



A Hybrid Machine Learning-Genetic Algorithm (ML-GA) to Predict Power Transformer Life

Rosena Shintabella^{1*} Catur Edi Widodo² Adi Wibowo³

^{1*}Master Program of Information System, School of Postgraduate Studies, Universitas Diponegoro, Indonesia

²Department of Physics, Faculty of Science and Mathematics, Universitas Diponegoro, Indonesia

³Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Indonesia

Email* : rosenasbella@gmail.com

Article details:

Published: 31th December 2023



Abstract

Power transformers are important elements of the electrical system, and it is important to estimate their remaining age with accuracy in order to plan maintenance efficiently. This research suggests an innovative method to improve power transformer life predictions by combining Machine Learning (ML) With Genetic Algorithms (GA). The method uses parameters

like temperature, current load conditions, and operational parameters like transformer oil for training engine learning models on time series data. Genetic algorithms are used to develop and choose the most relevant features for powerful predictions, therefore optimizing the model's performance and feature selection. The purpose of the ML-GA hybrid method is to solve the weaknesses of conventional methods while providing improved accuracy and adaptability to a changing operational environment. The experiments that suggest the efficiency of the suggested method on the real-world dataset are presented in this article. A comparison with current approaches demonstrates the benefits of the ML-GA Hybrid model in power transformer life prediction. According to MSE and RMSE, the resultant error level is actually close to 0, indicating that the ML-GA prediction model is fairly accurate. This innovative frame provides utilities and businesses useful tools to maximize maintenance plans, save downtime, and improve the life of power transformers, all of which improve the general reliability of the electrical power system.

Keywords: Genetic Algorithm, Machine Learning, ML-GA, Transformer Life.

To cite this article :

Rosena Shintabella, Catur Edi Widodo, A. W. (2023). A Hybrid Machine Learning-Genetic Algorithm (ML-GA) to Predict Power Transformer Life. *International Journal of Health, Education & Social (IJHES)*, 6(12), 26–38.



Introduction

The energy type that most affects a nation's economy is electrical. Electrical energy is significant in all industrial sectors because it is used in practically all activities (Panthi, 2021). Electrical energy is also significant to the housing and education industries. As a result, it is essential that electrical energy be distributed without any disruptions that could cost a nation financially (Huang et al., 2023). However, disruptions frequently occur in the electrical energy distribution network, hindering the flow of energy and costing the industrial sector money. The power transformer is one component of instruments that is essential to the process of distributing electrical energy from power plant to consumers (Alqudsi & El-Hag, 2019; Zhao et al., 2022).

Transformers have an important role in distributing electrical energy from power plant to consumers (Alqudsi & El-Hag, 2019). Therefore, the transformer is the most important electrical equipment, so it is hoped that it can be used longer and has a long life. The main factor that affects the performance and life of a transformer is aging (Zhao et al., 2022). Transformer aging is a decrease in the ability of an insulator due to heat (Huang et al., 2023). Excessive temperatures occur due to excessive load caused by increased load demand. This results in heat that is too high, resulting in decomposition which can accelerate the aging of the transformer. However, this can be minimized by using transformer life predictions that refer to electrical load demand.

To solve the transformer life prediction problem, conventional methods and artificial intelligence can be used. Machine learning is a type of artificial intelligence that allows computers to learn by training using large amounts of data without needing to program them explicitly (Alimi et al., 2020). Machine learning algorithms are capable of processing and studying data which can then make determinations or predictions. Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Logistic Regression (LR) are types of machine learning that can be used to build prediction models.

However, when the SVM prediction results were evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), SVM showed poor results (Mohamed et al., 2023; Suganthi et al., 2021). KNN produces an accuracy rate of 86% in the case of predicting electrical



energy transmission system disruptions (Omran & El Houbay, 2020). The KNN prediction results were evaluated using RMSE, the results were better than SVM, but the accuracy value of KNN was still below SVM (Srividya et al., 2018). Logistic regression produces high accuracy values and low error rates for power generation system prediction cases. However, if logistic regression is combined with other methods, the level of accuracy also increases (Vinayagam et al., 2022).

