ELKHA : Jurnal Teknik Elektro, Vol. 18 No.1, April 2026, pp. 65 - 73
ISSN: 1858-1463 (print), 2580-6807 (online)

Predicting Breakdown Voltage of Transformer Oil under Copper/Iron Contamination: A Comparative Study of Gradient vs Metaheuristic Training

Giovanni Dimas Prenata[1*)]

[1*)]Faculty of Intelligent Electrical and Informatics Technology, Universitas 17 Agustus 1945 Surabaya, Indonesia
Corresponding Email: [*)]gprenata@untag-sby.ac.id

Abstract – **Transformer oil functions as an insulating and cooling medium in high-voltage power systems, whose dielectric condition degrades over service life due to thermal aging, moisture ingress, and metallic contamination, leading to reduced Breakdown Voltage (BDV) and increased insulation failure risk that may necessitate oil regeneration, replacement, or indicate transformer end-of-life. Unlike Dissolved Gas Analysis (DGA), which evaluates transformer faults based on gas decomposition products, BDV directly reflects the dielectric strength of insulating oil and is more sensitive to particulate contamination such as Cu and Fe, making it more suitable for assessing material-level insulation degradation. This study investigates the influence of copper (Cu) and iron (Fe) particle contamination on BDV. It compares three Artificial Neural Network (ANN) training strategies for BDV prediction: gradient-based training (DFFNN-Pure), Genetic Algorithm optimization (DFFNN-GA), and Grey Wolf Optimizer-based training (DFFNN-GWO), using experimental data from 36 transformer oil samples obtained in accordance with IEC 60156:2018. The comparison represents a before-and-after modeling perspective on training strategy rather than on repeated physical testing. The results show that DFFNN-Pure achieved the highest prediction accuracy (R^2 = 0.996, RMSE = 0.296 kV, MAE = 0.238 kV), while DFFNN-GWO demonstrated stable convergence with competitive accuracy (R^2 = 0.971, RMSE = 0.886 kV), whereas DFFNN-GA exhibited unstable convergence and poor generalization. Unlike previous studies that primarily focus on transformer remaining useful life estimation at the system level, this work emphasizes predicting BDV of transformer oil at the material level under metallic contamination. It provides a systematic comparison between gradient-based and metaheuristic training within the same DFFNN framework, supporting non-destructive condition monitoring and predictive maintenance.**

Keywords: Breakdown Voltage (BDV), Artificial Neural Network (ANN), Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), metaheuristic optimization.

I. INTRODUCTION

Transformer oil serves a dual function as both an electrical insulating medium and a cooling agent in high-voltage power systems. The dielectric reliability of transformer oil is therefore critical to maintaining system stability, as insulation failure may lead to severe operational disturbances, irreversible damage, or catastrophic transformer outages. One of the most widely used indicators to evaluate transformer oil quality is the Breakdown Voltage (BDV), which represents the maximum electric stress the oil can withstand before electrical discharge occurs [1]. BDV testing is standardized under IEC 60156, in which an alternating electric field is gradually increased across a defined electrode gap until breakdown is observed, making BDV a fundamental criterion for assessing the suitability of both new and in-service transformer oils.

During service operation, transformer oil undergoes gradual degradation due to thermal-oxidative aging, moisture absorption, and contamination by metallic particles originating from copper windings and iron core components. Thermal and oxidative aging processes generate polar by-products such as organic acids and water, which increase oil conductivity and reduce its dielectric strength [2]. Moisture is particularly detrimental, as it enhances charge-carrier mobility and accelerates discharge-channel formation. In addition, suspended copper (Cu) and iron (Fe) particles distort the local electric field and promote particle-bridging mechanisms, significantly reducing BDV and accelerating insulation deterioration [3] [4]. These combined effects not only lower dielectric performance but also shorten the operational lifespan of both liquid and solid insulation systems.

To describe the probabilistic behavior of transformer oil breakdown, statistical models such as the Weibull distribution have been widely employed. The shape (β) and scale (η) parameters of the Weibull model characterize BDV variability and mean dielectric strength under different aging or contamination conditions [5]. Although effective for reliability estimation and failure probability analysis, Weibull-based approaches remain empirical and limited in capturing complex nonlinear interactions among physicochemical properties, contamination characteristics, and dielectric performance.

Artificial Neural Networks (ANNs) have increasingly been adopted to model nonlinear relationships between transformer oil condition parameters such as moisture content, contamination level, and physicochemical properties and BDV values. ANN-based models can learn

Manuscript received 2025-11-08; revised 2026-03-28; accepted 2026-03-30

complex input–output mappings directly from experimental data and have shown promising results in insulation diagnostics and condition assessment [6]. Unlike the study in [6], which focuses on BDV prediction under general operating conditions using machine learning models, this research specifically investigates the impact of metallic particle contamination (Cu and Fe) on BDV. It evaluates the effectiveness of different ANN training strategies within the same DFFNN architecture. However, the predictive accuracy and generalization capability of ANN models strongly depend on the training algorithm employed. Conventional gradient-based methods, including backpropagation, perform well on smooth error surfaces but are prone to premature convergence and local minima, particularly when datasets are small, noisy, or highly nonlinear [7].

