Scientific Journal of Informatics
Vol. 12, No. 1, Feb 2025

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Implementation of ANN Optimization with SMOTE and Backward Elimination for PCOS Prediction

Miftakhul Ilmiyah[1*], Mula Agung Barata[2], Pelangi Eka Yuwita[3]
[1, 2, 3]Department of Informatics Engineering, Faculty of Sains and Technology, Universitas Nahdlatul Ulama Sunan Giri, Indonesia

Abstract.
Purpose: Polycystic Ovary Syndrome (PCOS) is a health disease that targets the ovaries and is frequently disregarded by women, making it potentially fatal owing to delayed diagnosis and treatment. With the advent of current technology, machine learning and medical care may become associated with disease prediction. The purpose of the study is to predict PCOS using an Artificial Neural Network (ANN) Deep Learning algorithm combined with Synthetic Minority Oversampling Technique (SMOTE) for data balancing and backward elimination for feature selection, aiming to provide a more accurate diagnosis of PCOS with high accuracy from thoose combination.
Methods: ANN algorithm structure with three hidden layers, each with a ReLU activation function of 128, 64, and 32 neurons, a Dropout layer, an output layer with a sigmoid activation function, and an Adam learning rate.
Result: Using the SMOTE approach for data balance and backward elimination feature selection, the research attributes are reduced to 18. And ANN algorithm predicts PCOS disease achieve an accuracy of 92%.
Novelty: This study uses an ANN algorithm model combined with the SMOTE data balancing technique and a feature selection method using backward elimination. These methods and techniques have proven to have high accuracy. The results of this study are expected to be used as a more accurate diagnosis by medical professionals in predicting PCOS disease.

Keywords: Polycystic ovary syndrome, Deep learning, Artificial neural network, SMOTE, Backward elimination
Received March 2025 / **Revised** May 2025 / **Accepted** May 2025

INTRODUCTION

The ovary plays a crucial role in the female reproductive system, not only as the site for producing eggs or ova but also as the place where a fetus develops. However, ovarian diseases are often overlooked by women, such as Polycystic Ovary Syndrome (PCOS), a condition characterized by irregular menstruation, hormonal imbalances, and the formation of polycystic ovaries (PCO) or immature eggs within the ovaries[1]. PCOS occurs due to an increase in androgen levels, a male hormone produced by the ovaries[2]. PCOS disease include hormonal imbalance, irregular menstruation (amenorrhea), hyperandrogenism (increased androgen hormones), hirsutism (excessive hair growth in body areas), alopecia (hair loss), acne, and skin darkening[3].

The World Health Organization (WHO) estimates that approximately 117 million women worldwide have been affected by PCOS, accounting for about 3.5% of the global female population[4]. Many PCOS patients are unaware of their condition due to a lack of awareness[5]. If left untreated, PCOS can lead to serious health complications, making early detection essential to manage the condition before it progresses to more severe stages. With advancements in technology, the integration of computer science and healthcare has provided solutions for disease diagnosis, particularly through data mining and machine learning techniques.

Research on early disease detection using data mining methods has been widely conducted. One such method is predictive modeling, also known as forecasting, which is used to predict future outcomes or probabilities[6]. Several data mining algorithms, including Support Vector Machine (SVM), Neural Networks, and Linear Regression, have been employed for prediction.

Artificial Neural Networks (ANN) one of Neural Network in machine learning algorithm, particularly

*Corresponding author.
Email addresses: miftakhulilmiyah0503@gmail.com (Ilmiyah)*, mula.ab26@gmail.com (Barata), pelangi.ardata@gmail.com (Yuwita)
DOI: 10.15294/sji.v12i1.22886

effective for predictive tasks. ANN mimics the computational patterns of biological neural networks and consists of interconnected processes called neurons[7]. A study by Muhammad Resha et al. demonstrated the superiority of ANN in predicting tuberculosis, achieving an accuracy of 97.59%[8]. Similarly, Serin Wulandari et al. applied ANN to a stroke dataset and achieved an accuracy of 94.83%[9]. Research on the implementation of the ANN algorithm was conducted by Ryan Haris Bawafi et al in 2022 on the diagnosis of hepatitis[10]. Shows that this algorithm can be used to predict hepatitis well. In this study, the training data used came from the Tambak Health Center, Gresik Regency, as many as 20 data. with the ANN model built in the form of a single-layer perceptron.