For the purpose of feature selection, a metaheuristic algorithm is required in order to optimize machine learning performance. The most important features for prediction experiences have been identified with heuristic algorithms like genetic algorithms, that can also optimize the process of prediction as a whole (Delavar et al., 2019). In the field of machine learning for low-cost air quality sensor calibration, a concept of a genetic algorithm that works for the best combination of models and features has been proposed as well (Pyrih et al., 2019). The application of genetic algorithms in the feature selection process is expected to improve machine learning performance and provide prediction results with a good level of accuracy.

The integration of Machine Learning (ML) and Genetic Algorithms (GA) is a powerful approach that combines the learning capabilities of ML with the optimization strategies inspired by natural evolution in GA. Using Machine Learning (ML) in conjunction with Genetic Algorithms (GA) for prediction tasks can be a powerful approach to optimize models and improve their accuracy (Yi et al., 2023). Applying ML-GA for prediction tasks enables an automated and effective way to search for optimal configurations, improving the overall performance of machine learning models. In the case of transformer life prediction, there are several parameters that must be optimized for the success of the prediction model. ML-GA is a method that can solve case prediction problems with multiclass parameters (Lok et al., 2023). So this research will propose a model for predicting the life of electrical transformers using a machine learning approach that is optimized using a genetic algorithm in the feature selection process to produce an optimal prediction model.



Methods

This research uses transformer current and temperature data in the electricity distribution sector in Indonesia. The exchange rate is from 1 July 2022 to 30 June 2023. From this date range, 365 records were obtained in the time series which are presented in Figure 1.

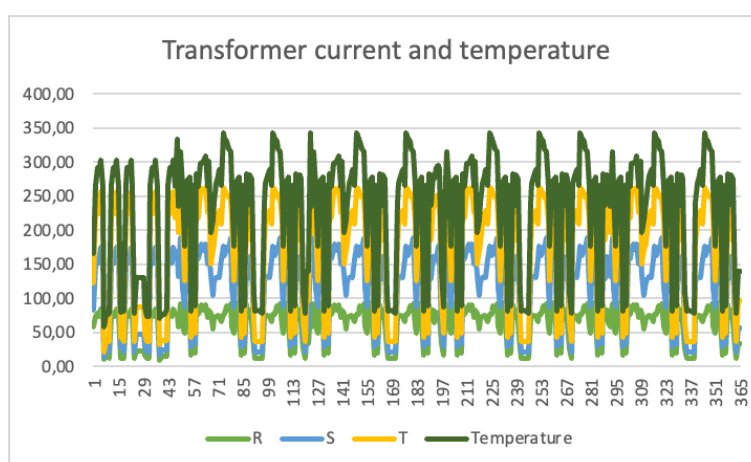


Figure 1. Transformer current and temperature data

In data processing it is divided into parts which fall into the separating data category. The stages of separating data include data exploration, data labeling and data cleaning. Data exploration is one of the initial stages in data preprocessing which aims to understand data, identify patterns, and gain initial insights before carrying out further analysis or building models. Data exploration helps identify problems or opportunities that affect transformer life reduction and to make decisions. After data exploration, the next step is the data labeling process, namely the process of adding labels or categories to the electricity distribution system disturbance data to identify or classify any data disturbances. The data cleaning stage aims to identify, correct or delete errors, discrepancies and anomalies in the data. In this research, data cleaning was carried out by deleting data that was not used in the training model. The overall research process is shown in Figure 2, starting from data processing, application of algorithms and evaluation.

