

## Short-Term Load Forecasting Based on Bayesian Ridge Regression Coupled with an Optimal Feature Selection Technique

Abdullah Ashraf\* and Shaikh Saaqib Haroon<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan

<sup>2</sup>Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan

\*(abdullah.ashraf@students.uettaxila.edu.pk)

(Received: 5 May 2023, Accepted: 30 May 2023)

(DOI: 10.59287/ijanser.787)

(1st International Conference on Contemporary Academic Research ICCAR 2023, May 17-19, 2023)

**ATIF/REFERENCE:** Ashraf, A.. & Haroon, S. S. (2023) Short-Term Load Forecasting Based on Bayesian Ridge Regression Coupled with an Optimal Feature Selection Technique. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(4), 435- 441.

**Abstract** –Load forecasting has been an important aspect in power system operations. The increase in the integration of different renewable energy resources during past decades has made it even more crucial as an accurate load forecast can be highly beneficial for the energy market as well as the ongoing economic dispatch and unit commitment problem. The increased influence of artificial intelligence and machine learning in electrical engineering has also caused an improvement in load forecasts immensely. This study presents a short-term load forecasting methodology using Bayesian Ridge Regression paired up with an optimal feature selection technique which is a combination of Coyote Optimization Algorithm and Quadratic Discriminant Analysis. The test systems used in the study are based on the historical load data obtained from publicly available API offered by PJM data miner 2 and the weather data obtained using Visual Crossing. Before the application of the feature selection technique, the features were engineered by lagging the weather and the load data. The results of this method are compared with multiple state-of-the-art load forecasting methods including Gradient Boosting Regressor, Random Forest Regressor, Ensemble from ElasticNet, and Bagging with Decision Tree. The proposed method proved to be superior as it showed a noticeable decrease in mean absolute percentage error and root-mean-square error.

**Keywords** – Short-term Load Forecasting, Machine Learning, BRR, Feature Selection, COA-QDA

### I. INTRODUCTION

The importance of load forecasting has always been high in the many aspects of the electrical power system including power generation planning, electricity supply design, operations and maintenance of power plants, and the electricity market [1]. The increase in the number of renewable energy resources (RERs) in the power system has

magnified the importance of load forecasting even further because of the many complications brought by them including the expenses of storage as well as their integration [2].

Load forecasting is divided into four different categories depending on the duration of time. Long-term forecasting has a duration of multiple years. Medium-term forecasting ranges from a few weeks

to a few months. Short-term has a range between a few hours to a few weeks while very short-term has a duration of a few seconds to a few minutes [3]. This research is centered on short-term load forecasting.

The solution for the problem of short-term load forecasting can be divided into two methods, conventional and intelligent. Conventional or traditional methods involve statistical studies like auto smoothing and regression [4] while intelligent methods are further divided into artificial neural networks [5], metaheuristic methods [6] as well as hybrid methods which are a combination of metaheuristic methods with statistical techniques or statistical techniques with intelligent methods [5][7]. Artificially intelligent techniques such as these have enabled the system to do multi-input and multi-layer forecasts [8]. The load data obtained from a power system is a time-series dataset with multiple statistical properties and patterns which may not be visible to the human eye but can be used to create accurate forecasting after converting the dataset into a multi-dimensional shape [9].

Hybrid forecasting techniques are proven to be more accurate compared to traditional techniques as presented in [10] where the hybrid model created by combining Gated Recurrent Units (GRU) with Convolutional Neural Networks (CNN) showed improved results compared to both GRU and CNN individually. H. Eskandari [11] used Long Short Term Memory (LSTM) and CNN to create an hourly forecast by having temperature, and season of the year into the dataset. The multi-dimension feature extraction separated this study from previously done short-term forecasting techniques. The resulting model showed superior results compared to some previous methods in the same area. B. Farsi et al. [12] introduced a combination of LSTM and CNN with a minor modification, making it parallel LSTM-CNN Network (PLCNet), the accuracy of the forecasting results was compared with other ML methods as well and it outperforms all of them in both runtimes and mean absolute percentage error (MAPE). A novel hybrid method was introduced by Z. Gao et al. [13] which was named RF-IPWOA-ELM comprised of Random Forest, an improved parallel whale optimization algorithm, and an extreme learning machine. The comparison of the proposed method with generalized regression neural network (GRNN) showed improved accuracy.

