Wiley
Applied Computational Intelligence and Soft Computing
Volume 2025, Article ID 5211419, 16 pages
https://doi.org/10.1155/acis/5211419

WILEY

Research Article

Predicting Residential Energy Consumption in South Africa Using Ensemble Models

David Attipoe,[1] **Donatien Koulla Moulla**,[1] **Ernest Mnkandla**,[1] **and Alain Abran**[2]

[1]*Department of Computer Science, University of South Africa, Johannesburg, Gauteng, South Africa*
[2]*Department of Software Engineering and Information Technology, École de Technologie Supérieure, Montréal, Quebec, Canada*

Correspondence should be addressed to Donatien Koulla Moulla; moulldk@unisa.ac.za

Received 10 November 2023; Revised 2 March 2025; Accepted 10 April 2025

Academic Editor: Pramita Mishra

This study presents ensemble machine learning (ML) models for predicting residential energy consumption in South Africa. By combining the best features of individual ML models, ensemble models reduce the drawbacks of each model and improve prediction accuracy. We present four ensemble models: ensemble by averaging (EA), ensemble by stacking each estimator (ESE), ensemble by boosting (EB), and ensemble by voting estimator (EVE). These models are built on top of Random Forest (RF) and Decision Tree (DT). These base predictor models leverage historical energy consumption patterns to capture temporal intricacies, including seasonal variations and rolling averages. In addition, we employed feature engineering methodologies to further enhance their predictive abilities. The accuracy of each ensemble model was evaluated by assessing various performance indicators, including the mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination R^2. Overall, the findings illustrate the efficiency of ensemble learning models in providing accurate predictions for residential energy consumption. This study provides valuable insights for researchers and practitioners in predicting energy consumption in residential buildings and the benefits of using ensemble learning models in the building and energy research domains.

1. Introduction

The growing global demand for electricity, driven by urbanization and industrialization, has increased the complexity of electricity distribution and management. South Africa, in particular, for more than a decade, has been facing the persistent problem of load shedding, which highlights the prevailing challenges in meeting the rising energy demand. Load shedding is the intentional and controlled temporary interruption of electricity distribution that is performed to better manage demand and prevent the power grid from breaking down. According to the International Energy Agency (IEA), residential buildings account for up to 32% of the overall energy consumption [1]. In South Africa, residential buildings account for up to 23% of the total energy consumption in the country [2]. With declining energy generation and an inability to meet the increasing demand [3], thus leading to extended periods of load shedding, there is a pressing need for innovative approaches to accurately predict and manage residential energy consumption. The detrimental impacts of load shedding on various sectors of the economy [4, 5], particularly residential buildings, have highlighted the importance of addressing energy consumption challenges. Accurate prediction of energy consumption patterns can help in mitigating these impacts and promoting sustainable energy management practices.

Although several studies have explored machine learning (ML) models for predicting energy consumption in residential and commercial buildings, it was observed that there is a lack of research specifically proposing and evaluating ensemble models for predicting residential energy consumption in South Africa. Figure 1 shows the South African electricity generation from 2004 to 2014.

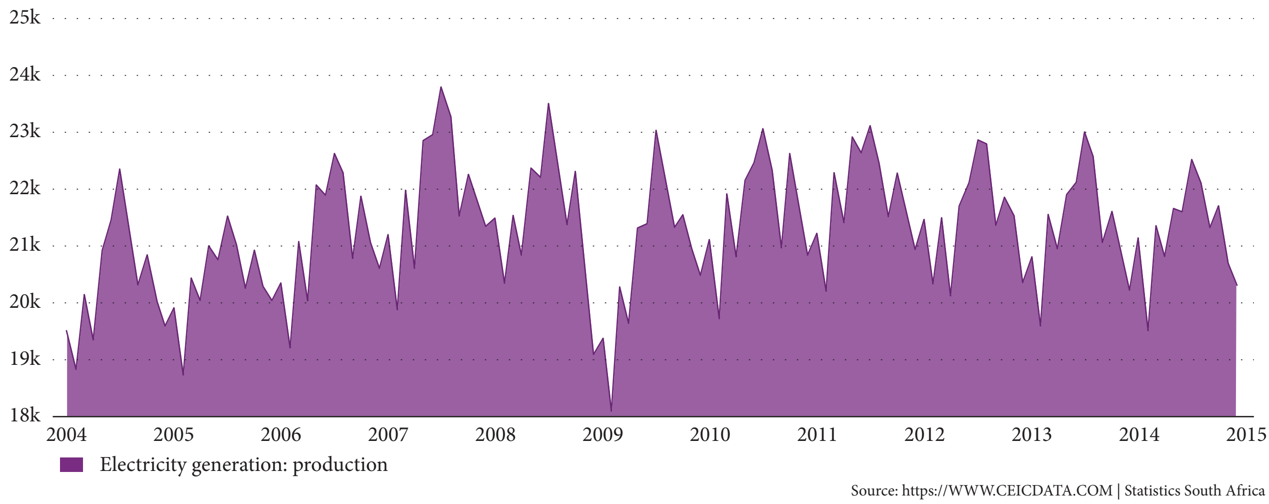


Figure 1: South African electricity generation (2004–2014) [3].

This study adopted predictive ensemble models to address the energy consumption problem. Building an ensemble model is a common method to improve the performance of the resulting model for regression tasks. Generally, it is well known that on average, an ensemble of individual predictors outperforms the single underlying predictor model. By combining the strengths of the base predictor models, the ensemble model can provide more accurate predictions for the underlying data. This study considered four ensemble models: ensemble by averaging (EA), ensemble by stacking each estimator (ESE), ensemble by boosting (EB), and ensemble by voting estimator (EVE). The base predictor models chosen in this study are the two tree models, Decision Tree (DT) and Random Forest (RF). RF can be viewed as an ensemble model, as it is essentially a collection of various DT models. Although one can only consider these ensemble models to provide a good prediction, we enhance their effectiveness by utilizing feature engineering techniques to improve learning. These techniques consider historical consumption patterns, temporal seasonality, and rolling average. This training strategy enhances the models to achieve significantly higher prediction accuracy.

1.1. Main Contributions, Novelty, and Findings of the Study. This study proposes and evaluates four ensemble models, namely, EA, ESE, EB, and EVE, for predicting residential energy consumption in South Africa. These ensemble models can mitigate drawbacks of each model and combine the strengths of individual models, such as RF and DT, to improve overall prediction accuracy. The study employs feature engineering techniques, including incorporating historical consumption patterns, temporal seasonality, and rolling averages, to enhance the predictive abilities of the ensemble models. These techniques aim to capture the temporal intricacies and trends present in the energy consumption data. We use the Domestic Electrical Load (DEL) dataset, which is recognized as the largest and most extensive study on residential energy consumption in Africa and covers diverse geographic regions, climatic zones, income groups, and dwelling structures in South Africa and Namibia, providing a representative sample for analysis.

To the best of our knowledge, this study is among the first to propose and evaluate ensemble models specifically for predicting residential energy consumption in South Africa. This study goes beyond traditional ML approaches by incorporating feature engineering techniques, such as incorporating historical consumption patterns and temporal seasonality, to enhance the predictive capabilities of the ensemble models.

This study will investigate which of the ensemble models, outperform the individual base models (RF and DT) in terms of prediction accuracy, as measured by various evaluation criteria such as mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). The study will investigate the effectiveness of ensemble models in providing accurate predictions for residential energy consumption in South Africa. This approach will provide valuable insights for researchers and practitioners in the field of building and energy research, highlighting the advantages or not of using ensemble models for predicting energy consumption in residential buildings.

The remainder of this paper is organized as follows. In Section 2, we present an overview of related work on predictive ML models and ensemble models used for predicting energy consumption in residential buildings. In Section 3, we present the proposed approach, including ML and ensemble models, dataset characterization, and a schematic framework. Section 4 presents the performance evaluation results, including the evaluation criteria and discussion of the results. Section 5 concludes the paper and provides directions for future research.

2. Related Work

Over the last decade, the application of ML models to predict energy consumption in residential and commercial buildings has attracted increasing interest from many researchers.

This stems from the fact that energy generation and consumption underpin the growth of the modern global economy. However, the amount generated does not match the consumption, especially in urban residential buildings.