To address these limitations, metaheuristic optimization algorithms have been introduced as alternative training strategies for neural networks. Algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Grey Wolf Optimizer (GWO) optimize network weights without relying on gradient information, thereby improving robustness to local minima [8]. Among these, GWO is inspired by the hierarchical leadership structure and cooperative hunting behavior of grey wolves, where solution candidates are guided by three dominant agents (α, β, and δ) to balance global exploration and local exploitation during optimization [9]. Previous studies have demonstrated that GWO-based training can improve convergence stability and prediction accuracy in both regression and classification problems, including ANN and Multi-Layer Perceptron (MLP) models [10].

Despite the growing use of ANN and metaheuristic methods in transformer diagnostics, comparative investigations focusing on gradient-based versus metaheuristic-based training strategies for BDV prediction under copper and iron contamination remain limited. In particular, systematic comparisons between DFFNN-Pure, DFFNN-GA, and DFFNN-GWO using the same experimental BDV dataset are rarely reported [11]. While previous studies [8] [10] have applied metaheuristic algorithms such as GA and GWO to optimize neural networks for general prediction and classification tasks, they do not specifically address BDV prediction under metallic contamination, nor evaluate the performance differences between gradient-based and metaheuristic training methods within a unified experimental dataset and network structure. Therefore, this study presents a comparative evaluation of three DFFNN training strategies gradient descent (DFFNN-Pure), Genetic Algorithm optimization (DFFNN-GA), and Grey Wolf Optimizer-based training (DFFNN-GWO for predicting the breakdown voltage of transformer oil contaminated with Cu and Fe particles. The results are expected to provide insights into algorithm performance, convergence behavior, and predictive accuracy, contributing to AI-based condition monitoring and predictive maintenance of power transformers.

The main novelty of this study lies in three key aspects. First, this research focuses specifically on material-level BDV prediction under controlled metallic contamination (Cu and Fe), which is less explored than system-level diagnostics such as DGA. Second, this study provides a systematic comparison of gradient-based and metaheuristic training methods within the same DFFNN architecture and dataset, ensuring a fair and controlled evaluation. Third, the integration of GWO and GA within a unified experimental framework enables detailed analysis of convergence behavior, stability, and predictive performance under nonlinear and small-sample conditions.

These contributions distinguish this work from previous studies that primarily apply ANN or metaheuristic optimization independently without direct comparative analysis under identical experimental conditions.

II. METHODOLOGY

A. Weibull Distribution for Breakdown

The Weibull distribution is widely used for reliability analysis and for predicting failure probability (breakdown) in insulating materials such as transformer oil. This distribution is expressed through the probability density function (PDF) and the cumulative distribution function (CDF) as follows:

$$f(v) = \frac{\beta}{\eta} \left(\frac{v}{\eta}\right)^{\beta-1} e^{-\left(\frac{v}{\eta}\right)^{\beta}} \quad \text{................ (1)}$$

$$F(v) = 1 - e^{-\left(\frac{v}{\eta}\right)^{\beta}} \quad \text{................ (2)}$$

where:

- v : breakdown voltage (observed value)
- β : shape parameter, which defines the form of the distribution curve
- η : scale parameter, representing the characteristic breakdown voltage
- f (v) : probability density function
- F (v) : cumulative probability of failure

The parameters β and η are typically estimated from experimental breakdown voltage data using the Maximum Likelihood Estimation (MLE) method or the linear regression approach based on the log–log transformation:

$$\ln(-\ln(1 - F(v_i))) = \beta \ln v_i - \beta \ln \eta \quad \text{........ (3)}$$

By plotting $\ln(-\ln(1 - F))$ $vs \ln v$, the slope of the resulting line, which corresponds to β while the intercept yields $\ln \eta$. The Weibull distribution helps quantify the probability that a sample of transformer oil will fail (break down) under a given voltage stress, and it is employed as a comparative feature or analytical tool in studies of electrical breakdown phenomena in transformer oil.

B. Margins Feedforward Neural Network (DFFNN): Fundamental Formulation

Figure 1 shows how to present a figure. The title should be located under the figure. The Figure must be easy to read and have a good resolution/quality. It should be noted that most readers may read the paper in black/white, although the online version is presented with the original color set from the final manuscript.

A Deep Feedforward Neural Network (DFFNN) consists of an input layer, one or more hidden layers, and an output layer. Each neuron computes the following:

$$z_j = \sum_i x_i w_{ij} + b_j \quad (4)$$

$$a_j = \sigma(z_j) \quad (5)$$

where:

- x_i : input signal (or activation from the previous layer)
- w_{ij} : weight connecting neuron
- b_j : bias of neuron j
- z_j : linear combination of the input signals
- a_j : activation of the neuron (output)
- σ : activation function : $\sigma(z) = \frac{1}{(1+e^{-z})}$

If two hidden layers are employed, the forward propagation process is executed sequentially through these layers until the final output $\hat{y}$. For regression tasks such as BDV prediction, the Mean Squared Error (MSE) is commonly used as the loss function:

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2 \quad (6)$$

The backpropagation process computes the gradient of the loss function with respect to the network weights as follows:

1. Output layer gradient:

$$\delta^{(L)} = \left(a^{(L)} - y\right).\sigma'(z^{(L)}) \quad (7)$$

2. Hidden layer gradient:

$$\delta^{(l)} = (W^{(l+1)})^T \delta^{(l+1)}.\sigma'(z^{(l)}) \quad (8)$$

3. Weight and bias update:

$$w^{(l)} := w^{(l)} - \eta a^{(l-1)} \delta^{(l)} \quad (9)$$

$$b^{(l)} := b^{(l)} - \eta \delta^{(l)} \quad (10)$$

where:

- η : learning rate
- $\sigma'(z)$: derivative of the activation function (for the sigmoid function: $\sigma(z)(1 - \sigma(z))$)

This method is efficient when the error surface is smooth, and the dataset is sufficiently large; however, it is prone to being trapped in local minima, particularly for small datasets or highly nonlinear error functions.