However, the application of ANN often faces challenges when the dataset used in research exhibits class imbalance. To address this issue, oversampling techniques such as the Synthetic Minority Oversampling Technique (SMOTE) can be employed. SMOTE increases the number of instances in the minority class by generating synthetic data based on existing minority class instances, thereby avoiding overfitting[11]. A study by Edi Sutoyo et al. combined ANN with SMOTE to analyze a television advertisement dataset, achieving an accuracy of 87.06%[12]. Also research related to the SMOTE technique was conducted by Reza Septian Pradana et al. The topic of the research conducted was modeling the NEET (Not in Education, Employment or Training) status of young people in Banten province in 2022[13]. Handling unbalanced data with the Synthetic Minority Oversampling Technique (SMOTE) in this study was proven to increase the accuracy of the final classification results. Researchers also recommend handling unbalanced data first before modeling using machine learning algorithms.

In addition to oversampling, feature selection plays a critical role in the application of algorithms to datasets. Feature selection reduces data dimensionality by retaining only the attributes relevant to the target class. One common feature selection method is backward elimination, which operates similarly to forward elimination but in reverse[14]. Backward elimination evaluates each feature and iteratively removes irrelevant ones to identify the optimal combination of attributes[15]. Attributes with significant values are selected for the model, with lower significance levels indicating stricter selection criteria. Research by Jasman Pardede demonstrated that backward elimination accelerates the training process[16]. Furthermore, a study by Mega Maulina comparing forward selection and backward elimination on the Fragile State Index (FSI) dataset found that backward elimination outperformed forward selection[14]. Given the high accuracy of ANN in previous studies across different datasets, it is a suitable algorithm for predicting PCOS. This research aims to implement the ANN algorithm combined with SMOTE to address class imbalance in PCOS datasets and backward elimination for feature selection. These techniques are expected to optimize prediction accuracy and improve the early detection of PCOS.

METHODS

This research builds an ANN algorithm model using the smote technique combined with backward elimination for feature selection. The research steps can be seen in Figure 1.

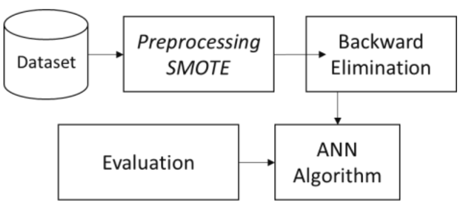

Figure 1 Research methods used

Dataset

This research collected a PCOS dataset from the Kaggle.com website, a public and open data source repository for research purposes, which was released by Shreyas Vedpathak in this link dataset. The dataset includes various clinical and biochemical parameters, making it a valuable resource for studying the prevalence and effects of PCOS in diverse populations. The dataset comprises 43 attributes and 541 records, with a target variable labeled "PCOS," which has two classes: "PCOS" (represented by 1) and "Non-PCOS" (represented by 0). This results in a total of 44 columns, including the target variable. Due to its detailed structure and open accessibility, this dataset offers a valuable resource for conducting predictive analysis and advancing research in the diagnosis of PCOS. Table 1 below outlines the attributes present in the dataset.

Table 1 Dataset's attributes and describtion

Atribut	Describtion	Atribut	Describtion
Sl. No	ID patient	Waist:Hip Ratio	A measurement that compares waist circumference to hip circumference
Patient File No.	ID patient on hospital	TSH (mIU/L)	Thyroid Stimulating Hormone levels in the blood
Age (yrs)	Patient age	AMH(ng/mL)	Amount of Anti-Mullerian hormone
Weight (Kg)	Body weight of patient	PRL(ng/mL)	Amount of prolactin hormone
Height(Cm)	Body height of patient	Vit D3 (ng/mL)	Vitamin D3 levels
BMI	Body mass index patient	PRG(ng/mL)	Progesterone levels measured in nanograms per milliliter
Blood Group	Blood type classification	RBS(mg/dl)	Blood sugar level measurement
Pulse rate(bpm)	Number of times the heart pumps blood	Weight gain(Y/N)	Increased body mass
RR (breaths/min)	Number of breaths per minutes	Hair growth(Y/N)	Significant hair growth on other parts of the body
Hb(g/dl)	Measurement of hemoglobin levels in the blood	Skin darkening (Y/N)	A condition where the skin becomes dark and thick
Cycle(R/I)	Menstrual cycle regular or irregular	Hair loss(Y/N)	A condition in which there is excessive hair loss
Cycle length(days)	Menstrual cycle on days	Pimples(Y/N)	The acne problem that occurs
Marriage Status (Yrs)	Refers to a patient's marital status	Fast food (Y/N)	Patient condition of consuming fast food or not
Pregnant(Y/N)	Whether the patient is pregnant or not	Reg.Exercise(Y/N)	Physical activity that is done regularly
No. of abortions	Amount of abortions ever done	BP _Systolic (mmHg)	Refers to systolic blood pressure
I beta-HCG(mIU/mL)	Human Chorionic Gonadotropin (hCG) hormone levels	BP _Diastolic (mmHg)	Refers to diastolic blood pressure.
II beta-HCG(mIU/mL)	Human Chorionic Gonadotropin (hCG) hormone levels	Follicle No. (L)	Number of follicles (fluid-filled sacs where eggs develop) in the left ovary
FSH(mIU/mL)	Follicle Stimulating Hormone (FSH) levels in the blood	Follicle No. (R)	Number of follicles in the right ovary
LH(mIU/mL)	LH (Luteinizing Hormone) levels in the blood	Avg. F size (L) (mm)	Number of average follicle size in the left ovary
FSH/LH	Follicle Stimulating Hormone and Luteinizing Hormone levels	Avg. F size (R) (mm)	Number of average follicle size in the right ovary
Hip(inch)	Hip size or hip circumference	Endometrium (mm)	Measuring the thickness of the endometrium
Waist(inch)	Waistline of patient	PCOS(Y/N)	Class of data

