INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

An Enhanced Ensemble of Long Short-Term Memory and Vector AutoRegression for Energy Consumption Forecasting

Anagha D Ananth[1], Sujatha Arun Kokatnoor[2]*

1 Student, Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore, 560074, India
2 Assistant Professor, Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore, 560074, India

Abstract

Objectives: To design and develop an enhanced ensemble model for residential energy consumption prediction using time series analysis. **Methods:** The system is consistent of an ensemble made from two time series models, which are later combined to produce a final prediction through the use of bagging techniques. The energy profile is built for this study using the data collected from a single-phase Minion Energy Monitor, installed in residential buildings, which is developed by Minion Labs India Private Limited. Samples are collected for a time of 7 days with a two second interval. This data is then restructured and normalized for it to fit the enhanced ensemble model of Long Short-Term Memory (LSTM) and Vector Autoregression (VAR) using bagging techniques and weighted average to obtain the predictions. **Findings:** The proposed model has produced an enhanced R2 score of 98.99% when compared to LSTM (74.85%), SVM (62.41%), VAR (82.914%) and ARIMA (93.152%) standalone models. It makes use of an analyzed lag variable to reduce computation complexity and resource utilization. Further, an ensemble technique is used to pick out the strengths of two models. An analysis performed showed that this architecture is 6.08% better than the algorithms for LSTM and VAR individually. **Novelty:** A data driven solution is proposed in this study through the enhancement of existing models to create an ensemble and thereby creating a stable structure that predicts values using a weighted average. The weighted average ensures that precise outputs are obtained by giving more importance to predictions that are closer to the actual values. The use of a lag variable further increases the efficiency, learning rate, dealing with non-linear features, less error and faster training of the proposed architecture. All these factors also aid in improving the accuracy for a time series data prediction.

Keywords: Energy Consumption Forecast; LSTM; VAR; Ensemble; Bagging; Artificial Neural Networks; Support Vector Machine

1 Introduction

A high percentage of energy consumption in India is directed by the utilization in both commercial and domestic buildings. The country has seen a constant rise in energy consumption in terms of percentage over the years and is said to grow by at least 15 times between the years 2010 and 2050[1]. Such high demands can create a dent in the environment's resources and also lead to a huge amount of CO_2 emissions. It has been found that for every 1% per capita increase in power consumption, there is a 7.85% increase in the effect of Carbon dioxide (CO_2) emissions over the long run[2]. Currently, the potential for energy savings ranges only between 17-42%[3]. There is hence a prospective of improving this number by being more aware of planning and implementing an efficient methodology that can help people manage their daily consumptions with reduced power wastage. Long term forecasting is hence necessary in order to supervise and make informed decisions about one's power consumption activities without any profligacies.

There are two major approaches for energy consumption prediction: the physical modeling approach and the data-driven approach. The first one deals more with physical and environmental factors like the weather, structure of the building, choice of construction materials, the surrounding environment, and insulations[4,5]. The second approach makes use of historical data that is generated from the energy meters to identify common trends. In this study, data-driven approaches are considered for analyzing energy utilization forecasting.

Existing data-driven models for power predictions include a wide variety of machine learning and deep learning models like AutoRegressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Random Forests (RF), and Artificial Neural networks (ANN). ARIMA is a commonly used univariate time series model for future predictions based on historical data[6]. However, this model has often been used only for short-term predictions[4]. This means that it is only able to retain information and patterns for a short period of time, thereby making it less efficient for long-term forecasting. ARIMA is also known to fall short in predicting values that are outside the expected limits. With SVR, there is a considerably large computation time, which makes it expensive for the analysis[7]. Though the accurate results that are produced using SVR model, the process of selecting a kernel for this model can be extremely strenuous as there is no set standard for doing so, which decreases the ease of optimizing the model. Using ANN model, forecasting is done only on a short-term basis without much retaining of information for future predictions. It also adds to slow convergence for especially long parameter lists[8]. Another limitation of this model is that it is unable to adjust to major changes that happen within the data due to power failures or load fluctuations. This is because an ANN model tends to perform well only for well-structured data with a uniform understanding[9,10]. Random Forest (RF) is generally said to be an ensemble with classification and regression trees as the base models. RF has proved to outperform ARIMA in certain use cases and is also known to be a good tool for attribute picking[11]. Although it is said to yield good results in terms of forecasting, the algorithm is highly complex for huge amounts of data, which increases the time as well as the cost of execution. It also lacks the involvement of occupant variables in order to improve the accuracy.