C. Genetic Algorithm (GA) for Weight Optimization

The Genetic Algorithm (GA) is an evolutionary optimization technique inspired by natural selection, employing genetic operations such as selection, crossover, and mutation. For optimizing the weights of an Artificial Neural Network (ANN), the following steps are typically implemented:

1. Chromosome representation: Each chromosome encodes a set of network weights and biases (all weights are flattened into a single gene array).
2. Fitness function: The fitness of each chromosome is evaluated based on the Mean Squared Error (MSE) or another error function; chromosomes producing lower MSE values are assigned higher fitness scores.
3. Selection: Parent chromosomes are selected based on their fitness probability using methods such as the roulette wheel or tournament selection.
4. Crossover: Genetic material from two parents is combined to generate offspring using single-point, multi-point, or uniform crossover strategies.
5. Mutation: Small random perturbations are introduced into the genes to maintain population diversity and avoid premature convergence.
6. Regeneration: The least fit chromosomes are replaced by the newly generated offspring from the crossover and mutation processes.

This iterative process continues until a stopping criterion is met, such as the MSE falling below a predefined threshold or the maximum number of generations is reached.

D. Grey Wolf Optimizer (GWO) for Network Training

The Grey Wolf Optimizer (GWO) is a swarm-based algorithm that simulates the social hierarchy and cooperative hunting behavior of grey wolves. Its fundamental principles are as follows:

1. Hierarchy: The three best-performing wolves are designated as α (alpha), β (beta), and δ (delta), while the remaining wolves are classified as ω (omega).
2. Search phase: The control coefficient a decreases linearly from 2 to 0 throughout iterations, shifting the search process from global exploration to local exploitation. For each agent i, its position is updated according to :

$$D_\alpha = |C_1 . X_\alpha - X_i|, X_1 = X_\alpha - A_1 . D_\alpha \quad (11)$$

$$D_\beta = |C_2 . X_\beta - X_i|, X_2 = X_\beta - A_2 . D_\beta \quad (12)$$

$$D_\delta = |C_3 . X_\delta - X_i|, X_3 = X_\delta - A_3 . D_\delta \quad (13)$$

$$X_i := \frac{X_1 + X_2 + X_3}{3} \quad (14)$$

where :

$$A_j = 2a\sigma r_1 - a, C_j = 2r_2, j \in \{\alpha, \beta, \delta\} \quad (15)$$

and r_1, r_2 are random numbers uniformly distributed within the range [0,1].

3. Evaluation: Each agent (wolf) represents a vector of ANN weights; its fitness is computed based on the Mean Squared Error (MSE), and the wolves are ranked as α, β, δ.
4. Iteration: The process is repeated until convergence or until the maximum number of iterations is reached.

The GWO algorithm provides an effective balance between exploration and exploitation and does not rely on gradient information, making it suitable for optimizing highly nonlinear objective functions. In the context of ANN training, each wolf represents a set of network weights, and their positions are iteratively updated according to the positions of the α, β, and δ wolves until convergence. The control parameter a decreases linearly from 2 to 0 over iterations to maintain the balance between global exploration and local exploitation. Cinar [12] demonstrated that GWO effectively minimizes prediction error on nonlinear datasets such as wind speed and achieves faster convergence than GA. Furthermore, Liu [13] proposed an Improved Grey Wolf Optimizer (IGWO), which accelerates the search for optimal solutions through dynamic adaptation of control parameters, demonstrating significant performance improvements in ANN parameter optimization tasks.

E. Research Design

This study aims to compare the performance of the Gradient-based training method (Backpropagation) with two metaheuristic training approaches, the Genetic Algorithm (GA) and the Grey Wolf Optimizer (GWO) in predicting the Breakdown Voltage (BDV) of transformer oil contaminated with copper (Cu) and iron (Fe) particles. The research adopts a quantitative experimental approach, carried out in systematic stages that include the collection of laboratory test data, data preprocessing, DFFNN model training using three distinct optimization strategies, and performance analysis based on statistical metrics.

1) DFFNN Pure

The DFFNN-Pure model is trained using a conventional gradient-based backpropagation algorithm. The process includes data normalization, forward propagation to generate BDV predictions, error computation using Mean Squared Error (MSE), and iterative weight updates using gradient descent. The training continues until convergence or the maximum number of epochs is reached. Model performance is evaluated using R^2, RMSE, and MAE.

