Research Article

Predictive Analysis of Monthly Electricity Consumption Using Rule-Based and Machine Learning Models

Fatma Yaprakdal[1], Serkan Balli[2]

Abstract— **Accurate monthly electricity consumption (EC) forecasting is essential for power providers to allocate resources efficiently, develop reasonable sales plans, and support the creation of reliable smart grids and precise demand-side management policies. Given that factors such as climate, population, and economic conditions can significantly impact EC, it is crucial to consider a wide range of variables in medium-term EC forecasts. This paper addresses a gap in the existing literature by evaluating the performance of the M5 rule model—a relatively underutilized technique—in comparison with popular machine learning (ML) models like Random Forest (RF) and Support Vector Machine (SVM). The motivation for selecting the M5 rule regression technique stems from its effective feature selection process, which is simpler and more straightforward than the complex feature selection methods employed by other models. Using an aggregated dataset from the Czech Transmission System Operator, the study applies these three regression techniques independently to forecast monthly EC. The results demonstrate that the M5 rule regression model outperforms both SVM and RF models for monthly forecasts, achieving an impressive correlation coefficient (R^2) value of 0.9063, compared to 0,8915 for SVM and 0,8598 for RF. The M5 rule model also identifies air temperature, relative humidity, and clear sky surface irradiance as the most influential features in predicting EC.**

Index Terms—**Monthly EC forecasting, M5 rule, SVM, RF, tree-based regression.**

I. INTRODUCTION

Various studies have been carried out, such as operational management, planning, and forecasting in power systems by applying artificial intelligence technologies to energy data [1]. Accurate forecasting of EC is essential to ensure the safe and stable operation of the ever-smartening grid, which predicts how much electricity will be consumed in the short-term, medium-term, and long-term. Forecasts of EC in the short-term period are used to predict a few hours to a few days ahead, medium-term EC forecasts are used to predict a week to a few months ahead, and long-term EC forecasts are used to predict a year ahead and more [2]. These forecasting horizons vary depending on the applications implemented in power systems [3]. In particular, the medium-term forecast is crucial for regulating and operating the power system, planning power plant generation, and formulating the energy marketing strategy [4].

However, population growth, government policies, technological progress, climate change, and the diversification of EC patterns lead to new characteristics and trends in medium-term electricity demand, making it challenging to predict EC accurately [5]. Therefore, many researchers are currently working to develop more accurate forecasting models. The literature on short-term EC forecasting is comprehensive and covers several decades. In comparison, less attention has been paid to medium-term EC forecasting [6].

The EC forecasting models have been developed using various statistical techniques, ML, and deep learning (DL) methods [7]. DL models are most effective when working with large datasets, while ML algorithms generally perform well with smaller datasets. Since monthly EC data only capture one data point per month, they are typically lower in volume than daily and hourly EC data. Tree-based models (TBMs) are particularly effective when dealing with null data, surpassing most other ML approaches in this scenario. One significant advantage of TBMs over linear models is their capability to handle outliers. When the features do not exhibit a linear relationship with the variable of interest, regression trees tend to outperform linear models. These robust models have proven effective in accommodating complex datasets [8]. TBMs use decision trees (DTs) to classify or calculate target variable values for regression or classification tasks. The fundamental tree-based model is the DT. The M5 rule method is powerful for addressing regression and classification problems using DT and model trees. During the DT construction, the M5 rule algorithm strategically selects features to be partitioned at each node based on reducing the variance of the variable of interest within the formed subsets. As a result, more informative features for predicting the target variable are more likely to be chosen for partitioning at higher levels of the tree. This process can be seen as a form of implicit feature selection, prioritizing useful features for prediction, which presents a significant advantage over other tree-based regression methods [9]. RF regression is an ensemble method that involves constructing multiple DTs and then averaging their predictions to improve accuracy and decrease overfitting. It is more resilient to overfitting when compared to a single DT and can effectively

[1] **Fatma Yaprakdal**, is with Department of Information Systems Engineering Mehmet Akif Ersoy University, Burdur, Türkiye, (fatmayaprakdal@mehmetakif.edu.tr). https://orcid.org/0000-0003-0623-1881

[2] **Serkan Ballı**, is with Department of Software Engineering Mehmet Akif Ersoy Universit, Burdur, Türkiye, (serkanballi@mehmetakif.edu.tr). https://orcid.org/0000-0002-4825-139X

handle a large number of features. Furthermore, it provides feature importance scores, which are valuable for understanding the most influential factors impacting the load. Nevertheless, it can be computationally intensive, particularly when utilizing many trees. For medium- to long-term load forecasting, this method is ideal as it considers multiple factors, including weather, time of year, and economic indicators that influence EC [10]. Another powerful ML model is the SVM, which is used to address predictive challenges. Initially, it was used for classification and pattern recognition. However, after the introduction of Vapnik's e-insensitive loss function, SVMs have been extended to solve nonlinear regression estimation problems. This extension has made SVMs highly effective in time series forecasting [11]. In summary, TBMs such as the M5 rule, RF, and SVM each offer distinct advantages for EC forecasting, with their strengths. The primary motivation behind using the M5 rule lies in its ability to perform implicit feature selection and its strength in handling non-linear relationships, offering a simpler yet effective approach compared to more complex ML techniques.

Novelty of this study stems from the application of the M5 rule algorithm to monthly EC forecasting, an area that is underrepresented in the literature. While traditional forecasting studies have relied on daily or hourly EC data, monthly data, though limited in volume, can present unique challenges and opportunities for prediction. This study explores explicitly how the M5 rule can handle lower-frequency data and efficiently predict future consumption trends based on historical patterns and climatic parameters. By conducting a thorough comparative analysis between the M5 rule, RF, and SVM, this work demonstrates the superior performance of the M5 rule in medium-term forecasting using monthly EC data and climatic parameters from the Czech Republic (2010-2022).

Furthermore, this study uses correlation maps to evaluate the relationships within the Czech Republic EC data across years and months and to assess the correlations of the dataset's parameters, including EC and all temperature-related parameters, with each other. The study analyzes the dataset using the widely embraced Minitab statistical software package, which is common in industrial and academic research communities. Additionally, open-source software such as Weka and Python, known for their flexibility and popularity, are used for implementing the ML models. Key contributions and innovations of this paper include:

1) Novel application of the M5 rule model: This paper is among the first to apply the M5 rule regression model to monthly EC forecasting, addressing a significant gap in the literature where tree-based models, particularly the M5 rule, have not been widely used for medium-term EC prediction.
2) Comparative evaluation: The study provides a systematic comparison of the M5 rule with widely-used ML models, RF, and SVM, using a dataset of monthly EC data. It shows how the M5 rule outperforms both RF and SVM in terms of forecasting accuracy as measured by various metrics consisting of mean absolute error (MAE), correlation coefficient (R2), relative absolute error (RAE), root mean squared error (RMSE), and root relative squared error (RRSE).
3) Feature selection advantage: The study highlights how the M5 rule's implicit feature selection process contributes to

better forecasting performance by prioritizing the most influential features for EC prediction, such as air temperature and humidity, in contrast to the more complex feature selection techniques used by other models.
4) Practical implications for energy companies: By accurately forecasting monthly EC and identifying key influencing factors, this study provides valuable insights for power management companies, enabling better planning for power generation, supply, and distribution. The findings suggest that M5 rule-based forecasting models can play a crucial role in optimizing grid operations, avoiding overproduction, and managing supply shortages.