Preprocessing SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) is an over-sampling method where data in the minority class is augmented using synthetic data derived from data replication in the minority class. If class imbalance in the data is ignored, the machine learning model will tend to focus on predicting the majority class and ignore the minority class. This will cause the model to underfitting or overfitting[17]. The SMOTE technique works by taking values in the minority class to be replicated into synthetic data by taking the k-nearest neighbor values. Mathematically, data generation in SMOTE can be formulated in equation 1.

$$x^b = x + u(x^k - x) \quad (1)$$

Where x^b or synthetic data is created by adding x (minority class data) with u (random number between 0 and 1) and multiplying it by the result of subtracting the nearest neighbor data or x^k with the minority class data.

Backward elimination

Backward elimination is a method for selecting features or attributes that are not relevant to the target class. Backward elimination is a method that can be used to eliminate insignificant attributes from the model[15]. This technique is a wrapper-type feature selection strategy that involves inserting all predictor variables into a linear regression model. The process of backward elimination is shown in Figure 2.

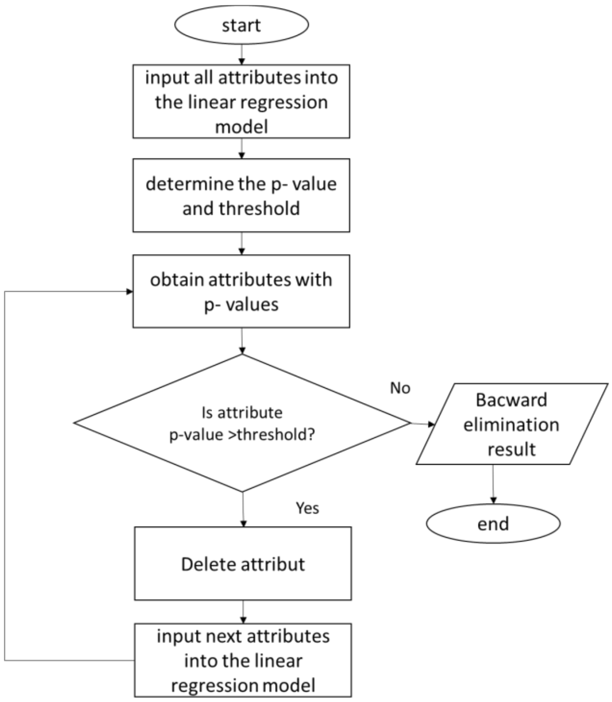

Figure 2 Process of backward elimination

First, all attributes are input into the linear regression model. Next, the p-values and the threshold are determined. The process then involves evaluating each attribute and its corresponding p-value to see if it exceeds the threshold. If the p-value is greater than the threshold, the attribute is removed. This process is repeated until only the most significant attributes remain. The threshold value is set at 0.5, and attributes are selected through a backward elimination process based on their p-values.
Backward elimination is performed to identify the most relevant attributes for the target class in the PCOS dataset by examining the p-values of each attribute. Attributes with p-values exceeding a predefined threshold are eliminated. This process is repeated iteratively until the final set of attributes is determined. If put into a mathematical equation like equation 2.

$$y = b_0 + b_1 * x_1 + b_2 * x_2 \dots . + b_n * x_n \quad (2)$$

Where:

- y = dependent variable
- x = independent variable
- b_n = regression coefficient

Artificial neural network (ANN)
An Artificial Neural Network (ANN) is a technique that solves issues by simulating how the human brain functions. ANN makes decisions by identifying actions based on historical data[8]. This algorithm keeps trying to encourage the learning process to make a choice on the output layer as it has characteristics with human biological neural networks[18]. Between the input and output layers, or it can be referred to as the hidden layer, are a number of tiny units called neurons which make up the network on an ANN[19]. And each neuron has a weight that can influence decisions on the next layer. As visualized in figure 3.