Several studies on predictive ML models of energy consumption and their possible applications in optimizing energy usage in residential buildings have been conducted to address consumption issues. Unlike commercial buildings, which generally have monitoring systems in place (such as sensors) to record energy consumption and provide more granular datasets for researchers, energy consumption data from residential buildings lack granularity [6]. In [7], the authors considered the aggregated dataset from [8] and adopted ML models, including RF, DT, Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost), to predict the hourly energy consumption of South African residential buildings. Both RF and DT provided the highest accuracy for the prediction of residential energy consumption, which are time-consuming for training. In [9], an artificial neural network (ANN) model was used to improve the accuracy of energy consumption predictions during the initial design phases of residential buildings by employing parametric modeling techniques. In addition, they introduced an automated platform that enables the analysis, modeling, and simulation of building energy consumption, with a focus on both accuracy and performance.

Cheng et al. [10] implemented ensemble models using eight individual models: multiple linear regression (MLR), autoregressive integrated moving average (ARIMA), support vector regression (SVR), RF, multilayer perceptron (MLP), boosting tree (BT), multivariate adaptive regression splines (MARS), and K-nearest neighbors (kNN) to forecast the next-day energy consumption and peak demand of the tallest building in Hong Kong. While the dataset used was large and the quality of the data was well documented, the ensemble models presented low prediction accuracy, with MAPE of 2.32% and 2.85% for the next-day energy consumption and peak power demand, respectively.

Wahid et al. [11] used MLP and RF for the classification of residential buildings in terms of energy consumption, and reported that MLP outperformed RF in terms of prediction accuracy. Priyadarshini et al. [12] also presented an ML-based ensemble model for predicting energy consumption in smart homes using DT, RF, and XGBoost: while the authors provided a useful contribution to the field of energy consumption prediction, they did not discuss the quality of the data.

In a recent study to improve the forecasting potential in the early schematic design phase, Olu-Ajayi et al. [13] explored various ML models and found that deep neural network (DNN) outperformed ANN, Gradient Boosting (GB), support vector machine (SVM), RF, KNN, DT, Stacking, and logistic regression (LR). However, this novel improvement was time-consuming for training. Rahman et al. [14] proposed an ensemble model based on the Mahalanobis distance to predict the energy consumption of a smart home using a combination of ARIMA, Recurrent Neural Network (RNN), and multivariate and univariate linear regression models. They reported that the ensemble model performed better than individual prediction models. However, the generalizability of these findings to other smart home systems is limited, because the authors used data collected from a single smart home system. Moreover, the authors did not report the quality of their data.

Hosseini and Farad [15] analyzed residential energy consumption with the objective of forecasting the various factors that mostly influence energy usage in buildings, including overall height, roof area, surface area, and relative compactness. To achieve this prediction, they employed DT, RF, and KNN. Their findings showed that RF was the best model compared with the DT and KNN models in terms of prediction accuracy. Furthermore, Konhäuser et al. [16] implemented 12 ML models, including standalone models, as well as both homogeneous and heterogeneous ensemble learning models, with the aim of increasing the accuracy of predicting building energy consumption in the residential sector.

Wang et al. [17] investigated a novel model called Ensemble Bagging Trees (EBTs) to predict hourly building energy usage. To train their model, the authors used data obtained from meteorological systems, building-level occupancies, and meters. Although the authors discussed the quality of the data collected and highlighted that the proposed EBT model outperformed the Classification and Regression Tree (CART) model in predicting the hourly electricity demand of the test building, training was time consuming. Wang [18] applied a Deep Learning model to predict the energy consumption of four types of public buildings in China and found that the model performed well in terms of MAPE and RMSE; however, it was also time-consuming.

Pinto et al. [19] introduced three ensemble learning models, namely gradient boosted regression trees, RF, and an adaptation of AdaBoost, to forecast electricity consumption 1 hour ahead, using real data from an office building as a case study. The authors highlighted that the adapted AdaBoost model outperformed the other two models in terms of prediction accuracy; however, they did not report the quantity and quality of their data. Luo et al. [20] introduced a model known as GA-DFNN, which leverages genetic algorithms (GAs) to design an optimal architecture of deep feedforward neural network (DFNN). This model was used to predict the day-ahead hourly and week-ahead daily electricity consumption of a real-world campus building in the United Kingdom. Their research was based on data collected over a period of 1 year and 6 months. Although the GA-DFNN outperformed the reference models in terms of prediction accuracy, namely, a single-layer feedforward neural network, DFNN models with different architectures, long-term-short-memory (LSTM) neural network model, and temporal convolutional network (TCN) model, it is time-consuming.

Amiri et al. [21] applied XGBoost model to predict the energy usage of residential and commercial buildings in Philadelphia. Their study provided energy consumption prediction for the year 2015, which served as a reference point, and for the year 2045, which was constructed by considering economic factors such as income and employment trends. Shi et al. [22] explored the utilization of ML models in building energy management based on studies published between 1998 and 2020. They presented an integrated framework and highlighted the development trends in ML-Building Energy Management, making a valuable contribution to the existing knowledge in this field.

Liu et al. [23] explored how feature selection techniques, including filter, wrapper, and embedded methods, improve the prediction accuracy of XGBoost, LightGBM, and RF models for predicting building energy consumption. The authors used an energy consumption dataset of 478 healthcare buildings in China, and their findings indicated that the wrapper method gave the best result in terms of prediction accuracy using XGBoost model, but required more computation time.

Iftikhar et al. [24] proposed a novel ensemble learning approach for predicting monthly electricity consumption in Pakistan. The study divides electricity consumption into deterministic and stochastic components, using multiple regression and ML. The heterogeneous ensemble model outperformed other models, achieving the lowest error metrics. However, the study is limited to Pakistan.

Moon et al. [25] used ensemble learning models including RF, Gradient Boosting Machine, and CatBoost, combined with explainable artificial intelligence methods such as Shapley Additive Explanations, to improve residential building electricity consumption forecasting accuracy. The authors used the university residential complex and appliances energy prediction datasets. The Gradient Boosting Machine model performed best, while CatBoost is most effective. The study found that ensemble models outperformed deep learning models in handling noisy and high-variability data, with historical consumption patterns and temperature–humidity index being key predictors. However, one of the limitations includes high error rates.

The study in [26] predicted energy consumption in U.S. residential buildings using ML algorithms and the residential energy consumption survey dataset. It developed separate prediction models for apartments and single-family houses, revealing key features influencing energy use intensity. The LightGBM-based model performed best for apartments, while CatBoost-based model performed best for single-family houses. The study highlights the need for separate prediction models for different building types.

Kumaraswamy et al. [27] have developed a hybrid neural network model that combines LSTM networks with Feed Forward Neural Networks (FFNN) to improve energy predictions. The model is integrated with the Stationary Wavelet Transform (SWT) method to reduce instability and increase data dimensionality, leading to better forecasting accuracy. However, the model's performance depends on the quality and quantity of data used for training and validation, and its complexity may lead to challenges in model interpretability and computational resources. The model's applicability for long-term forecasting is limited, and a detailed comparison with baseline models is needed to fully understand its scope and applicability.

Compared with other works, this study concentrates on the predictive modeling of energy consumption in South African residential buildings using ensemble models with real-world data. While the authors in [7, 28, 29] proposed different techniques for predicting energy consumption of residential building, this study investigates whether or not the ensemble models proposed outperform their individual models in terms of prediction accuracy.

Table 1 presents some of the recent studies on ML-based predictive energy consumption, their strengths and weaknesses. In summary, researchers have proposed several ML models for predicting energy consumption in residential buildings. However, to the best of our knowledge, no study has proposed ensemble learning models for the energy consumption in residential buildings in South Africa. Table A1 in Appendix A presents the nomenclature of the abbreviation used in this study.

Furthermore, in reviewing the related literature, we noted a critical short coming in how the quality and quantity of the underlying dataset are reported. In this study, we provide an overview of the extensive consumption dataset used, and adopt the ensemble to predict the consumption.

3. Approach

This section provides a brief overview of the ML models used as base predictor models to build an ensemble. For a more detailed description, refer to [30, 31] and other related studies therein. We explored the energy consumption dataset collected over 2 decades in [8].