The structure of this paper is organized as follows: Section 2 reviews related works and highlights the distinctions between these existing studies and the proposed approach. Section 3 demonstrates how the M5 rule addresses limitations found in the current literature. Section 4 details the methodology for monthly EC forecasting, supported by a flowchart. Section 5 gives the results of the various monthly EC forecasting models and discusses them in detail. Section 6 provides the concluding remarks of the paper.

II. RELATED WORK

The field of medium-term power forecasting has undergone extensive study, with various approaches developed to tackle this complex issue. These include similarity-based methods, traditional statistical techniques, and advanced ML-based models [12-14].

Recent advancements have incorporated ML models due to their effectiveness in capturing complex patterns. Recent literature has demonstrated the great potential of neural networks (NNs) in a range of forecasting applications, showcasing their ability to model highly nonlinear relationships. For example, [15] introduced four artificial neural network (ANN) models for predicting monthly EC, showing that the ANN outperformed the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. Similarly, a recent study [16] has demonstrated the effectiveness of NNs in predicting complex time series data, such as agricultural commodity prices. This study highlighted the ability of a nonlinear autoregressive NN to capture intricate patterns in price forecasts, with improved performance over traditional methods. These studies illustrate the power of NNs for tackling nonlinearities in forecasting, but also indicate that NN-based models are often better suited for either short-term or long-term forecasting rather than medium-term horizons. In line with this, [17] developed an XLG-LR fusion model that combined extreme gradient boosting (XGBoost), gradient boosting DTs (GBDTs), and light gradient boosting machine (LightGBM). This fusion model demonstrated superior performance over other neural network models like Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Temporal Convolutional Networks (TCN) for short- and medium-term forecasting. However, as [18] highlights, recent advancements in NN architectures still face challenges in achieving precise medium-term forecasts. Specifically, [18] applied nonlinear autoregressive NNs to forecast daily crude oil prices and monthly natural gas and heating oil prices, achieving accurate predictions with low root mean square errors. This underscores the potential of NNs for energy forecasting, but optimizing medium-term forecasts remains an open challenge. In another study, [19] explored the

potential of ML approaches for district-level EC forecasting over medium and long-term periods. This research applied adaptive boosting, multivariate linear regression (LR), and nonlinear autoregressive exogenous multivariable inputs. Additionally, NN-based models have been instrumental in capturing dynamic patterns in energy systems, as evidenced by a study on their application in forecasting complex price trends such as carbon emission allowance prices [20]. By leveraging nonlinear autoregressive NNs, these models have demonstrated accurate and stable performance in predicting price fluctuations, highlighting their broader applicability in energy-related forecasting. Furthermore, a comparison of four ML approaches—adaptive neuro-fuzzy inference system (ANFIS), ANN, fuzzy time series (FTS), and least squares SVMs (LSSVMs)—revealed that FTS performed best in most countries studied [21]. While these models have proven effective for forecasting in particular domains, a more robust solution is needed for medium-term forecasting that can handle both nonlinearities and dynamic changes in energy systems across diverse regions.

Regression models are also powerful tools in this field. For example, LR models have been effective for medium-term and long-term EC prediction [22], with Mean Absolute Percentage Error (MAPE) values consistently below 3.8% for forecasts within a year. [23] developed a multiple LR model using calendar and weather-related variables to forecast monthly electricity demand up to one month in advance. However, despite their effectiveness, LR models struggle with nonlinearity, a challenge that becomes even more apparent in complex energy systems. Recent studies such as [24], which focuses on China's weekly wholesale price index for yellow corn, and [25], which examines the regional steel price index in Northeast China, demonstrate GPR's exceptional ability to model nonlinear patterns, making it particularly valuable in the context of medium-term and long-term energy forecasting. Furthermore, recent work [26] has highlighted GPR's effectiveness in predicting pre-owned housing price indices across major Chinese cities, reinforcing its applicability in capturing complex market dynamics. Despite these advancements, the integration of GPR with other models, especially for medium-term EC forecasting, has not been widely explored. Meanwhile, studies such as [27-28] applied LR, time series models, and SVM to predict daily and monthly EC in Italy. [29] utilized multi-output support vector regression for medium-term EC forecasting in North America and Australia, incorporating a memetic algorithm based on the Firefly algorithm for feature extraction.

These studies emphasize the diverse range of methods available for EC forecasting, with each technique offering specific strengths and addressing particular forecasting scenarios. However, none have fully addressed the need for a unified approach capable of seamlessly integrating multiple forecasting techniques to handle medium-term predictions with high accuracy. The use of TBMs like the M5 rule has also proven highly effective for both regression and classification tasks in forecasting [8]. The M5 rule has demonstrated strong performance across various applications. For instance, [30] compared SVR, multiple LR, J48 pruned tree, and M5 Rules for forecasting marine diesel engine fuel consumption. M5 Rules provided the most accurate predictions, as evidenced by high correlation scores and low error metrics. Similarly, [31] used the M5 rule and other ML methods to predict agricultural residue production, achieving high accuracy. [32] employed M5 algorithms to analyze the effects of elevated temperatures on recycled concrete properties, showcasing its versatility. In short-term EC forecasting, the M5 rule has also shown promise. [33] created an ensemble ML model combining ANN, SVR, and M5 Rules to predict building energy use, effectively forecasting for the next 24 hours. Additionally, [34] compared the M5P algorithm, a modified version of the M5 tree algorithm, with the Bagging and J48 algorithms for day-ahead short-term load forecasting. The results indicated that M5P consistently outperformed both Bagging and J48 algorithms. [35] evaluated day-ahead EC forecasting models using various algorithms, including M5 Rules, which demonstrated superior performance.

