Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 4, 2025
ISSN : 2477-5126, DOI:10.30591/jpit.v10i4.9449

Enhancing PCOS Classification with Weighted Loss-Based Neural Network on Imbalanced Data

Amanda Maslakhah [1], Lukman Hakim [2]*
[1,2]Informatics Engineering Study Program, Yudharta Pasuruan University, Jl.Yudharta, Pasuruan, 67162, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-07-26
Revised 2025-09-30
Accepted 2025-10-21

Abstract *– Polycystic Ovary Syndrome (PCOS) represents a multifaceted endocrine–metabolic condition that poses a significant risk to reproductive health in women of childbearing age. The disorder is influenced by various contributing factors and is commonly associated with clinical features such as disrupted ovulation, hormonal imbalance due to excess androgens, and morphological changes in the ovaries. In automated PCOS classification, a major limitation arises from the disproportionate distribution of data samples, in which instances without PCOS considerably outnumber affected cases. This imbalance tends to bias predictive models toward the dominant class, thereby reducing the detection capability for minority instances and increasing the likelihood of missed PCOS diagnoses. To address this issue, this study proposes the incorporation of a Weighted Loss Function into a Neural Network-based classification framework aimed at improving sensitivity to PCOS cases. The research workflow comprises data preprocessing, neural network architecture construction, integration of class-weighted loss, and systematic experimentation across multiple architectural designs and training configurations. The experimental findings demonstrate that applying a Weighted Loss Function with manually assigned class weights of 1:2, a learning rate of 0.001, five hidden layers, and 50 training epochs delivers optimal classification performance. Under these settings, the model achieves high values across evaluation metrics, including precision, recall, F1-score, and overall accuracy, reaching up to 99%. The results confirm that the proposed approach effectively mitigates majority-class bias and enhances the model's ability to identify PCOS cases. This improvement is further reinforced through careful hyperparameter tuning and comprehensive experimental evaluation..*

Keywords: *Classification; Imbalanced Data; Neural Network; Polycystic Ovary Syndrome (PCOS); Weighted Loss.*

Corresponding Author:

Amanda Maslakhah

Email: amndmslkh@gmail.com

Abstrak *– Sindrom Ovarium Polikistik (PCOS) adalah kelainan endokrin dan metabolik kompleks yang berdampak signifikan terhadap kesuburan wanita usia subur. PCOS bersifat multifaktorial dan ditandai dengan berbagai manifestasi klinis, seperti gangguan ovulasi, kadar hormon androgen yang tinggi, dan adanya ovarium polikistik. Salah satu hambatan teknis yang dihadapi dalam proses identifikasi PCOS adalah adanya ketimpangan frekuensi data. Hal ini terlihat dari jumlah data pasien tanpa PCOS yang melampaui representasi data pasien dengan PCOS, sebuah situasi yang menantang efektivitas sistem klasifikasi dalam membedakan kedua kategori tersebut secara presisi. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga mengakibatkan rendahnya recall kelas minoritas dan banyak kasus PCOS yang tidak terdeteksi. Penelitian ini mengusulkan penerapan Weighted Loss Function pada model Neural Network untuk meningkatkan sensitivitas model dalam mendeteksi PCOS. Proses penelitian meliputi preprocessing, perancangan model, penerapan Weighted Loss Function, dan pengujian dengan berbagai arsitektur dan parameter pelatihan. Hasil pengujian menunjukkan bahwa kombinasi Weighted Loss dengan bobot manual 1:2, learning rate 0,001, lima layer, dan 50 epoch menghasilkan kinerja terbaik, dengan presisi, recall, F1-Score, dan akurasi mencapai 99%. Temuan ini menunjukkan bahwa metode yang diusulkan efektif dalam meningkatkan akurasi dan sensitivitas model terhadap kelas minoritas. Kinerja tinggi ini didukung oleh optimasi hiperparameter yang cermat dan penggunaan pembobotan kelas, yang secara efektif mengurangi bias terhadap kelas mayoritas dan meningkatkan kemampuan model untuk mendeteksi kasus minoritas (PCOS). Hasil ini diperoleh dari desain dan evaluasi eksperimen sistematis di berbagai konfigurasi.*

Kata Kunci*: Data Tidak Seimbang, Klasifikasi, Neural Network, Polycystic Ovary Syndrome (PCOS), Weighted Loss.*

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) functions as a complex metabolic-endocrine disorder with substantial repercussions for fertility and reproductive health in the female population of reproductive age. This condition is often marked by intricate hormonal imbalances that necessitate advanced diagnostic approaches[1], [2], [3]. PCOS is widely acknowledged as a primary endocrinological disorder affecting women globally, with clinical data suggesting a frequency between 5.6% and 21.3% within the reproductive-age demographic[4]. Furthermore, international statistics provided by the World Health Organization (WHO) corroborate this

widespread occurrence, projecting a global prevalence range between 6% and 26%. In certain regions, such as India, the prevalence is reported to be even higher, varying from 9.13% to 36%. Beyond reproductive complications, PCOS is a major contributor to psychological health issues, with approximately 28–39% of affected women experiencing depressive symptoms and 11–25% suffering from anxiety disorders [5].

According to the Rotterdam consensus, a definitive diagnosis of Polycystic Ovary Syndrome is confirmed when a patient manifests a minimum of two out of three specific clinical indicators: ovulatory irregularity, signs of hyperandrogenism (either biochemical or clinical), and polycystic ovaries as identified through morphology[6], [7]. PCOS is widely acknowledged as a multifactorial condition with diverse clinical manifestations that extend beyond ovarian morphology, encompassing hormonal imbalances, irregular or absent ovulation, and other reproductive abnormalities [8]. Due to this wide spectrum of symptoms, no single diagnostic criterion can fully capture all clinical presentations of PCOS [9]. Moreover, despite its substantial public health impact, limited understanding of the underlying pathophysiological mechanisms continues to hinder the development of effective prevention strategies and targeted therapeutic interventions [10].

