

=====

**INTELLIGENT MACHINE LEARNING APPROACH FOR ELECTRICITY  
LOSS ANALYSIS AND FORECASTING IN POWER TRANSMISSION  
NETWORKS**

*DSc (Phys.-Math.), Prof. Uteuliev N.U*

**Bazarbaeva U.X.**

*Master's student*

*Nukus State Technical University*

**ARTICLE  
INFORMATION**

**ABSTRACT:**

**ARTICLE HISTORY:**

*Received: 13.04.2026*

*Revised: 14.04.2026*

*Accepted: 15.04.2026*

**KEYWORDS:**

*electricity loss;  
power transmission  
networks; machine  
learning; random  
forest; cascade model;  
regression and  
classification; IQR;  
adaptive algorithms;  
energy systems;  
interpretable artificial  
intelligence*

*This study presents an intelligent machine learning framework for electricity loss analysis in power transmission networks. The proposed approach combines a two-stage cascade prediction algorithm with an adaptive post-processing mechanism based on interquartile range (IQR) statistics. The regression model estimates electricity losses, while the classification model identifies their causes using the predicted loss value as an additional feature. An adaptive post-processing algorithm ensures consistency with physical constraints and generates interpretable explanations for each prediction. The proposed framework improves prediction accuracy, provides reliability estimation, and enhances model transparency, making it suitable for practical applications in energy systems.*

Electricity losses in transmission networks represent a significant technical and economic problem for modern power systems. Reducing these losses is essential for improving the efficiency of electricity distribution and ensuring reliable grid operation. Traditional

approaches to modeling electricity losses rely mainly on physical equations describing electrical processes in conductors. Although these models are useful, they often fail to capture complex nonlinear relationships and temporal patterns that occur in real operational data. With the development of data-driven technologies, machine learning methods have become an effective tool for modeling complex energy systems. However, in most existing studies regression models for predicting the magnitude of losses and classification models for identifying their causes are developed independently. As a result, the classification model cannot use information about the predicted amount of electricity loss, even though the magnitude of losses is often directly related to their cause.

To overcome this limitation, this research proposes a new machine learning architecture based on a cascade structure. In the proposed approach, the predicted electricity loss obtained from a regression model is used as an additional input feature for the classification model. This mechanism allows information to flow between the two models and improves the accuracy of identifying the causes of electricity losses. In addition to the cascade prediction mechanism, an adaptive post-processing algorithm is introduced. The algorithm automatically calculates statistical boundaries using interquartile range statistics and generates natural language explanations describing the reasons for the predicted losses. Such interpretability is particularly important for practical decision-making in energy systems.

### 1. Mathematical Model of Electricity Losses

To formally describe the transmission network, the observation object is defined as the  $i$ -th transmission line during the  $t$ -th observation period. Let

$$L = \{l_1, l_2, \dots, l_n\}$$

denote the set of transmission lines, and

$$T = \{t_1, t_2, \dots, t_m\}$$

denote the set of observation periods.

Each observation is represented by a vector of input variables describing operational and environmental conditions. The main variables include:

$E_{gen}$ – generated or input energy (kWh)

$E_{del}$ – delivered energy (kWh)

$L_{km}$ – transmission line length (km)

$Q_{avg}$ – average current load (A)

$T_{amb}$ – ambient temperature (°C)

$H_{rel}$ – relative humidity (%)

$d_w$  – day of week

$M$  – month

$d_y$  – day of year

$id_{line}$  – line identifier

According to the law of conservation of energy, the observed electricity loss can be calculated as

$$E_{lost} = E_{gen} - E_{del}$$

The loss ratio is defined as

$$\rho = \frac{E_{lost}}{E_{gen}}$$

Thermal energy losses in conductors follow Joule’s law:

$$P = I^2R$$

where  $I$  is the current and  $R$  is the resistance of the transmission line.

