Improving Electricity Load Forecasting Precision with Multimodal Time Series Analysis

Chauhan Hardik Deepakbhai VIT BHOPAL M.TECH Student , Dr.Virendra Singh kuswah VIT BHOPAL Assistant Professor

***Abstract*—This paper explores the important problem of using multimodal time series analysis to increase the accuracy of energy load forecasts. For effective power system management and cost-effective power generation, accurate power load forecasting is crucial. The document emphasizes the difficulties in precisely projecting long-term power demand as well as the major financial ramifications of overbuilding or subpar infrastructure. It talks about the several models used for power load forecasting, including AI technologies like Back Propagation and conventional statistical techniques. The file also describes how the integration of model properties analysis, structure and fusion strategy optimization, optimal model preference selection, and predictive performance evaluation into the evolutionary process for developing an intelligent decision-making scheme can improve the accuracy of electricity load forecasting through multimodal time series analysis. Furthermore,The benefits of ensemble learning techniques—such as Random Forests, Evolutionary Algorithms, and Generalized Boosted Regression Models—as well as the addition of outside variables—such as meteorological variables—to improve power load prediction accuracy are discussed in the file. All things considered, this file offers insightful information about the discipline of power load forecasting and the methods employed to increase its precision, making it essential reading for anybody with an interest in this crucial sector.**

***Index Terms*—electricity load forecasting, multimodal time series analysis, predictive performance evaluation, model properties analysis, structure optimization, fusion strategy optimization, optimal model preference selection, evolutionary algorithms, random forests, generalized boosted regression models, external factors, weather variables, AI models, Back Propagation, power system management, economic power generation, power load predictions.**

I. INTRODUCTION

IN 2023, the global landscape of electricity consumption will continue to evolve, with a notable 0.2The inability to predict long-term power load accurately can lead to either an excess or insufficient number of power generation facilities, both of which have significant economic implications. Overbuilding may result in electricity wastage, affecting economic decision-making, while inadequate infrastructure can jeopardize daily life by causing power shortages. Consequently, modern power companies focus on forecasting power load to optimize power supply, informing the strategic expansion of power generation facilities. Various models have been employed for power load forecasting, ranging from traditional statistical methods to artificial intelligence (AI) approaches. Classical statistical models, such as the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA), have demonstrated success in certain scenarios. However, their effectiveness is limited to stationary stochastic processes, prompting the exploration of AI models.AI models, including Back Propagation (BP), Artificial Neural Networks (ANN), and Recurrent Neural Networks (RNN), have gained prominence due to their data-driven nature. While traditional BP neural networks may struggle with time series problems, RNNs, and specifically Long Short-Term Memory (LSTM) models, exhibit improved performance by addressing both short-term and long-term dependence. The introduction of the Transformer model, with its Attention mechanism, has further advanced time series prediction capabilities. Researchers have explored models like Autoformer, enhancing the Transformer's ability to handle power consumption data with seasonality and other factors. Recognizing the complexity of power load data, ensemble learning approaches, such as Evolutionary Algorithms, Random Forests, and Generalized Boosted Regression Models, have been integrated to improve prediction accuracy. Noteworthy studies highlight the benefits of ensemble learning, emphasizing superior performance compared to individual models. Additionally, external factors like weather variables have been incorporated to enhance the precision of power load predictions, illustrating the impact of multivariate prediction on overall forecasting accuracy. In summary, the evolving landscape of electricity consumption in 2023 emphasizes the critical importance of accurate long-term power load forecasting. This involves a nuanced exploration of traditional and AI-based models, alongside ensemble learning approaches, to meet the dynamic challenges presented by the global demand for electricity.

II. LITERATURE SURVEY

[1]This study introduces a new method for long-term power load forecasting using the LSTM-Informer with Ensemble Learning. This study underscores the importance of accurate forecasting for efficient power distribution and cost reduction, emphasizing the challenges associated with long-term predictions. This paper addresses the limitations of traditional short-term models and proposes an ensemble approach that combines LSTM for short-term correlations and the Informer model for long-term dependencies. In addition, the paper conducts a literature review comparing the effectiveness of the LSTM model with transformer structure models. It concludes by highlighting the advantages of the LSTM-Informer with Ensemble Learning over other models such as Informer, Autoformer, and Reformer, offering insights into the complexities of power load forecasting.