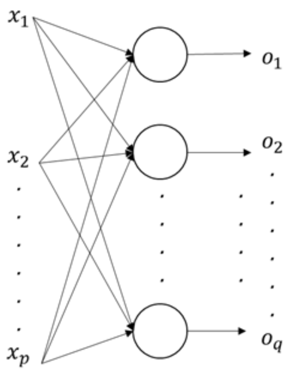

Figure 3 ANN algorithm flow [20]

And the calculation of the linear combination of input with weight(w) is added with bias formulated with equation 3.

$$net_k^{(h)} = \sum_{i=1}^{1} w_{ij}^{(h)}\, x_i^{(h-1)} + b_j^{(h)} \tag{3}$$

For j = 1, 2, …, l, where l represents the total number of neurons in layer h. The outcomes will be used as input for the activation function of choice. The activation function is responsible for delivering output to each neuron; multiple activation functions can be applied depending on the issue to be addressed. We used the ReLU activation function for the hidden layer, which was used to store the value 0 obtained from the forward pass[17], a dropout layer for regularization, and the Sigmoid function for the output layer with binary classification. Main components used to build a model using the ANN algorithm are:

1. Neuron (Node)
   In ANN, the neuron is the main component that receives input, performs computation, and generates output. Each neuron has a weight, which is used to calculate the output.
2. Layers
   There are three types of layers in a neural network:
   a. Input Layer
      The first layer that receives input data. It an be formulated with the following equation 4:

$$z = \sum_{i=1}^{n} (w_i \,.\, x_i) + b \tag{4}$$

   Where:
   - z : The calculated input value.
   - w_i : The weight for the -i input.
   - x_i : Input to i
   - b : Bias.
   - n : The number of inputs.

   b. Hidden Layer
      The layer between the input and output layers that performs non-linear transformation on the data.
   c. Output Layer
      The last layer that generates the output or predicted result.
3. Activation Function
   The activation function is the function that determines the output of a neuron. Some types of activation functions include:
   a. ReLU (Rectified Linear Unit)
      The ReLU activation function is used in hidden layers. Its purpose is to filter out zero values produced during the forward pass[17]. It can be formulated in equation 5:

$$f(x) = max(0, x) \tag{5}$$

   Where x is the value received during the forward pass.

b. Sigmoid
   The sigmoid activation function is highly suitable for outputs that are continuous or measurable data[21].
c. Softmax
   The softmax function is used for models with multi-class outputs and returns the probability of each class, with the target class having the highest probability[22].
d. Loss Function
   This function is used to measure how well the model predicts the data. Some commonly used loss functions include BinaryCrossentropy. Binary cross entropy is a loss function applied to binary classification use cases[23]. This function is used for target variables with two possible outcomes, 0 or 1. This function measures the model's performance in classification tasks where the output is a probability value between 0 and 1. The goal is to minimize the loss function during training.

4. Optimizer
   An optimizer is used to update weights and biases during training. Some commonly used optimizers include:
   a. Adam (AdaptiveMomentEstimation)
      An optimization technique for gradient descent. This method is highly efficient when working with large problems involving a lot of data or parameters because it requires less memory[24].

Evaluation

The final step is to evaluate the research results using a confusion matrix. Confusion Matrix is a very important tool in evaluating the performance of a classification model, especially in the context of binary classification problems, although it can also be used for multiclass classification[25]. The main purpose of the Confusion Matrix is to provide a clear picture of how well the model is classifying data based on predicted and actual results, by showing accuracy, precision, and recall values. If it is determined that the model built using Artificial Neural Networks (ANN) produces less than optimal accuracy, the model can be improved by adding or removing hidden layers, adding dropouts, and so on to achieve the best results. Furthermore, conclusions can be drawn after conducting an evaluation and will be used as the final results of the research. Structure of confusion matrix is as shown in Figure 4.

		Actual value	
Prediction value	Positive	**TP**	**FN**
	Negative	**FP**	**TN**
		Positive	Negative

Figure 4 Confusion matrix

With explanation:
a. TP (True Positive): Data that has an actual value of 1 is predicted to be 1
b. TN (True Negative): Data that has an actual value of 0 is predicted to be 0
c. FN (False Negative): Data that has an actual value of 1 but is predicted to be 0
d. FP (False Positive): Data that has an actual value of 0 but is predicted to be 1

Accuracy refers to the proportion of correct predictions. It analyzes how accurately a classifier predicts a condition[26]. Then the formula is obtained to find the overall accuracy value with equation 6.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

RESULTS AND DISCUSSIONS

This chapter explains about dataset preprocessing, the implementation of the first scenario of Artificial Neural Network (ANN) algorithm and the second scenario, Artificial Neural Network (ANN) algorithm combined with the SMOTE technique and the backward elimination feature selection method.