In recent years, deep learning, with its rapid development, has become a fairly researched topic in forecasting [14]. This type of machine learning refers to creating stacks of multiple nonlinear network layers which enables the model to create a nonlinear mapping of the data along with feature abstraction. It also relies on stochastic optimization techniques to attain the computation from the dataset.

D. Lu et al. [15] introduced an advanced load forecasting method which is a combination of three ML methods, namely Least Absolute Shrinkage and Selection Operator (LASSO), Bayesian Ridge Regression with Principal Component Analysis (PCA) as a feature selection technique. The proposed method was designed to help the system operator create a day-ahead schedule for the units in the system. The integration of smart metering systems in the power system in recent years has made it possible to get data from residential loads and has made optimization of load shaving as well as demand response of such loads possible through their short-term load forecasting [16][17].

Load forecasting is dependent on multiple external factors including temperature, humidity, UV index, and precipitation [11]. The increase in external factors in turn causes an increase in the number of features that will be responsible for the forecasting of the load. In [18] the number of features was increased by creating lagged data version of the load while in [15], along with the lagged version of the load, the lagged version of some weather parameters was also produced which resulted in an improvement in the forecasting. The increase in the number of features has made it a necessity to have a feature selection technique paired with the forecasting model which can be fast and would be able to determine and choose the most important features from the entire dataset. In [19], P. Matrenin makes use of feature selection methods in time series-based models like XGBoost and SVR for medium load forecasting. The conclusion was that the most over-fitting resilient model was a combination of four linear regressions.

H. Naseri [20], proposed a novel feature selection technique called Coyote Optimization Algorithm-Quadratic Discriminant Analysis (COA-QDA) to minimize the data required to make accurate forecasting. The technique outperformed LASSO used by [15] along with some other commonly used methods.

This study uses BRR paired with COA-QDA, presented in [20] which has shown its superiority compared to LASSO used in [15]. The conclusion showed improved results compared to the other models [15].

## II. MATERIALS AND METHOD

The study follows the underlying steps:

1. Collection of historical load data of 2021 from Pennsylvania, New Jersey, and Maryland Interconnection (PJM) [21].

2. Acquisition of weather conditions and features from Visual Crossing [22].

3. Feature engineering of the load and weather data acquired above to include the lagged load and weather data.

4. Feature selection using COA-QDA.

5. Development and training of the BRR model.

The workflow of this study is presented in Fig. 1.

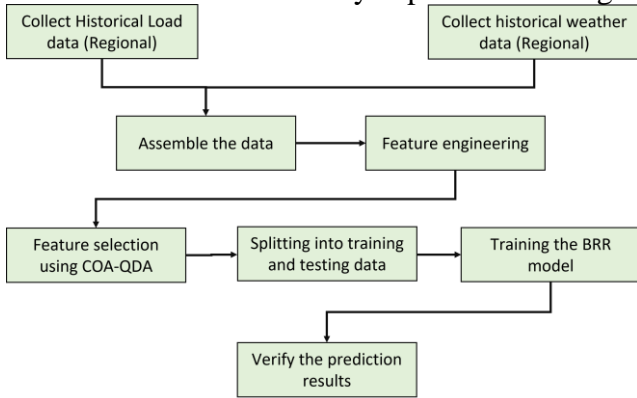


Fig. 1: Workflow of the proposed method

The key aspects of the proposed method are presented in the upcoming sub-section.

### A. COA-QDA

This feature selection technique first introduced in [1] is a hybrid of COA which is a metaheuristic swarm intelligence algorithm [2] coupled with QDA.

A particular number of solution vectors are applied in COA to reach the optimal solution of a problem. Each of these vectors has a single value of the dependent variable of the problem. The independent variables of the vector are represented as the social behavior of the coyote. These independent variables are shown in (1).

$$\text{var}_c^{h, \text{iter}} = v = (v_1, v_2, v_3, \dots, v_N) \quad (1)$$

Here,  $\text{var}$  is the independent variable of coyote  $c$  in the herd  $h$  for the iteration  $\text{iter}$ .  $N$  is the number of dimensions and  $v$  is the value of the variable.

Coyotes have a hierarchical structure in their herd according to their fitness. The coyote with the highest fitness is ranked alpha and thus the solution vector with the highest fitness will be considered the optimal solution. The equation used to calculate the alpha vector is presented in (2).

$$a^{h, \text{iter}} = \left\{ \text{var}_c^{h, \text{iter}} \mid_{c=\{1,2,3,\dots,N\}} \min f(\text{var}_c^{h, \text{iter}}) \right\} \quad (2)$$

Here,  $a$  is the alpha of the herd  $h$ , at iteration  $\text{iter}$ .