Convolutional Neural Network (CNN) and LSTM models solve the problem of finding exploratory variables using linear regression[12]. Through literature review, it is observed that this model performs consistently well. However, there is a need to spend time on selecting the best hyperparameters based on the nature of the given data. It also has a considerably higher training time, thereby increasing the cost. A hybrid model of Group Method of Data Handling (GMDH) and Least Square Support Vector Machine (LSSVM), also called GLSSVM, to overcome the time efficiency condition faced in standalone models are applied for measuring energy consumption[13,14]. The proposed hybrid structure has shown the potential to outperform the individual constituent models, especially for a time-series application. A limitation to this model is that it has only proved to perform extremely well for a restricted dataset of finite value. Flexibility is low, and the study has not been performed on a larger dataset.

An Improved Sine Cosine Optimization Algorithm based LSTM (ISCOA-LSTM) is introduced[15], where certain benefits like quick convergence speed and improved accuracy of SCOA are used to tune the hyperparameters of an LSTM model to enhance predictions for energy consumption values. This model proves to be useful for short-term, medium-term as well as long-term predictions for time series analysis. However, factors such as real-time disturbances in data and timely updates of the parameters have not yet been studied. These are a few aspects that could be extended for further study.

Building energy prediction values are different for various profiles within the building. A single model developed for this purpose consumes a lot of computational resources. A framework called Multiple Electric Energy Consumption forecasting in a smart building using Transfer Learning and LSTM, denoted by MEC-TLL is used[16]. This structure makes use of the Silhouette analysis as well as K-means to distribute data into different clusters, which is then passed onto the MEC algorithm for training by making use of LSTM models to learn the data. The analysis of the outputs has suggested that this framework takes less computation time and also has improved predictability when compared with traditional machine learning models. It is yet to be tested for real-time applications, and hence there is no uncertainty about performance for industry-standard datasets.

Three different base models - Support Vector Machine (SVM), Back Propagation Neural Network (BPNN), and Extreme Learning Machine (ELM) are put together to form an integrated structure to predict energy utilization values with the least amount of forecasting errors [17]. This has been used to overcome the inaccuracy of the standalone models in terms of prediction accuracy with respect to data from multiple sources. An advantage to this stacked integration is that it has been tested for multiple different datasets and has shown a constant rise in performance for different data. A limitation, however, is that it has only been validated for short-term predictions, and has no verified results of good performance for long-term data analysis.

Studies have been done to compare the performance of single prediction models to ensemble models. They show high reliability on prediction values for the ensemble creations along with flexibility and stability [18]. This is because the multiple base models that are used in the process are chosen to cancel out the disadvantages of each and also throw light on the strength of each model to take care of different aspects of the data being given. Moreover, if one base model has failed, the remaining ones can still be used to give close to accurate results. Such hybrid models are also seen to perform well in identifying abnormalities and in the application of early fault detection [5,19].

Despite the extensive study that has been done, there is often a lack of importance that is given to the lag variable that needs to be considered to give more relevant outputs. Also, current models in existence tend to use the entire dataset for training purposes every time the model is required to forecast values. This can lead to an unnecessarily high cost of time and effort with a lack of flexibility in terms of the data being collected and used for analysis.

The objective of this study is hence to create an enhanced version of an ensemble consisting of LSTM and VAR as the base models in order to overcome the challenges mentioned above. This is done by making use of historical data that is generated from the energy meter to identify common trends. This system then aims at analyzing the data using visualizations and makes predictions for the future, which can pose a helpful tool to enable people to understand their energy utilization patterns. This helps a person closely monitor and single out an anomaly for early fault detections as well, thereby allowing an individual to track and manage his daily energy consumptions.