Preprocessing and SMOTE

First, we need to clean the dataset to obtain data that is ready to be used for the ANN algorithm process. Preprocessing stages carried out include:

Data cleaning

In the data cleaning phase, the presence of missing values and 'NaN' entries in the attributes 'Marriage Status (Yrs)' and 'Fast Food (Y/N)' was systematically investigated. To address these missing values, the median of each respective column was used to impute the empty entries. Furthermore, an analysis of the target class distribution revealed an imbalance, with 364 instances classified as 'No PCOS (0)' and 177 instances classified as 'Yes PCOS (1)'. This disparity in class distribution could introduce bias into predictive modeling, potentially affecting the accuracy of the results if not appropriately managed. A graphical representation of the data distribution is provided in Figure 5.

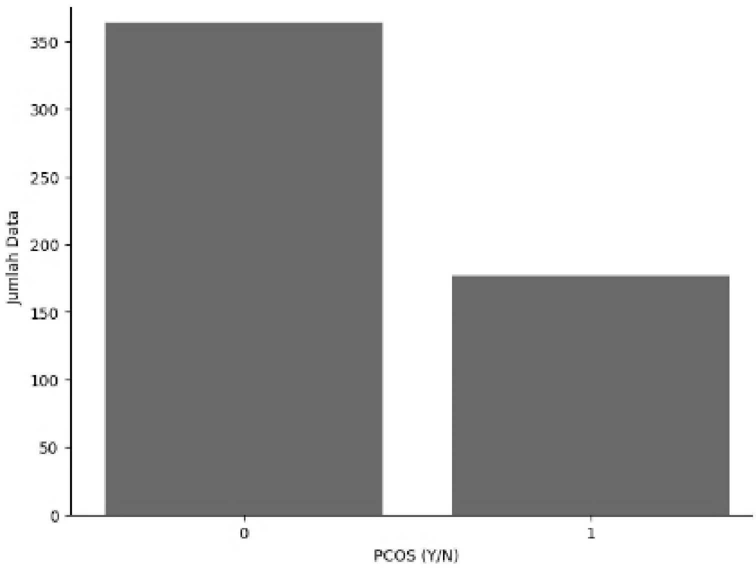

Figure 5 Distribution of dataset

Data Selection

This process picks data based on its relevance to the study being conducted. The researchers eliminated the 'Sl. No.' and 'Patient File No.' sections since they were unrelated.

Data transformation

Data is transformed into a mining-ready format using summary or aggregate methods, followed by standardization using a standard scaler to ensure consistent results. This is done to keep the range of values between measurements in acceptable ranges.

SMOTE

After applying the Synthetic Minority Oversampling Technique (SMOTE), the research dataset, which initially consisted of 541 records, increased to 728 records. This technique effectively balanced the target classes, ensuring an equal distribution between the "PCOS" and "Non-PCOS" categories. By generating synthetic data for the minority class, SMOTE addressed the issue of class imbalance, thereby enhancing the dataset's suitability for training predictive models and improving the accuracy of the analysis. The results can be seen in figure 6.

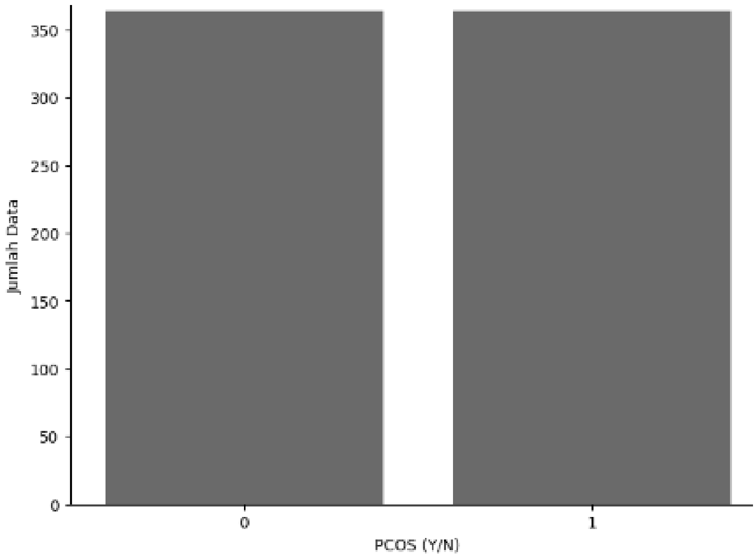

Figure 6 Distribution of dataset after applying SMOTE

Backward elimination

The following stage is backward elimination, in which the results of the linear regression process are removed if the p value is more than 0.5, and the final results of the attributes used for this research are shown in the table 2.