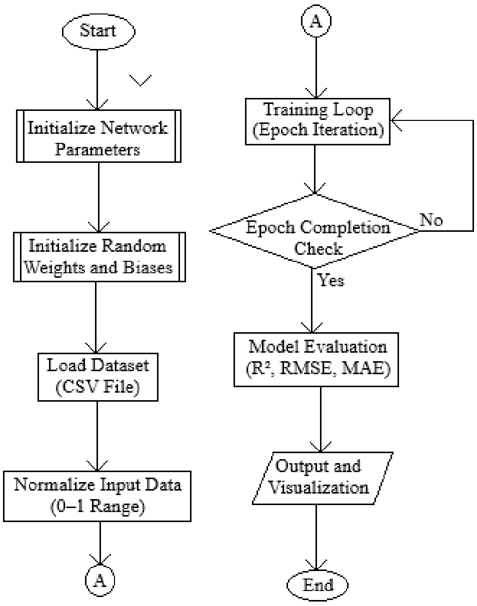

Figure 1. Flowchart of DFFNN Pure

Figure 1 illustrates the workflow of the DFFNN-Pure model based on gradient descent learning. The process begins with data loading and normalization, followed by forward propagation to estimate the breakdown voltage (BDV). The prediction error is then propagated backward to update network weights using a fixed learning rate. This iterative process continues until the convergence criterion or maximum epoch is reached. The flowchart emphasizes the dependence of gradient-based learning on local error gradients, which may limit convergence performance when the error surface is highly nonlinear.

2) DFFNN-GA

The DFFNN-GA model employs a Genetic Algorithm to optimize network weights. Each chromosome represents a set of weights and biases, and fitness is evaluated based on MSE. Through iterative selection, crossover, and mutation, the population evolves toward optimal solutions. The best chromosome is selected as the final model and evaluated using R^2, RMSE, and MAE.

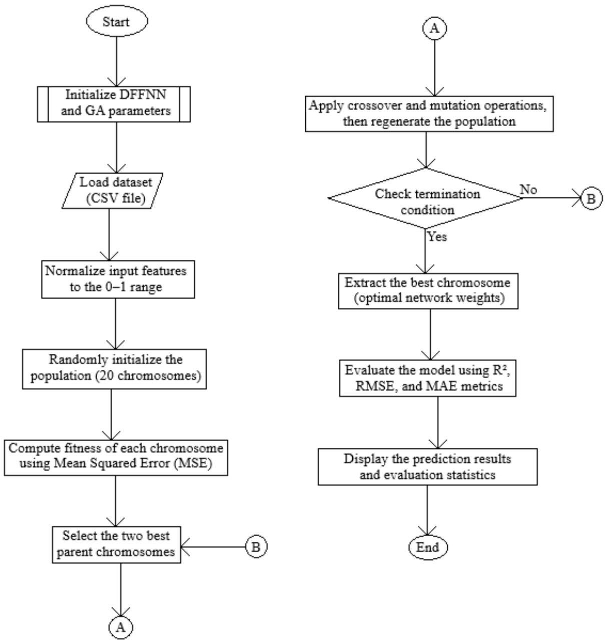

Figure 2. Flowchart of DFFNN-GA

Figure 2 presents the operational flow of the DFFNN-GA model, in which network weight optimization is performed using a Genetic Algorithm. Instead of direct gradient updates, the model encodes neural weights as chromosomes and evaluates their fitness using the Mean Squared Error (MSE). Through iterative selection, crossover, and mutation processes, the population evolves toward improved solutions. This figure highlights the stochastic nature of GA, which enables global exploration but may introduce convergence instability when population diversity is insufficient.

3) DFFNN-GWO

The DFFNN-GWO model integrates the Grey Wolf Optimizer to optimize weights. Each wolf represents a candidate solution, and the best three wolves (α, β, δ) guide the search process. The algorithm iteratively updates positions to balance exploration and exploitation until convergence. The final model is evaluated using R^2, RMSE, and MAE.

Figure 3 depicts the training mechanism of the DFFNN-GWO model, which integrates the Grey Wolf Optimizer into neural network training. Each wolf represents a candidate weight configuration, and the hunting behavior, guided by the α, β, and δ wolves drives the optimization process. The linearly decreasing control parameter ensures a balance between global exploration and local exploitation. This flowchart demonstrates how GWO effectively navigates complex error landscapes without relying on gradient information.

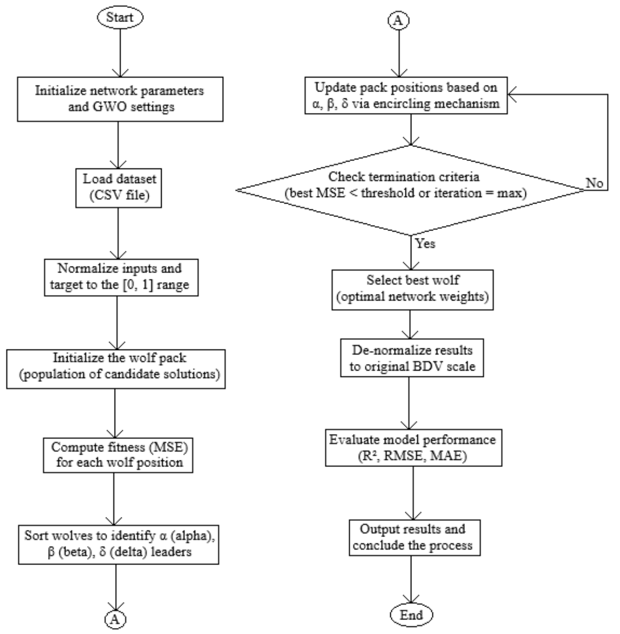

Figure 3. Flowchart of DFFNN-GWO

F. Research Data

The dataset used in this study was obtained from breakdown voltage tests of transformer oil conducted in accordance with IEC 60156:2018. The samples were subjected to controlled contamination using copper (Cu) and iron (Fe) particles. The dataset comprises 36 test sets compiled in the file MinyakTrafo.xlsx, with each sheet representing a distinct testing condition, such as before and after purification, and with varying levels of metallic contamination. The consolidated results from all sheets were merged into a single file for processing and analysis.