This paper proposes a novel data-driven solution by building a time series prediction ensemble model that integrates the advantageous properties of LSTM and VAR using enhanced ensemble bagging techniques. A lag value of the past 48 hours is taken into account to predict values for the next three days, to increase the accuracy. Further, Principal Component Analysis (PCA) techniques have been used for cleaning the data to enhance the outputs. The processed data is passed through an ensemble of LSTM and VAR using a weighted average bagging technique to enhance the outputs. It considers the results produced by both models individually along with the results of the ensemble model and compares their performance using the Mean Squared Error (MSE) value. The proposed model has proven to be more benign in terms of increased precision as well as decremented computational resource utilization.

The structure of the paper is organized as follows. Section 2 discusses the data collection and preprocessing techniques used along with the architecture of the proposed LSTM-VAR ensemble model. Section 3 presents the empirical results, which is followed by Section 4 to conclude the paper.

2 Proposed Methodology

Figure 1 is an outline of the research being done in this paper. Initially, data is collected from the energy meters and is fed into the preprocessing function to clean and extract only the required information. This processed data is then passed through the created ensembles of LSTM and VAR for training after the models are individually optimized for best parameter usage. Once the models have been trained, the testing data is passed through the ensemble to validate the performance of the proposed model.

2.1 Corpus Collection

The data used for this study is collected from a single-phase Minion Energy Monitor, which is developed by Minion Labs India Private Limited. It is installed in a commercial building to collect data from households. The acquired data is not publicly available and is collected for the interest of this study. The dataset contains 7 attributes, namely Device ID, Current, Active Power, Voltage, Power Factor, Frequency, and Timestamp. The dataset has been collected over a period of one week, with a time interval of two seconds. Information pertaining to multiple devices and their energy consumption activity for 7 days are recorded. Active power is chosen as the target variable to be predicted for power consumption.

2.2 Data Processing

Missing and inconsistent values are found by calculating the difference between two consecutive rows of data and by comparing the resultant difference against the specified threshold. These values are then superseded by the result obtained by an Iterative Imputer that uses Bayesian Ridge Regression. Once a consistent dataset is formed, it is then reshaped to have the timestamp column of power consumption transformed into the index of the data frame. The data is also checked for stationarity, to make sure that statistical values like mean, variance, and autocorrelation do not change over time. A 0.8 fraction of the entire data is extracted to form the training subset, with the remaining falling under the test category. Each of these subsets is then normalized using the equation (1).
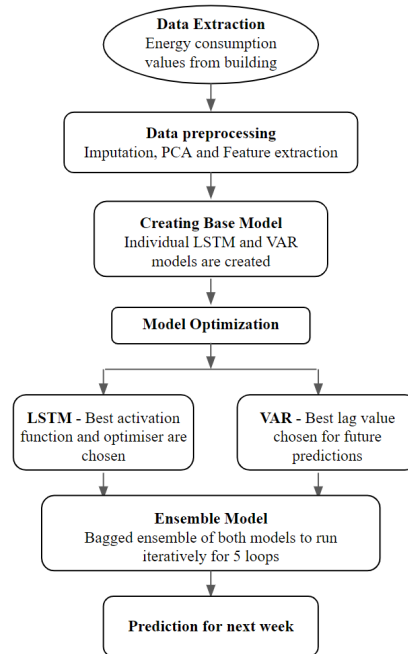
$$N = \frac{(x - trainx[mean])}{trainx[std]} \tag{1}$$

Fig 1. Architectural Diagram for Energy Consumption Forecasting

where N is the normalized score, x represents the values in the dataset, train[mean] and train[std] is the mean and standard deviation of the values for each column respectively. This is done to scale the data for a more radical transformation.

Since the data collected is fairly large, Principal Component Analysis (PCA) is performed to reduce the size of the dataset without compromising the necessary information. This in turn reduces the computation time as well as prevents overfitting of data. PCA has been performed using the decomposition submodule within the sci-kit learn library provided by Python.

2.3 Proposed Enhanced Ensemble Model

Ensembling is a technique used in machine learning where multiple base models are integrated to create a single model that can perform better. This is due to the fact that an integrated model tends to overcome the limitations of each model and only strengthens the performance by combining their independent capabilities. They are known to reduce the need for a heavily set of tuned parameters for optimization. Moreover, they are known to be robust by predicting more accurate results as well as maintaining stability across multiple datasets. Most ensemble techniques are a two-step process:

1. Create and identify the connections for the base model. The models chosen can either be connected in parallel or series.
2. Pick one valid scheme to evaluate the results produced. A few techniques that are often used include Max voting, Weighted average, Stacking, Boosting, Bagging, and Blending.