3.1. ML and Ensemble Models. Here, we introduce the 2 ML models and four ensemble models used in this study. As mentioned earlier, the ML models form the basis of the proposed ensemble models. In [7], the authors introduced several ML models to address the South African residential energy dataset, and it was shown that RF and DT outperformed other boosted tree-based models. We briefly outline the models as follows: DT, RF, EA, ESE, EB, and EVE.

3.1.1. DTs. A DT is a supervised ML algorithm that uses a tree-like structure to make predictions. It splits the dataset into different branches based on different features and creates a tree of decisions that leads to the final prediction. Each internal node of the tree represents a feature, each branch represents a possible value for that feature, and each leaf node represents a class label or prediction. DTs are known for their interpretability and their ability to handle both categorical and numerical data. The process of constructing the tree recursively partitions the data into subsets based on feature values with the objective of minimizing the MSE of the predictions. Figure 2, presents an example of an algorithm for building DT.

3.1.2. RF. RF is an ML model that constructs a collection of decision trees and aggregates their predictions to enhance accuracy and mitigate overfitting. The construction of each tree involves the use of a random subset of the training data,

Table 1: Summary of the related work on machine learning-based predictive energy consumption in residential buildings.

References	Machine learning models	Strengths of the study	Weaknesses of the study
Moulla et al. [7]	RF, DT, XGBoost, AdaBoost	Higher prediction accuracy	Time-consuming for training
Elbeltagi and Wefki [9]	ANN	Handle non-linear data.	The authors did not report the quantity and quality of the data collected.
Fan et al. [10]	MLR, ARIMA, SVR, RF, MLP, BT, MARS, and kNN	Large dataset and the quality of the data was well-documented.	Lower prediction accuracy.
Wahid et al. [11]	MLP, RF	Perform well with small dataset.	The dataset is relatively small, require higher computational speed with large dataset, lower prediction accuracy.
Priyadarshini et al. [12]	DT, RF, and XGBoost	Higher prediction accuracy.	The authors did not discuss the quality of the data, time-consuming for training.
Olu-Ajayi et al. [13]	DNN, ANN, GB, SVM, RF, KNN, DT, stacking, and LR	Perform well with large dataset.	Time-consuming for training.
Rahman et al. [14]	ARIMA, RNN, Multivariate and Univariate linear regression models.	Adapted for time series forecasting and regression tasks.	The dataset is relatively small. The authors did not report the quality of the data collected.
Hosseini and Farad [15]	DT, RF, and KNN	Perform well with large dataset.	Lower prediction accuracy. The dataset is relatively small.
Konhäuser et al. [16]	SVR, MLP, KNN, RCV, DT, ABR, BGR, RF, ETR, XGB, STR, and AVR	Large dataset, the authors discuss the quality of the data collected.	Lower prediction accuracy.
Wang et al. [17]	EBT	Higher prediction accuracy and the quality of the data was well-documented.	Time-consuming for training.
Pinto et al. [19]	Gradient boosted regression trees, RF, and an adaptation of AdaBoost	Higher prediction accuracy	The authors did not report the quality of the data collected.
Luo et al. [20]	GA-DFNN, DFNN, LSTM, and TCN	Higher prediction accuracy.	Time-consuming for training.
Liu et al. [23]	XGBoost, LightGBM, and RF	Higher prediction accuracy.	Require more computation time.

Input: an attribute-valued dataset D
1. Tree = {}
2. if D is 'true' or other stopping criteria met then
3. terminate
4. end if
5. for all attribute $a \in D$ do
6. compute information-theoretic criteria if we split on a
7. end
8. a_{best} = Best attribute according to above computed criteria
9. Tree = Create a decision node that tests a_{best}
10. D_v = Induced sub-datasets from D based on a_{best}
11. for all D_v do
12. $Tree_v = C4.5\ (D_v)$
13. Attach $Tree_v$ to the corresponding branch of Tree
14. end for
15. return Tree

Figure 2: Example of decision tree algorithm (C4.5) [30].

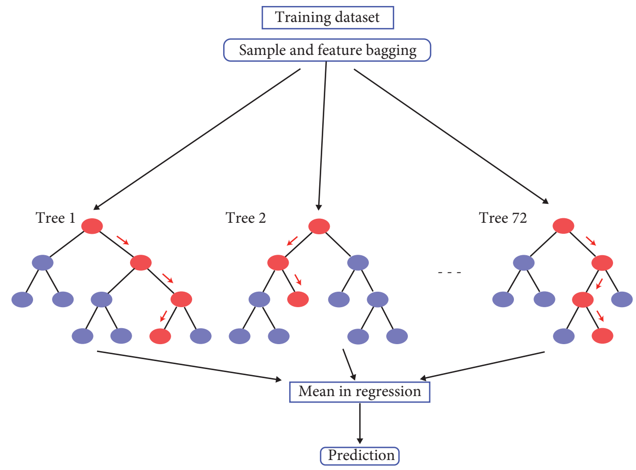


Figure 3: RF algorithm as described by Breiman in 2001.

a technique known as bagging [32], as an example of a bagging algorithm. In addition, a random subset of features is employed, which is referred to as feature bagging. The final prediction is determined by calculating the average or taking a majority vote among the predictions generated by each individual tree.

Figure 3 illustrates Breiman's 2001 description of the RF algorithm. In essence, RFs are ensembles of binary decision trees. Each node in the decision tree represents a condition based on a single characteristic. This condition is chosen to partition the dataset into two subsets, ensuring that samples with similar characteristics are grouped together. RFs are observable, unaffected by scaling and other feature transformations, resistant to the inclusion of irrelevant features, and capable of estimating the importance of features via mean decrease in impurity (MDI).

3.1.3. Ensemble Models. As mentioned previously, ensemble learning is a robust ML model that combines multiple base models to produce the best possible output. It has become highly popular owing to its exceptional ability to generalize [10]. These models frequently exhibit superior performance compared with their constituent individual models. There are three reasons for this [33]. First, the training data may not provide sufficient information to select the optimal model; hence, integrating models with comparable performances could be a more favorable option. In addition, ensembles can mitigate the limitations of the individual search processes.

Furthermore, it is important to note that in practical applications, the existence of a true target function may be uncertain or nonexistent. In this case, ensembles can offer a reasonably accurate approximation, leading to improved generalization performance [10, 33].

Ensemble learning has been applied in diverse domains, including face recognition, medical diagnosis, and gene expression analysis. In this study, we reviewed four ensemble models: EA, ESE, EB, and EVE. Each ensemble model and its specific uses are described as follows:

- EA integrates predictions from many base models by averaging their results, making it one of the most straightforward ensemble models. By combining several models, one may reduce the effect of errors in individual models and obtain a more robust and precise prediction [30]. In our case, we take the average of the outputs of the two base models RF and DT.
- A more advanced ensemble model is the ESE, which is often known as stacked generalization. This involves using predictions from various base models to train a metamodel (typically boosting tree models). The metamodel learns to make predictions based on base model predictions, which are interpreted as new features. Stacking tends to be more accurate than averaging because it often captures the complex relationships between the predictions of base models.
- EB prioritizes the fixing errors made by prior models. It operates by sequentially training the base models, with each new model assigning more weight to instances incorrectly predicted by the prior models.
- Another simple ensemble model is the EVE, which integrates predictions from many models by determining the majority votes or weighted averages. Voting ensembles can be divided into two categories: hard voting and soft voting. In hard voting, each model casts one "vote," and the prediction with the most votes win out as the result. Soft voting selects the class with the highest average probability by averaging the probabilities predicted by each model.

Although we could have considered several other ensemble models, we chose these models based on the rationales, advantages, and limitations outlined in Table 2.

In summary, EA provides a simple and computationally efficient way to combine the base models (RF and DT), reducing individual model errors and variance. The ESE captures complex relationships between the base model predictions, potentially improving accuracy. The EVE provides stable predictions. The EB iteratively focuses on

Table 2: Rationales, advantages, and limitations of the chosen ensemble models.

Ensemble method	Rationale	Advantages	Limitations
Averaging (EA)	Simple and reduces variance.	Easy to implement, computationally efficient.	May not capture complex relationships and treats all models equally.
Stacking (ESE)	Captures complex relationships.	Can lead to improved performance and leverages strengths of different models.	Computationally intensive, requires meta-model tuning, and is sensitive to correlations,
Boosting (EB)	Corrects errors iteratively.	Effective with weak learners, and adapts to difficult instances.	Sensitive to noisy data and requires hyperparameter tuning.
Voting (EVE)	Leverages collective wisdom.	Easy to implement and provides stable predictions.	May not capture complex relationships, equal weight to all models.

correcting mistakes made by previous models, adapting to difficult instances.