QDA is a statistical supervised classification technique that proceeds with the assumption that Gaussian distribution is applicable each class to model the likelihood of each feature in the dataset. The feature vector is assumed to be multivariate in the group with a given mean vector and a specific covariance matrix.

COA-QDA maximizes the accuracy of the prediction by selecting the optimal features of the dataset.

### B. Bayesian Ridge Regression

After the feature selection is complete, load forecasting is achieved using BRR. It is a linear combination of multiple nonlinear kernels with a ridge, which is a penalty for over complicating the model.

The regression part of this technique is to find an appropriate set of weights to express the correspondence between one or more target values with input data, which may contain 1 to  $N$  dimensions of variables. As shown in (3):

$$y(x, w) = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad (3)$$

Here,  $w$  is the weights,  $x$  is the input data, and  $y$  is the target value. Weights are calculated as  $(w_1, w_2, w_3, \dots, w_n)^T$  and input data  $x$  is  $(x_1, x_2, x_3, \dots, x_n)$ . Equation (3) possesses many limitations so an extended version of it is introduced in (4).

$$y'(x, w) = w_0 + w_1 \phi(x_1) + w_2 \phi(x_2) + \dots + w_n \phi(x_n) = w^T \phi(x) \quad (4)$$

Here  $\phi(0)$  is initialized to 1 and  $\phi$  is a nonlinear function which can either be Gaussian or sigmoidal.

The objective error function with the inclusion of the ridge part in BRR is presented in (5).

$$\min \frac{1}{2} \sum_{i=1}^N (t_i - w^T \phi(x_i))^2 + \frac{\lambda}{2} w^T w \quad (5)$$

Here  $w^T w$  is the ridge regression,  $\lambda$  is the gamma distribution chosen during the fitting of the model.

$\sigma_N^2$  is the variance of the model, and is shown in (6) as:

$$\sigma_N^2(x) = \sigma_1^2 + \phi(x)^T S_N \phi(x) \quad (6)$$

Here,  $S_N$  is the covariance for every weight.

In Bayesian optimization techniques, regularization parameters are included in the prediction procedure. These parameters are not hard-coded but are tuned based on the data being worked with.

This can be done by presenting some uninformative and obscure priors over the hyperparameters of the model. The regularization method used in ridge regression is equivalent to finding a maximum posterior Gaussian estimation under a prior over the coefficients with precision. The value of  $\lambda$  is not fixed as well and is treated as a random variable to be estimated based on the data. The output is also assumed to be Gaussian distribution and is again treated as a random variable.

Using BRR has multiple advantages, some of which are its capability to adapt to the data and its flexibility to include regularization parameters. The disadvantage of using BRR is that it is time-consuming.

### III. EXPERIMENTAL SETUP

The simulation of this research was carried out on a free version of Google Colab on a 2VCPU @ 2.2GHz with 13GB RAM which proved to be a viable alternative for deep learning applications to actual hardware [24].

#### A. Data description

The load data used in this study is obtained from PJM for the year 2021. PJM comprises multiple substations and the selected stations for this study are in Table 1. The load data for all the selected substations is shown in Fig. 2.

Table 1. Load Areas in Test Cases

Sr #	Abbreviation	Load Area
1	AP	Allegheny Power Systems
2	DAY	Dayton Power and Light Corporate
3	DEOK	Duke Energy Ohio and Kentucky Corporation
4	EKPC	East Kentucky Power Cooperative

5	PEPCO	Potomac Electric Power Co.
6	RECO	Rockland Electric Co.

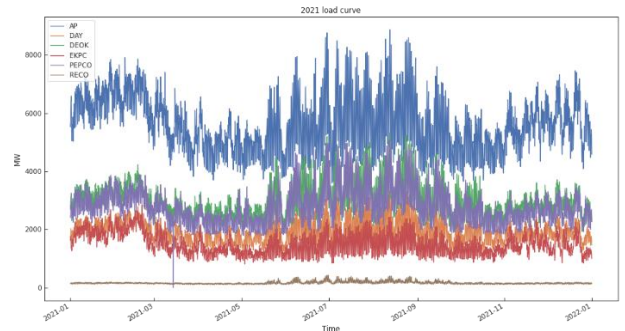


Fig. 2. Load curve for the year 2021

The weather data obtained was obtained from Visual Crossing for specific locations which are part of the load areas of PJM. The data was composed of temperature, feels-like, dew, humidity, precipitation, precipitation probability, snow, snow depth, wind speed, wind direction, sea level pressure, visibility, solar radiation, and UV index.