[2]This paper proposes a multimodal feature extraction and fusion deep neural network (DNN) approach for short-term load forecasting. This approach integrates empirical mode

decomposition, similar day methods, and DNNs to extract comprehensive information from the input data, including raw load sequences, electricity price data, and day and hour information. The proposed scheme is tested on actual data from the electricity market in Singapore and compared with other available models. The results show that the proposed scheme outperforms other models in terms of accuracy and effectiveness in load forecasting.

[3]This paper presents a study on short-term load forecasting using a multimodal evolutionary algorithm and a random vector functional link network based on ensemble learning. The proposed approach improves the accuracy and intelligence of short-term load forecasting systems, which can greatly benefit modern power systems management and economic power generation. This study introduces a novel multimodal evolutionary algorithm based on comprehensive weighted vector angle and shift-based density estimation, which has strong discrimination ability to find more tradeoff multimodal solutions in high-dimensional domains. The proposed approach also integrates predictive performance evaluation, model properties analysis, structure and fusion strategy optimization, and optimal model preference selection into the evolutionary process for building an intelligent decision-making scheme. The efficacy of the proposed approach is demonstrated on 15 complex large-scale multimodal multi-objective problems.

[4]This paper presents a study on short-term load forecasting using a multimodal evolutionary algorithm and random vector functional link network-based ensemble learning. The proposed approach improves the accuracy and intelligence of short-term load forecasting systems, which can greatly benefit modern power systems management and economic power generation. This study introduces a novel multimodal evolutionary algorithm based on comprehensive weighted vector angle and shift-based density estimation, which has strong discrimination ability to find more tradeoff multimodal solutions in high-dimensional domains. The proposed approach also integrates predictive performance evaluation, model properties analysis, structure and fusion strategy optimization, and optimal model preference selection into the evolutionary process for building an intelligent decision-making scheme. The efficacy of the proposed approach is demonstrated on 15 complex large-scale multimodal multi-objective problems.

[5]This study proposes a novel method for short-term forecasting of individual residential load, which is a crucial aspect of demand side response (DSR). This method combines k-means clustering and deep learning to extract the similarity and patterns of residential load, resulting in highly accurate predictions at the individual level. The experimental results demonstrate that the proposed method outperforms a published benchmark method, making it a promising approach for improving the efficiency and reliability of power network operations.

III. Methodology

A. Dataset overview

This research leverages a comprehensive dataset derived from the UK Power Networks' Low Carbon London project, encompassing smart meter readings and weather data for 5,567 households in the London area between November 2011 and February 2014. The dataset provides a rich source for investigating energy consumption patterns, enabling advancements in understanding and optimizing electricity usage. The dataset comprises 19 files, encompassing information on households, half-hourly smart meter measurements, daily aggregated statistics, acorn group details, and weather data from the darksky API. The diversity of data types allows for a holistic exploration of energy consumption behaviors.

B. Data acquisition

1) Historical Electricity Load Data: : From the data set, daily dataset.zip is taken, consisting of data of each customer with their id and energy consumption parameter from which day, energy-sum, and LCLid will be selected. With this data, a new dataset named energy.csv will be created, in which these parameters will be collected for all blocks into one file.

2) Weather Data: :Weather data consists of various features like temperaturemax , dewpoint , cloudcover, windspeed , pressure , visibility , humidity , uvindex ,moonphase which we have taken by using darkskyapi.We have taken only important parameter which are given above and dropped remaining features.