PCOS is also characterized as a syndrome driven by excessive androgen production [11]. In women, androgens are primarily synthesized by the ovaries and adrenal cortex, with additional contributions from peripheral tissues [12]. Reproductive manifestations commonly associated with PCOS include polycystic ovarian morphology, menstrual irregularities such as amenorrhea and oligomenorrhea, and reduced fertility [13]. The syndrome disrupts normal follicular development, causing ovarian follicles to arrest at early growth stages and preventing their maturation and ovulation [14].

In addition to reproductive dysfunction, PCOS is frequently accompanied by a wide range of comorbid conditions, including hirsutism, acne, cardiovascular complications, increased susceptibility to breast and endometrial cancers, infertility, pregnancy-related complications, and metabolic disorders such as obesity, insulin resistance, hyperinsulinemia, and type 2 diabetes mellitus [15], [16], [17]. The clinical heterogeneity of PCOS is further influenced by interactions between genetic predisposition and environmental factors [18], including lifestyle-related factors such as diet and obesity, which play a critical role in shaping disease manifestation and severity [19], [20].

Given the broad clinical impact of PCOS, early identification and timely intervention are essential to reduce long-term complications. Ultrasonography is among the most commonly used diagnostic tools, enabling the assessment of ovarian follicle number and size through imaging techniques. Although effective, this method requires high-quality imaging, considerable expertise, and can be time-consuming, which may limit its accessibility and reliability in large-scale screening scenarios [21].

With rapid advancements in computational technologies, numerous studies have focused on developing predictive models for PCOS detection. Research reported in [5] introduced an automated PCOS and mental health prediction framework utilizing Fuzzy TOPSIS and Support Vector Machine (SVM) techniques. Alternative methodologies for PCOS identification have been explored, such as the integration of Gabor wavelet features with Competitive Neural Networks (CNN) for analyzing ultrasound scans[6]. Beyond standard neural networks, the study in [22] investigated the application of complex deep learning models, including LSTM and hybrid CNN–LSTM configurations. Additionally, the research presented in [23] showcased the effectiveness of a diagnostic system that utilizes a combination of Random Forest and Artificial Neural Networks (ANN).

Despite these advancements, one of the most critical challenges in PCOS prediction remains the issue of data imbalance. In many datasets, the number of non-PCOS samples significantly exceeds that of PCOS cases, causing predictive models to become biased toward the majority class. Consequently, models may achieve high overall accuracy while exhibiting poor recall for PCOS cases, resulting in a high rate of false negative predictions. Addressing this imbalance is therefore essential to improve the detection of individuals genuinely affected by PCOS.

Several studies have investigated strategies to mitigate class imbalance and enhance classification performance. In [24], Adaptive Synthetic Sampling (ADASYN) was applied prior to Random Forest classification, leading to an improved recall rate of up to 92%. However, this approach remains susceptible to overfitting due to extensive synthetic data generation. The study in [25] combined Synthetic Minority Over-sampling Technique (SMOTE) with focal loss in a Convolutional Neural Network framework, achieving disease detection accuracy exceeding 93% in cassava plant datasets. In [26], the Radial-Based Undersampling (RBU) algorithm was introduced and demonstrated strong performance when paired with a Naive Bayes classifier, although its effectiveness declines in datasets containing a large number of safe minority samples or outliers. Meanwhile, [27] integrated Support Vector Machine (SVM) with a cost-sensitive learning strategy, achieving an F1-score of 89.7%. Although stable, this approach requires careful parameter tuning to achieve optimal performance.

To overcome the limitations of existing imbalance-handling techniques, this study proposes the integration of a Weighted Loss Function within a Neural Network-based classification framework. To improve the detection rate of the minority class (PCOS), a cost-sensitive learning approach is implemented by imposing

heavier penalties for misclassifications. This strategy significantly boosts the model's sensitivity, thereby minimizing the incidence of false-negative results. Incorporating class weights into the Cross Entropy Loss function allows the algorithm to recalibrate its learning trajectory based on the actual class frequency, fostering a more equitable and robust predictive framework.

Despite the growing body of research on PCOS classification, many existing models continue to suffer from majority-class bias and insufficient sensitivity toward minority-class instances. Oversampling and undersampling techniques, including SMOTE and ADASYN, have been widely adopted, but they often introduce overfitting and compromise model generalization. This study contributes to the field of medical informatics by: (i) developing a Neural Network model that employs a Weighted Loss Function to address the challenges of skewed PCOS data distribution; (ii) performing an extensive ablation study and hyperparameter optimization—encompassing training epochs, hidden layers, and weight ratios—to identify the optimal model configuration; and (iii) demonstrating that the proposed approach yields high-precision results (up to 99% accuracy), specifically improving the model's ability to detect positive PCOS cases, which is a fundamental requirement for accurate medical screening.

II. METHOD

This study aims to classify Polycystic Ovary Syndrome (PCOS) using a Neural Network method integrated with a Weighted Loss Function to address class imbalance. Several stages were carried out, including data collection and preprocessing, model architecture design, implementation of the Weighted Loss function, and model performance evaluation Figure 1. The implementation and training of the model were conducted using the Google Colab platform.