#### B. Data preprocessing

The time was split into days, months, and weeks and their sine and cosine values were obtained which presented improved results [3]. The 24-hour lagged version of the data was created which increased the number of features from 15 to 408 features.

This feature engineering increased the amount of data and was then passed into the COA-QDA feature selection technique which decreased the number of features to a quarter of the original.

#### C. Evaluation metrics

The results which are generated through this study, along with their comparison with other widely researched studies are evaluated by their root-mean-square error and mean absolute percentage error. MAPE is shown in (7) is the evaluation of regression loss in prediction problems and is represented as:

$$MAPE(y, y_{\text{predicted}}) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{|y_i - y_{\text{predicted}_i}|}{\max(\epsilon, |y_i|)} \quad (7)$$

Here,  $y$  is the test data to be validated by  $y_{\text{predicted}_i}$ ,  $n$  is the number of observations.

RMSE is defined as a measure of the distance between the prediction of a point with its true value presented in (8). To remove the case where the prediction of two points is equal and opposite to

their true values, a square is placed so that the sum may not add up to zero. RMSE is represented as:

$$RMSE(y, y_{\text{predicted}}) = \sqrt{\frac{\sum_{i=0}^n (y_{\text{predicted}_i} - y_i)^2}{n}} \quad (8)$$

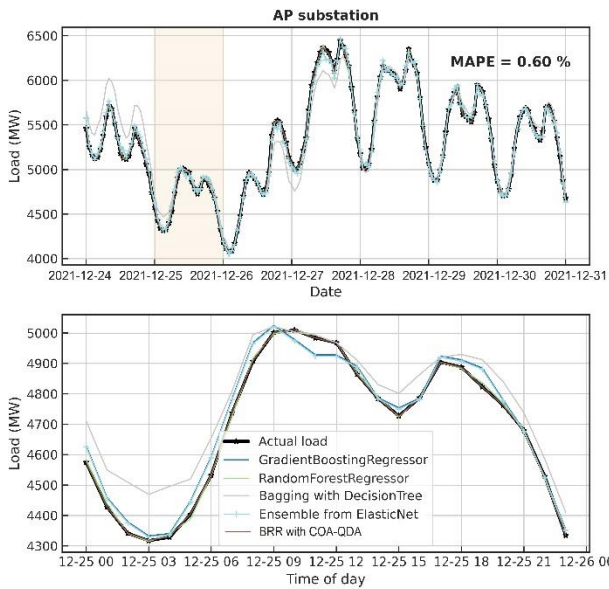
IV. RESULTS AND DISCUSSION

The test case used is of the year 2021 from PJM and is compared to Gradient Boosting Regressor,

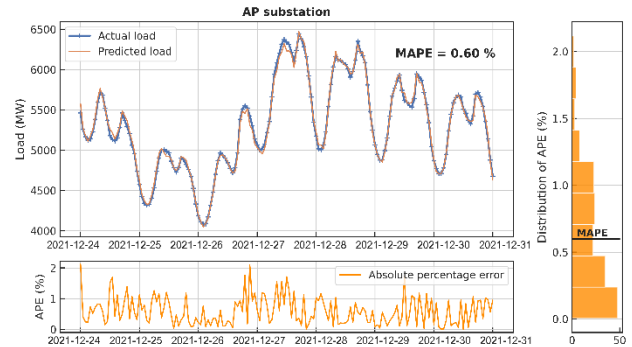
Random Forest Regressor, Bagging with Decision Tree, and Ensemble from ElasticNet. The comparison results for all the substations are recorded in Table 2. The worst-case result obtained is of AP substation which is at MAPE 0.6%. In Fig. 3 and Fig. 4, the comparison between this research and others for the forecasting of the last week of 2021 (from December 24th to December 31st) along with a magnified version of December 25th.