Several authors have considered ensemble models, other than those listed in Table 2. The objective of this study was to explore a diverse range of ensemble models to determine the model that best suits the data without being exhaustive.

3.2. Dataset and Schematic Framework

3.2.1. Dataset Description. This study utilized the DEL dataset for South Africa [8, 34] to evaluate the performance of the four ensemble learning models. The dataset includes metered household electricity consumption data covering a diverse population sample that encompasses urban, informal (township settlements), and rural environments; five climatic zones; a wide range of income groups; households that have recently been electrified; those that have had electricity for a long time; and various types of dwelling structures. This dataset was collected in South Africa and Namibia in [8, 34]. This dataset is widely recognized as the largest and most extensive study on residential energy consumption in Africa [8]. We examined the DEL metering hourly data, which is an aggregation of consumption current (Amps) data over 1 hour. Although the dataset spans 20 years and has undergone multiple annual validations and testing procedures to ensure reliability [8], this study only considered the span from 2004 to 2008. This choice is based on the size of the dataset and the fact that electricity generation in South Africa peaked in 2007 (see Figure 1) [3].

The National Rationalized Specification Load Research (NRSLR) program and DEL study dataset contributed significantly to the electrification of South African households. In particular, this study influenced power system design, improved load specifications, and the development of new technologies for national energy provider (Eskom) and municipalities. This dataset is the most comprehensive electricity usage dataset collected in Southern Africa. In terms of geographical factors, Figure 4 shows the spread of the data-collection process across Southern Africa.

Furthermore, from 1994 to 2014, the DEL dataset comprised granular electricity meter readings obtained at 5-min intervals and household surveys that gathered socioeconomic information about metered households and certain non-domestic entities in South Africa and Namibia [8, 34]. Information such as household income, number of people in household, location, number of employed individuals, and so forth, were recorded.

3.2.2. Schematic Framework. As mentioned earlier, residential electricity consumption data lack granularity and thus an aggregation model must be adopted to structure the data for research. The aggregation of the DEL dataset followed a well-structured data processing regime to remove all invalid readings and missing values [8, 34]. The original data collected was a 5-min interval electricity metering dataset, and the observations were later aggregated to hourly values. Details of the aggregation model are outlined in [8]. Figure 5 illustrates the schematic framework adopted to build the

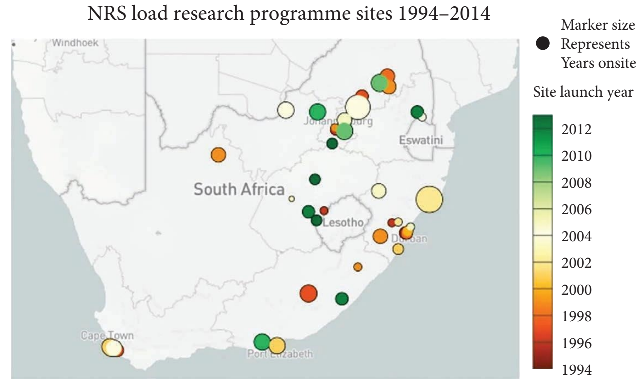


Figure 4: Map view of DEL study data collection sites, 1994–2014.

prediction model. To train the model, the initial 80% of the data were chosen as the training dataset, while the remaining 20% were used to assess the accuracy and effectiveness of the model.

The adopted schematic framework involves several steps. Algorithm 1 presents the steps involved, along with a detailed explanation of the purpose and reasoning behind each step, for predicting residential energy consumption.

This algorithm outlines the key steps involved in the proposed schematic framework for predicting residential energy consumption using ensemble models. It starts with obtaining and preprocessing the DEL dataset, followed by splitting the data into training and testing sets. Feature engineering techniques are then applied to enhance the predictive capabilities of the models. The ensemble models are trained using the training set, and their performance is evaluated on the testing set using various evaluation criteria. Finally, the best-performing ensemble model is selected and deployed for predicting future residential energy consumption patterns, providing valuable insights for energy management and decision-making processes.

3.2.3. Feature Extraction and Engineering. While the authors in [8] performed a lot of work on data preprocessing, we transformed the data to suit our needs by combining the various years (2004–2008) together and extracted the relevant features. The extracted DEL dataset contained approximately 19 million rows and five columns. The first column indicates RecorderID, and the second column documents ProfileID. The third to final columns give a date and time (Datefield), indicate units read (Unitsread), and indicate whether the details captured are valid (Valid), respectively. Further preprocessing was performed to remove noise from the data prior to model training, which reduced the dataset to approximately 17 million rows.

To capture and understand the temporal dependencies within the dataset, a feature engineering technique incorporating the values of lag1 and lag2 was performed. This enables the model to incorporate the most recent historical data of the time series to make accurate predictions [11]. A rolling mean feature was also implemented, and all features in the test and training split data were normalized. Normalizing the features helps to maintain consistency while enabling the interpretation of feature importance.

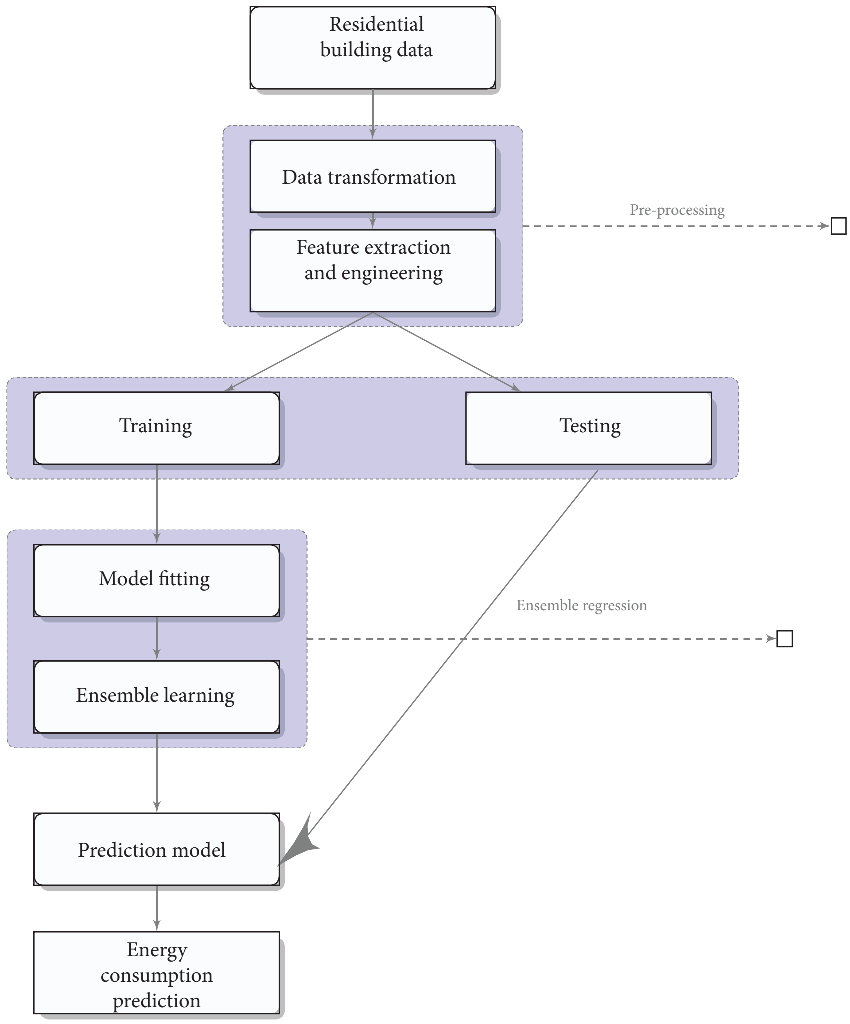


Figure 5: Schematic framework of predicting energy consumption.

Consequently, the overall performance of the model improves. Table 3 presents a summary of the dataset indicating the spread and count of Unitsread and ProfileID.