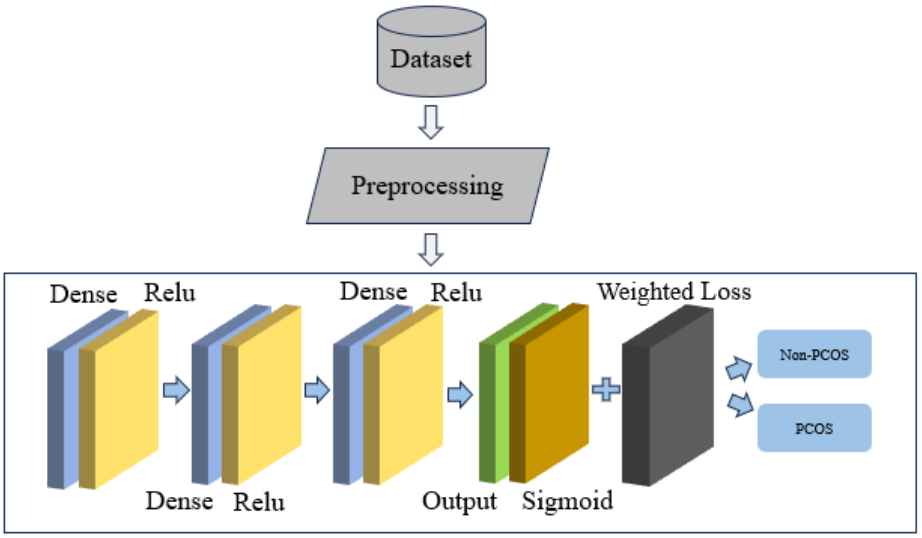

Figure 1. Research Stages

A. Dataset

The empirical data for this research was retrieved from the Kaggle repository, specifically the 'PCOS Diagnosis Dataset' [28], which was accessed on March 08, 2025. Although the explicit licensing terms were not specified on the platform, the dataset provides a robust collection of 1,000 clinical records. These entries are structured into five distinct input features serving as predictors, alongside a single target label for classification purposes. The target label is PCOS Diagnosis. The PCOS Diagnosis label is a binary variable indicating the diagnostic status of each patient, where a value of "0" represents a non-PCOS patient and a value of "1" indicates a patient diagnosed with PCOS. These five features were chosen because: (1) they are clinically relevant according to Rotterdam criteria and supporting studies[6], [7]; (2) they represent the main variables from the original dataset without derived features; (3) initial analysis showed moderate correlations with the target label; (4) the small number of features made feature selection unnecessary; and (5) retaining all features preserves the medical context and prevents loss of important information.A detailed explanation of the dataset attributes is presented in TABLE 1

TABLE 1
DATASET ATTRIBUTE EXPLANATION

Attribute	Description	Range
Age	Age of female patient	18 - 45
BMI	Body Mass Index	18.1 – 35.0
Menstrual Irregularity	Menstrual disorders	0 or 1
Testosterone Level(ng/dL)	Blood testosterone levels	20.0 – 99.8
Antral Follicle Count	Number of antral follicles in the ovaries	5 - 29
PCOS Diagnosis	No PCOS or PCOS	0 or 1

No records were excluded since the dataset contained no missing values, and all entries were within reasonable ranges (e.g., Age 18–45, BMI 18.1–35.0). Outlier checks using *z-score* (>3 or <−3) and boxplots revealed no extreme outliers, so all data were retained. The dataset was partitioned into training (80%) and testing (20%) subsets utilizing a stratified splitting approach to ensure that the inherent class proportions were maintained across both sets. For experimental reproducibility, a fixed random seed of 42 was applied during the split. The data exhibits a notable class imbalance, consisting of 199 PCOS instances (19.9%) and 801 non-PCOS records (80.1%). This disproportionate distribution necessitated the implementation of a Weighted Loss Function to prevent the model from developing a bias toward the majority class. The class allocation is graphically represented in Figure 2.

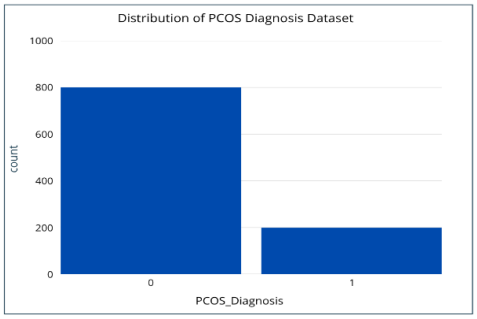

Figure 2. Distribution of Number of Classes

B. Preprocessing

Data preprocessing is a crucial first step to ensure the dataset is fit for use in model training. Because the dataset contains no missing values, no missing value handling is required. Data normalization is implemented via the StandardScaler framework to standardize the feature space, thereby neutralizing potential biases arising from disparate value intervals. The methodology strictly applies the fit operation only to the training data to calculate the necessary scaling constants (mean and standard deviation). By doing so, the integrity of the evaluation is maintained when these parameters are used to transform the training, validation, and test sets, ensuring that no information from the testing phase influences the model's preprocessing. This ensures that information from the validation and test data does not influence the normalization process, thus preventing potential data leakage.