While existing research explores various ML and statistical methods for EC forecasting, each approach exhibits specific strengths and limitations. Models such as ANN, XLG-LR fusion, and FTS have achieved success in particular forecasting scenarios but often excel in either short-term or long-term predictions rather than medium-term forecasting. The proposed approach, leveraging the M5 rule, offers a distinct advantage by bridging the gap between statistical and ML methods. Unlike other models, the M5 rule effectively captures complex and non-linear patterns in medium-term EC forecasting, simplifies feature selection, and integrates multiple forecasting techniques into a cohesive framework. This integration allows the M5 rule to provide more accurate and robust forecasts across diverse timeframes and data characteristics, addressing the limitations observed in previous models and offering a more comprehensive solution for medium-term power forecasting.

Despite these advancements, research specifically focusing on the application of the M5 rule for medium-term EC forecasting remains limited. Further investigation is needed to fully understand its potential and benefits in enhancing forecasting accuracy in this domain.

III. METHODOLOGY

This section provides a comprehensive overview of the data analysis process, detailing both the rule-based and ML models used. It also includes a thorough performance evaluation. Fig.1. visually represents the flowchart of the EC forecasting process implemented in this study. The initial step in data processing involves addressing missing values by either imputing them with calculated estimates or removing the corresponding entries. Additionally, any extreme data points that could potentially distort the results are identified and removed from the dataset. The data is resampled from a daily to a monthly format to better suit the forecasting needs. Following data preprocessing, the dataset is divided into two subsets: 75% for training the ML models and 25% for testing their performance. Various ML algorithms are then applied to predict monthly EC based on the preprocessed dataset. The evaluation of model performance involves a comprehensive analysis using multiple metrics, including MAE, R2, RAE, RMSE, and RRSE. These metrics contribute to a detailed understanding of the models' performance and support the creation of comprehensive performance reports. Furthermore, the meticulous identification of the models' strengths and weaknesses and the formulation of actionable improvement recommendations enhance the interpretation and understanding of the performance results.
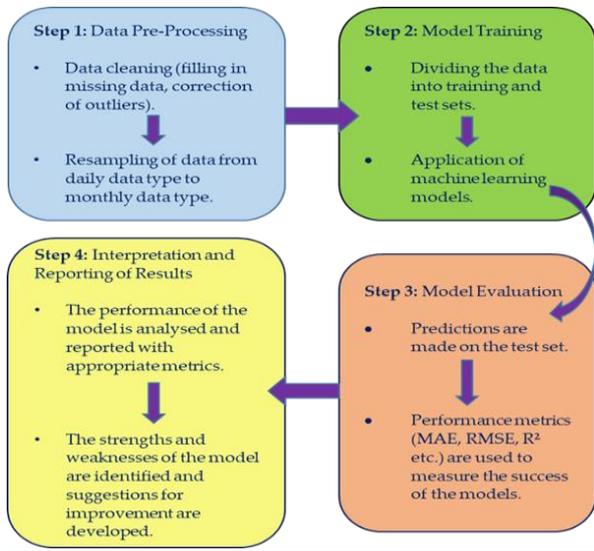
Fig.1. Flowchart of the monthly EC forecasting process

IV. DATA DESCRIPTION AND ANALYSIS

A. Description of Dataset

The dataset used in this study is provided by ČEPS, a joint-stock company and the sole transmission system operator in the Czech Republic. This dataset comprises aggregated EC data for the Czech Republic from 2010 to 2022, initially recorded at an hourly resolution. For the purposes of medium-term forecasting, this dataset is resampled to a monthly resolution. Climatic parameters relevant to the study are sourced from the "The Power Data Access Viewer" website. Table 1 provides a detailed listing and explanation of each dataset parameter used in this study. In this table, Wh/m2 represents watt-hour per square meter, C° represents Celsius degree, % represents percentage, mm/h represents a millimeter per hour, kPa represents kilopascal, m/s represents a meter per second, and MW represents megawatt.

Table I. Output power of a pv system under different irradiance levels.

Parameters	Descriptions in ML models
Date and time information	DateTime
Clear Sky Surface Shortwave Downward Irradiance (Wh/m^2)	CLRSKY_SFC_SW_DWN
Temperature at 2 Meters (C°)	T2M
Dew/Frost Point at 2 Meters (C°)	T2MDEW
Wet Bulb Temperature at 2 Meters (C°)	T2MWET
Relative Humidity at 2 Meters (%)	RH2M
Precipitation Corrected (mm/hour)	PRECTOTCORR
Surface Pressure (kPa)	PS
Wind Speed at 10 Meters (m/s)	WS10M
Wind Direction at 10 Meters (Degrees)	WD10M
Wind Speed at 50 Meters (m/s)	WS50M
Wind Direction at 50 Meters (Degrees)	WD50M
Montly EC (MW)	Load

B. Distribution of the EC data by month

Fig.2. illustrates the distribution of the monthly average EC dataset for the Czech Republic. Analysis of this dataset reveals several key observations:

1) Seasonality of EC: The data demonstrates a pronounced seasonal pattern. High levels of EC are typically observed during the winter months (e.g., January and February) and summer months (e.g., July and August). This seasonal variation is likely due to increased heating demands in winter and cooling demands in summer.

2) Cooling Demand: Despite the general trend of high EC in summer, the dataset shows that the EC in July, for example, is not as elevated as might be expected. This suggests that cooling demand during these summer months may not be as high as anticipated.

3) Lower Consumption in Transitional Seasons: EC levels tend to be lower during the spring and autumn months (e.g., April, May, September, and October). This reduction in consumption during these transitional periods is consistent with decreased heating and cooling needs.
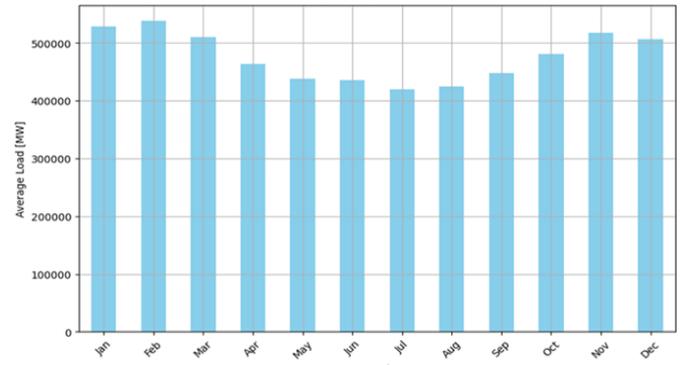