The proposed ensemble has been depicted in Figure 2. The model incorporates multiple alternate layers of the LSTM and VAR models. They are connected in parallel for introducing the concept of non-linearity so that the output layers are not a direct function of the data being fed into the model. Bagging has been used to create the ensemble model as it is known to reduce variance and also make the outputs independent of the noise in the data. It is a mixture of two unique steps, Base Modeling and Aggregation.
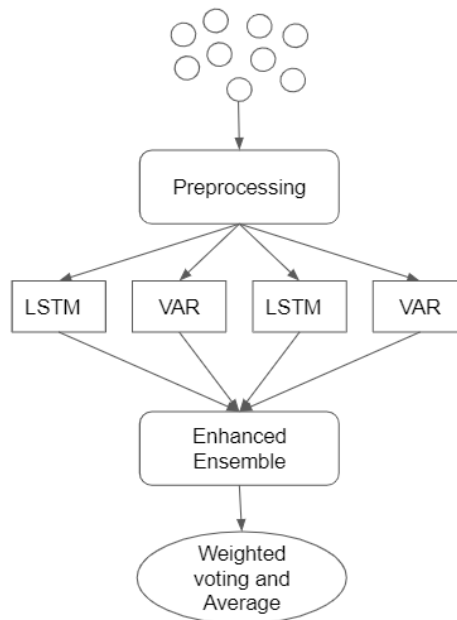

Fig 2. Proposed Ensemble Architecture for LSTM-VAR

Base Modeling: The base models have been chosen due to their already validated performance with time series analysis and multivariate datasets. Data has been reshaped individually for both models to optimize their performance. All prediction models are then run on the sample input data for a specified number of times. This study has run the ensemble for five iterations to create five models with subsets of LSTM and VAR. Only five iterations are done so that the complexity is not increased but at the same time, the accuracy is not compromised. An increase in the iteration count past five did not show a huge varying difference with respect to the results.

Aggregation: The final outputs predicted are combined to form a single array of forecasted values using a well-known existent algorithm like weighted average or maximum voting. Since this is a numerical regression problem, the aggregate is considered as the mean of all predicted values given by the base models as shown in equation (2).

$$prediction = \left(\frac{1}{n}\right) \sum_{i=1}^{n} pred_i \tag{2}$$

A stacked LSTM of two layers has been used as the base model to store information about the attributes and their contribution to power consumption patterns. LSTM networks are developed to solve the problem of long-term dependencies. It is one of the most commonly used models for time series analysis due to its advantageous property of being able to retain memory for longer durations when compared to a standard Recurrent Neural Network (RNN). This has been achieved by making use of four neural network structures within the single repeating module that is present in a simple RNN. It comprises multiple memory blocks called cells, which are responsible for retaining the information learned. Manipulations to these cells are done through three different gates, forget gate, input gate, and output gate. These gates contain a Sigmoid activation function that helps the structure understand what parts of the data need to be remembered and what can be ignored. The outputs are binary in the form of 1 and 0 where 1 corresponds to data being allowed to pass through the gate and 0 corresponds to data being blocked by the gate. The equations for the three gates are shown in equations (3), (4), and (5).

$$i_t = \sigma\left(w_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{3}$$

$$f_t = \sigma \left(w_f \left[h_{t-1}, x_t \right] + b_f \right) \tag{4}$$

$$o_t = \sigma \left(w_o \left(h_{t-1}, x_t \right] + b_o \right) \tag{5}$$

where i_t, f_t and o_t represent the input, forget and output gates respectively, σ represents the Sigmoid function, w_x represents the weight of the neuron at that gate(x), h_{t-1} is the output of the previous LSTM block, x_t is the input at the current timestamp and b_x is the bias at the specified gate.