Exploratory Data Analysis (EDA) revealed that patient ages were concentrated between 22–32 years, BMI values most frequently ranged from 22–30 kg/m², and PCOS-positive patients tended to have higher BMI, testosterone levels, and antral follicle counts compared to non-PCOS patients. Pearson correlations with the PCOS label were: BMI (0.42), Antral Follicle Count (0.39), Testosterone Level (0.36), Menstrual Irregularity (0.31), and Age (0.12). The independence of the predictor variables was validated through a VIF analysis, where all values remained within the acceptable limit of less than 5. Such results demonstrate that multicollinearity is negligible within the dataset, supporting the reliability of the subsequent neural network training phase. The normalization process is performed using the following equation

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

In this equation, x represents the raw feature value, while μ and σ correspond to the mean and standard deviation of the feature, respectively. This Z-score transformation ensures that each input variable is recalibrated to follow a distribution with a zero mean and unit variance. By implementing this approach, the classification

algorithm can process inputs on a uniform scale, which significantly enhances numerical stability and accelerates the convergence of the model during the training phase.The feature values before and after normalization are presented in TABLE 2 and TABLE 3.

TABLE 2
DATA BEFORE NORMALIZATION

Atribut	Data 1	Data 2	Data 3	Data 4	Data 5
Age	24	37	32	28	25
BMI	34.7	26.4	23.6	28.8	22.1
Menstrual_Irregularity	1	0	0	0	1
Testosterone_Level(ng/dL)	25.2	57.2	92.7	63.1	59.8
Antral_Follicle_Count	20	25	28	26	8

TABLE 3
DATA AFTER NORMALIZATION

Atribut	Data 1	Data 2	Data 3	Data 4	Data 5
Age	-0.918	0.618	0.027	-0.446	-0.800
BMI	1.685	0.003	-0.549	0.481	-0.870
Menstrual_Irregularity	0.942	-1.062	-1.062	-1.062	0.942
Testosterone_Level(ng/dL)	-1.510	-0.132	1.406	0.127	-0.016
Antral_Follicle_Count	0.354	1.066	1.494	1.207	-1.340

The dataset was also split into input features and target labels to facilitate the training process. Considering the class imbalance present in the Polycystic Ovary Syndrome (PCOS) dataset, a Weighted Loss Function approach was employed. This technique was integrated into the Neural Network architecture with the aim of reducing the model's bias toward the majority class and enhancing its ability to detect patterns in the minority class. As a result, the predictions are expected to be more accurate and balanced.

C. Network Architecture

In this study, a Neural Network model was utilized to address the issue of data imbalance commonly encountered in medical diagnosis tasks. Neural Networks were chosen for their flexibility and capability to learn complex non-linear patterns, making them appropriate for binary classification on medium-sized medical datasets. The model was developed in Google Colab using TensorFlow and the Keras API, which support GPU acceleration, provide a flexible interface, and facilitate hyperparameter optimization. To identify the optimal architecture and training parameters for handling class imbalance and improving classification performance, several configurations were explored by adjusting the number of layers, learning rate, and training epochs.

After testing several Artificial Neural Network configurations with the different combinations of layers, neurons, learning rate, and epochs, the best performing model featured an input layer of five neurons, followed by five hidden layers consisting 128, 64, 32, 16, and 8 neurons. The neural network architecture utilizes the Rectified Linear Unit (ReLU) activation function across all hidden layers to mitigate gradient-related issues, while a sigmoid activation is applied to the single-neuron output layer for binary classification. The model was trained using the Adam optimizer with a predefined learning rate of 0.001 and a batch size of 16. To prevent overfitting and ensure clinical relevance, the training process incorporated Dropout and L2 regularization, along with an EarlyStopping mechanism. This mechanism was configured to monitor recall with a patience of 5 epochs, ensuring that the final model parameters were restored to the state that achieved the highest sensitivity toward PCOS cases. The final model had 11,777 trainable parameters. In general, the output of a Neural Network can be computed using equation.

$$y = f\left(\sum_{i=1}^{n} w_i \cdot x_i + b\right) \tag{2}$$

Where y denotes the output, x_i represents the input, w_i is the weight, b is the bias, and f is the activation function used. ReLU was applied to the hidden layers, whilw the output used a sigmoid activation because it is appropriate for binary classification. One of the approaches used is the Weighted Loss Function, which assigns different weights to each class during model training. These weights can be determined automatically based on the data distribution or manually according to specific considerations. The weights are calculated using equation.

$$L = -w_1 \cdot y \cdot \log(p) - w_0 \cdot (1 - y) \cdot \log(1 - p) \tag{3}$$

Where L represents the loss value, y is the ground truth label, p is the predicted probability, and w_1 and w_0 are the weights assigned to each class. This weighting method prevents the model does not focus solely on the majority class, but also improves its sensitivity to the minority class, which is often overlooked. In addition to automatic weighting, the weights can also be manually assigned. For example, to increase sensitivity toward the minority class, a weight of $w_1 = 2$ can be assigned to the positive class and $w_0 = 1$ to the negative class. The values of w_1 and w_0 can be adjusted based on the level of class imbalance or specific requirements of the classification model. Applying the weight values gives equation.

$$L = -2 \cdot y \cdot \log(p) + 1 \cdot (1 - y) \cdot \log(1 - p) \quad (4)$$

The Neural Network model implemented in this study was designed with several architectural variations by combining different numbers of neurons in the hidden layers to identify the most optimal configuration. The training process was carried out using the Adam optimizer with various learning rate values to achieve the best performance.