Combining these relationships with the energy variables leads to the following formulation:

$$E_{lost} = \rho_{spec} \cdot L_{km} \cdot Q_{avg}^2 \cdot \Delta t + \varepsilon$$

Where  $\rho_{spec}$  is the resistivity of the conductor,  $\Delta t$  is the observation period duration,  $\varepsilon$  represents additional losses influenced by environmental conditions such as temperature and humidity.

Traditional physical models describe a static state and do not capture temporal patterns in electricity transmission. However, electricity demand and operational conditions vary over time. To incorporate temporal dynamics, several calendar-based features are introduced:

$$d_w \in \{0, 1, \dots, 6\}$$

$$M \in \{1, 2, \dots, 12\}$$

$$d_y \in \{1, \dots, 365\}$$

These variables allow the model to capture weekly and seasonal patterns in electricity transmission.

The final feature vector used by the machine learning model is expressed as

$$x_{aug} = [x_1, x_2, \dots, x_d]$$

Where  $d = 10 + k$ , and  $k$  represents the number of encoded categorical line identifiers generated by OneHotEncoder.

For statistical analysis of model features, the following statistics are calculated for each numerical variable  $f$ :

$$IQR = Q_3 - Q_1$$

Based on the boxplot rule, adaptive boundaries are defined as

$$T_{high} = Q_3 + \beta \cdot IQR$$

$$T_{extreme} = Q_3 + \gamma \cdot IQR$$

where  $\beta$  and  $\gamma$  are adjustable hyperparameters.

## 2. Two-Stage Cascade Prediction Algorithm

The main contribution of this research is the proposed Two-Stage Cascade Prediction Algorithm (TSCPA), which integrates regression and classification models in a unified framework.

The algorithm consists of two connected pipelines:

$$P_{reg}: x \rightarrow \hat{E}_{lost}$$

$$P_{clf}: \tilde{x}_{aug} \rightarrow y$$

The extended feature vector used by the classification model includes the predicted electricity loss:

$$\tilde{x}_{aug} = [x, \hat{E}_{lost}]$$

This predicted value acts as a bridge feature connecting the regression and classification stages.

In the first stage, electricity losses are predicted using a Random Forest Regressor. Random Forest is an ensemble learning method consisting of multiple decision trees. The regression output is defined as

$$\hat{E}_{lost} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where  $B$  is the number of trees and  $T_b(x)$  is the prediction of the  $b$ -th tree.

The regression model is trained using the Mean Squared Error loss function:

$$MSE = \frac{1}{N_{val}} \sum_{i=1}^{N_{val}} (y_i - \hat{y}_i)^2$$

In the second stage, a Random Forest classifier determines the probable cause of electricity loss. The classifier outputs a probability vector

$$p(y|x) = [p_1, p_2, \dots, p_k]$$

The final classification decision is determined by majority voting among all trees.

To estimate prediction reliability, a new metric based on tree variance is proposed. Individual tree predictions are represented as

$$\hat{E}_b(x)$$

The variance of predictions is

$$\sigma_{\text{pred}}^2 = \text{Var}(\hat{E}_b(x))$$

The reliability function is defined as

$$\text{reliability}(x) = \frac{1}{1 + \frac{\sigma_{\text{pred}}}{\sigma_{\text{train}}}}$$

and the final confidence score is

$$\text{con } f_{\text{reg}} = 50 + 50 \cdot \text{reliability}(x)$$

If the predictions of all trees are similar, the variance approaches zero and the reliability approaches one, indicating a highly reliable prediction.

### 3. Adaptive Post-Processing and Interpretation

Machine learning models may sometimes produce predictions that contradict known physical constraints. To address this issue, this study proposes an adaptive post-processing algorithm called IQR-Adaptive Domain-Informed Postprocessing (IADIP). During training, statistical characteristics of each feature are stored, including the median, quartiles, and interquartile range. These statistics are used to define adaptive thresholds for different operational conditions.