The VAR model is used as a multivariate time series forecasting model where a group of time-dependent variables is modeled as a linear combination of the previous value. Since it makes use of multiple independent variables, it is created as a system of equations for each variable being predicted. Each of these equations makes use of an endogenous lag variable to identify a deterministic trend. Suppose a_1 and a_2 are the variables taken into consideration for forecasting. Then, the value of $a_1(t)$ will be calculated using the previous values of a_1 and a_2, as shown in equations (6) and (7).

$$a_1(t) = x_1 + w_1 11 * a_1(t-1) + w_1 12 * a_2(t-1) + e_1(t-1) \tag{6}$$

$$a_2(t) = x_2 + w_1 21 * a_1(t-1) + w_1 22 * a_2(t-1) + e_2(t-1) \tag{7}$$

where x_1 and x_2 are constants, $w_1 pq$ ($w_1 11$, $w_1 12$, $w_1 21$ and $w_1 22$) are the coefficients and e_1 and e_2 are the error terms. This enables the model to understand and apply the associations between several variables. However, before this model can be estimated, it is important to identify certain parameters like stationarity and cointegration of variables. Statistical tests like the Augmented Dickey-Fuller test (ADF) can be performed and if found, they need to be processed to make the data readily available for the model during the time of forecasting.

The results from the standalone LSTM and VAR models have been taken to iteratively create an ensemble using bagging techniques. This newly created ensemble produces the final prediction by taking the weighted mean of base models. This can be expressed through equation (8).

$$\sum_{i=1}^{n} \frac{(w_{n*} pred_n)}{w_n} \tag{8}$$

Where, w_n is the vector of weights being used. The weights that are used are adjusted according to the observations made through the performance of the time series models being used. Hence out-of-bag observations are considered in the next iteration in order to make sure more accurate results are obtained. This additional parameter of adding weights to the average makes sure that only relevant observations are made through the reduction of erroneous predictions in every consecutive iteration. It also allows each ensemble member to contribute to the final prediction through the use of its individual advantages.

The ensemble has been tuned in the following ways to provide the most effective results. The chosen activation function is the Rectified Linear activation function (ReLU) due to its property of being able to process stochastic gradient descent with the additional characteristic of performing backpropagation for error correction. It is also known to circumvent easy saturation. The return sequence parameter for the LSTM has been set to 'True' so that the output of each hidden layer is made visible to convert the result into a three-dimensional input for the next layer. To avoid overfitting, a callback method has been considered while fitting the model to the data, which will automatically terminate the training process once the specified threshold has been reached. The VAR model used in the ensemble made use of the application of the Akaike Information Criterion (AIC) to identify 31 as the best order for lag and 3.644 as the AIC score. Once these parameters have been taken into account, the data is fit to the VAR model, and then predictions are made. Data that is preprocessed is passed through both of these models, where probabilities for each prediction are calculated. They trigger each neuron in the LSTM network based on a threshold value, thereby giving more weight to certain neurons in order to predict accurate values.
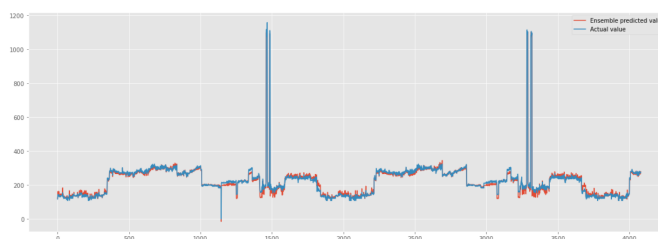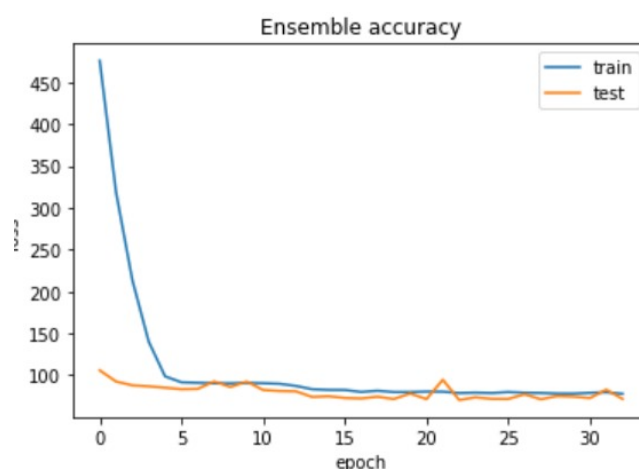
3 Results and Discussion