We considered temporal trends within the dataset, including features such as hour, day of the week, day of the year, quarter, month, and year of consumption. Training and testing of the models were performed using these features. The feature importance is presented in Table 4.

We also observed that, after training, testing, and validation, the predictions were not as accurate; thus, we employed feature engineering techniques to enhance the model prediction. As mentioned in the study, the lag features include past values of the target variable. In the code, the lag features, "lag_1" and "lag_2" were created by shifting the electricity consumption, Unitsread, by one time step. We further considered the moving average to capture trends and seasonality in the data so that the initial feature importance may not have been captured. In the code, a rolling mean feature was calculated to smooth out fluctuations in consumption over time. After validation, these engineered features aim to improve the prediction results of the models.

4. Performance Evaluation

4.1. Settings of the Study. In this study, the ensemble models were trained using features extracted from the DEL dataset using Python libraries. Table 5 provides the information on the environment used in this study. The data and code used to support the findings of this study are available from the corresponding author upon reasonable request.

4.2. Evaluation Criteria. To evaluate the performance of the four ensemble models outlined in Section 3.1.3, this study considered the following four criteria commonly used in statistical analyses: the MAE, MSE, MAPE, and R^2. MAE is a commonly used criterion in ML to quantify the average absolute deviation between predicted and actual values [35]. Moreover, MAE assigns equal weights to all errors. In addition, the mean squared deviations between the expected and actual values are quantified using MSE. The algorithm assigns a higher penalty to larger errors than to smaller errors [35].

Input: D ▷ DEL dataset
Output: Predicted residential energy consumption
1. Obtain the DEL dataset
2. Preprocess the dataset
    2.1. Remove invalid readings, missing values, and outliers
    2.2. Aggregate data to a suitable time interval (hourly)
3. Split the preprocessed dataset into {D1, D2} ▷ Training and testing sets
    3.1. Assign 80% of the data to D1 ▷ Training set
    3.2. Assign 20% of the data to D2 ▷ Testing set
4. Perform feature engineering on D1
    4.1. Incorporate historical consumption patterns
    4.2. Include temporal seasonality features (hour, day of week, day of year, quarter, month, and year)
    4.3. Calculate rolling averages to capture trends and seasonality
5. Train the ensemble ML models on D1
    5.1. Train EA model
    5.2. Train ESE model
    5.3. Train EB model
    5.4. Train EVE model
6. Evaluate the performance of the trained models on D2
    6.1. Calculate MSE
    6.2. Calculate MAE
    6.3. Calculate MAPE
    6.4. Calculate R^2
7. Select the best-performing ensemble model based on the evaluation criteria
8. Deploy the selected ensemble model for predicting residential energy consumption
    8.1. Use the trained model to predict future energy consumption patterns
    8.2. Provide insights and support energy-related decision-making processes

ALGORITHM 1: Schematic Framework for Predicting Residential Energy Consumption.

TABLE 3: Summary of the DEL dataset.

Statistics	ProfileID	Unitsread
Count	18,890,940	18,890,940
Mean	1,004,008	2.64
Standard deviation	1359.01	5.18
Minimum	1,001,634	0.00
25%	1,002,766	0.13
50%	1,004,101	0.97
75%	1,005,272	2.75
Maximum	1,006,659	98.99

TABLE 4: Feature importance of the ensemble models.

Feature importance	
Hour	0.068654
Day of week	0.014288
Day of year	0.059465
Quarter	0.000000
Month	0.000000
Year	0.857593

TABLE 5: The platform environment of the study.

Parameter	Value
OS	Red Hat Enterprise Linux 7.4.
CPU	• 128 standard compute nodes with dual Intel Xeon E5-2690 v4 CPUs (2.6 GHz, 14 cores per socket, 28 cores per node) • Supports AVX2 instruction set for optimized performance
RAM	• 256 GB RAM per standard compute node • 2 "Fat" nodes with 1 TB RAM each (64 cores per fat node)
Libraries	sklearn, imblearn, xgboost, matplotlib.pyplot, numpy, pandas, os

In addition, MAPE provides the average absolute percentage difference between the predicted and actual values. This is particularly useful when errors are represented as the ratio of true values. The R^2 score quantifies the percentage of the variance in the dependent variable that can be accounted for by the independent variables. The output is a numerical value ranging from 0 to 1, with a value of 1 indicating a perfect fit. The four evaluation criteria are defined in Equations (1)–(4), respectively, as follows:

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n}, \tag{1}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i}, \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}, \tag{4}$$

where y_i is the actual measurement, $\hat{y}_i$ the predicted value, $\overline{y}_i$ the mean of the actual target values, and n the number of measurements.

The relevance and limitations of MSE, MAE, MAPE, and R^2 in this context are presented next:

4.2.1. Relevance

- MSE and MAE: both capture the difference between the predicted and actual energy consumption. Lower values indicate a better prediction. MSE penalizes larger errors more heavily, whereas MAE focuses on the average magnitude of errors.
- MAPE: this is useful when dealing with data containing significant fluctuations, as it expresses errors as a percentage of actual consumption. This allows for a fair comparison across different consumption levels.
- R^2: indicates how well the predicted values align with the actual trends. A higher R^2 value suggests a strong correlation between the predicted and actual values.

4.2.2. Limitations

- MSE: sensitive to outliers. A single large error can significantly inflate the MSE, potentially masking an otherwise accurate prediction.
- MAE: does not consider the error direction. An underestimation by the same amount as an overestimation will have the same MAE, which may not be ideal.
- MAPE: not suitable for cases where actual consumption values are close to zero, as it can lead to division by zero errors.
- R^2: a high R^2 can occur even with a consistent under- or overestimation by the model. This only reflects the strength of the linear relationship.

Wilcoxon and analysis of variance (ANOVA) statistical tests were performed to ensure the quality of the base models (RF and DT). We ran the tests for three and five years of consumption data, that is, 2004–2006 and 2004–2008, respectively. Table 6 presents the results of Wilcoxon and ANOVA tests for the selected periods.

We can observe a higher p value (typically more than 0.05) for both the Wilcoxon and ANOVA tests for the two groups of data (2004–2006 and 2004–2008, respectively). Because the p value is greater than the level of significance (p value > 0.05) for the Wilcoxon and ANOVA tests, the null hypothesis is accepted for both statistical tests, and we can conclude that there is no statistically significant difference in performance between the RF and DT models.

To validate the performance comparisons between the RF and DT models, we considered the Bland–Altman plot test. The Bland–Altman plot compares the performance of the RF and DT models by plotting the difference between their predictions against the average of their predictions. Figure 6 presents the Bland–Altman plot test.

Figure 6 shows a horizontal line at 0, indicating perfect agreement between the models. The majority of the data points are clustered around this line, suggesting that the two models generally agree in their predictions. However, there is some spread of data points, with some exhibiting large differences from the line of perfect agreement. This indicates that while the models often agree, there are also cases where their predictions diverge.

The spread of data points appears to increase as the average prediction value increases, suggesting that the discrepancy between the models becomes more pronounced for higher prediction values.

In addition, there seems to be a slight negative trend in the data points, implying that the RF model tends to predict slightly lower values compared with the DT model for higher average predictions.

Overall, the Bland–Altman plot suggests that while the RF and DT models generally agree in their predictions, there are instances where their performance diverges, especially for higher prediction values.

4.3. Results and Discussion. In this section, we present the results of the ensemble models. Owing to the size of the dataset, we ran the algorithms for each model using the University of South Africa (UNISA) high performance computing (HPC) system. As mentioned earlier, we were interested in the accuracy of the presented predictive models. We ran the models for three and five years of consumption data, that is, 2004–2006 and 2004–2008, respectively. The selected periods are shown in Figure 7. Figure 7(a) shows the data related to the time intervals from 2004 to 2006 and 2004 to 2008. Figure 7(b) presents the electricity consumption for the week from October 1 to 8, 2004, and the training and testing data split.

Table 6: Wilcoxon and ANOVA tests results.