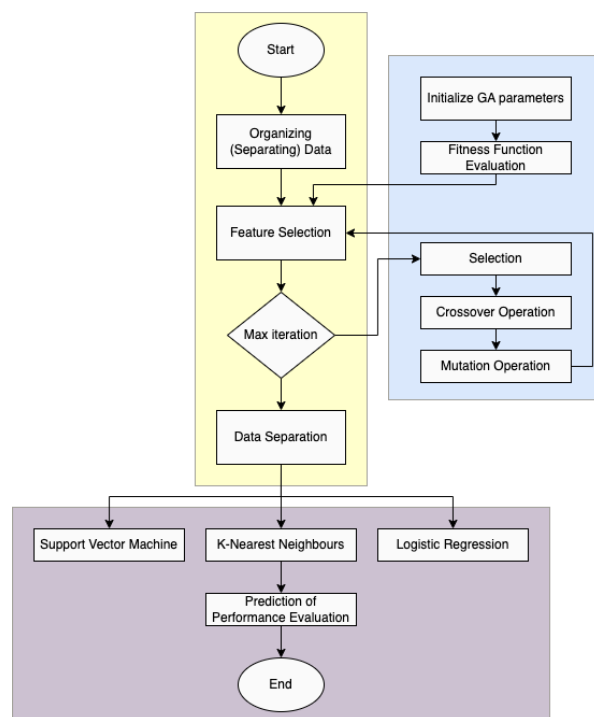


Figure 2. Research process

Genetic Algorithms (GA) are used to optimize the process of feature selection with the aim of obtaining relevant features that influence classification results. Feature selection is used to reduce data dimensions and irrelevant features before entering the machine learning training process. The application of the algorithm in feature selection starts by running a function of random population initialize. The next process runs fitness function which will produce the best parent genes (highest accuracy). Selection is the process for find best parental genes will occur depending on the n-parent parameters. After the selection process, it will be included through the crossover and mutation function in each gene. Crossover are created by combining genes from the two strongest parents obtained at random. Mutation are achieved by inserting randomly selected genes into the offspring of the cross. A new generation is created by selecting the most suitable parents from the previous generation and applying crossover and mutation, this process is repeated for n number of generations. After getting the best results from GA process, the next process is the machine learning model training process.



The machine learning algorithms that will be used in this research are Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Logistic Regression (LR). The model that will be built is an individual machine learning algorithm model and combined with GA. SVM-GA is a combination of supervised learning methods that produce mappings of input and output functions from a set of training data. KNN-GA is a method for classifying objects based on learning data that is closest to the object. Learning data is projected into a multidimensional space, where each dimension represents a feature of the data. LR-GA is a classification method that extends the concept of logistic regression to solve the problem of probability outcomes from multiclasss with one or more independent variables.

In the machine learning field, model evaluation is important to ensure that the machine learning model has good performance. There are several matrices that are used to measure the performance of machine learning models and compare them with other machine learning models. A matrix is a measure or indicator used to measure the performance of a machine learning model. Matrices are used to measure how well a machine learning model fulfills its objectives, such as predicting targets with accuracy or separating data into classes correctly. Matrices can measure how well a machine learning model predicts targets or classifies data. This can determine whether a machine learning model is working well or needs to be improved. Matrices that are often used in machine learning evaluations for predictions include accuracy, precision, recall, f1-score, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

Accuracy is the ratio of correct predictions (positive and negative) to the overall data. Accuracy measures the percentage of correct predictions out of all predictions (Alfebi & Anasanti, 2023; Ariyanti et al., 2023). Accuracy values range between 0 and 1, with 1 indicating a perfect prediction and 0 indicating a completely incorrect prediction. In the accuracy formula TP = true positive, TN = true negative, FP = false positive and FN = false negative.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Precision is the ratio of true positive predictions compared to the overall positive predicted results. Precision measures the ratio of correct positive predictions out of all positive predictions.



Precision provides information about how often a machine learning model makes correct positive predictions (Alfebi & Anasanti, 2023; Ariyanti et al., 2023). The precision value ranges between 0 and 1. In the precision formula $P = \text{true positive}$ and $FP = \text{false positive}$.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall is the ratio of true positive predictions compared to the total number of true positive data. Recall measures the ratio of true positive data found to all positive data (Alfebi & Anasanti, 2023; Ariyanti et al., 2023). Recall provides information about how well the machine learning model found all positive data. in the precision formula $TP = \text{true positive}$ and $FN = \text{false negative}$.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1 score is a measure of balance between precision and recall, with a high value indicating a good balance between the two metrics. The best F1-Score value is 1.0 and the worst value is 0 (Alfebi & Anasanti, 2023; Ariyanti et al., 2023). In representation, if the F1-Score has a good score, it indicates that our classification model has good precision and recall.

$$\text{F1 Score} = 2 \left(\frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{precision}} \right) \quad (3)$$

Mean Squared Error (MSE) is used to measure the average square of the difference between the prediction and the target. MSE can also be used to estimate the inaccuracy of the target machine model (Suganthi et al., 2021). in the MSE formula $y' = \text{predicted value}$, $y = \text{actual value}$ and $n = \text{data}$.