Table 2 Atribut after proses backward elimination

No.	Atribut	p-value	No.	Atribut	p-value	No.	Atribut	p-value
1	Cycle(R/I)	0.000	7	Hip(inch)	0.001	13	Hair loss(Y/N)	0.004
2	Skin darkening (Y/N)	0.000	8	Waist(inch)	0.001	14	LH(mIU/mL)	0.010
3	Fast food (Y/N)	0.000	9	Waist:Hip Ratio	0.001	15	Pimples(Y/N)	0.010
4	Follicle No. (L)	0.000	10	Weight gain(Y/N)	0.001	16	AMH(ng/mL)	0.013
5	Follicle No. (R)	0.000	11	hair growth(Y/N)	0.001	17	Cycle length(days)	0.015
6	Pregnant(Y/N)	0.001	12	Age (yrs)	0.002	18	Reg.Exercise(Y/N)	0.043

This method ensures that only the most statistically significant attributes are retained, thereby optimizing the dataset for model training and improving the accuracy of the predictive analysis.

Artificial neural network (ANN)

Layer (type)	Output Shape
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
dense_1 (Dense)	(None, 64)
dropout_1 (Dropout)	(None, 64)
dense_2 (Dense)	(None, 32)
dropout_2 (Dropout)	(None, 32)
dense_3 (Dense)	(None, 1)

Figure 7 Structure model of ANN

At this point, in the first scenario researcher only applied the ANN algorithm to the dataset without improving data balance or feature selection. The layers utilized have the same structure as the approach previous to the improvement of the SMOTE methodology and backward elimination on the PCOS illness dataset. The training and testing data are compared 40:60, with a model structure consisting of one input layer based on the number of attributes, three hidden layers with 128, 64, and 32 neurons, each layer using the ReLU activation function, three dropout layers weighing 0.8, and one output layer with a sigmoid activation function. To achieve best performance, the model trains 100 times and updates weights and biases using ADAM optimizer with learning rates 0.001. The layer structure is shown in the figure 7.

Evaluation

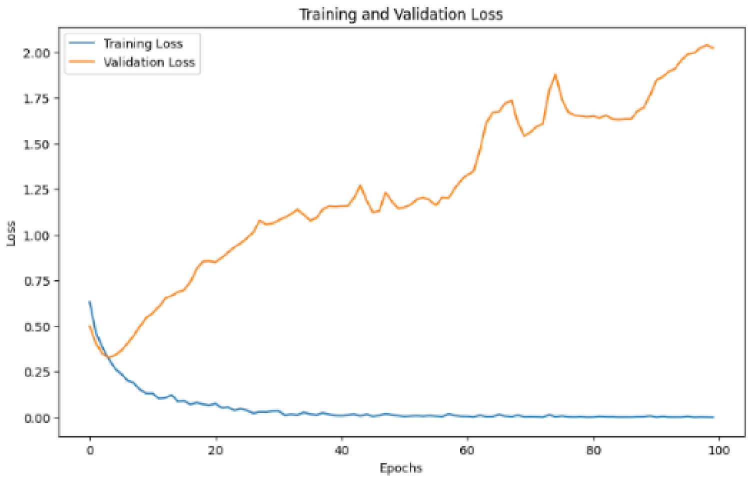

Figure 8 ANN model training

Figure 9 ANN model training after optimation

The training process in the first scenario depicted in Figure 8 illustrates that the validation data value expands during the training process, whereas the training data value continues to drop before stabilizing. Both graphs demonstrate a considerable value range of 0 to 2.0, indicating that the model is overfitting. Meanwhile, after adding the SMOTE technique and backward elimination feature selection displayed in Figure 9, the validation loss and training loss values have a range of values 0 - 0.6 and the validation loss value decreases approaching the training loss so that it can be concluded that the model built has good performance.

	precision	recall	f1-score	support
0	0.89	0.91	0.90	148
1	0.79	0.77	0.78	69

Figure 10 Resume of training process

	precision	recall	f1-score	support
0	0.92	0.92	0.92	144
1	0.93	0.92	0.92	148

Figure 11 Resume of training model after optimization

With an unbalanced distribution of target classes (support) where class (0) has 148 data and class (1) has 69 data that shown in figure 10. And with an balanced distribution of target classes (support) where class (0) has 144 data and class (1) has 148 data that shown in figure 11.

Then In the first scenario for evaluation using confusion matrix obtained accuracy result of 86%. With dataset that has unbalanced target value and using 44 attributes in total. As demonstrated in the figure 12. And the second scenario involves the utilization of ANN algorithm with SMOTE approach and backward elimination to select features from PCOS dataset. And the results are shown in Figure 13.