Fig.2. The monthly average Load [MW] in the Czech Republic

C. Correlation Map of EC by Years and Months

Essential insights into the dynamics of EC in the Czech Republic can be derived from the correlation map presented in Fig.3. The map reveals notable fluctuations in EC over the years, which may be attributable to various factors, including changes in energy demand patterns, population growth, technological advancements, and improvements in energy systems. The correlation map indicates that higher EC values are typically observed during the winter and summer months. This seasonal variation is likely due to increased heating demands in winter and cooling demands in summer. Conversely, EC tends to be lower during the spring and autumn months, reflecting reduced heating and cooling needs during these transitional periods. This analysis offers valuable insights for energy forecasting and planning within smart grids. For instance, understanding these seasonal patterns can assist in balancing energy generation and consumption by optimizing generation capacity to match fluctuating energy demands.
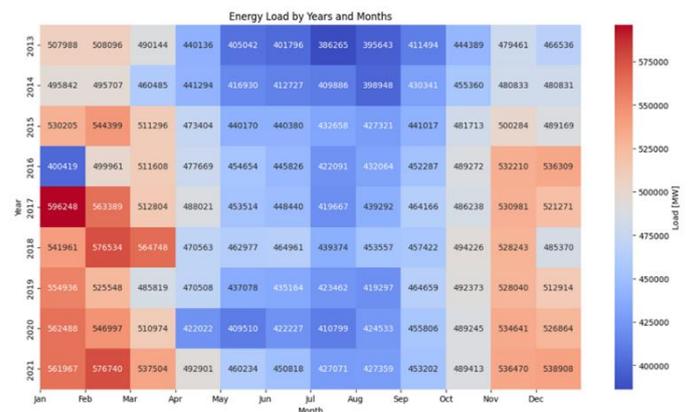

Fig.3. Republic Correlation map of EC values (Load[MW]) by years and months for the Czech Republic

D. Correlation Map Between All Variables

The correlation map in Fig.4. reveals several key relationships among the variables analyzed. There is a strong negative correlation between EC and the wet thermometer temperature,

dew point temperature, and air temperature. This suggests that as temperatures increase, EC generally declines. Conversely, EC exhibits a positive correlation with wind speed at both 10 meters and 50 meters, as well as with relative humidity. This indicates that higher wind speeds and increased humidity are associated with higher EC. Additionally, there is a strong negative correlation between EC and the shortwave downward irradiance of the clear sky surface, suggesting that increased solar radiation correlates with decreased EC. EC shows weaker negative correlations with total precipitation and surface pressure. Examining the correlations among the non-EC variables, we find that the wet thermometer temperature, dew point temperature, and air temperature are strongly correlated with the clear sky surface shortwave downward irradiance. Each of these temperature-related parameters also exhibits strong positive correlations with each other. For example, the air temperature has a strong positive correlation with both the clear sky surface shortwave downward irradiance and the wet thermometer and dew point temperatures. On the other hand, wind speeds at 10 meters and 50 meters, as well as relative humidity, display a strong negative correlation with the clear sky surface shortwave downward irradiance. However, not all parameters exhibit consistent negative correlations with others, and some correlations among the remaining parameters are relatively weak.
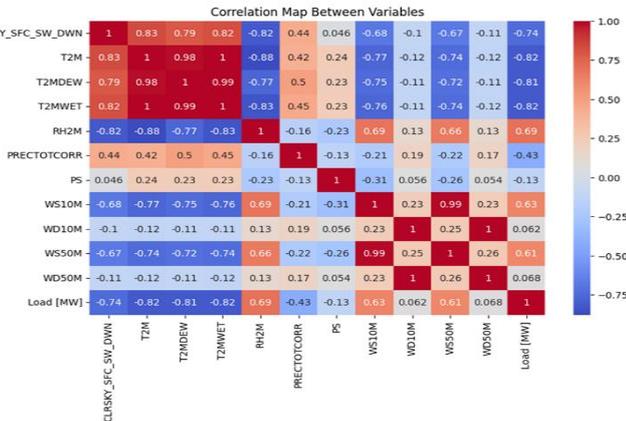

Fig.4. Correlation map between all variables of EC dataset of the Czech Republic

E. Scatter Charts

Fig.5., Fig.6., Fig.7., and Fig.8. present scatter plots that explore the relationships between several key weather variables-humidity, solar radiation, air temperature, wind speed-and EC. The analysis of these scatter plots yields the following observations:

1) *Load [MW] versus T2M (Air Temperature)*:   Fig.5. demonstrates that as air temperature increases, EC generally decreases. This negative correlation becomes more pronounced at higher temperatures.

2) *Load [MW] versus RH2M (Relative Humidity)*: Fig.6. reveals a positive correlation between relative humidity and EC. As relative humidity rises, EC tends to increase as well.

3) *Load [MW] versus WS10M (Wind Speed at 10 Meters)*: In Fig.7., there is a visible correlation between wind speed at 10 meters and EC. However, this relationship is less pronounced compared to the correlations observed with other variables.

4) *Load [MW] versus CLRSKY_SFC_SW_DWN (Clear Sky Surface Solar Radiation)*: In Fig.8. indicates that EC decreases as solar radiation increases, reflecting a strong negative correlation between these two variables.

These scatter plots provide a clear visual representation of how each weather variable influences EC, reinforcing the insights obtained from the correlation analysis.
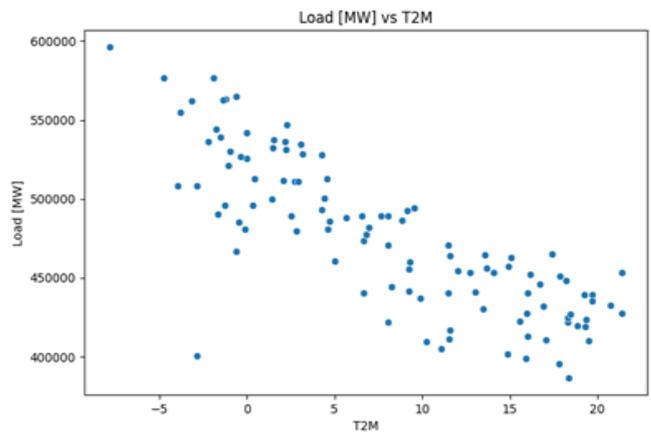

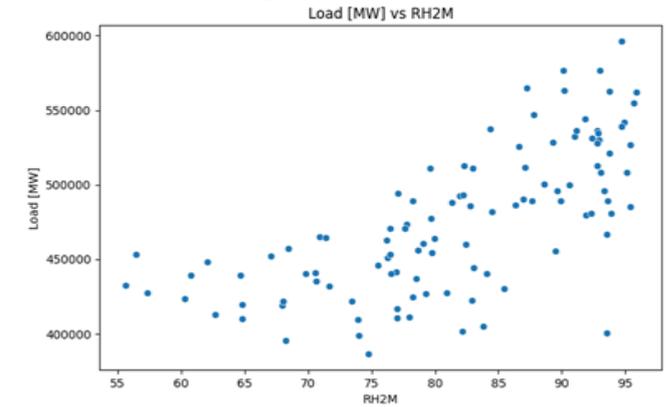Fig.5. Scatter plot of Load[MW] versus T2M