$$\text{MSE} = \sum \frac{(y'-y)^2}{n} \quad (4)$$

Root Mean Square Error (RMSE) is the sum of the squared errors or the difference between the actual value and the predetermined predicted value. RMSE evaluation is calculated based on the average value of the sum of the squared errors between the forecast results and the actual value. The value is used to measure the amount of error produced by the prediction model (Suganthi et al., 2021). A prediction model approaches the actual value variation if the RMSE calculation results are low. in the RMSE formula $y' = \text{predicted value}$, $y = \text{actual value}$ and $n = \text{data}$.



$$\text{RMSE} = \sqrt{\sum \frac{(y' - y)^2}{n}} \quad (5)$$

Results and Discussion

This research uses the Python programming language, where the input data comes from transformer current and temperature data in the electricity distribution sector in Indonesia. Predictions in this research using machine learning algorithm such as SVM, KNN and LR combined with a metaheuristic algorithm, namely genetic algorithm. The results of machine learning evaluation will be taken into concern in comparing results of individual ML and ML-GA. In table 4 is the result of a comparison of individual machine learning evaluations.

Table 1. Result of a comparison of individual machine learning evaluations

	Accuracy	Precision	Recall	F1 Score	Average
SVM	0.90	0.91	0.90	0.90	0.90
KNN	0.92	0.91	0.92	0.91	0.91
LR	0.82	0.77	0.82	0.79	0.80

The results presented in table 1 the average findings for accuracy, precision, recall, and f1 score indicate that the KNN method is better than SVM and LR. Therefore, without applying GA to hybridize, a machine learning method that uses learning data that is closest to the object to classify objects produces the most accurate results.

The three selected models show different results when the model is then used to predict the age of an electrical transformer using GA, which will produce an estimate of the age of the transformer so that it can work according to its capacity. The LR-GA model proved to be the most suitable among the three. The optimized output prediction must at most produce an evaluation average of 100% because it will not meet realistic assumptions when calculating the number of disturbances that cause the life of the transformer to be reduced. Another thing to consider is that the output used in this study does not include things that do not affect the life of the transformer.



Temperature and R, S, and T phase current are important variables in preserving transformer quality. In addition, because unsuitable transformer oil can end in temperature instability, it is important to consider the oil's quality.

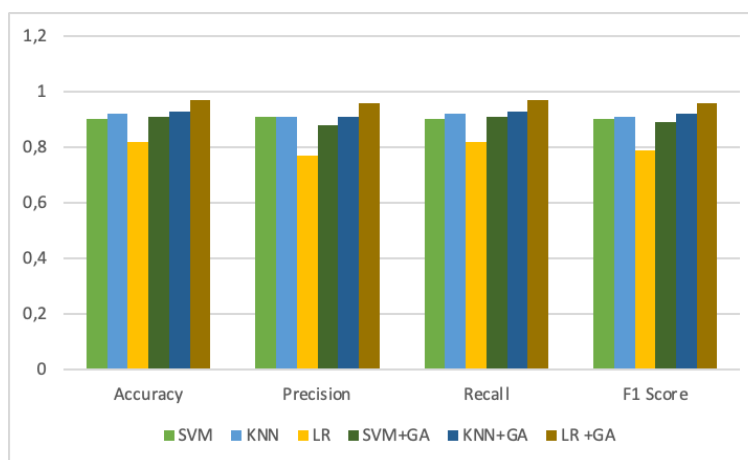


Figure 3. Comparison of evaluation individual ML and ML-GA

When setting up the GA, the following settings were applied with the population size of the genetic algorithm being 50, and the best 64 parents were selected for each run, with random crossovers among individual input columns and a mutation rate of 25%. As can be seen in Figure 3, when using a combination of the SVM-GA, KNN-GA and LR-GA algorithms it shows that the average evaluation achieves the best results. The results of the combination of the three machine learning models are better than the results of the individual models. In this context, the ML-GA model meets the expectations of this work when viewed from the values of accuracy, precision, recall and f1 score.