Statistical tests	Base model	
	RF vs. DT (2004–2006)	RF vs. DT (2004–2008)
Wilcoxon signed-rank test	p value = 0.25 (p value > 0.05)	p value = 0.38 (p value > 0.05)
One-way ANOVA test	p value = 0.99 (p value > 0.05)	p value = 0.99 (p value > 0.05)

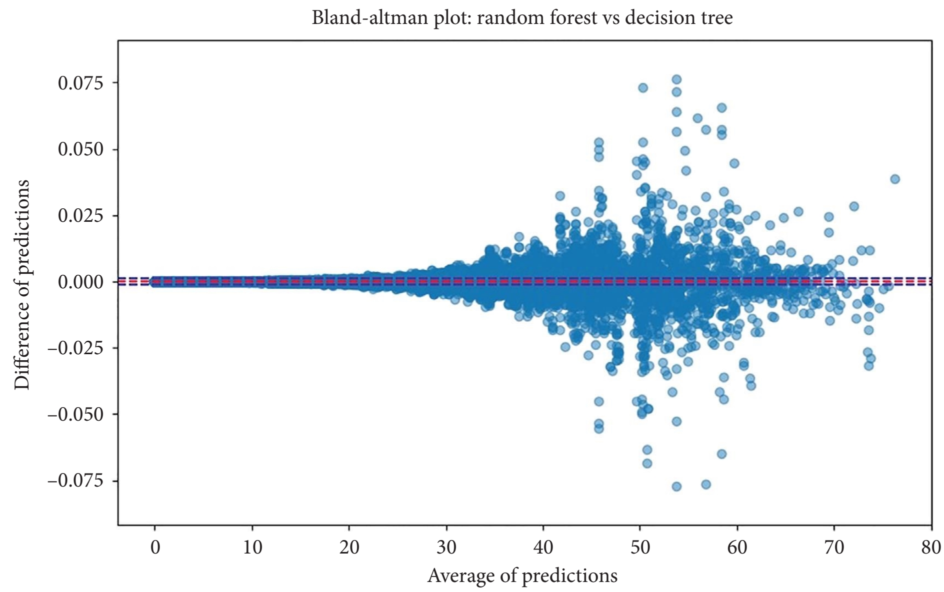
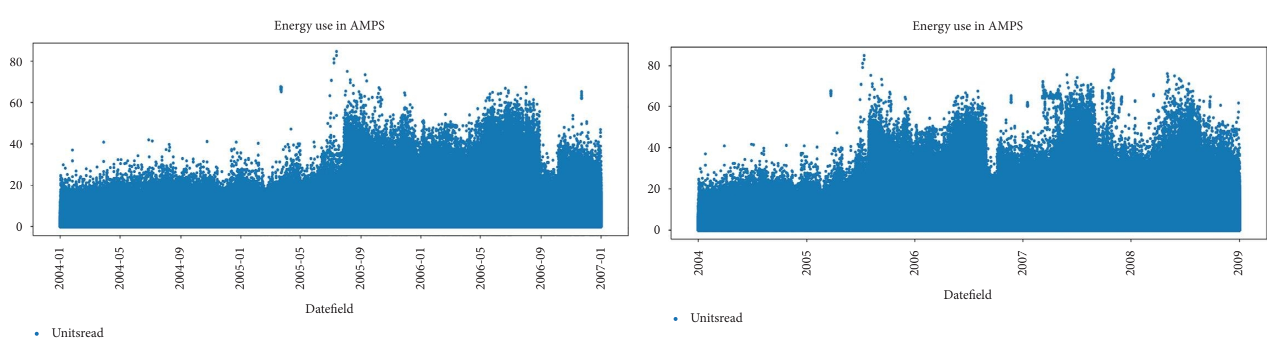
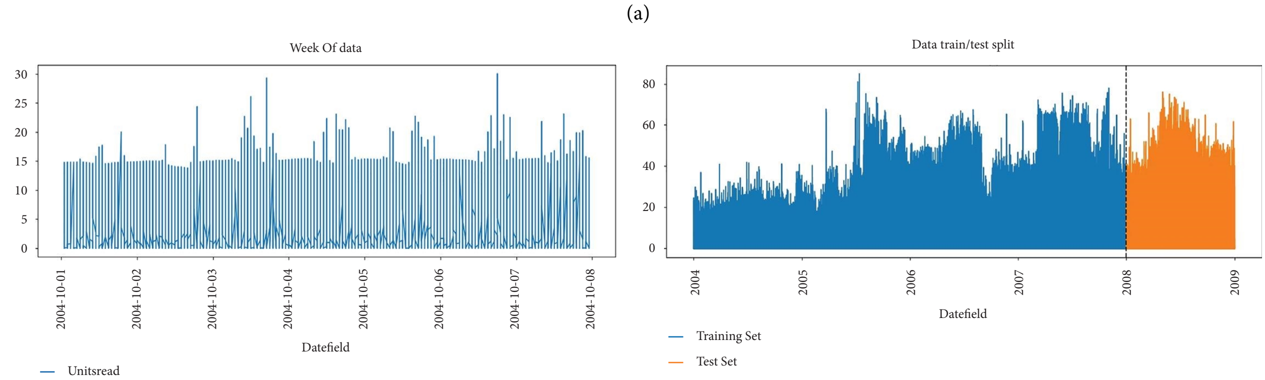


Figure 6: Bland–Altman plot test.



(a)



(b)

Figure 7: Energy consumption data showing the various weekly, hourly, and yearly consumptions from 2004 to 2008. (a) Energy consumption data from 2004 to 2006 and 2004 to 2008. (b) Weekly (first week of October 2004, that is 10-01-2004) and training and testing split.

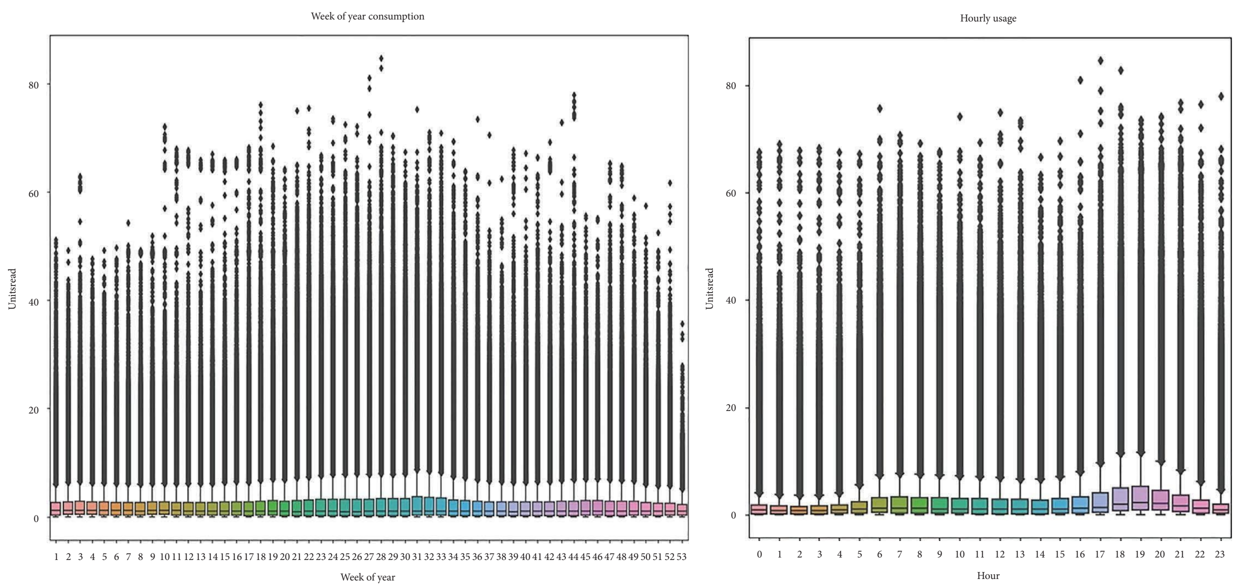


Figure 8: Week of year and hourly energy consumption from 2004 to 2008.

Furthermore, Figure 8 shows the relationship between the week of the year and hourly consumption of the data over the course of the year. This visual representation allows for the observation and analysis of any potential patterns or trends that may exist between these two variables.