D. Model Training and Validation

The dataset was partitioned into two distinct subsets, with 80% of the observations designated for model training and the remaining 20% reserved as an independent testing set. To ensure the development of a robust classifier, the training phase involved the evaluation of various hyperparameter configurations. This systematic tuning was essential for identifying the optimal model architecture and achieving peak predictive performance. One of the main parameters tuned was the learning rate, which was manually adjusted to 0.00001, 0.0001, 0.0005, 0.001, 0.005, and 0.01, alongside the default learning rate of the Adam optimizer as a baseline. The number of training epochs was also varied at 10, 25, 50, 75, and 100 to observe convergence behavior. The classification effectiveness of the proposed model was quantitatively measured through accuracy, precision, recall, and the F1-score. Given the inherent class imbalance within the dataset, recall was designated as the primary metric of interest to ensure maximum sensitivity in PCOS detection. Furthermore, a stratified k-fold cross-validation scheme was implemented to obtain a more reliable and generalized estimation of the model's performance, mitigating the risk of bias associated with a single train-test split. The scaler was fitted only on the training subset in each fold and then applied to the validation subset to avoid data leakage. The final performance was reported as mean ± standard deviation across folds.

E. Evaluation

In this stage, the model outputs were examined to evaluate howthe use of the Weighted Loss Function influenced the classification results. The assessment was carried out with the help of a Confusion Matrix to calculate the following metrics: accuracy, precision, recall, and F1-score. Accuracyreflects the proportion of correct predictions across both the positive and negative classes. Precision represents the share of positive prediction that are truly correct. Recall assesses how well the model identifies positive samples out of all actual positive instances. The F1-score is defined as the harmonic mean of precision and recall, offering a balanced indicator that accounts for both false positives and false negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Here, TP (True Positive) refers to the number of correctly identified positive samples, TN (True Negative) represents correcly identified negative samples, FP (False Positive) corresponds to incorrectly predicted positive cases, and FN (False Negative) refers to incorrectly predicted negative cases. These four metrics were used to evaluate the performance of the Neural Network across the tested configurations, which included varying the numbers of epochs, layers, and learning rates. The evaluation aimed to idetermine the most effective combination of settings that delivered the best model performance.

III. RESULTS AND DISCUSSION

This study conducted a series of model evaluations using the Neural Network algorithm with several configurations: (1) Neural Network without Weighted Loss Function, (2) Neural Network with Weighted Loss Function, and (3) variations in learning rate, number of layers, and epochs. To ensure robustness, k-fold cross-validation was applied. Results showed that the baseline Neural Network without weighting tended to be biased toward the majority class, achieving 0.95 accuracy but only 0.85 recall for the minority class. When employing the Weighted Loss Function, the model's performance improved substantially, achieving 0.98 accuracy, 0.97 precision, and 0.98 recall.

In addition, comparative experiments were carried out with Logistic Regression, SVM with class-weight, Random Forest, and XGBoost, as well as supplementary tests using Neural Network with focal loss and ablation settings, including cases without normalization or weighting, focal loss application, and SMOTE applied only during training. These comparisons further confirmed the effectiveness of the Weighted Loss Function in addressing class imbalance. A detailed summary of the experimental results is presented in TABLE 4.

TABLE 4
PERFORMANCE COMPARISON OF NEURAL NETWORK-BASED MODELS AND OTHER CLASSIFIERS

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.88	0.66	0.92	0.76
SVM (class-weight)	0.95	0.82	0.98	0.90
Random Forest	0.95	0.81	0.99	0.89
XGBoost	0.99	1.00	0.99	0.99
Neural Network (Baseline)	0.95	0.92	0.85	0.88
NN + Weighted Loss Function	0.98	0.97	0.98	0.98
NN + Focal Loss	0.94	0.75	0.75	0.75
NN (no normalized)	0.69	0.29	0.73	0.42
NN + SMOTE (train only)	0.97	0.95	0.92	0.93

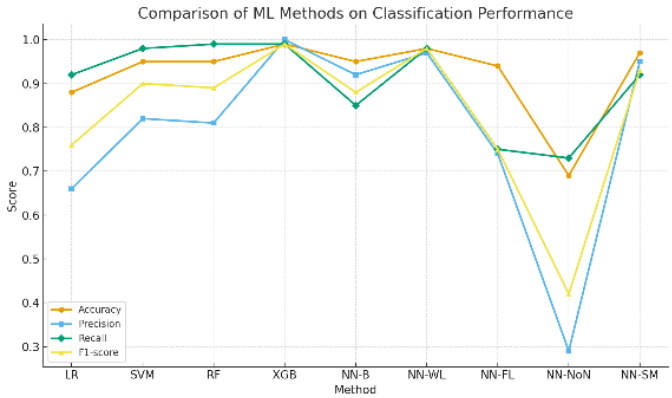

Figure 3. Comparison Of ML Method Of Classification Performance

Additional experiments were conducted by applying manual weight variations in the Weighted Loss Function, including ratios of 1:2, 1:3, 1:4, 1:5, 1:2.5, and 3:7. The results showed that the 1:2 and 3:7 weight combinations produced the best performance, with precision of 0.96, recall of 0.98, F1-score of 0.97, and accuracy of 0.98. In contrast, the 1:5 weight combination led to a decrease in precision to 0.90, even though recall remained high. The 1:2 and 3:7 weights demonstrated optimal performance because they maintained a good balance between penalizing classification errors and preserving the model's generalization ability.

Overly extreme weights, such as 1:5, caused the model to overemphasize the minority class, resulting in a drop in precision due to incorrect predictions on the majority class. This highlights the importance of selecting proportional weights to ensure that the model remains both accurate and fair toward both classes. A complete summary of the manual weight variation results is presented in TABLE 5.