Table 2. Detailed Comparison of Proposed Method

Load areas	Gradient Boosting Regressor [4]		Random Forest Regressor [4]		Bagging with Decision Tree [4]		Ensemble from ElasticNet [5]		BRR with COA-QDA	
	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)
AP	41.00	1.60	10.81	2.15	151.72	3.10	40.96	1.62	1.52	0.60
DAY	11.89	1.55	4.34	1.20	33.09	2.72	11.43	1.50	0.57	0.52
DEOK	19.80	1.62	9.45	1.27	37.74	2.20	10.06	1.30	0.71	0.32
EKPC	9.36	0.71	6.62	1.35	54.97	3.83	9.36	0.61	0.86	0.59
PEPCO	23.16	1.68	5.40	1.17	33.88	1.85	10.55	0.83	0.56	0.33
RECO	10.38	0.83	3.58	0.78	15.87	1.77	5.77	0.84	0.10	0.40



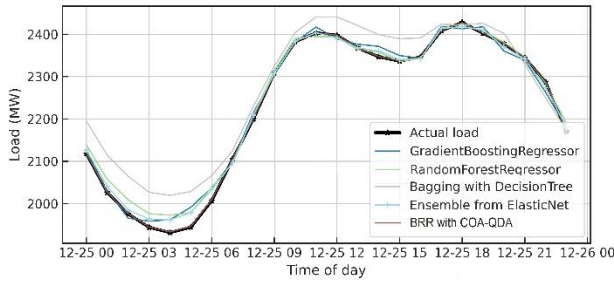
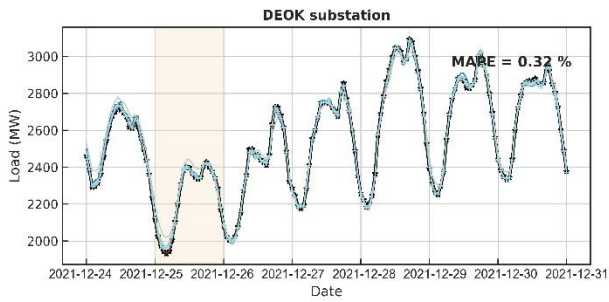
(a) Comparison of multiple techniques with the proposed study



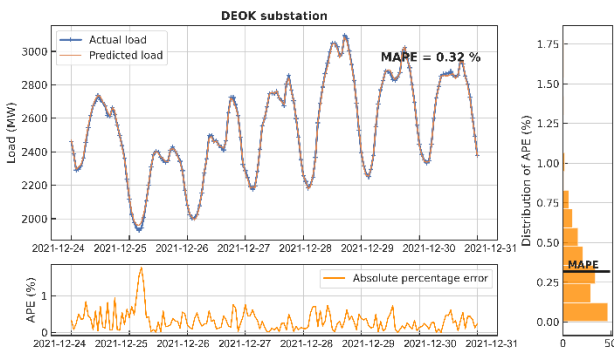
(b) APE and MAPE

Fig. 3. Worst-case forecast

The absolute percentage error (APE) and the MAPE are also presented. The results from this study can be used to perform unit commitment and economic dispatch problems. This forecasting method, because of its high accuracy can also be used for short-term planning.



(a) Comparison of multiple techniques with the proposed study



(b) APE and MAPE

Fig. 4. Best-case forecast

## V. CONCLUSION

The study proposes a short-term load forecasting method which, when coupled with showed improved results compared to other widely researched techniques. The method combines BRR with an optimal feature selection technique named COA-QDA. The worst-case result obtained through this study was 0.60% MAPE which is an improvement when compared with multiple other techniques for the same test case. The use cases for this research include short-term generation planning and economic dispatch of power resources.

## REFERENCES

[1] G. Memarzadeh and F. Keynia, "Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm," *Electr. Power Syst. Res.*, vol. 192, no. July 2020, p. 106995, 2021, doi: 10.1016/j.epsr.2020.106995.

[2] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y.

Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019, doi: 10.1109/TSG.2017.2753802.

[3] C. Tian, J. Ma, C. Zhang, and P. Zhan, "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," *Energies*, vol. 11, no. 12, 2018, doi: 10.3390/en11123493.

[4] E. Yiğit, U. Özkaya, Ş. Öztürk, D. Singh & H. Gritli "Automatic detection of power quality disturbance using convolutional neural network structure with gated recurrent unit. Mobile" *Information Systems*, 2021, 1-11.

[5] J. Lee and Y. Cho, "National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?," *Energy*, vol. 239, p. 122366, 2022, doi: 10.1016/j.energy.2021.122366.

[6] D. H. Tran, D. L. Luong, and J. S. Chou, "Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings," *Energy*, vol. 191, p. 116552, 2020, doi: 10.1016/j.energy.2019.116552.

[7] M. Zulfiqar, K. A. A. Gamage, M. Kamran, and M. B. Rasheed, "Hyperparameter Optimization of Bayesian Neural Network Using Bayesian Optimization and Intelligent Feature Engineering for Load Forecasting," *Sensors*, vol. 22, no. 12, p. 4446, 2022, doi: 10.3390/s22124446.