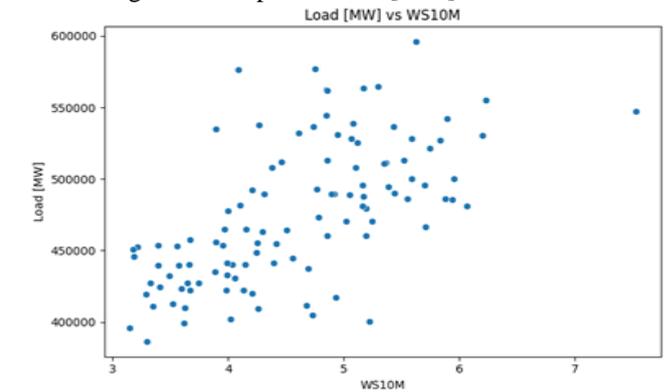Fig.6. Scatter plot of Load[MW] versus RH2M

Fig.7. Scatter plot of Load[MW] versus WS10M

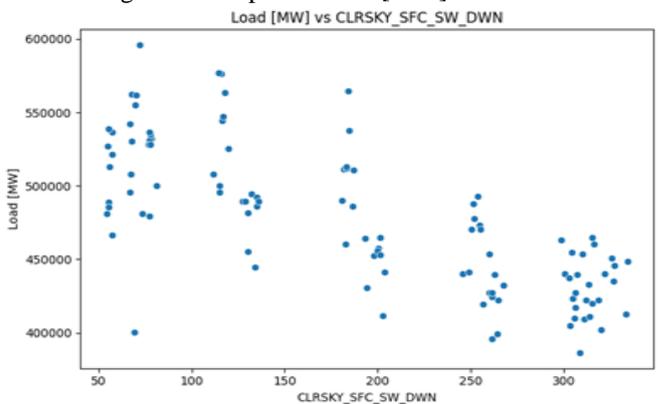Fig.8. Scatter plot of Load[MW] versus CLRSKY_SFC_SW_DWN

V. MODELLING

This section provides a detailed explanation of the mathematical principles and operational logic underpinning the ML techniques utilized in this study. It also describes the methodology for evaluating the accuracy of the models' performance. Data analysis is conducted using the Minitab statistical software package, renowned for its user-friendly

interface, robust analytical tools, and effective graphical representations. Minitab is extensively utilized in both industrial and academic research due to its comprehensive capabilities [36]. In addition to Minitab, open-source software tools such as Weka and Python are employed for implementing the ML models. Weka offers a broad range of ML algorithms and data preprocessing tools within a versatile architecture, facilitating the development of complex data mining procedures with various basic learning tools and algorithms [31]. Python, which has gained considerable popularity for ML analysis, is supported by prominent libraries such as Scikit-Learn, further enhancing its utility in ML applications [37].

A. M5 Rule Model

The DTs are powerful models that use a binary tree structure to make predictions about the value of a dependent variable based on independent variables. In a model tree, each leaf contains a specifically designed LR model to forecast the target variable. The M5 DT algorithm, developed by Quinlan in 1992, constructs a regression sequence by iteratively splitting the sample space using tests on single attributes that maximize the variance in the target space. The process involves calculating the standard deviation reduction (SDR) using the formula [9]:

$$SDR = sd(T) - \sum \frac{|T_i|}{T} sd(T_i) \qquad (1)$$

Here, T represents the set of instances reaching the node, T_i denotes the subset of instances with a possible outcome, and sd indicates the standard deviation.

As the DT is developed, a multiple LR model is established for each internal node by using the specific node data and all the attributes involved in the tests within the sub-tree of that node. During pruning, each subtree is carefully analyzed to address the issue effectively. Pruning begins when the estimated error for the linear model at the root of a subtree matches or falls below the predicted error of the subtree. Smoothing is employed to mitigate abrupt changes between adjacent linear models within the branches of the pruning tree [33].

B. SVM Model

The SVM method, originally developed by Vapnik and which uses statistical learning techniques, is widely used for classification tasks and can also be leveraged as SVRs for regression problems. The primary goal of the SVR method is to minimize prediction error by identifying a function that closely approximates the training dataset [38]. This involves maximizing the flatness of the function and reducing the risk of getting stuck in local minima. The function for approximating the continuous function is represented as:

$$f(x) = \big(w.\phi(x)\big) + b \qquad (2)$$

Here, x denotes the training dataset, w symbolizes the vector, b stands for the intercept, and Φ represents a nonlinear transfer function. The goal is to find the best values for "w" and "b" that minimize the regression risk (Rreg), which is given by:

$$R_{reg}(f) = C \sum_{i=0}^{l} \Psi\big(f(x_i) - Y_i\big) + \frac{1}{2}\|w\|^2 \qquad (3)$$

Assuming a complexity of 0.1, denoted by $\Psi(.)$, the vector w may be transformed over the data points as illustrated by equation 4.

$$w = \sum_{i=1}^{l} (a_i - a_i^*) + \phi(x_i) \qquad (4)$$

By re-expressing Equation (2) using Equation (4), a revised form is obtained in Equation 5:

$$f(x) = \sum_{i=1}^{l} (a_i - a_i^*)\big(\phi(x_i).\phi(x)\big) + b$$
$$= \sum_{i=1}^{l} (a_i - a_i^*)\, k(x_i, x) + b \qquad (5)$$

In Equation 5, the dot product is reformulated as a kernel function (k), ensuring that it occurs within a multidimensional domain using low-dimensional inputs having an uncertain Φ transformation. The study has employed various kernel functions, with the polynomial kernel function delivering the most effective performance.

$$k(x_i, x) = [(x_i * x) + c]^d \qquad (6)$$

The 'c' value affects the model's margin size. A large 'c' value reduces training error but may lead to overfitting. Conversely, a small 'c' value can cause underfitting but will result in higher training errors. An increase in polynomial degree 'd' can result in a higher training error. Adjusting the degree helps control the complexity of the model. Higher degrees lead to more complex models that fit the data better, while lower degrees yield simpler models that are still effective in fitting the data [29].

C. RF Model

In the current study on monthly average EC prediction, the RF is used as a tool for regression. The RF handles regression and classification tasks, utilizing DTs as base models. Its foundation on regression trees, and its advantages over single trees, are leveraged. RF incorporates bagging with a random subspace method to deliver more stable and smoother outcomes. Bagging works by averaging multiple noisy yet nearly unbiased base models, effectively reducing variance. Being inherently noisy and low-biased models, trees are well-suited for bagging, particularly when they are deeply developed. The random subspace method plays a vital role in enhancing diversity among trees by constraining them to operate on varied random subsets of the comprehensive predictor space. This is achieved by selecting a random subset of predictors at each tree node. Moreover, each tree in the forest is constructed from a bootstrap sample of the original dataset, contributing to increased diversity. Incorporating random predictors in the bagged trees' nodes allows us to decorrelate the trees and enhance prediction accuracy, while simultaneously minimizing model variance [39].