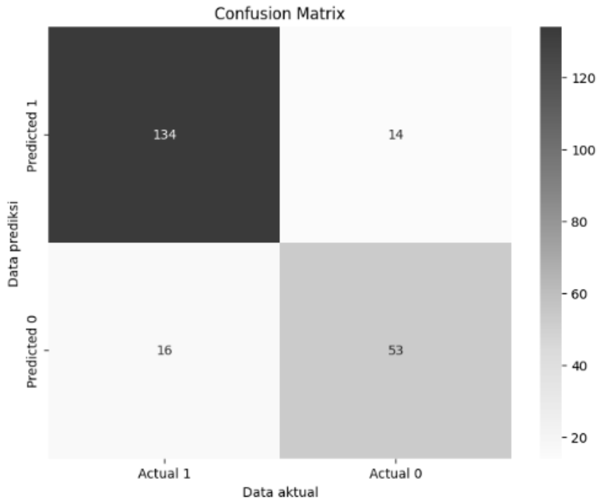

Figure 12 Confusion matrix scenario I

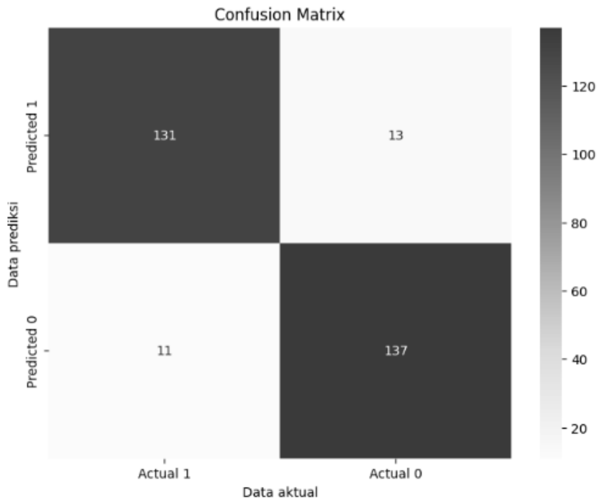

Figure 13 Confusion matrix scenario II

Finally based on Figures 12 and 13, the accuracy obtained based on the calculation using equation 6.

$$Accuracy\ (figure\ 12) = \frac{134 + 53}{134 + 53 + 14 + 16} = \frac{187}{217} = 0.86$$

$$Accuracy\ (figure\ 13) = \frac{131 + 137}{131 + 137 + 13 + 11} = \frac{268}{292} = 0.92$$

CONCLUSION

The results of this research show that the ANN algorithm can be applied to predict PCOS, with an accuracy of 86%. However, this accuracy is affected by the imbalance in the target class and the large number of attributes in the dataset. By incorporating data balancing techniques using SMOTE and feature selection through backward elimination, the accuracy of the ANN model improved to 92%. Initially, Scenario 1 was tested using the same model structure, without any optimization. This model experienced overfitting during training. The addition feature selection method of backward elimination helped identify the most significant attributes related to PCOS, allowing the model to perform optimally. A total of 18 attributes were selected through the backward elimination process, including Cycle (R/I), Skin darkening (Y/N), Fast food (Y/N), Follicle No. (L), Follicle No. (R), Pregnant (Y/N), Hip (inch), Waist (inch), Waist: Hip Ratio, Weight gain (Y/N), Hair growth (Y/N), Age (yrs), Hair loss (Y/N), LH (mIU/mL), Pimples (Y/N), AMH (ng/mL), Cycle length (days), and Regular Exercise (Y/N). The findings of this research are expected to assist medical professionals in diagnosing PCOS and provide the public with knowledge about the symptoms and indications of PCOS. It is hoped that future research will explore different ANN model structures and compare other techniques for data balancing and feature selection to further improve accuracy.

REFERENCES

[1] F. Novianti and N. Ulinnuha, "DATA REKAM MEDIS PCOS MENGGUNAKAN SVM FEATURE SELECTION USING GENETIC ALGORITHM IN PCOS MEDICAL RECORD DATA CLASSIFICATION USING SVM," vol. 9, no. 1, pp. 9–19, 2024.

[2] S. Bharati, P. Podder, and M. R. H. Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms," no. June, pp. 5–7, 2020.

[3] D. A. Sari, E. Y. Kurniawati, and M. A. Ashari, "Skrining Dan Determinan Kejadian Sindrom Ovarium Polikistik (Sopk) Pada Remaja," *J. Ilmu Kebidanan*, vol. 9, no. 2, pp. 102–106, 2023, doi: 10.48092/jik.v9i2.211.

[4] K. Nisa, B. P. Dewa, and S. K. Wibisono, "Model Comparison of Random Forest and Logistic Regression Algorithms in PCOS Disease Detection," vol. 4, no. 1, pp. 73–79, 2023.

[5] K. M. Hoeger, A. Dokras, and T. Piltonen, "Update on PCOS : Consequences , Challenges , and Guiding Treatment," vol. 106, no. 3, 2021, doi: 10.1210/clinem/dgaa839.

[6] M. Barata, "PENINGKATAN PERFORMA ALGORITMA C4.5 DENGAN METODE CHI-SQUARE PADA KLASIFIKASI TEH MURNI MENGGUNAKAN PERANGKAT ELECTRONIC NOSE," 2022.