[8] H. Zang *et al.*, "Residential load forecasting based on LSTM fusing self-attention mechanism with pooling," *Energy*, vol. 229, p. 120682, 2021, doi: 10.1016/j.energy.2021.120682.

[9] N. Andriopoulos *et al.*, "Short term electric load forecasting based on data transformation and statistical machine learning," *Appl. Sci.*, vol. 11, no. 1, pp. 1–22, 2021, doi: 10.3390/app11010158.

[10] L. Wu, C. Kong, X. Hao, and W. Chen, "A Short-Term Load Forecasting Method Based on GRU-CNN Hybrid Neural Network Model," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/1428104.

[11] H. Eskandari, M. Imani, and M. P. Moghaddam, "Convolutional and recurrent neural network based model for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 195, no. February, p. 107173, 2021, doi: 10.1016/j.epsr.2021.107173.

[12] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[13] Z. Gao, J. Yu, A. Zhao, Q. Hu, and S. Yang, "A hybrid

- method of cooling load forecasting for large commercial building based on extreme learning machine,” *Energy*, vol. 238, p. 122073, 2022, doi: 10.1016/j.energy.2021.122073.
- [14] Y. Xuan *et al.*, “Multi-Model Fusion Short-Term Load Forecasting Based on Random Forest Feature Selection and Hybrid Neural Network,” *IEEE Access*, vol. 9, pp. 69002–69009, 2021, doi: 10.1109/ACCESS.2021.3051337.
- [15] D. Lu, D. Zhao, and Z. Li, “Short-term nodal load forecasting based on machine learning techniques,” *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 9, pp. 1–24, 2021, doi: 10.1002/2050-7038.13016.
- [16] L. Li, C. J. Meinrenken, V. Modi, and P. J. Culligan, “Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features,” *Appl. Energy*, vol. 287, no. January, p. 116509, 2021, doi: 10.1016/j.apenergy.2021.116509.
- [17] M. A. Shah *et al.*, “Short - term Meter Level Load Forecasting of Residential Customers Based on Smart Meter ’ s Data,” 2020.
- [18] A. Kumar, B. Yan, and A. Bilton, “Machine Learning-Based Load Forecasting for Nanogrid Peak Load Cost Reduction,” *Energies*, vol. 15, no. 18, p. 6721, 2022, doi: 10.3390/en15186721.
- [19] P. Matrenin, M. Safaraliev, S. Dmitriev, S. Kokin, A. Ghulomzoda, and S. Mitrofanov, “Medium-term load forecasting in isolated power systems based on ensemble machine learning models,” *Energy Reports*, vol. 8, pp. 612–618, 2022, doi: 10.1016/j.egy.2021.11.175.
- [20] H. Naseri, E. O. D. Waygood, B. Wang, Z. Patterson, and R. A. Daziano, “A novel feature selection technique to better predict climate change stage of change,” *Sustain.*, vol. 14, no. 1, pp. 1–23, 2022, doi: 10.3390/su14010040.
- [21] Pennsylvania-New Jersey-Maryland Interconnection, “PJM Data Miner 2,” 2022. <https://dataminer2.pjm.com/list>
- [22] V. C. Corporation, “Visual Crossing Weather (2021) [data-retrieved],” 2022. <https://www.visualcrossing.com/>
- [23] J. Pierezan and L. Dos Santos Coelho, “Coyote Optimization Algorithm: A New Metaheuristic for Global Optimization Problems,” *2018 IEEE Congr. Evol. Comput. CEC 2018 - Proc.*, 2018, doi: 10.1109/CEC.2018.8477769.
- [24] T. Carneiro, R. Victor, M. Da, T. Nepomuceno, G. Bian, and V. H. C. D. E. Albuquerque, “Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications,” pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.
- [25] D. A. Otchere, T. O. A. Ganat, J. O. Ojero, B. N. Tackie-Otoo, and M. Y. Taki, “Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions,” *J. Pet. Sci. Eng.*, vol. 208, no. December 2020, p. 109244, 2022, doi: 10.1016/j.petrol.2021.109244.
- [26] Y. Liang, J. Zhao, D. Sampath Kumar, and D. Srinivasan, “Real-time and consistent sparse estimation of power system distribution factors using online adaptive elastic-net,” *Int. J. Electr. Power Energy Syst.*, vol. 142, no. PB, p. 108361, 2022, doi: 10.1016/j.ijepes.2022.108361.