For each of K trees, denoted by k = 1, ..., K, a tree T is created by recursively dividing the input space at each node down to a minimum leaf size, using a bootstrap pattern. The data is split at each node based on a random selection of p out of n predictors. The optimal split is identified by maximizing the reduction in the mean squared error (MSE) across all potential splits and cut points. After growing all K trees in this manner, the RF predictor is given by equation (7):

$$\hat{f}_K(x) = \frac{1}{K} \sum_{k=1}^{K} T(x; \Theta_k) \qquad (7)$$

where x represents the input pattern, and Θ_k describes the characteristics of the kth tree in terms of the splitting predictors, cut points, and values of the terminal nodes [40].

D. Performance Evaluation of All Models

The performance of all studied models is evaluated using several error metrics: R^2, MAE, RMSE, RAE, and RRSE.

In regression analysis, the R² value is employed to measure the proportion of variance in the dependent variable that is predictable from the independent variable(s). It effectively quantifies the degree of correlation between the predicted model and the actual observations. R² ranges from 0 to 1, with a value of 1 indicating a perfect fit of the model to the data. A higher R² value signifies a better fit. The formula for calculating R² is provided in Eq. (8).

$$R^2 = \frac{\sum_{i=1}^{n}(y_t - y')^2}{\sum_{i=1}^{n}(y_p - y_t)^2} \qquad (8)$$

where n shows all data points, y_p is the value of the forecast, y_t is the real value, y' is the mean of the independent data.

Equation (9) presents the formula for calculating the MAE, which measures the average absolute difference between the actual and predicted values. A lower MAE indicates more accurate predictions, as it reflects a smaller discrepancy between the observed and forecasted values.

$$MAE = \frac{\sum_{i=1}^{n}|y_p - y_t|}{n} \qquad (9)$$

The RMSE is a widely used statistical measure for evaluating model performance. Equation (10) illustrates the calculation of RMSE. This metric quantifies the square root of the average squared differences between the actual and predicted values. A lower RMSE value signifies a better fit of the model to the data, indicating that the predictions are closer to the observed values [39].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_p - y_t)^2}{n}} \qquad (10)$$

Equation (11) calculates the Relative Absolute Error (RAE). This metric is derived by dividing the sum of the absolute errors by the sum of the absolute differences between the mean of the observed values and the actual values. A lower RAE indicates better model performance, as it reflects smaller relative errors in the predictions compared to the baseline of mean value predictions.

$$RAE = \frac{\sum_{i=1}^{n}|y_p - y_t|}{n\sum_{i=1}^{n}|y' - y_t|} \qquad (11)$$

The Relative Root Mean Squared Error (RRSE) normalizes the sum of squared errors by comparing it to the sum of squared errors from a simple predictor. This normalization helps assess model performance relative to a baseline model. Equation (12) shows how to calculate the RRSE [41].

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_p - y_t)^2}{\sum_{i=1}^{n}(y' - y_t)^2}} \qquad (12)$$

VI. RESULTS

A. Performance Results Of The Models

The performance results for predictions made using the M5 rule, SVM, and RF models are summarized in Table 2. The table presents comparisons across various performance evaluation metrics, including R², MAE, RMSE, RAE, and RRSE.

Table II. Monthly ec forecasting performance evaluation results of all models

Models	Evaluation Metrics				
	R2	MAE score	RMSE	% RAE	% RRSE
M5 Rule	0.91	16772.29	22287.94	43.12	45.74
RF	0.86	21390.68	26658.89	54.99	54.71
SVM	0.90	17477.57	23114.17	44.93	47.44

B. Feature Selection Results of The M5 Rule Model

The proposed M5 rule model builds the DT and selects the features to be partitioned at each node based on the criterion of reducing the variance of the target variable within the subsets generated by the partitioning, as mentioned before. Features that are more informative for predicting the target variable are more likely to be selected for partitions in the upper parts of the tree. At the end of these processes, the proposed model has established the following rule, which is an implicit form of selecting features useful for predicting within this mathematical formula as follows.

$$Load = -134.8188 * CLRSKY_{SFC_{SW_{DWN}}} - 4859.48 * T2 - 942.8733 * RH2M + 617286.7669 \qquad (13)$$

According to this formula, clear sky surface shortwave downward irradiance, temperature at 2 m and relative humidity at 2 m are the features that most influence the estimation of monthly EC among all features in the Czech dataset. This highlights the model's ability to effectively identify key variables that play a critical role in predicting monthly EC.

VII. DISCUSSION

The M5 rule algorithm outperforms all other models based on the given performance metrics as seen in Table 2. For a detailed interpretation of the prediction models according to their performance evaluation metrics, refer to Fig.9.
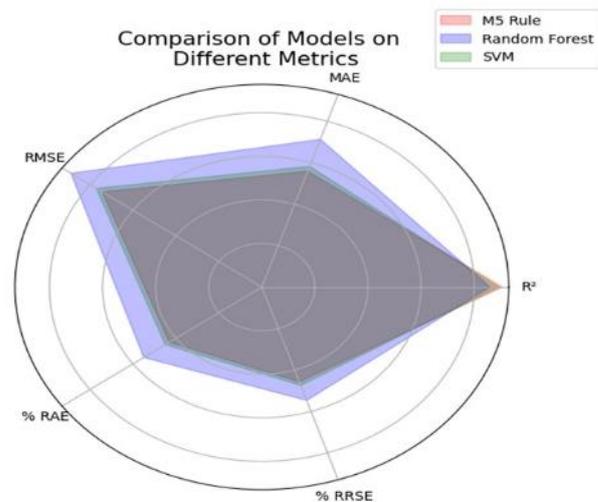

Fig.9. Comparison of all models performed on error metrics

A. Performance Evaluation of Forecasting Models

R² Error Metric: The M5 rule model demonstrates the highest performance in terms of the R² value, indicating the best agreement between predicted and actual values. Specifically, the M5 rule model explains 90.63% of the variance in the dataset, reflecting its strong predictive accuracy. In comparison, the SVM model accounts for 89.15% of the variance, and the RF model explains 85.98% of the variance. This shows that the M5 rule model has a marginally better ability to explain the variability in the data compared to the other models.

MAE and RMSE Error Metrics: The MAE and RMSE metrics measure the accuracy of the predictions, with MAE reflecting the average magnitude of errors and RMSE giving more weight to larger errors. The M5 rule model exhibits the lowest values for both MAE and RMSE, signifying that its predictions are closer to the true values. Conversely, the RF and SVM models show higher error values, with the RF model particularly notable for having the highest RMSE. This reinforces the superior predictive accuracy of the M5 rule model.