Table 1. List of variables used in the research

Variable	Description	Unit
X_1	Operating time of the oil	hours
X_2	Testing temperature	°C
X_3	Moisture content	ppm
X_4	Copper (Cu) concentration	ppm
X_5	Iron (Fe) concentration	ppm
X_6	Purification condition (0 = unpurified, 1 = purified)	-
Y	Breakdown Voltage (BDV)	kV

Table 1 summarizes the input and output variables employed in this study. The selected input features represent operational, thermal, and contamination-related parameters that directly influence transformer oil dielectric strength. The output variable, BDV, serves as a quantitative indicator of insulation performance. This structured variable selection ensures that both physical and chemical degradation factors are incorporated into the predictive model.

The dataset consists of several input variables (features) and one output variable: the average breakdown voltage measured in kilovolts (kV). Before model training, all data were normalized to the range [0, 1] to ensure numerical stability and improve convergence during training.

G. DFFNN Model Architecture

The primary model employed in this study is a Deep Feedforward Neural Network (DFFNN) configured as follows:

1. Input layer: 6 neurons (corresponding to the number of input features)
2. Hidden layer: 1-2 layers with 8-12 neurons (empirically tuned)
3. Output layer: 1 neuron (predicted BDV value)
4. Activation function: Sigmoid
5. Loss function: Mean Squared Error (MSE)

The initial weights were randomly initialized, and three distinct training methods were implemented:

Table 2. The differences between each method

Method	Brief Description
DFFNN-Pure (Gradient Descent)	Conventional backpropagation training using a fixed learning rate.
DFFNN-GA	Weight optimization using a population-based Genetic Algorithm involving selection, crossover, and mutation.
DFFNN-GWO	Weight training using the Grey Wolf Optimizer with three hierarchical roles (α, β, δ) that iteratively update weight positions during optimization.

Table 2 compares the fundamental characteristics of the three training strategies applied to the DFFNN model. Gradient-based training relies on deterministic error minimization, whereas GA and GWO utilize population-based search mechanisms. The table highlights that metaheuristic approaches do not rely on gradient information, making them more suitable for highly nonlinear and multimodal optimization problems.

H. Training and Validation Procedure

The training process was carried out through the following steps:

1. Data Preparation:
   All numerical data were cleaned of missing values and normalized. The dataset was divided into 80% training and 20% testing data.
2. Model Training:
   a. Each model was trained for up to 5000 iterations or until the MSE converged.
   b. For DFFNN-GA and DFFNN-GWO, the metaheuristic parameters were defined as follows: population size: 10 individuals (wolves or chromosomes), maximum generations: 5000 and convergence threshold: MSE < 0.001.
3. Model Evaluation:
   After training, all models were tested using the test dataset. Predicted BDV values were compared with the actual experimental results.
4. Performance Metrics:
   The models were evaluated based on three primary performance measures:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad \text{................ (16)}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad \text{................ (17)}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \text{..............} (18)$$

I. *Experimental Scheme*

The experiments were conducted through the following steps:

1. Training the DFFNN-Pure model:
   The model was trained using the gradient-based backpropagation algorithm with a constant learning rate of 0.01.
2. Training the DFFNN-GA model:
   The network weights and biases were optimized through a population of chromosomes using selection–crossover–mutation mechanisms. The fitness function was defined as the Mean Squared Error (MSE).
3. Training the DFFNN-GWO model:
   The weights and biases were optimized through the collaborative behavior of the α, β, and δ wolves. The control parameter "a" decreased linearly from 2 to 0 over 5000 iterations.
4. Performance Evaluation:
   The performance of the three models was compared based on MSE, RMSE, MAE, and R^2.

Table 3. List of variables used in the research

Parameter	DFFNN-Pure	DFFNN-GA	DFFNN-GWO
Number of Iterations	5000	5000	5000
Population Size	–	10	10
Learning Rate	0.01	–	–
Mutation Rate	–	0.3	–
MSE Threshold	0.001	0.001	0.001
Activation Function	Sigmoid	Sigmoid	Sigmoid

Table 3 outlines the hyperparameters used for each training method. To ensure a fair comparison, the maximum number of iterations and convergence thresholds were kept identical across all models. Differences in learning mechanisms, such as the learning rate in DFFNN-Pure and the mutation rate in DFFNN-GA, reflect the inherent characteristics of each optimization approach.

J. *Training and Validation Procedure*

To improve the robustness and reliability of the experimental results, each model (DFFNN-Pure, DFFNN-GA, and DFFNN-GWO) was executed multiple times (10 independent runs) using different random initializations. The final performance metrics were reported as the mean and standard deviation values of R^2, RMSE, and MAE.

In addition, a k-fold cross-validation approach (k = 5) was applied to evaluate the models' generalization capability. The dataset was randomly partitioned into five subsets, with each subset used once as test data while the remaining subsets were used for training. This process was repeated until all subsets had been used as testing data.

This validation strategy ensures that the reported results are not dependent on a single data split and provides a more reliable assessment of model performance, particularly considering the limited dataset size.