In the case of predictions, a further evaluation is required to determine the accuracy of the implemented prediction model. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are important metrics in the context of prediction tasks, especially in regression analysis. Both metrics provide insights into the accuracy and goodness of fit of a predictive model. MSE



and RMSE values are considered good if they are close to 0. Figure 4 is the result of a comparison of MSE and RMSE values in research on electrical transformer life prediction.

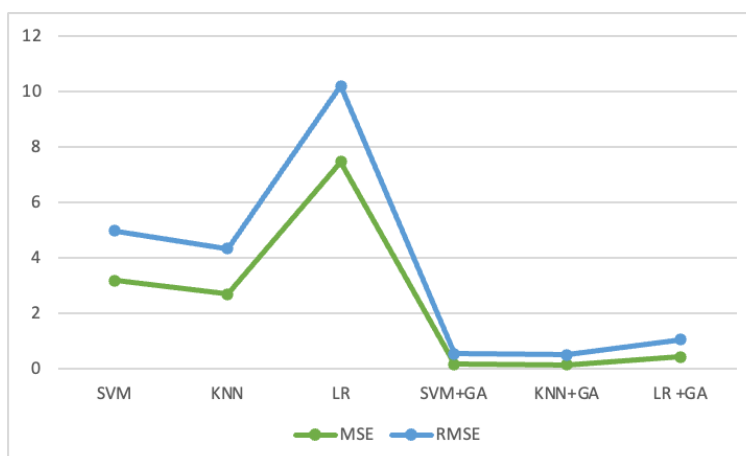


Figure 4. Comparison of MSE and RMSE values for individual ML and ML-GA

There are 6 prediction models built with 3 individual ML models and 3 ML-GA models. The ML-GA model is proven to be the most suitable compared to individual models with an average value close to 0. Optimized output predictions are more accurate because there are minimal errors. The individual ML model value is far from 0, which means the method is not very accurate for the electrical transformer life prediction model. Figure 5 is an illustration of the comparison of the original value with the predicted results using ML-GA. in the table randomly using the comparison results of 25 data from a total of 365 data. The results do not have a significant difference because the model that was built is quite accurate with an average evaluation of close to 1 with an error value close to 0.

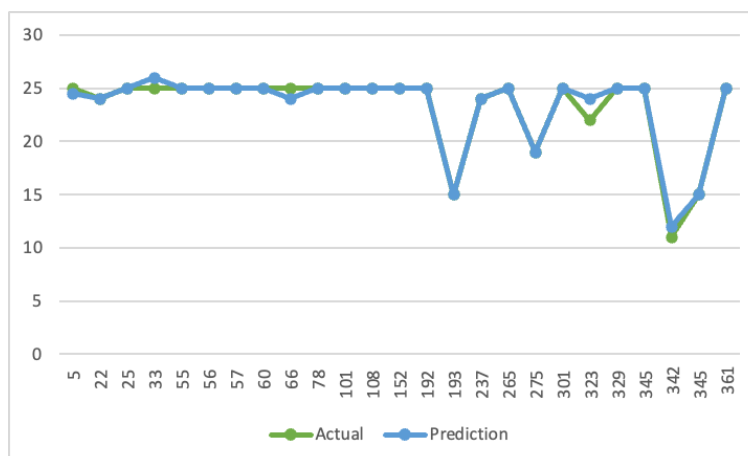


Figure 5. Comparison of actual and predicted ML-GA values

Conclusion and Recommendation

In conclusion, the combination of machine learning and genetic algorithms (hybridization) presents an exciting approach to intricate problems by providing an adaptable and versatile framework that enhances the effectiveness and performance of machine learning models. ML-GA combined can provide meaningful results with an average increase of 16.5% in electrical transformer life prediction. ML-GA is a suggested method for prediction issues since it is also quite accurate in the prediction scenario, as shown by the reporting of MSE and RMSE values that are near to 0. Future developments in this field are anticipated to improve and broaden the applicability of hybrid ML-GA techniques across a range of subjects.



