stratified sampling to maintain class balance. Performance metrics included:

* **Accuracy:** The ratio of correctly predicted instances to total instances.

**Software and Tools**

* **Programming Language:** Python
* **Libraries:** TensorFlow, Scikit-learn, OpenCV, Matplotlib, Seaborn

**Results and Discussion**

**Performance Comparison**

1. **Neural Networks (NN):** Achieved an accuracy of 85% with moderate recall but required significant tuning for optimal performance.
2. **Convolutional Neural Networks (CNN):** Outperformed all other models with 93% accuracy, highlighting the importance of spatial features in MRI-based PCOS detection.
3. **K-Nearest Neighbors (KNN):** Performed adequately (82% accuracy) but was sensitive to feature scaling and parameter selection.
4. **Support Vector Machines (SVM):** Delivered a balanced performance (88% accuracy) with high precision but computationally expensive during training.

**Key Observations**

* CNNs demonstrated superior performance due to their ability to learn hierarchical features directly from images.
* Feature engineering significantly influenced the performance of KNN and SVM, emphasizing the need for domain-specific expertise.
* Computational costs varied, with NN and CNN requiring GPUs for efficient training, whereas KNN and SVM were less resource-intensive.

**Limitations and Future Work**

* **Integration of Clinical Data:** Combining MRI features with clinical and biochemical markers could improve diagnostic accuracy.
* **Explainability:** Enhancing model interpretability remains a critical challenge for adoption in clinical settings.

**Conclusion**

This study demonstrated the feasibility of using machine learning for automated PCOS detection via MRI images. Among the four models, CNNs emerged as the most effective, leveraging their ability to extract spatial and hierarchical features. While classical models like KNN and SVM showed promise, their reliance on manual feature extraction posed limitations. Future research should focus on integrating multimodal data and improving model interpretability to bridge the gap between research and clinical application.

**References**

[**https://www.researchgate.net/profile/Vaidehi- Thakre/publication/348627784\_PCOcare\_PCOS\_Detection\_and\_Prediction\_using\_Machine\_Learning\_Algorithms/links/60dd359392851ca9449b5cef/PCOcare-PCOS-Detection-and-Prediction-using-Machine-Learning-Algorithms.pdf**](https://www.researchgate.net/profile/Vaidehi-%20Thakre/publication/348627784_PCOcare_PCOS_Detection_and_Prediction_using_Machine_Learning_Algorithms/links/60dd359392851ca9449b5cef/PCOcare-PCOS-Detection-and-Prediction-using-Machine-Learning-Algorithms.pdf)

[**https://ieeexplore.ieee.org/abstract/document/10214584/**](https://ieeexplore.ieee.org/abstract/document/10214584/)

[**https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2024.1298628/full**](https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2024.1298628/full)

[**https://www.mdpi.com/1660-4601/21/3/354**](https://www.mdpi.com/1660-4601/21/3/354)

[**https://www.mdpi.com/1660-4601/21/3/354**](https://www.mdpi.com/1660-4601/21/3/354)