The loss ratio defined earlier

$$\rho = \frac{E_{\text{lost}}}{E_{\text{gen}}}$$

is used as the main metric for identifying abnormal conditions. Depending on its value relative to adaptive thresholds, the loss can be classified as normal, high, or extreme. Another innovation of the algorithm is the automatic generation of explanations for each prediction. The influence score of feature  $f$  for observation  $j$  is calculated as

$$\text{score}(f, j) = \text{importance}_{\text{clf}}(f) \cdot \frac{|x_{f,j} - \text{med}_f|}{\text{IQR}_f}$$

The two features with the highest scores are selected as the main factors influencing the prediction. Based on these values, the system generates an explanation describing the cause of electricity loss in natural language. For example, the generated explanation may include the predicted cause of the loss, the estimated loss value, the predicted loss ratio, and the model confidence level. Such explanations increase the transparency of the model and make it easier for energy specialists to interpret the results.

The proposed framework therefore provides not only accurate predictions but also interpretable insights into the factors contributing to electricity losses.

This study proposed an intelligent machine learning framework for analyzing electricity losses in power transmission networks. The framework combines a two-stage cascade prediction algorithm with an adaptive post-processing mechanism based on statistical analysis. The cascade architecture integrates regression and classification models by transferring predicted electricity loss as an additional feature for the classification stage. This approach improves the accuracy of identifying the causes of electricity losses, particularly in cases involving abnormal or extreme loss values. A new reliability estimation method based on the variance of predictions in Random Forest ensembles was also introduced. This metric allows quantitative assessment of prediction confidence. Furthermore, the adaptive post-processing algorithm automatically determines statistical thresholds using IQR statistics and generates human-readable explanations for model predictions.

The proposed approach demonstrates that combining physical knowledge, statistical analysis, and machine learning techniques can significantly improve the analysis and interpretation of electricity losses in modern power transmission systems.

## **References**

1. Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
2. Jerome H. Friedman, Trevor Hastie, Robert Tibshirani, *The Elements of Statistical Learning*, Springer, 2009.
3. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2019.
4. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
5. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016.
6. Scott Lundberg, Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *NeurIPS*, 2017.
7. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *KDD*, 2016.
8. International Energy Agency, *World Energy Outlook*, 2022.
9. World Bank, "Reducing Technical and Non-Technical Losses in Power Systems," 2020.

- 
10. IEEE, "IEEE Guide for Electric Power Distribution Reliability Indices," IEEE Std 1366, 2012.
  11. Zhang, Y., Wang, J., and Li, X., "Electricity Loss Analysis Using Machine Learning Techniques," Energy Reports, vol. 8, pp. 1023–1035, 2022.
  12. Wang, H., Chen, Z., and Liu, Y., "Data-Driven Methods for Power Loss Prediction in Smart Grids," IEEE Access, vol. 8, pp. 124567–124578, 2020.
  13. Gholami, A., Shekari, T., and Aminifar, F., "Energy Loss Reduction in Power Systems Using Data Analytics," Electric Power Systems Research, vol. 189, 2020.
  14. Chicco, G., Napoli, R., and Piglione, F., "Distribution System Loss Analysis: A Review," Electric Power Systems Research, vol. 81, no. 12, pp. 2298–2310, 2011.
  15. Mutanen, A., Ruska, M., and Repo, S., "Customer Classification and Load Profiling," IEEE Transactions on Power Systems, vol. 26, no. 3, pp. 1755–1763, 2011.
  16. Ahmad, T., Chen, H., "Short-Term Load Forecasting Using Machine Learning Methods," Energy, vol. 102, pp. 42–52, 2016.
  17. Fan, S., Hyndman, R. J., "Short-Term Load Forecasting Based on Temperature Effects," Journal of Business & Economic Statistics, vol. 30, no. 3, pp. 456–464, 2012.
  18. Box, G. E. P., Cox, D. R., "An Analysis of Transformations," Journal of the Royal Statistical Society, 1964.
  19. Tukey, J. W., Exploratory Data Analysis, Addison-Wesley, 1977.
  20. Montgomery, D. C., Introduction to Statistical Quality Control, Wiley, 2019.