[7] C. B. Handoko, C. Sri, and K. Aditya, "Penerapan Teknik SMOTE dalam Mengatasi Imbalance Data Penyakit Diabetes Menggunakan Algoritma ANN Application of SMOTE Technique in Addressing Imbalanced Diabetes Disease Data Using the ANN Algorithm," vol. 14, no. 105, pp. 13–20, 2025.

[8] M. Resha and S. Syamsu, "Prediksi Penyebaran Kasus Tuberkulosis dengan metode Artificial Neural Network dan Multi-Layer Perceptron di kota makassar," 2021.

[9] S. Wulandari, Y. I. Mukti, and T. Susanti, "OPTIMALISASI PREDIKSI PENYAKIT STROKE MENGGUNAKAN," vol. 8, no. 2, pp. 1826–1833, 2024.

[10] R. H. Bawafi and U. Chotijah, "SISTEM PREDIKSI DIAGNOSA PENYAKIT HEPATITIS MENGGUNAKAN METODE ARTIFICIAL NEURAL NETWORK (ANN) SINGLE LAYER PERCEPTRON STUDI KASUS PADA," vol. 04, no. 2, pp. 85–103, 2022.

[11] R. Peranginangin, E. J. G. Harianja, I. K. Jaya, and B. Rumahorbo, "Penerapan Algoritma Safe-Level-Smote Untuk Peningkatan Nilai G-Mean Dalam Klasifikasi Data Tidak Seimbang," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 1, pp. 67–72, 2020, doi: 10.46880/jmika.vol4no1.pp67-72.

[12] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," vol. 6, no. 3, pp. 379–385, 2020.

[13] R. S. Pradana, R. Nooraeni, B. Pusat, S. Kabupaten, and R. Logistik, "PENERAPAN SMOTE PADA DATA TIDAK SEIMBANG DALAM PEMODELAN STATUS NEET PENDUDUK USIA MUDA DI PROVINSI BANTEN TAHUN 2022 THE APPLICATION OF SMOTE TO IMBALANCED DATA ON NEET STATUS MODELLING OF YOUTH IN BANTEN PROVINCE 2022," vol. 18, pp. 91–104, 2023.

[14] M. Maulina, M. Wijaya, Yunita, and N. Fitriyati, "View of PERBANDINGAN METODE BACKWARD DAN FORWARD PADA SELEKSI MIXED-EFFECTS MODEL.pdf." 2024.

[15] H. Nugroho, G. E. Yuliastuti, and A. Firman, "Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Backward Elimination Diabetes Melitus Diagnosis Classification Using the Naive Bayes Method With Feature Selection Backward Elimination," *J. Ilm. NERO*, vol. 8, no. 2, p. 2023, 2023.

[16] J. Pardede and N. Desita, "View of Hepatitis Identification using Backward Elimination and Extreme Gradient Boosting Methods.pdf." 2024.

[17] M. Handayani, M. Riandini, and Z. Zakarias, "Perbandingan Fungsi Optimasi Neural Network Dalam Klasifikasi Kelayakan Calon Suami," *J. Inform.*, vol. 9, no. 1, pp. 78–84, 2022, doi: 10.31294/inf.v9i1.12318.

[18] M. R. Ghufron and M. J. Vikri, "PENERAPAN ALGORITMA NEURAL NETWORK BACKPROPAGATION UNTUK PROTOTYPE SISTEM KONTROL POMPA AIR BERBASIS MIKROKONTROLER," 2023.

[19] J. Zupan, "Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them," *Acta Chim. Slov.*, vol. 41, 1994.

[20] N. Ye, *Data Mining*.

[21] S. Narayan, "The Generalized Sigmoid Activation Function :," *Intell. Syst.*, vol. 82, no. 99, pp. 69–82, 1997.

[22] L. O. A. S. Sagala, "Klasifikasi Cats dan Dogs dengan Metode CNN dalam Fungsi Aktivasi Relu , Sigmoid , Softmax , Softplus , Softsign , dan Selu," *Pros. Semin. Nas. Inform. Bela Negara di UPN "Veteran" Jawa Timur*, vol. 2, no. 1, pp. 1–6, 2023.

[23]. Usha Ruby Dr.A, "Binary cross entropy with deep learning technique for Image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, 2020, doi: 10.30534/ijatcse/2020/175942020.

[24] R. N. Singarimbun, E. B. Nababan, and O. S. Sitompul, "Adaptive Moment Estimation to Minimize Square Error in Backpropagation Algorithm," *2019 Int. Conf. Comput. Sci. Inf. Technol. ICoSNIKOM 2019*, 2019, doi: 10.1109/ICoSNIKOM48755.2019.9111563.

[25] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification : A measure driven view," *Inf. Sci. (Ny).*, vol. 507, pp. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.

[26] D. Valero-carreras, J. Alcaraz, and M. Landete, "Computers and Operations Research Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. April 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.