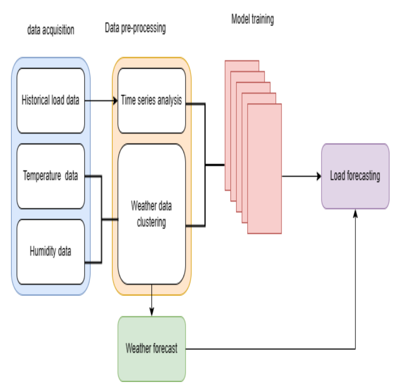


Fig. 1. Process Flow

C. Data Pre-processing

1) weather data clustering: :Now that process of merging weather data to energy data is completed , The process of selecting the feature from weather data is required. To do this process I will now check Relationship of weather conditions with electricity consumption via Graphical and Statistical method like Correlation matrix.

- Energy has high positive correlation with humidity and high negative correlation with temperature.
- Dew Point, UV Index display multicollinearity with Temperature, hence discarded
- Cloud Cover and Visibility display multicollinearity with Humidity, hence discarded
- Pressure and Moon Phase have minimal correlation with Energy, hence discarded
- Wind Speed has low correlation with energy but does not show multicollinearity

[23]:	avg_energy	temperatureMax	dewPoint	cloudCover	windSpeed	pressure	visibility	humidity	uvIndex	moonPhase
avg_energy	1.000000	-0.846965	-0.755901	0.241779	0.149624	-0.028851	-0.246404	0.361237	-0.733171	-0.031716
temperatureMax	-0.846965	1.000000	0.865038	-0.333409	-0.153602	0.118933	0.259108	-0.404899	0.696497	0.003636
dewPoint	-0.755901	0.865038	1.000000	-0.025207	-0.092212	-0.028121	0.042633	0.055514	0.486692	-0.008239
cloudCover	0.241779	-0.333409	-0.025207	1.000000	0.170235	-0.101079	-0.330177	0.480056	-0.248695	-0.062126
windSpeed	0.149624	-0.153602	-0.092212	0.170235	1.000000	-0.344354	0.281088	-0.042391	-0.152634	-0.023273
pressure	-0.028851	0.118933	-0.028121	-0.101079	-0.344354	1.000000	-0.012508	-0.250941	0.100774	0.038462
visibility	-0.246404	0.259108	0.042633	-0.330177	0.281088	-0.012508	1.000000	-0.578130	0.240485	0.062813
humidity	0.361237	-0.404899	0.055514	0.480056	-0.042391	-0.250941	-0.578130	1.000000	-0.533919	-0.013997
uvIndex	-0.733171	0.696497	0.486692	-0.248695	-0.152634	0.100774	0.240485	-0.533919	1.000000	0.012833
moonPhase	-0.031716	0.003636	-0.008239	-0.062126	-0.023273	0.038462	0.062813	-0.013997	0.012833	1.000000

Fig. 2. Process Flow

Fig. 3. Process Flow

Now that unwanted features are dropped I will perform K-means clustering to form the cluster of remaining weather data.for that i will perform elbow technique to find optimal number of clusters.

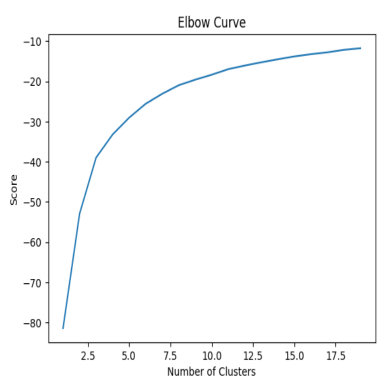


Fig. 4. Process Flow

2) Time series analysis for electrical load data: Now that Pre-processing for weather data is finalised I will proceed with pre-processing Electrical load Data.The nature of data is varying with respect to time so to process this data I will use Time series analysis. Time Series Analysis : Time-ordered data points are the subject of time series analysis, a statistical method. It entails analyzing sequential data's patterns and trends in order to forecast outcomes or identify underlying structures. Observations or measurements made at successive points in time are commonly used to create time series data. In time series analysis, the following are some essential ideas and methods: Time series components include:

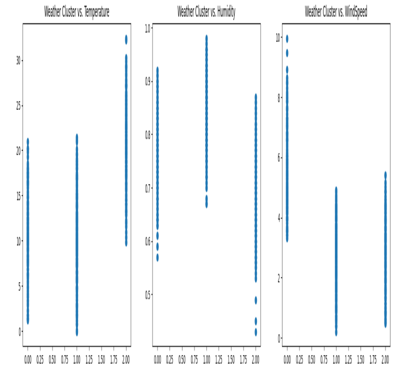


Fig. 5. Process Flow

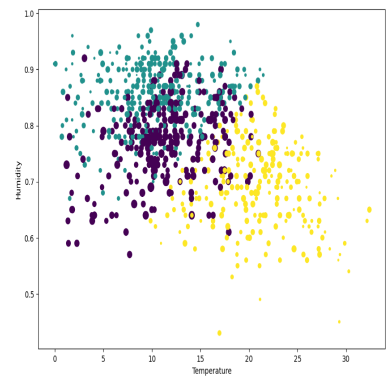


Fig. 6. Process Flow

1) Trend: The long-term movement or direction in the data.
2) Seasonality: Repeating fluctuations or patterns at regular intervals.
3) Cycle: Long-term undulating patterns that are not necessarily regular.
4) Irregularity/Noise: Random fluctuations that cannot be attributed to the above components.

so now i will Seasonal decompose the energy consumption data to its trend's and residual plot. To perform Time series forecasting I will do test's like Autocorrelation and Partial Autocorrelation(ACF PACF) and Dickey Fuller's Test to get optimal parameter.

D. Model Training

1) SARIMAX: Modeling and predicting time-dependent data while taking into account autoregressive (AR), moving average (MA), and exogenous (X) components is known as time series analysis with SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables). SARIMAX is an excellent tool for extracting meaningful patterns from time series data because it captures temporal dependencies, seasonality, and external influences. This allows for reliable insights into complicated temporal behaviors and

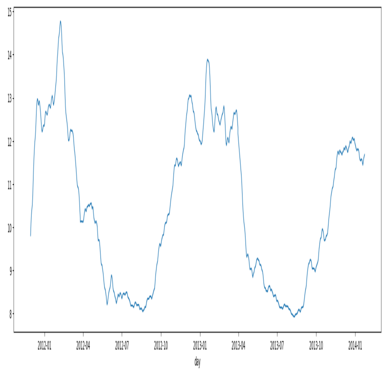


Fig. 7. Process Flow

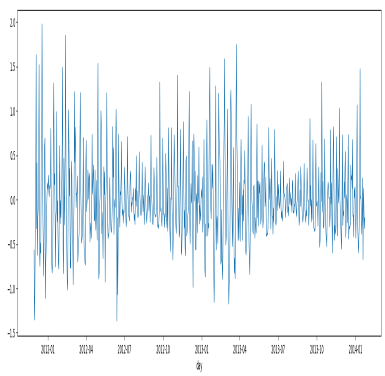


Fig. 8. Process Flow

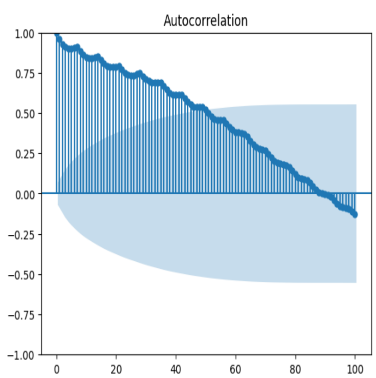


Fig. 9. Process Flow

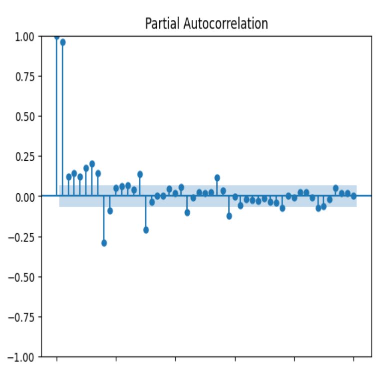


Fig. 10. Process Flow

accurate forecasts. Because of its adaptability to different situations, it is a significant tool in a variety of sectors, including epidemiology, finance, economics, and climate research, where understanding and forecasting temporal trends are essential for decision-making and planning. For this study I will take weather data as exogenous variable and avg energy as variable which has to be predicted and with help of statsmodels library I will import time series analysis SARIMAX module. I will break the data into 7:3 to as Train and Test Data Run the model. now that Model is trained I will perform prediction