III. RESULTS AND DISCUSSION

This study compared three training strategies for the Deep Feedforward Neural Network (DFFNN)-Gradient Descent (DFFNN-Pure), Genetic Algorithm (DFFNN-GA), and Grey Wolf Optimizer (DFFNN-GWO) to predict the Breakdown Voltage (BDV) of transformer oil contaminated with copper (Cu) and iron (Fe). All models were trained using BDV test data obtained under IEC 60156 standards, which were consolidated from 36 experimental spreadsheets into a single dataset (MinyakTrafo_Final36.csv). The dataset included metal concentration parameters, average BDV test results, and two Weibull distribution parameters (β and η) that represent dielectric reliability. Increasing Fe contamination led to a greater reduction in BDV compared to Cu, as Fe exhibits catalytic properties that accelerate oil oxidation. This phenomenon alters the molecular structure and lowers the critical electric field strength, consistent with the dielectric degradation theory of hydrocarbon-based insulating oils.

The MSE convergence curves in Figure 6 reveal distinct optimization characteristics among the three methods. The DFFNN-Pure model stabilized at MSE ≈ 0.11 after 3000 iterations, showing a monotonic yet relatively slow convergence. This behavior is typical of gradient-based algorithms that depend on the derivative of the error function. When the error surface is non-convex, gradients tend to approach zero at local minima, causing the learning process to stagnate before reaching the global optimum.

The DFFNN-GA model exhibited significant MSE oscillations due to the stochastic nature of crossover and mutation processes. Although GA initially explored a wide solution space, the small population size (10 individuals) caused premature convergence in suboptimal regions. In contrast, the DFFNN-GWO model displayed a smooth exponential decline in MSE from an initial value of 14 to 0.00255 by the final iteration. This dynamic reflects an adaptive balance between exploration and exploitation through the linearly decreasing parameter a (2 → 0). The α–β–δ leadership mechanism facilitated cooperative learning: α guided global exploration, β maintained directional stability, and δ refined local exploitation steps. This behavior aligns with the findings of Mirjalili [4] and Cinar et al. [6], who reported that GWO maintains greater population diversity and exhibits superior convergence compared to GA.

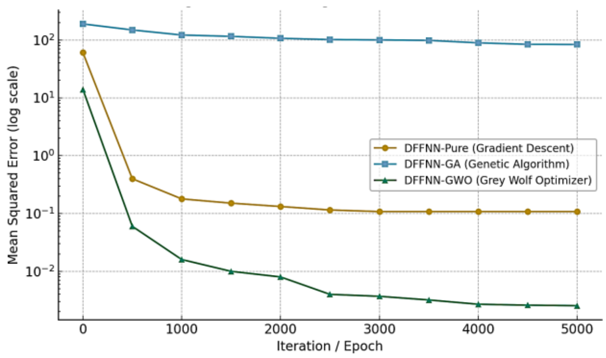

Figure 4. MSE convergence curve of three DFFNN training models

The DFFNN-Pure model produced predictions that closely followed the trend of the actual data, achieving $R^2 = 0.996$, RMSE = 0.296, and MAE = 0.238. In contrast, the DFFNN-GA model exhibited poor performance with $R^2 = -2.26$ and a high error value (RMSE = 9.15), indicating that the evolutionary process failed to identify an effective weight configuration due to excessive exploration and insufficient elitism. Meanwhile, the DFFNN-GWO model demonstrated the best trade-off between accuracy and stability, with $R^2 = 0.971$, RMSE = 0.886, and MAE = 0.760.

Figure 4 illustrates the relationship between the predicted and actual BDV values. The blue points (DFFNN-Pure) lie almost entirely along the ideal diagonal line, indicating high prediction accuracy. The green points (DFFNN-GWO) show slight dispersion but remain close to the ideal line. In contrast, the orange points (DFFNN-GA) deviate significantly, indicating poor agreement between the predicted and actual BDV values.

Table 4. Comparison of predictive performance among the three DFFNN models

Model	R^2	RMSE	MAE
DFFNN-Pure	0.996	0.296	0.238
DFFNN-GA	–2.26	9.149	9.050
DFFNN-GWO	0.971	0.886	0.760

Table 4 quantitatively compares the predictive performance of the three models using R^2, RMSE, and MAE metrics. The results indicate that DFFNN-Pure achieves the highest accuracy, while DFFNN-GWO provides a strong balance between accuracy and stability. Conversely, the poor performance of DFFNN-GA suggests that genetic optimization is less effective for small experimental datasets.

To ensure the statistical reliability of the results, each model was evaluated over 10 independent runs. The average performance and standard deviation are summarized as follows:

1. DFFNN-Pure: $R^2 = 0.996 \pm 0.002$, RMSE = 0.296 ± 0.015, MAE = 0.238 ± 0.012
2. DFFNN-GA: $R^2 = -2.10 \pm 0.45$, RMSE = 9.15 ± 1.20, MAE = 9.05 ± 1.10
3. DFFNN-GWO: $R^2 = 0.971 \pm 0.010$, RMSE = 0.886 ± 0.050, MAE = 0.760 ± 0.040

The results indicate that DFFNN-Pure consistently achieves the highest accuracy with low variance, while DFFNN-GWO demonstrates stable and reliable convergence. In contrast, DFFNN-GA exhibits high variability, indicating unstable optimization performance.

Figure 5 shows the relationship between actual BDV values and predictions obtained from the three DFFNN models. The DFFNN-Pure predictions closely follow the ideal diagonal line, indicating high prediction accuracy. The DFFNN-GWO model exhibits slight dispersion but maintains a strong linear relationship with the actual values. In contrast, the DFFNN-GA predictions deviate significantly, confirming its limited generalization capability under the given dataset conditions.