TABLE 5
RESULTS OF WEIGHTED LOSS FUNCTION COMPARISON

Bobot Manual	Precision	Recall	F1-Score	Accuracy
1:2	0.96	0.98	0.97	0.98
1:3	0.93	0.97	0.95	0.96
1:4	0.93	0.98	0.95	0.96
1:5	0.90	0.97	0.93	0.95
1:2.5	0.93	0.97	0.95	0.96
3:7	0.96	0.98	0.97	0.98

In addition, experiments were carried out to examine how different learning rate values influenced the model performance. The findings indicated that a learning rate of 0.001 yielded the highest performance, achieving a precision of 0.93, a recall of 0.98, an F1-score of 0.96, and an accuracy of 0.97.

Conversely, using a very small learning rate such as 0.00001 led to a significant drop in performance. The learning rate of 0.001 provided optimal results because it was large enough to accelerate the learning process while still being small enough to maintain training stability.

A learning rate that is too small, such as 0.00001, results in a very slow training process and prevents the model from effectively identifying patterns in the data, leading to poor performance. On the other hand, a learning rate that is too large can cause the model to overshoot the optimal point in the loss function, resulting in instability and failure to converge properly. A complete summary of the learning rate experiments is presented in TABLE 6.

TABLE 6
RESULTS OF LEARNING RATE COMPARISON

LR	Precision	Recall	F1-Score	Accuracy
0.00001	0.40	0.50	0.44	0.80
0.0001	0.89	0.96	0.92	0.94
0.001	0.93	0.98	0.96	0.97
0.01	0.94	0.94	0.94	0.96
0.005	0.92	0.97	0.94	0.96

Further investigations into the network's structural design revealed that performance gain was tied to the number of layers utilized. The experiments showed that an architecture consisting of five layers achieved the highest diagnostic accuracy of 0.99. At this depth, the model demonstrated a robust balance between precision (0.98) and recall (0.99), with a corresponding F1-score of 0.98, highlighting its capability in accurately classifying PCOS instances.

Adding more layers up to five improved performance because it allowed the model to have greater capacity to capture complex patterns in the data. A deeper architecture enables the extraction of more abstract and hierarchical feature representations. However, it is important to note that increasing the number of layers beyond a certain point, especially without proper regularization, can lead to overfitting. Nevertheless, in this experiment, the five-layer configuration still delivered excellent performance. The detailed results of this experiment are presented in TABLE 7.

TABLE 7
RESULTS OF LAYER COMPARISON

Layer	Precision	Recall	F1-Score	Accuracy
2	0.93	0.98	0.96	0.97
3	0.97	0.97	0.97	0.98
4	0.97	0.99	0.98	0.98
5	0.98	0.99	0.98	0.99

In the final phase of experimental evaluation, the research investigated the impact of training duration—specifically the number of epochs—on the model's predictive capacity. The empirical findings demonstrated that extending the training up to 50 epochs significantly augmented model performance. At this iteration threshold, the system reached its peak efficacy, yielding a precision of 0.98 and a recall of 0.99, while the F1-score and overall accuracy stabilized at 0.98 and 0.99, respectively.

Training for 50 epochs provided an optimal balance between underfitting and overfitting. At this point, the model had learned sufficiently from the data without overfitting. Although extending the number of epochs to 100 still resulted in high performance metrics, a slight increase in precision was accompanied by a decrease in recall, which may indicate the onset of mild overfitting where the model starts to overly adapt to the training data. The complete results of the epoch variation experiments are presented in TABLE 8.

TABLE 8
RESULTS OF EPOCH COMPARISON

Epoch	Precision	Recall	F1-Score	Accuracy
10	0.97	0.96	0.96	0.97
25	0.96	0.98	0.97	0.98
50	0.98	0.99	0.98	0.99
75	0.98	0.96	0.97	0.98
100	0.99	0.97	0.98	0.99

Based on the entire series of experiments conducted, it can be concluded that the optimal configuration was achieved using an Artificial Neural Network model with the following settings: implementing a Weighted Loss Function, with automatic weighting, a learning rate of 0.001, five hidden layers, and 50 training epochs.

This configuration produced the highest performance, with a precision of 0.99, a recall of 0.97, an F1 score of 0.98, and an accuracy of 0.99. These results indicate that this combination effectively increases the model's sensitivity to minority classes while maintaining a high overall accuracy level. A detailed summary of the best results across all experiments is presented in TABLE 9 and TABLE 10 .

TABLE 9
MODEL EVALUATION WITH CLASSIFICATION METRICS

Best Performance	Accuracy	Precision	Recall	F1-Score
Train-Test	0.99	0.98	0.99	0.98
K-Fold	0.97±0.01	0.92±0.04	0.95±0.05	0.93±0.02

TABLE 10
MODEL EVALUATION WITH ADDITIONAL METRICS

Best Performance	ROC-AUC	PR-AUC	Brier Score	Specificity	Balanced Acc
Train-Test	0.99	0.99	0.01	0.98	0.99
K-Fold	0.99±0.001	0.98±0.007	0.01±0.007	0.98±0.01	0.96±0.02