Table 1 lists the first eight values that are predicted along with their respective forecasted values by LSTM and the Ensemble. As it is seen, there is a considerably large variation for the values being predicted for an individual LSTM model with an average difference of 9.3 against the actual values. It is also observed that the predictions being made by the LSTM base model are not consistent and stable over time. The enhanced results, on the other hand, have proven to be more congruous with the original test values.

Table 1. Comparison of Predicted Values against Actual Test Values

Test Dataset Values	LSTM Predicted Values	Enhanced Predicted Values
116	122.01	120.80
126	131.37	129.07
132	139.52	136.14
143	160.40	156.82
134	139.52	137.14
142	154.73	151.20
143	160.40	158.82
133	139.52	137.14

3.1 Error Metrics

To evaluate the performance of the proposed ensemble, the forecasting error has been calculated. This is done through the use of the MSE value, which is an average of the squares of all the errors present. It is seen that the modeling performance is low for this structure in contrast to the plot shown in Figure 3 (x-axis: Time Taken and y-axis: Energy Consumed) which depicts the predicted results of the ensemble model when compared with the actual values in the validation set. The proposed model has manifested its reliability as the predicted values are centered close to its original values. The LSTM model individually provides an MSE value of 84.848. The proposed model has efficiently reduced the MSE to 79.685 with a difference of 5.163, which makes it 6.08% better than the existing base architectures. Figure 4 is a representation of the test and train validation with respect to the loss and epoch count for one iteration.

Fig 3. Ensemble Predicted Values vs. Original Values Plot

Fig 4. Ensemble Accuracy Values vs. Original Values Plot

Analysis has also been done to compare the proposed model with existing studies that have used different models. SVM and Autoregression are two commonly used models that have been implemented previously in studies for this application. They

have proven to be dependable for predictions. But as seen in Figure 5, the created enhanced LSTM model has a much lesser MSE value when compared to those of SVM and Autoregression. This improves the credibility of the proposed architecture when compared with existing studies.

R2 score, also known as the coefficient of determination, is often used to describe the percentage of variance in the dependent attribute in accordance with the target attribute. A high R2 score suggests more accuracy. Similarly, another metric called Explained Variance explains the amount of data being lost while approximating the model. The lower the value, the less accurate the model. Figure 6 and Figure 7 are a representation of the R2 scores and Explained Variance respectively in comparison with the enhanced model. It is seen that the ensemble has outperformed both of the other models in terms of these metrics and can be considered reliable for all future applications towards energy prediction for domestic uses.
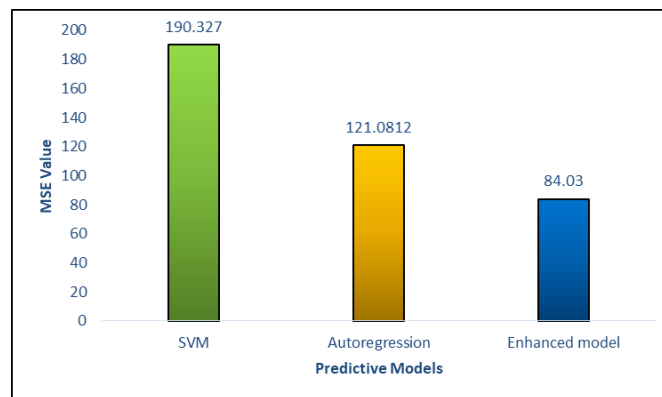