% RAE and % RRSE Error Metrics: The RAE and RRSE percentages offer insights into how the model's errors compare to those of a simple predictor. The M5 rule model achieves the lowest values for both RAE and RRSE, indicating relatively smaller prediction errors. In contrast, the RF model exhibits higher error percentages, suggesting less consistent prediction performance. The RF model's higher error percentages suggest less consistent prediction performance, potentially due to its inability to generalize as effectively as the M5 rule.

B. Consistency with Feature Selection and Correlation Analysis

The feature selection results from the M5 rule algorithm align with the correlation map shown in Figure 4. The analysis reveals a positive association between EC and air temperature at a 2-meter height, along with a negative correlation with relative humidity and clear sky surface shortwave downward irradiance. These findings are consistent with the factors identified as most influential in the M5 rule model's monthly EC forecast formula, further validating the model's predictive performance and feature selection process.

C. Advantages of the M5 Rule Model

The superior performance of the M5 rule algorithm compared to other models can be attributed to several factors:

1) *Advanced Tree Model*: The M5 rule integrates both regression and classification rules, enabling it to capture complex interactions and linear relationships effectively.
2) *Handling Complex Data Relationships*: By employing LR models at each terminal node, the M5 rule model excels at modeling intricate relationships in the data.
3) *Robust Data Handling*: The model's robust structure allows it to manage data distribution, anomalies, and noise effectively.
4) *Effective Feature Selection*: The M5 rule model is proficient in selecting and emphasizing the most informative features, which is crucial for its overall success in prediction.

Overall, the M5 rule model's advanced methodology ccontribute to its superior performance in forecasting monthly EC. This combination of strengths enables the M5 rule model to provide reliable and accurate forecasts, even when dealing with complex, multi-variable datasets.

VIII. CONCLUSION

In this study, we focus on the monthly forecasting of EC, a topic that has been relatively underexplored compared to short-term EC forecasts. To address this gap, we apply tree-based regression analysis using ML models, specifically the M5 rule, SVM, and RF, to aggregate EC data from the Czech Republic over 12 years, with monthly resolution. Among the models evaluated, the M5 rule algorithm demonstrates the best performance across several robust error metrics, including R^2, MAE, RMSE, % RAE, and % RRSE. The R^2 metric, which indicates the degree to which the forecasts align with actual values, shows that the M5 rule achieves an R^2 score of approximately 0.91 for monthly forecasts of multi-variable EC. This highlights its effectiveness in capturing the variance in the dataset. The M5 rule algorithm also offers advantages in feature selection. Unlike other ML models that treat feature selection as a separate step - often introducing additional computational complexity and time - the M5 rule integrates feature selection directly within its framework. This not only simplifies the modeling process but also enhances the efficiency of identifying the most significant features affecting predictions. Given these advantages, the M5 rule model is expected to make significant contributions to both academic research and practical applications in energy management, planning, and operational studies of power systems. Its utility in supporting the transition to smart grids is particularly noteworthy.

However, while this study demonstrates the strengths of the M5 rule for medium-term forecasting, there are some limitations that should be considered. One potential limitation is the reliance on data from a single country, which may affect the generalizability of the findings to other regions with different energy consumption patterns and data characteristics. Namely, the homogeneity of economic, climatic, and energy consumption conditions in this dataset may restrict the model's generalizability to regions with different grid compositions, renewable energy integration levels, or consumer behaviors. Furthermore, while the study highlights the effectiveness of the M5 rule in multivariable EC forecasting, the model's performance could vary with the inclusion of additional features, such as socio-economic factors, or data from other domains like weather forecasts. Expanding the dataset to include a broader range of variables could help further assess the model's robustness and versatility in different contexts. Another limitation is the model's focus on monthly predictions, which, while valuable for certain operational and planning applications, may not fully capture the dynamics of daily or weekly energy fluctuations. Future studies could explore the M5 rule's performance in shorter-term forecasting or compare it with other models designed for higher-frequency data. Looking ahead, several potential avenues for future research can further enhance the model's capabilities. Future work could extend this study by applying the M5 rule model to datasets with a greater number of features, such as incorporating weather data, demographic changes, or policy impacts, to evaluate its performance in feature selection and prediction accuracy more comprehensively. Moreover, the development of ensemble models that combine the M5 rule with other advanced techniques, such as GPR or LSTM networks, could offer deeper insights into its strengths and identify potential areas for improvement.

In summary, while this study contributes to the understanding of medium-term EC forecasting using the M5 rule, it also opens the door for further research that explores the model's adaptability to broader datasets, higher-frequency forecasting, and integration with other forecasting techniques. These future efforts will help refine the model's capabilities and enhance its practical applications in energy systems and beyond.

REFERENCES

[1] A. Cabrera, L. G. B. Ruiz, D. Criado-Ramón, C. D. Barranco, and M. C. Pegalajar, Application of fuzzy and conventional forecasting techniques to predict energy consumption in buildings, *Int. J. Intell. Syst.*, vol. 2023, 2023, Art. no. 4391555, pp. 1–12.

[2] M. Jawad *et al.*, Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed, *J. Eng.*, 2018, pp. 721–729.

[3] S.-M. Jung, S. Park, S.-W. Jung, and E. Hwang, Monthly electric load forecasting using transfer learning for smart cities, *Sustainability*, vol. 12, no. 15, p. 6364, 2020.

[4] S. R. Khuntia, J. L. Rueda, and M. A. van der Meijden, Forecasting the load of electrical power systems in mid- and long-term horizons: A review, *IET Gener. Transm. Distrib.*, vol. 10, no. 15, pp. 3971–3977, 2016.

[5] T. Gao, D. Niu, Z. Ji, and L. Sun, Mid-term electricity demand forecasting using improved variational mode decomposition and extreme learning machine optimized by sparrow search algorithm, *Energy*, vol. 261, p. 125328, 2022.

[6] N. B. Behmiri, C. Fezzi, and F. Ravazzolo, Incorporating air temperature into mid-term electricity load forecasting models using time-series regressions and neural networks, *Energy*, vol. 278, p. 127831, 2023.

[7] L. Di Persio and N. Fraccarolo, Energy consumption forecasts by gradient boosting regression trees, *Mathematics*, vol. 11, no. 4, p. 1068, 2023.

[8] A. Mystakidis *et al.*, Energy forecasting: A comprehensive review of techniques and technologies, *Energies*, vol. 17, no. 6, p. 1662, 2024.

[9] R. Srivastava, A. N. Tiwari, and V. K. Giri, Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India, *Heliyon*, vol. 5, no. 11, Art. no. e02692, 2019.