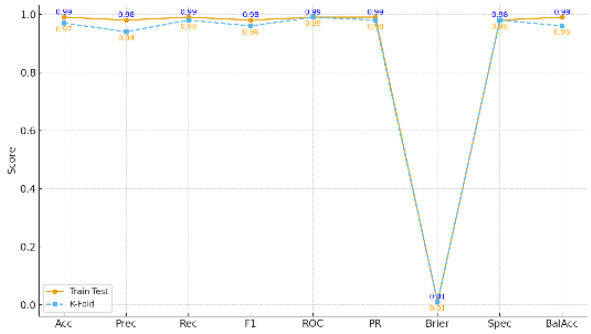

Figure 4. comparison of train-test and K-Fold results

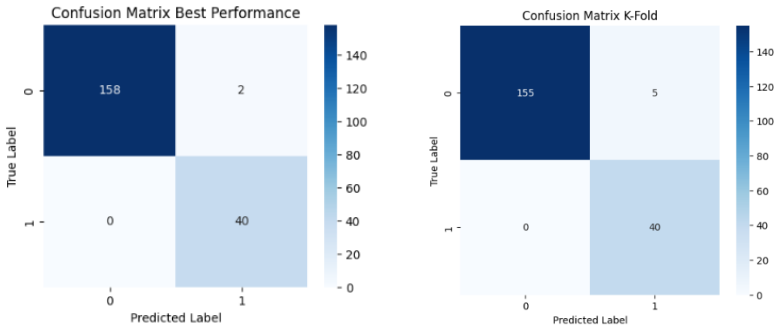

Figure 5. Comparison of Confusion Matrix Train-Test and K-fold

Testing on the final configuration demonstrated the best performance compared to all previous experiments. This is evident from the accuracy values of 0.99, precision of 0.98, recall of 0.99, and F1-score of 0.98 in the train-test scheme, and stable performance in the k-fold evaluation with an accuracy of 0.97 and an F1-score of 0.96. These results were achieved because the model used a Neural Network with five hidden layers, 50 epochs, a learning rate of 0.001, and the application of an automatic Weighted Loss Function that effectively balanced the majority and minority classes. This combination made the model not only highly accurate overall but also sensitive in recognizing fewer PCOS cases without sacrificing predictions in the healthy class.

Furthermore, the results in Table 10 reinforce the superiority of this configuration, as near-perfect ROC-AUC (0.99) and PR-AUC (0.98) values confirm the model's capability to discriminate between PCOS and non-PCOS classes despite the imbalanced dataset. A very low Brier Score (0.01) indicates that the predicted probabilities are well-calibrated, which is critical in medical decision-making where probability confidence plays an essential role. High specificity (0.98) and balanced accuracy (0.96) also highlight that the model maintains equilibrium in identifying both positive and negative cases.

Overall, this configuration can be considered the most optimal setting because it combines a sufficiently deep network architecture, stable training parameters, and an appropriate loss weighting strategy to produce high and consistent performance. In addition, the integration of both standard classification metrics (accuracy, precision, recall, F1-score) and advanced metrics (ROC-AUC, PR-AUC, Brier Score, specificity, and balanced accuracy) demonstrates that the model is not only robust in terms of overall accuracy but also reliable in handling imbalanced medical data. This ensures that the model achieves a good trade-off between sensitivity and specificity, reducing the likelihood of undetected PCOS cases while maintaining accuracy in identifying healthy patients.

IV. CONCLUSION

The problem of data imbalance in the classification of Polycystic Ovary Syndrome (PCOS) can be effectively addressed by implementing a Weighted Loss Function in an Neural Network model. The results show that a NN with Weighted Loss is capable of providing good classification performance and a balance between sensitivity and specificity, making it more suitable for clinical contexts. Comparison with several other algorithms such as Logistic Regression, Random Forest, SVM, and XGBoost confirms that although XGBoost appears to have very high performance, this likely reflects overfitting to a relatively limited dataset, rather than improved generalization ability. These findings strengthen the relevance of a NN-based approach with Weighted Loss for imbalanced medical data scenarios. However, this study has limitations such as the limited number of features and the lack of external validation on an independent dataset, so the generalizability of the findings to a broader population is uncertain. For future research, external validation using data from other sources or hospitals, as well as exploring decision curves and model calibration, is recommended to ensure the predictive reliability and clinical value of the developed model.

REFERENCE

[1] S. D. P. Núñez Arroba, L. K. Sailema López, and G. A. Zúñiga Cárdenas, "Analysis of polycystic ovary syndrome and infertility using PRISMA 2020 literature review," *Data and Metadata*, vol. 3, Dec. 2024, doi: 10.56294/dm2024.239.

[2] H. Guan *et al.*, "Integrated bioinformatics and network pharmacology to explore the therapeutic target and molecular mechanisms of Bailing capsule on polycystic ovary syndrome," *BMC Complement Med Ther*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12906-023-04280-6.

[3] M. S. Khan Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee, and R. Hasan, "Improved Sampling and Feature Selection to Support Extreme Gradient Boosting For PCOS Diagnosis," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA: IEEE, Jan. 2021, pp. 1046–1050. doi: 10.1109/ccwc51732.2021.9375994.

[4] L. Doretto, F. C. Mari, and A. C. Chaves, "Polycystic Ovary Syndrome and Psychotic Disorder," *Front. Psychiatry*, vol. 11, p. 543, Jun. 2020, doi: 10.3389/fpsyt.2020.00543.

[5] A. Kodipalli and S. Devi, "Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM," *Front. Public Health*, vol. 9, p. 789569, Nov. 2021, doi: 10.3389/fpubh.2021.789569.

[6] R. M. Dewi, Adiwijaya, U. N. Wisesty, and Jondri, "Classification of polycystic ovary based on ultrasound images using competitive neural network," *J. Phys.: Conf. Ser.*, vol. 971, p. 012005, Mar. 2018, doi: 10.1088/1742-6596/971/1/012005.

[7] L. Jiang *et al.*, "Diagnostic value of anti-Müllerian hormone combined with androgen-levels in Chinese patients with polycystic ovary syndrome," *Gynecological Endocrinology*, vol. 39, no. 1, Dec. 2023, doi: 10.1080/09513590.2023.2206927.