References

- Alfebi, F. H., & Anasanti, M. D. (2023). *Improving Cardiovascular Disease Prediction by Integrating Imputation, Imbalance Resampling, and Feature Selection Techniques into Machine Learning Model*. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 17(1), 55. <https://doi.org/10.22146/ijccs.80214>
- Alimi, O. A., Ouahada, K., & Abu-Mahfouz, A. M. (2020). *A Review of Machine Learning Approaches to Power System Security and Stability*. In *IEEE Access* (Vol. 8, pp. 113512–113531). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2020.3003568>
- Alqudsi, A., & El-Hag, A. (2019). *Application of machine learning in transformer health index prediction*. *Energies*, 12(14). <https://doi.org/10.3390/en12142694>
- Ariyanti, A. P., Mazdadi, M. I., Farmadi, A., Muliadi, M., & Herteno, R. (2023). *Application of Extreme Learning Machine Method With Particle Swarm Optimization to Classify of Heart Disease*. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 17(3), 281. <https://doi.org/10.22146/ijccs.86291>
- Delavar, M. R., Gholami, A., Shiran, G. R., Rashidi, Y., Nakhaeizadeh, G. R., Fedra, K., & Afshar, S. H. (2019). *A novel method for improving air pollution prediction based on machine learning approaches: A case study applied to the capital city of Tehran*. *ISPRS International Journal of Geo-Information*, 8(2). <https://doi.org/10.3390/ijgi8020099>
- Huang, W., Sun, Y., Zhou, Y., Cheng, X., Xiang, Z., Shao, C., Hu, B., & Xie, K. (2023). *A hybrid ensemble learning approach for the aging-dependent reliability prediction of power transformers*. *Energy Reports*, 9, 219–230. <https://doi.org/10.1016/j.egy.2023.04.046>
- Lok, J. Y., Tsai, W. H., & Cheng, I. C. (2023). *A hybrid machine learning-genetic algorithm (ML-GA) model to predict optimal process parameters of nanoporous Cu for CO₂ reduction*. *Materials Today Energy*, 36. <https://doi.org/10.1016/j.mtener.2023.101352>
- Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V., & Hussain, A. (2023). *A hybrid mental health prediction model using Support Vector Machine, Multilayer*



- Perceptron, and Random Forest algorithms. Healthcare Analytics, 3.*
<https://doi.org/10.1016/j.health.2023.100185>
- Omran, S., & El Houbay, E. M. F. (2020). *Prediction of electrical power disturbances using machine learning techniques. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2987–3003.* <https://doi.org/10.1007/s12652-019-01440-w>
- Panthi, M. (2021). *Identification of disturbances in power system and DDoS attacks using machine learning. IOP Conference Series: Materials Science and Engineering, 1022(1).* <https://doi.org/10.1088/1757-899X/1022/1/012096>
- Pyrih, Y., Tchaikovskiy, I., Kaidan, M., & Pleskanka, M. (2019). *Research of Genetic Algorithms for Increasing the Efficiency of Data Routing. IEEE. 978-1-7281-2399-8/19*
- Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). *Behavioral Modeling for Mental Health using Machine Learning Algorithms. Journal of Medical Systems, 42(5).* <https://doi.org/10.1007/s10916-018-0934-5>
- Suganthi, S. T., Vinayagam, A., Veerasamy, V., Deepa, A., Abouhawwash, M., & Thirumeni, M. (2021). *Detection and classification of multiple power quality disturbances in Microgrid network using probabilistic based intelligent classifier. Sustainable Energy Technologies and Assessments, 47.* <https://doi.org/10.1016/j.seta.2021.101470>
- Vinayagam, A., Veerasamy, V., Tariq, M., & Aziz, A. (2022). *Heterogeneous learning method of ensemble classifiers for identification and classification of power quality events and fault transients in wind power integrated microgrid. Sustainable Energy, Grids and Networks, 31.* <https://doi.org/10.1016/j.segan.2022.100752>
- Yi, Y. K., Anis, M., Jang, K., & Kim, Y. J. (2023). *Application of machine learning (ML) and genetic algorithm (GA) to optimize window wing wall design for natural ventilation. Journal of Building Engineering, 68.* <https://doi.org/10.1016/j.jobbe.2023.106218>
- Zhao, L., Zhu, Y., & Zhao, T. (2022). *Deep Learning-Based Remaining Useful Life Prediction Method with Transformer Module and Random Forest. Mathematics, 10(16).* <https://doi.org/10.3390/math10162921>