[10] P. Geurts, D. Ernst, and L. Wehenkel, Extremely randomized trees, *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[11] H. Wei-Chiang, Electric load forecasting by support vector model, *Appl. Math. Model.*, vol. 33, no. 5, pp. 2444–2454, 2009.

[12] P. Bunnoon, K. Chalermyanont, and C. Limsakul, Mid term load forecasting of the country using statistical methodology: Case study in Thailand, in *Proc. Int. Conf. Signal Process. Syst.*, vol. 1, 2009, pp. 924–928.

[13] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer, 2008.

[14] D. Oliveira, E. Meira, C. Oliveira, and F. Luiz, Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods, *Energy*, vol. 144, pp. 776–788, 2018.

[15] C. Hamzaçebi, H. A. Es, and R. Çakmak, Forecasting of Turkey's monthly electricity demand by seasonal artificial neural network, *Neural Comput. Appl.*, vol. 31, pp. 2217–2231, 2019.

[16] B. Jin and X. Xu, Wholesale price forecasts of green grams using the neural network, *Asian J. Econ. Bank.*, 2024.

[17] Q. Ji *et al.*, Short- and medium-term power demand forecasting with multiple factors based on multi-model fusion, *Mathematics*, vol. 10, p. 2148, 2022.

[18] B. Jin and X. Xu, Price forecasting through neural networks for crude oil, heating oil, and natural gas, *Meas.: Energy*, vol. 1, no. 1, p. 100001, 2024.

[19] T. Ahmad and H. Chen, Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment, *Energy*, vol. 160, pp. 1008–1020, 2018.

[20] B. Jin and X. Xu, Carbon emission allowance price forecasting for China Guangdong carbon emission exchange via the neural network, *Global Finance Rev.*, vol. 6, no. 1, p. 3491, 2024.

[21] M. H. L. Lee *et al.*, A comparative study of forecasting electricity consumption using machine learning models, *Mathematics*, vol. 10, p. 1329, 2022.

[22] E. Barakat, Modeling of nonstationary time-series data. Part II. Dynamic periodic trends, *Int. J. Electr. Power Energy Syst.*, vol. 23, pp. 63–68, 2001.

[23] F. Apadula, A. Bassini, A. Elli, and S. Scapin, Relationships between meteorological variables and monthly electricity demand, *Appl. Energy*, vol. 98, pp. 346–356, 2012.

[24] B. Jin and X. Xu, Forecasting wholesale prices of yellow corn through the Gaussian process regression, *Neural Comput. Appl.*, vol. 36, no. 15, pp. 8693–8710, 2024.

[25] B. Jin and X. Xu, Predictions of steel price indices through machine learning for the regional northeast Chinese market, *Neural Comput. Appl.*, vol. 36, no. 33, pp. 20863–20882, 2024.

[26] B. Jin and X. Xu, Pre-owned housing price index forecasts using Gaussian process regressions, *J. Model. Manag.*, 2024.

[27] M. De Felice, A. Alessandri, and F. Catalano, Seasonal climate forecasts for medium-term electricity demand forecasting, *Appl. Energy*, vol. 137, pp. 435–444, 2015.

[28] Z. Hu, Y. Bao, R. Chiong, and T. Xiong, Mid-term interval load forecasting using multioutput support vector regression

with a memetic algorithm for feature selection, *Energy*, vol. 84, pp. 419–431, 2015.

[29] O. Yuksel, M. Bayraktar, and M. Sokukcu, Comparative study of machine learning techniques to predict fuel consumption of a marine diesel engine, *Ocean Eng.*, vol. 286, p. 115505, 2023.

[30] B. Singh and A. K. Jana, Forecast of agri-residues generation from rice, wheat and oilseed crops in India using machine learning techniques, *Environ. Res.*, vol. 245, p. 117993, 2024.

[31] A. Srivastav *et al.*, Predictive analysis of recycled concrete properties at elevated temperatures using M5 pruned rule classifiers, *Asian J. Civ. Eng.*, vol. 25, pp. 2623–2640, 2024.

[32] N. T. Ngo *et al.*, An ensemble machine learning model for enhancing the prediction accuracy of energy consumption in buildings, *Arab. J. Sci. Eng.*, vol. 47, pp. 4105–4117, 2022.

[33] A. K. Srivastava *et al.*, A day-ahead short-term load forecasting using M5P machine learning algorithm along with elitist genetic algorithm and random forest-based hybrid feature selection, *Energies*, vol. 16, p. 867, 2023.

[34] M. Lydia and G. Edwin Prem Kumar, Soft computing models for forecasting day-ahead energy consumption, *Mater. Today Proc.*, vol. 58, pp. 473–477, 2022.

[35] S. A. Lesik, *Applied Statistical Inference with MINITAB*, 2nd ed. U.S.: Chapman and Hall, 2018.

[36] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*, 2nd ed. UK: Packt Publishing, 2017.

[37] A. S. Ahmad *et al.*, A review on applications of ANN and SVM for building electrical energy consumption forecasting, *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, 2014.

[38] L. Lang, L. Tiancai, A. Shan, and T. Xiangyan, An improved random forest algorithm and its application to wind pressure prediction, *Int. J. Intell. Syst.*, vol. 36, no. 8, pp. 4016–4032, 2021.

[39] G. Dudek, A comprehensive study of random forest for short-term load forecasting, *Energies*, vol. 15, p. 7547, 2022.

[40] E. Elbasi, A. E. Topcu, and S. Mathew, Prediction of COVID-19 risk in public areas using IoT and machine learning, *Electronics*, vol. 10, p. 1677, 2021.

BIOGRAPHIES

Fatma Yaprakdal Fatma Yaprakdal obtained her Bachelor's degree in Electrical Engineering in 2009 from Yildiz Technical University (YTÜ), followed by an M.Sc. degree in Energy Institute/ Istanbul Technical University in 2012, and a Ph.D. degree in Electrical Engineering in 2020 from YTÜ. She currently serves as an Assistant Professor at the Information Systems Engineering department of Mehmet Akif Ersoy University, Burdur. Her research focuses on Power Systems, Artificial Intelligence, Machine Learning and intelligent systems.

Serkan Ballı Serkan Ballı obtained his Bachelor's degree in Statistics and Computer Sciences from Muğla Sıtkı Koçman University in 2001. He completed his M.Sc. degree in Statistics and Computer Sciences at the same university in 2005. He received his Ph.D. degree in Computer Engineering from Ege University in 2010, with a dissertation on the design and implementation of hybrid intelligent decision support systems. He currently serves as a Professor at the Software Engineering Department of Mehmet Akif Ersoy University, Burdur. His research focuses on Artificial Intelligence, Machine Learning, intelligent systems, cybersecurity solutions, and advanced information systems, with a particular interest in leveraging AI technologies to solve real-world problems.