[8] H.-J. Guan, L.-Q. Pan, H. Song, H.-Y. Tang, and L.-S. Tang, "Predictors of pregnancy after intrauterine insemination in women with polycystic ovary syndrome," *J Int Med Res*, vol. 49, no. 5, May 2021, doi: 10.1177/03000605211018600.

[9] R. Galagan, S. Andreiev, N. Stelmakh, Y. Rafalska, and A. Momot, "AUTOMATION OF POLYCYSTIC OVARY SYNDROME DIAGNOSTICS THROUGH MACHINE LEARNING ALGORITHMS IN ULTRASOUND IMAGING," *Appl. Comput. Sci.*, vol. 20, no. 2, pp. 194–204, Jun. 2024, doi: 10.35784/acs-2024-24.

[10] A. Chen-Patterson *et al.*, "Distinct Reproductive Phenotypes Segregate With Differences in Body Weight in Adolescent Polycystic Ovary Syndrome," *Journal of the Endocrine Society*, vol. 8, no. 2, Jan. 2024, doi: 10.1210/jendso/bvad169.

[11] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, and G. K. Hegde, "A Distinctive Explainable Machine Learning Framework for Detection of Polycystic Ovary Syndrome," *ASI*, vol. 6, no. 2, p. 32, Feb. 2023, doi: 10.3390/asi6020032.

[12] J. Zhang, M. Pan, Q. Zhang, F. Wang, and F. Qu, "Association of Tumor Markers with Total Testosterone in the Serum of Women with Polycystic Ovary Syndrome," *Clin. Exp. Obstet. Gynecol.*, vol. 49, no. 7, p. 162, Jul. 2022, doi: 10.31083/j.ceog4907162.

[13] K. Soucie, T. Samardzic, K. Schramer, C. Ly, and R. Katzman, "The Diagnostic Experiences of Women With Polycystic Ovary Syndrome (PCOS) in Ontario, Canada," *Qual Health Res*, vol. 31, no. 3, pp. 523–534, Feb. 2021, doi: 10.1177/1049732320971235.

[14] S. Bharati, P. Podder, and M. R. H. Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms".

[15] G. Alessandri *et al.*, "Disclosing α-lactalbumin impact on the intestinal and vaginal microbiota of women suffering from polycystic ovary syndrome," *Microbial Biotechnology*, vol. 17, no. 10, p. e14540, Oct. 2024, doi: 10.1111/1751-7915.14540.

[16] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," *IEEE Access*, vol. 10, pp. 97610–97624, 2022, doi: 10.1109/access.2022.3205587.

[17] Poorani. B and R. Khilar, "Classification of PCOS Using Machine Learning Algorithms Based on Ultrasound Images of Ovaries," in *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Chennai, India: IEEE, Apr. 2023, pp. 1–7. doi: 10.1109/iconstem56934.2023.10142359.

[18] 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, p. 1506, Apr. 2023, doi: 10.3390/diagnostics13081506.

[19] F. J. Barrera *et al.*, "Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review," *Front. Endocrinol.*, vol. 14, p. 1106625, Sep. 2023, doi: 10.3389/fendo.2023.1106625.

[20] S. Dhinakaran, C. Thangavel, S. S, and H. V S, "PCOS Perception analysis prediction using Machine learning algorithms," in *2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, MANGALORE, India: IEEE, Dec. 2022, pp. 260–265. doi: 10.1109/icraie56454.2022.10054279.

[21] S. C. R. Nandipati, C. XinYing, and K. K. Wah, "Polycystic Ovarian Syndrome (PCOS) Classification and Feature Selection by Machine Learning Techniques," vol. 9, 2020.

[22] R. Ahmad, L. A. Maghrabi, I. A. Khaja, L. A. Maghrabi, and M. Ahmad, "SMOTE-Based Automated PCOS Prediction Using Lightweight Deep Learning Models," *Diagnostics*, vol. 14, no. 19, p. 2225, Oct. 2024, doi: 10.3390/diagnostics14192225.

[23] N.-N. Xie, F.-F. Wang, J. Zhou, C. Liu, and F. Qu, "Establishment and Analysis of a Combined Diagnostic Model of Polycystic Ovary Syndrome with Random Forest and Artificial Neural Network," *BioMed Research International*, vol. 2020, no. 1, p. 2613091, Jan. 2020, doi: 10.1155/2020/2613091.

[24] R. Ferdiansyah Putra, K. Fithriasari, and S. Puteri Rahayu, "Random Forest With Adaptive Synthetic Sampling- Nominal Continuous and Cost Sensitive Learning on Imbalanced Binary Classification," in *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, Balikpapan, Indonesia: IEEE, Jul. 2024, pp. 708–713. doi: 10.1109/ICSINTESA62455.2024.10748121.

[25] G. Sambasivam and G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks," *Egyptian Informatics Journal*, vol. 22, no. 1, pp. 27–34, Mar. 2021, doi: 10.1016/j.eij.2020.02.007.

[26] M. Koziarski, "Radial-Based Undersampling for imbalanced data classification," *Pattern Recognition*, vol. 102, p. 107262, Jun. 2020, doi: 10.1016/j.patcog.2020.107262.

[27] A. Iranmehr, H. Masnadi-Shirazi, and N. Vasconcelos, "Cost-sensitive support vector machines," *Neurocomputing*, vol. 343, pp. 50–64, May 2019, doi: 10.1016/j.neucom.2018.11.099.

[28] "PCOS Diagnosis Dataset." Accessed: Jul. 26, 2025. [Online]. Available: https://www.kaggle.com/datasets/samikshadalvi/pcos-diagnosis-dataset