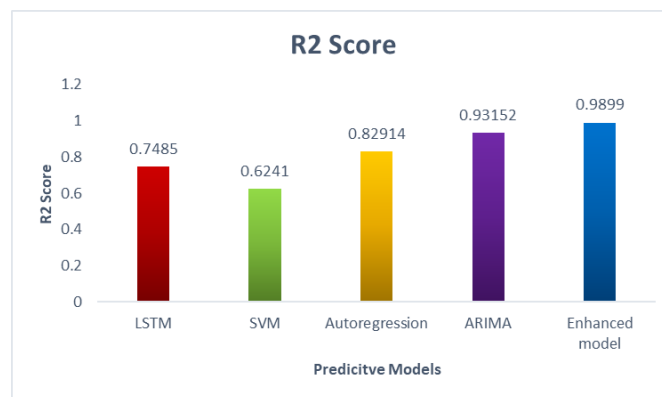**Fig 5.** MSE Value Comparison of Proposed Enhanced Model with Standalone Models

Fig 6. R2 Score Comparison of Proposed Enhanced Model with Standalone Models

An observed limitation is the lack of variation in data with respect to the parameters selected for training. This study can be extended for a future scope with the addition of several other parameters corresponding to energy consumption for multiple datasets to check its consistency.

4 Conclusion

This research study has proposed a data-driven, deep learning ensemble of LSTM and VAR in order to overcome the gaps identified in existing approaches for energy consumption prediction. The time-series data is analyzed and normalized to fit the needs of the model, which is then passed through the ensemble to learn and extract useful information. A lag variable is then introduced to increase the efficiency and decrease computation time in predicting energy consumption values. The predicted values are plotted to identify the MSE and evaluate the model's performance. The paper concluded with a comparison of indifference in error for the standalone models against the created ensemble. It is seen that the ensemble performed
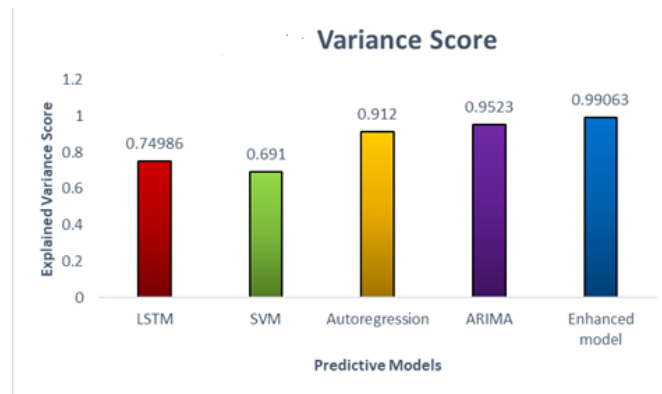
Fig 7. Variance Score Comparison of Proposed Enhanced Model with Standalone Models

more efficiently and stably than the standard models such as LSTM, SVM and VAR, thereby making it more reliable for future applications. The proposed enhanced model gave an improved R2 score of 98.99% when compared to LSTM (74.85%), SVM (62.41%), ARIMA (93.15%) and VAR (82.914%) standalone models. LSTMs are affected by various random weight initializations and are hence quite comparable to a Feed-Forward Neural Network. LSTMs are susceptible to overfitting and the dropout technique is hard to use to stop this problem. Dropout is a strategy for regulating inputs and recurrent LSTM connections which are likely to be excluded from weight updates and activation during training a network. In addition, LSTM requires the operation of 4 linear layers per cell for every time-phase sequence. Linear layers need to calculate big quantities of the memory bandwidth, since the system doesn't have enough memory bandwidth to feed the computer units, because of all these reasons, LSTM performed less. SVM took a long training time and led to a huge error percentage because of which it performed less. VAR imposes a one-way relationship. The variables of the forecast are influenced by the prediction variables, but not vice versa. Due to this reason, VAR too performed less. ARIMA models can be extremely precise and dependable under the right conditions and with enough data. It is observed during experimentation process that, the model's parameters are to be defined manually, making obtaining the best fit a lengthy trial-and-error process. Also, the model is strongly reliant on the consistency and differencing of past data. To ensure that the model produces reliable results and forecasts, it is critical to ensure that data is collected accurately and over a lengthy period of time. Due to these reasons, ARIMA model performed less when compared to the proposed enhanced model.

The future scope for this study includes testing the model for various datasets and taking into consideration a number of other parameters that contribute to an energy profile. Further, an empirical analysis can be done to analyse the time complexity of the model with respect to other models that have been used for similar problem statements.