Tables 7 and 8 present the prediction performance (hourly data) for the four ensemble models presented in Section 3.1.3, as well as the underlying ML base models. Table 7 shows the performance of each model in relation to four criteria: MSE, MAE, MAPE, and R^2. The best score is marked in bold text. As mentioned previously, we chose RF and DT as the base models because of their prediction accuracy in [6]. Table 7 presents the performance for the 3 years of data, from 2004 to 2006. From Table 7, the ensemble models EA and EVE outperformed the base models RF and DT, respectively. As observed, the prediction error was very close to zero, which indicates that these models performed well (with very high prediction accuracy) for the energy consumption dataset. The R^2 of course shows how well the model fits the data. Furthermore, we can observe the challenges of EB and ESE in accurately predicting consumption data. It is important to note that EB is not able to fit the data, and further analysis is required to improve the accuracy of the model.

In Table 8, regarding the MSE, the ensemble models (EA and EVE) outperformed the base models (RF and DT). This case was slightly different for MAE and MAPE, where the base models slightly outperformed the ensemble models (EA and EVE). All of these models present very good predictions of the data. In general, tree-based ensemble models outperformed boosted and stacking ensemble models with respect to the electricity consumption dataset. We note that this may not necessarily be the case for other datasets [16, 17].

4.4. Limitations of the Study. Limitations refer to influences or shortcomings that are beyond researchers' control and place restrictions on the methodology and analysis of research data [36]. The limitations of this study related to the research problem under investigation are as follows:

- In this study, we focused only on the accuracy of the ensemble models without considering computational efficiency.
- More elaborate exploratory data analysis related to the utilized dataset could be provided.
- The ensemble models used default hyperparameters. In future work, these hyperparameters can be optimized to improve prediction accuracy.
- The specific characteristics of the South African residential energy consumption dataset could be further elaborated.
- More statistical tests could be performed on the data to ensure the quality of the proposed models.

Although the proposed ensemble models showed promising results in predicting residential energy consumption with the DEL dataset, there are some important limitations to consider regarding the representativeness of this dataset and potential biases.

The representativeness of the DEL dataset of the entire South African population should be examined. The DEL dataset covers a diverse range of urban, informal (township), and rural environments, as well as various climatic zones and income groups. However, the specific sampling methodology and the extent to which different demographic groups are represented in the dataset are not clearly delineated. If certain regions or socioeconomic segments of the population are underrepresented or overrepresented, the

Table 7: Performance of the prediction models for three years of data (2004–2006).

	Base model			Ensemble model		
Criteria	DT	RF	EB	EA	ESE	EVE
MSE	9.534e – 06	3.965e – 06	18.662	2.009e – 07	3.162	**1.824e – 07**
MAE	5.940e – 05	3.667e – 05	3.055	2.689e – 05	0.341	**2.664e – 05**
MAPE	1.858e – 03	1.334e – 03	5496	1.355e – 04	5.601	**1.34e – 04**
R^2	**0.99**	**0.99**	0.43	**0.99**	0.90	**0.99**

Note: Table presents the performance for each model in relation to the four criteria, MSE, MAE, MAPE, and R_2, and the best score is marked in bold text.

Table 8: Performance of the prediction models for five years of data (2004–2008).

	Base model			Ensemble model		
Criteria	DT	RF	EB	EA	ESE	EVE
MSE	3.112e – 07	3.561e – 07	18.332	**6.014e – 08**	5.181	6.08e – 8
MAE	7.795e – 06	**5.885e – 06**	3.673	9.412e – 06	0.327	9.464e – 06
MAPE	**2.263e – 04**	2.582e – 04	11,137	3.014e – 04	5.724	3.014e – 04
R^2	**0.99**	**0.99**	0.53	**0.99**	0.87	**0.99**

Table shows the performance for each model in relation to the four criteria, MSE, MAE, MAPE, and R_2, and the best score is marked in bold text.

predictive models may exhibit biases and fail to capture the consumption patterns accurately for those groups.

While the timeframe was chosen due to the peak in electricity generation in South Africa in 2007, it is essential to consider potential changes in residential energy consumption patterns over time.

5. Conclusion and Future Work

We introduced an array of ensemble ML models that have been employed to predict electricity consumption patterns observed in residential buildings in South Africa. We conducted a comparative analysis of various ensemble models with the underlying base machine learning models, including RF and DT to improve the prediction of consumption patterns in a previous study. The ensemble models considered in this study were averaging, stacking, voting, and boosting models. The findings of this study indicate that, in general, the averaging and voting ensemble models improve the predictive ability of the underlying base models with respect to the electricity data provided.

Predicted hourly national consumption can offer useful insights to firms, individuals, and government officials, which can help them make well-informed decisions. These insights include peak time identification and consumption patterns at the local and national levels. By leveraging this data, more accurate and comprehensive projections of energy consumption can be generated, enabling stakeholders to take preventative steps. In addition, this study contributes to the continuing conversation on sustainable energy management by fusing recent advances in machine learning with actual energy-related problems. In addition, this work offers practitioners and academics useful information on how to estimate the consumption of energy in residential buildings and the advantages of applying ensemble learning models in the building and energy research fields.

The practical implications of this research are significant. By accurately predicting residential energy consumption patterns, stakeholders such as energy providers, policymakers, and households can make informed decisions to optimize energy usage and mitigate issues like load shedding. Accurate predictions enable energy providers to better manage supply and demand, reducing the risk of grid failures and prolonged power outages. Furthermore, households can leverage these predictions to implement energy-saving measures during peak demand periods, potentially leading to cost savings and contributing to a more sustainable energy landscape. Policymakers can also use these insights to develop targeted initiatives and incentives for promoting energy efficiency in the residential sector, aligning with national goals for energy security and environmental sustainability.

In future work, we plan to investigate the quantitative data used in greater depth. We also plan to conduct further analyses by considering other attributes (features) such as appliance usage, features related to buildings (square footage, floors per building, etc.), and energy tariffs. We also plan to explore the potential of incorporating ensemble and integrated ensemble models to enhance the prediction accuracy of this study. In addition, we will consider incorporating other relevant factors, such as weather conditions and occupancy patterns, to further refine our analysis. Furthermore, we plan to adopt techniques that can prescribe IoT solutions based on the DEL dataset. We plan to add a comparison of standard single models with ensemble models in terms of computational efficiency. In future work, we plan to investigate the performance criteria used in a much deeper manner. Moreover, we plan to conduct further performance evaluations by considering additional criteria relevant to the energy consumption prediction.

By addressing the limitations of this study, future research can further refine and improve the accuracy and generalizability of ensemble ML models for predicting residential energy consumption, contributing to more informed decision-making processes and effective energy management strategies.

Appendix A: List of Abbreviation

Table A1: Nomenclature of the abbreviation used.

ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
BT	Boosting tree
CART	Classification and regression tree
DEL	Domestic electrical load
DFNN	Deep feedforward neural network
DNN	Deep neural network
DT	Decision tree
EA	Ensemble by averaging
EB	Ensemble by boosting
EBT	Ensemble bagging trees
ESE	Ensemble by stacking each estimator
ETR	Extra trees regressor
EVE	Ensemble by voting estimator
GB	Gradient boosting
GA	Genetic algorithm
IEA	International energy agency
IoT	Internet of things
kNN	k-nearest neighbors
LR	Logistic regression
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression splines
MDI	Mean decrease in impurity
ML	Machine learning
MLP	Multi-layer perceptron
MLR	Multiple linear regression
MSE	Mean squared error
NRSLR	National Rationalized Specification Load Research
RF	Random Forest
RNN	Recurrent neural network
SVM	Support vector machine
SVR	Support vector regression
TCN	Temporal convolutional network
XGB	Extreme gradient boosting
XGBoost	Extreme gradient boosting

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research was supported by the University of South Africa under Grant no. 409000.

References

[1] International Energy Agency, "Africa Energy Outlook 2019," (2019), https://www.iea.org/reports/africa-energy-outlook-2019.

[2] Enerdata, "South Africa Energy Information 2022," (2022), https://www.enerdata.net/estore/energy-market/south-africa/.

[3] "Electricity, Gas and Water Supply Industry," (2021), https://www.statssa.gov.za/publications/Report-41-01-02/Report-41-01-022019.pdf.

[4] J. L. Erero, "Impact of Loadshedding in South Africa: A CGE Analysis," *Journal of Economics and Political Economy* 10, no. 2 (2023): 78–94.