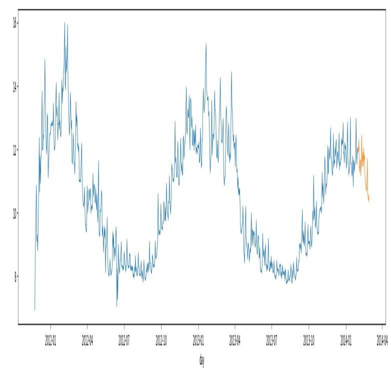


Fig. 11. Process Flow

using the model.

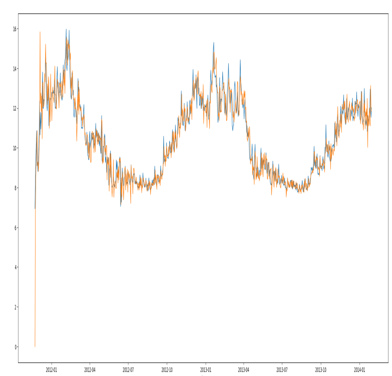
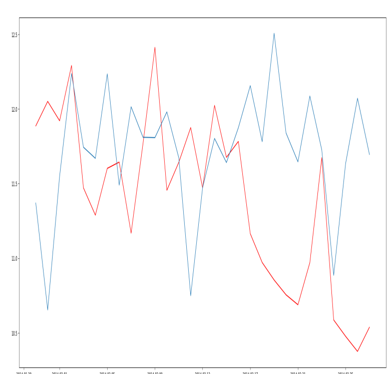


Fig. 12. Process Flow



IV. Results

A. Evaluation parameter

1) Mean Squared Error: it is a common metric used to measure the average squared difference between the actual and predicted values in a regression problem. It is a way to quantify the accuracy of a model's predictions.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

2) Mean Absolute Percentage Error: it is a metric used to evaluate the accuracy of forecasts, particularly in time series analysis. MAPE expresses the average percentage difference between predicted and actual values. The formula for Mean Absolute Percentage Error (MAPE) is:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100$$

3) Mean Absolute Error: it is a metric commonly used to evaluate the accuracy of predictions in a regression problem. It measures the average absolute difference between the predicted and actual values. The formula for Mean Absolute Error (MAE) is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

4) Root Mean Squared Error: it is a commonly used metric to measure the average magnitude of the errors between predicted and actual values in a regression problem. RMSE is a variation of the Mean Squared Error (MSE), and it is particularly useful because it has the same units as the variable being predicted. The formula for Root Mean Squared Error (RMSE) is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

```
MAE: 0.6346071037474794
MAPE: 5.71123002919292
Mean Squared Error: 0.6643815704300448
Root Mean Squared Error: 0.8150960498186975
```

Fig. 13. Process Flow

V. Conclusion

The accuracy of the SARIMAX model in predicting both short- and long-term electricity loads is confirmed by this investigation. The model demonstrates adaptability across contexts by skillfully capturing consumption dynamics through the incorporation of external variables and autoregressive features. Taking demand influences into account improves adaptability when external factors are included.