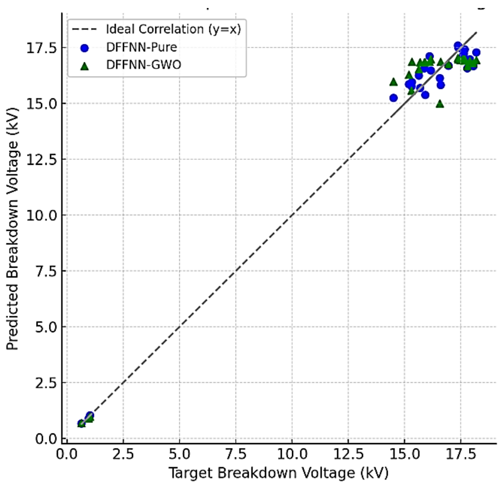

Figure 5. Relationship between actual BDV values and predicted results of the three DFFNN models

The reduction in BDV caused by metallic particles exhibits a highly nonlinear behavior. When voltage is applied, the metallic particles suspended in the oil act as local centers of electric-field intensification. Charge accumulation around these particles increases the field enhancement factor (β-f), thereby reducing the effective breakdown voltage according to the following relation:

$$E_c = \frac{E_O}{1+\beta_f} \quad (19)$$

Such relationships are difficult to represent analytically, making machine learning–based approaches highly relevant. Gradient-based models (DFFNN-Pure) can approximate smooth continuous functions but tend to lose sensitivity to complex cross-interactions among nonlinear parameters. In contrast, GWO optimizes network weights collectively across the entire solution space rather than locally, which explains why DFFNN-GWO produces a more uniform approximation surface and exhibits greater resistance to overfitting.

Furthermore, because GWO does not require derivative information, it is more robust to experimental noise and measurement uncertainty. These findings support the results reported by Liu [13], indicating that population-based metaheuristics are more adaptive to non-convex data structures and small sample sizes than gradient-based optimization methods.

From the overall findings, three key points can be highlighted:

1. Numerical stability - GWO demonstrates stable and predictable convergence behavior, whereas GA tends to oscillate due to its random mutation mechanism.
2. Resilience to small datasets - DFFNN-GWO achieves low error rates with only 36 training samples, indicating efficient global search performance.
3. Practical application - DFFNN-GWO can be implemented as a soft-sensor system to predict the dielectric condition of transformer oil without requiring direct destructive testing, serving as a crucial component of modern Transformer Health Index assessment systems.

Overall, these results indicate that DFFNN-GWO is the most efficient and reliable method for predicting the breakdown voltage of transformer oil contaminated with metallic particles. The model not only excels in optimization stability but also demonstrates strong generalization capability for nonlinear physical phenomena and limited experimental datasets.

From a practical perspective, the proposed model can be implemented as a soft-sensor system for transformer condition monitoring. By integrating measurable input parameters such as temperature, moisture content, and metallic contamination levels obtained from online sensors, the trained DFFNN model can estimate BDV in real time without requiring destructive laboratory testing.

This approach complements existing diagnostic techniques such as Dissolved Gas Analysis (DGA), which focuses on fault detection at the system level. In contrast, the proposed BDV prediction model provides material-level insight into insulation degradation, enabling early detection of dielectric weakening and supporting predictive maintenance strategies in modern smart grid systems.

IV. Conclusion

This study presented a comparative evaluation of gradient-based and metaheuristic training strategies for predicting the Breakdown Voltage (BDV) of transformer oil contaminated with copper and iron particles, addressing the challenge of modeling nonlinear dielectric behavior with limited experimental data. Experimental results demonstrate that the DFFNN-Pure model achieves the highest numerical accuracy under smooth error conditions. In contrast, the DFFNN-GWO model provides the best balance between prediction accuracy, convergence stability, and robustness against nonlinear contamination effects. In contrast, the DFFNN-GA model exhibits unstable convergence and poor generalization, making it unsuitable for BDV prediction using small datasets. Overall, the findings confirm that metaheuristic-based training particularly the Grey Wolf Optimizer effectively solves the BDV prediction problem by enhancing learning stability and reliability, thereby supporting non-destructive condition monitoring and predictive maintenance of power transformers. Despite the promising results, this study has several limitations. The dataset used consists of only 36 samples, which may limit the models' generalization capability. In addition, the evaluation is conducted under controlled laboratory conditions and may not fully represent real-world transformer operating environments. Future work should focus on expanding the dataset, incorporating additional diagnostic parameters such as dissolved gas analysis (DGA), and validating the model using real-time field data to enhance its practical applicability.

References

[1] International Electrotechnical Commission (IEC), "Iec 60156," *61010-1 © Iec2001*, vol. 3, p. 13, 2018.

[2] B. Cao, J. W. Dong, and M. H. Chi, "Electrical breakdown mechanism of transformer oil with water impurity: Molecular dynamics simulations and first-principles calculations," *Crystals*, vol. 11, no. 2, 2021.

[3] L. Cheng *et al.*, "and Breakdown Characteristics of Natural Ester and," 2020.

[4] I. M. Y. Negara, D. Fahmi, D. A. Asfani, I. G. N. Satriyadi Hernanda, M. Wahyudi, and M. N. Ibrahim, "Effect of floating metallic particles in pre-breakdown and breakdown characteristics of oil transformer under dc voltage," *Energies*, vol. 14, no. 12, 2021.