[5] M. V. Mabunda, R. M. Mukonza, and L. R. Mudzanani, "The Effects of Loadshedding on Small and Medium Enterprises in the Collins Chabane Local Municipality," *Journal of Innovation and Entrepreneurship* 12, no. 1 (2023): 57–20, https://doi.org/10.1186/s13731-023-00327-7.

[6] R. E. Edwards, J. New, and L. E. Parker, "Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study," *Energy and Buildings* 49 (2012): 591–603, https://doi.org/10.1016/j.enbuild.2012.03.010.

[7] D. K. Moulla, D. Attipoe, E. Mnkandla, and A. Abran, "Predictive Model of Energy Consumption Using Machine Learning: A Case Study of Residential Buildings in South Africa," *Sustainability* 16, no. 11 (2024): 4365, https://doi.org/10.3390/su16114365.

[8] T. Wiebke, "Domestic Electrical Load Metering," *Hourly Data 1994-2014* (2014).

[9] E. Elbeltagi and H. Wefki, "Predicting Energy Consumption for Residential Buildings Using ANN through Parametric Modeling," *Energy Reports* 7 (2021): 2534–2545, https://doi.org/10.1016/j.egyr.2021.04.053.

[10] C. Fan, F. Xiao, and S. Wang, "Development of Prediction Models for Next-Day Building Energy Consumption and Peak Power Demand Using Data Mining Techniques," *Applied Energy* 127 (2014): 1–10, https://doi.org/10.1016/j.apenergy.2014.04.016.

[11] F. Wahid, R. Ghazali, A. S. Shah, and M. U. Fayaz, "Prediction of Energy Consumption in the Buildings Using Multi-Layer Perceptron and Random Forest," *International journal of advanced science and technology* 101 (2017): 13–22, https://doi.org/10.14257/ijast.2017.101.02.

[12] I. Priyadarshini, S. Sahu, R. Kumar, and D. Taniar, "A Machine-Learning Ensemble Model for Predicting Energy Consumption in Smart Homes," *Internet of Things* 20 (2022): 100636–100718, https://doi.org/10.1016/j.iot.2022.100636.

[13] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building Energy Consumption Prediction for Residential Buildings Using Deep Learning and Other Machine Learning Techniques," *Journal of Building Engineering* 45 (2022): 103406–103413, https://doi.org/10.1016/j.jobe.2021.103406.

[14] S. Rahman, M. G. Rabiul Alam, and M. Mahbubur Rahman, "Deep Learning Based Ensemble Method for Household Energy Demand Forecasting of Smart Home," in *Proceedings of the 2019 22nd International Conference on Computer and Information Technology (ICCIT)* (Dhaka, Bangladesh: IEEE, March 2020), 1–6.

[15] S. Hosseini and R. H. Fard, "Machine Learning Algorithms for Predicting Electricity Consumption of Buildings," *Wireless Personal Communications* 121, no. 4 (2021): 3329–3341, https://doi.org/10.1007/s11277-021-08879-1.

[16] K. Konhäuser, S. Wenninger, T. Werner, and C. Wiethe, "Leveraging Advanced Ensemble Models to Increase Building Energy Performance Prediction Accuracy in the Residential Building Sector," *Energy and Buildings* 269 (2022): 1–15.

[17] Z. Wang, Y. Wang, and R. S. Srinivasan, "A Novel Ensemble Learning Approach to Support Building Energy Use

Prediction," *Energy and Buildings* 159 (2018): 109–122, https://doi.org/10.1016/j.enbuild.2017.10.085.

[18] Y. Wang, "Application of Deep Learning Model in Building Energy Consumption Prediction," *Computational Intelligence and Neuroscience* 2022 (2022): 1–9, https://doi.org/10.1155/2022/4835259.

[19] T. Pinto, I. Praça, Z. Vale, and J. Silva, "Ensemble Learning for Electricity Consumption Forecasting in Office Buildings," *Neurocomputing* 423 (2021): 747–755, https://doi.org/10.1016/j.neucom.2020.02.124.

[20] X. J. Luo, L. O. Oyedele, A. O. Ajayi, et al., "Genetic Algorithm-Determined Deep Feedforward Neural Network Architecture for Predicting Electricity Consumption in Real Buildings," *Energy and AI* 2 (2020): 100015–100018, https://doi.org/10.1016/j.egyai.2020.100015.

[21] S. S. Amiri, M. Mueller, and S. Hoque, "Investigating the Application of a Commercial and Residential Energy Consumption Prediction Model for Urban Planning Scenarios with Machine Learning and Shapley Additive Explanation Methods," *Energy and Buildings* 287 (2023): 1–17.

[22] Q. Shi, C. Liu, and C. Xiao, "Machine Learning in Building Energy Management: A Critical Review and Future Directions," *Frontiers of Engineering Management* 9, no. 2 (2022): 239–256, https://doi.org/10.1007/s42524-021-0181-1.

[23] X. Liu, H. Tang, Y. Ding, and D. Yan, "Investigating the Performance of Machine Learning Models Combined with Different Feature Selection Methods to Estimate the Energy Consumption of Buildings," *Energy and Buildings* 273 (2022): 112408–112412, https://doi.org/10.1016/j.enbuild.2022.112408.

[24] H. Iftikhar, J. Zywiołek, J. L. López-Gonzales, and O. Albalawi, "Electricity Consumption Forecasting Using a Novel Homogeneous and Heterogeneous Ensemble Learning," *Frontiers in Energy Research* 12 (2024): 1442502, https://doi.org/10.3389/fenrg.2024.1442502.

[25] J. Moon, M. Maqsood, D. So, S. W. Baik, S. Rho, and Y. Nam, "Advancing Ensemble Learning Techniques for Residential Building Electricity Consumption Forecasting: Insight from Explainable Artificial Intelligence," *PLoS One* 19, no. 11 (2024): e0307654, https://doi.org/10.1371/journal.pone.0307654.

[26] X. Cui, M. Lee, C. Koo, and T. Hong, "Energy Consumption Prediction and Household Feature Analysis for Different Residential Building Types Using Machine Learning and SHAP: Toward Energy-Efficient Buildings," *Energy and Buildings* 309 (2024): 113997, https://doi.org/10.1016/j.enbuild.2024.113997.

[27] S. Kumaraswamy, K. Subathra, S. Geeitha, G. Ramkumar, A. S. M. Metwally, and M. Z. Ansari, "An Ensemble Neural Network Model for Predicting the Energy Utility in Individual Houses," *Computers & Electrical Engineering* 114 (2024): 109059.

[28] F. Dinmohammadi, Y. Han, and M. Shafiee, "Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms," *Energies* 16, no. 9 (2023): 3748–3823, https://doi.org/10.3390/en16093748.

[29] I. U. Haq, A. Ullah, S. U. Khan, et al., "Sequential Learning-Based Energy Consumption Prediction Model for Residential and Commercial Sectors," *Mathematics* 9, no. 6 (2021): 605–617, https://doi.org/10.3390/math9060605.

[30] J. D. Wichard and M. Orgorzalek, "Time Series Prediction with Ensemble Models," in *IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)* (2004).

[31] C. Merkwirth and J. Wichard, "Entool – A Mathlab Toolbox for Ensemble Modeling" (2002).

[32] Y. Zhu, C. Xie, G.-J. Wang, and X.-G. Yan, "Comparison of Individual, Ensemble and Integrated Ensemble Machine Learning Methods to Predict China's SME Credit Risk in Supply Chain Finance," *Neural Computing & Applications* 28, no. S1 (2016): 41–50, https://doi.org/10.1007/s00521-016-2304-x.

[33] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proceedings of the International Workshop of Multiple Classifier Systems*, 1857 (Berlin, Heidelberg: Springer, 2000), 1–15, https://doi.org/10.1007/3-540-45014-9_1.

[34] Eskom, Stellenbosch University, and University of Cape Town, *Domestic Electrical Load Metering Data 1994-2014* (Johannesburg: Eskom, Cape Town: UCT, Stellenbosch: US [producers], 2014).

[35] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (O'Reilly, 2019).

[36] A. Filippova, E. Trainer, and J. D. Herbsleb, "From Diversity by Numbers to Diversity as Process: Supporting Inclusiveness in Software Development Teams with Brainstorming," in *Proceedings of the 39th International Conference on Software Engineering* (Buenos Aires, Argentina, 2017), 152–163, https://doi.org/10.1109/icse.2017.22.