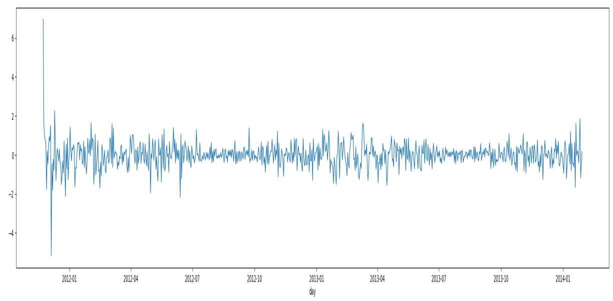


Fig. 14. Process Flow

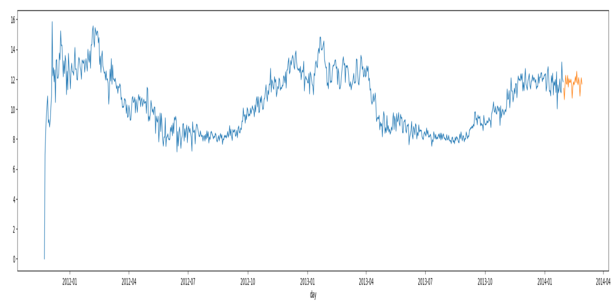


Fig. 15. Process Flow

SARIMAX performs better when validated against historical data, as seen by its noteworthy metrics (MAE: 0.63, MAPE: 5.71, MSE: 0.66, RMSE: 0.82). These findings enable utilities to make well-informed choices about the distribution of resources and grid management.

SARIMAX is a useful tool for all parties involved in the energy industry since it offers accurate projections that are essential for streamlining processes and guaranteeing a steady flow of electricity. This study emphasizes how important it is to use meteorological data and SARIMAX to do advanced time series analysis in order to improve energy management techniques and support the sustainability of power systems.

References

[1] N. Ahmad, Y. Ghadi, M. Adnan and M. Ali, "Load Forecasting Techniques for Power System: Research Challenges and Survey," in IEEE Access, vol. 10, pp. 71054-71090, 2022, doi: 10.1109/ACCESS.2022.3187839.

[2] Wang, K.; Zhang, J.; Li, X.; Zhang, Y. Long-Term Power Load Forecasting Using LSTM-Informer with Ensemble Learning. Electronics 2023, 12, 2175. https://doi.org/10.3390/electronics12102175

[3] Wu, K., Gu, J., Meng, L. et al. An explainable framework for load forecasting of a regional integrated energy system based on coupled features and multi-task learning. Prot Control Mod Power Syst 7, 24 (2022). https://doi.org/10.1186/s41601-022-00245-y

[4] Ibrahim, B.; Rabelo, L.; Gutierrez-Franco, E.; Clavijo-Buritica, N. Machine Learning for Short-Term Load Forecasting in Smart Grids. Energies 2022, 15, 8079. https://doi.org/10.3390/en15218079

[5] Y. Hong, Y. Zhou, Q. Li, W. Xu and X. Zheng, "A Deep Learning Method for Short-Term Residential Load Forecasting in Smart Grid," in IEEE Access, vol. 8, pp. 55785-55797, 2020, doi: 10.1109/ACCESS.2020.2981817

[6] F. Han, T. Pu, M. Li and G. Taylor, "Short-term forecasting of individual residential load based on deep learning and K-means clustering," in CSEE Journal of Power and Energy Systems, vol. 7, no. 2, pp. 261-269, March 2021, doi: 10.17775/CSEEJPES.2020.04060.

[7] Short-term load forecasting using multimodal evolutionary algorithm and random vector functional link network based ensemble learninghttps://doi.org/10.1016/j.apenergy.2020.116415

[8] Share of Electricity in Total Final Energy Consumption. Available online: https://yearbook.enerdata.net/electricity/shareelectricity- final-consumption.html (accessed on 1 January 2022).

[9] Tang, L.; Wang, X.; Wang, X.; Shao, C.; Liu, S.; Tian, S. Long-term electricity consumption forecasting based on expert prediction and fuzzy Bayesian theory. Energy 2019, 167, 1144–1154.

[10] Kim, T.Y.; Cho, S.B. Predicting the household power consumption using CNN-LSTM hybrid networks. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Madrid, Spain, 21–23 November 2018; Springer: Cham, Switzerland, 2018; pp. 481–490.

[11] Fan, D.; Sun, H.; Yao, J.; Zhang, K.; Yan, X.; Sun, Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy 2021, 220, 119708.