[5] X. Wang and M. Jin, "Convolutional Domain Adaptation Network for Fault Diagnosis of Thermal System under Different Loading Conditions," *Chinese Control Conf. CCC*, vol. 2020–July, no. 3, pp. 4193–4197, 2020.

[6] M. Hussain, M. Siddique, F. H. Malik, Z. M. Haider, and G. A. Hussain, "Machine Learning-Based Prediction of Breakdown Voltage in High-Voltage Transmission Lines Under Ambient Conditions," *Eng*, vol. 7, no. 1, p. 36, 2026.

[7] T. Kamyab, H. Daealhaq, A. M. Ghahfarokhi, F. Beheshtinejad, and E. Salajegheh, "Combination of Genetic Algorithm and Neural Network to Select Facial Features in Face Recognition Technique," *Int. J. Robot. Control Syst.*, vol. 3, no. 1, pp. 50–58, 2023.

[8] M. A. Taha, M. M. Saafan, and S. M. Ayyad, "Revisiting natural selection: evolving dynamic neural networks using genetic algorithms for complex control tasks," *Artif. Intell. Rev.*, vol. 58, no. 11, 2025.

[9] C. Bi *et al.*, "Optimizing a Multi-Layer Perceptron Based on an Improved Gray Wolf Algorithm to Identify Plant Diseases," *Mathematics*, vol. 11, no. 15, pp. 1–36, 2023.

[10] A. I. S. Azis, , Budy Santoso, and J. Jeffry, "Penerapan Grey Wolf Optimizer Dalam Pelatihan Multi Layer Perceptron Untuk Menangani Masalah Klasifikasi Dan Regresi," *Adv. Comput. Syst. Innov. J.*, vol. 2, no. 3, pp. 108–118, 2025.

[11] R. Tiwari, P. S. Agrawal, P. N. Belkhode, J. V. L. Ruatpuia, and S. L. Rokhum, "Hazardous effects of waste transformer oil and its prevention: A review," *Next Sustain.*, vol. 3, no. January, p. 100026, 2024.

[12] A. C. Cinar and N. Natarajan, "An artificial neural network optimized by grey wolf optimizer for prediction of hourly wind speed in Tamil Nadu, India," *Intell. Syst. with Appl.*, vol. 16, no. October, p. 200138, 2022.

[13] Y. Hou, H. Gao, Z. Wang, and C. Du, "Improved Grey Wolf Optimization Algorithm and Application," *Sensors*, vol. 22, no. 10, pp. 1–19, 2022.

[14] K. Y. Lau, M. S. Mohamad, and A. A. Suleiman, "Breakdown characteristics of transformer oil with different contaminants," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 22, no. 5, pp. 2766–2773, 2015.

[15] M. Duval, "Dissolved gas analysis: It can save your transformer," IEEE Electrical Insulation Magazine, vol. 5, no. 6, pp. 22–27, 1989.

[16] S. M. Gubanski et al., "Modern diagnostics of transformer insulation," CIGRE Technical Brochure, Paris, France, 2008.

[17] A. Beroual, M. Zahn, A. Badent, K. Kist, A. J. Schwab, H. Yamashita, K. Yamada, and S. M. Gubanski, "Breakdown voltage of insulating liquids," IEEE Electrical Insulation Magazine, vol. 14, no. 2, pp. 6–15, 1998.

[18] A. A. Suleiman, N. Bashir, and M. K. Hassan, "Effect of moisture and particles on transformer oil breakdown voltage," Energies, vol. 14, no. 5, pp. 1–15, 2021.

[19] I. M. Y. Negara, G. D. Prenata, and M. Fahmi, "Effect of floating metallic particles on breakdown characteristics of transformer oil," Energies, vol. 14, no. 3, pp. 1–12, 2021.

[20] L. Cheng, X. Wang, and Y. Li, "Breakdown characteristics of insulating oil with metallic particles," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 4, pp. 1254–1261, 2020.

[21] A. N. N'cho, M. Fofana, and I. Fofana, "Influence of conductive particles on dielectric breakdown of insulating oil," Electric Power Systems Research, vol. 146, pp. 175–182, 2017.

[22] S. K. Ghosh, A. Das, and S. Dalai, "Influence of metallic contamination on transformer insulation performance," IET Generation, Transmission & Distribution, vol. 12, no. 10, pp. 2405–2412, 2018.

[23] S. A. Khan, M. A. Khan, and A. Ahmad, "Prediction of transformer oil breakdown voltage using artificial neural networks," Energy Reports, vol. 6, pp. 325–332, 2020.

[24] A. S. Aljohani, M. A. Khan, and A. M. Alghamdi, "Machine learning-based prediction of transformer oil condition," IEEE Access, vol. 9, pp. 123456–123465, 2021.

[25] M. Rafiq, S. Riaz, and M. A. Khan, "Artificial intelligence-based condition monitoring of power transformers," IEEE Access, vol. 8, pp. 12345–12356, 2020.

[26] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, vol. 69, pp. 46–61, 2014.

[27] X. S. Yang, Nature-Inspired Metaheuristic Algorithms, 2nd ed. Oxford, U.K.: Elsevier, 2010.

[28] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proc. IEEE Int. Conf. Neural Networks (ICNN), Perth, Australia, 1995, pp. 1942–1948.