References

1) Yu S, Tan Q, Evans M, Kyle P, Vu L, Patel PL. Improving building energy efficiency in India: State-level analysis of building energy efficiency policies. *Energy Policy*. 2017;110:331–341. Available from: https://dx.doi.org/10.1016/j.enpol.2017.07.013.
2) Jayanthakumaran K, Verma R, Liu Y. CO2 emissions, energy consumption, trade and income: A comparative analysis of China and India. *Energy Policy*. 2012;42:450–460. Available from: https://dx.doi.org/10.1016/j.enpol.2011.12.010.
3) Tulsyan A, Dhaka S, Mathur J, Yadav JV. Potential of energy savings through implementation of Energy Conservation Building Code in Jaipur city, India. *Energy and Buildings*. 2013;58:123–130. Available from: https://dx.doi.org/10.1016/j.enbuild.2012.11.015.
4) Sen P, Roy M, Pal P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy*. 2016;116:1031–1038. Available from: https://dx.doi.org/10.1016/j.energy.2016.10.068.
5) Chwieduk DA. Towards modern options of energy conservation in buildings. *Renewable Energy*. 2017;101:1194–1202. Available from: https://dx.doi.org/10.1016/j.renene.2016.09.061.
6) Liu C, Wu K, Tsao M. Energy Efficient Information Collection with the ARIMA Model in Wireless Sensor Networks. *GLOBECOM'05 IEEE Global Telecommunications Conference*. 2005;5:2470–2474. Available from: https://doi.org/10.1109/GLOCOM.2005.1578206.
7) Grolinger K, Capretz MAM, Seewald L. Energy Consumption Prediction with Big Data: Balancing Prediction Accuracy and Computational Resources. *2016 IEEE International Congress on Big Data (BigData Congress)*. 2016;p. 157–164. doi:10.1109/BigDataCongress.2016.27.
8) Biswas MAR, Robinson MD, Fumo N. Prediction of residential building energy consumption: A neural network approach. *Energy*. 2016;117:84–92. Available from: https://dx.doi.org/10.1016/j.energy.2016.10.066.
9) Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings*. 2005;37(12):1250–1259. Available from: https://dx.doi.org/10.1016/j.enbuild.2005.02.005.
10) Banihashemi S, Ding G, Wang J. Developing a Hybrid Model of Prediction and Classification Algorithms for Building Energy Consumption. *Energy Procedia*. 2017;110:371–376. Available from: https://dx.doi.org/10.1016/j.egypro.2017.03.155.

11) Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S. Random Forest based hourly building energy prediction. *Energy and Buildings*. 2018;171:11–25. Available from: https://dx.doi.org/10.1016/j.enbuild.2018.04.008.

12) Kim TY, Cho SB. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*. 2019;182:72–81. Available from: https://dx.doi.org/10.1016/j.energy.2019.05.230.

13) Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*. 2014;33:102–109. Available from: https://dx.doi.org/10.1016/j.rser.2014.01.069.

14) Fang X, Gong G, Li G, Chun L, Li W, Peng P. A hybrid deep transfer learning strategy for short term cross-building energy prediction. *Energy*. 2021;215:119208. Available from: https://dx.doi.org/10.1016/j.energy.2020.119208.

15) Somu N, R GRM, Ramamritham K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy*. 2020;261:114131. Available from: https://dx.doi.org/10.1016/j.apenergy.2019.114131.

16) Le T, Vo MT, Kieu T, Hwang E, Rho S, Baik SW. Multiple Electric Energy Consumption Forecasting Using a Cluster-Based Strategy for Transfer Learning in Smart Building. *Sensors*. 2020;20(9):2668. Available from: https://dx.doi.org/10.3390/s20092668.

17) Wang H, Huang Y, Shi M, Liu S. Short-term load forecasting model based on multi-model integration. *Journal of Physics: Conference Series*. 2020;1549(5):052007. Available from: https://dx.doi.org/10.1088/1742-6596/1549/5/052007.

18) Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*. 2017;75:796–808. Available from: https://dx.doi.org/10.1016/j.rser.2016.10.079.

19) Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*. 2014;127:1–10. Available from: https://dx.doi.org/10.1016/j.apenergy.2014.04.016. doi:10.1016/j.apenergy.2014.04.016.