A TIME-SERIES FORECASTING OF POWER CONSUMPTION AND FEATURE EXTRACTION IN AGRICULTURE SECTOR USING MACHINE LEARNING

Megha Sharma,* Namita Mittal,* Anukram Mishra,** and Arun Gupta**

Abstract

Energy plays a vital role in the economic development, growth, and productivity of any country. Commercial, industrial, residential, and agriculture sectors require reliable and adequate energy services. The paper emphasises the significance of accurate electric power forecasting in the agriculture sector in India to avoid power outages and impact on productivity. The paper aims to forecast medium-term load using time series models and identify important features for power prediction in agriculture. The time series model is used to find the peak season in the year, while the XGBoost technique is used to identify the feature importance and load forecasting. Statistical approaches, such as correlation matrix and scatter plots, are also used for feature extraction. The results of the study show that the addition of exogenous and endogenous data on the historical load improves the accuracy in terms of mean absolute percentage error (MAPE) and R-squared (R^2). The research demonstrates the potential of using machine learning techniques to enhance the accuracy of medium-term agricultural electrical load forecasting.

Key Words

Agriculture sector, electrical power, time series-based load forecasting, XGBoost

1. Introduction

Energy production and availability are critical to a country's economy as they have a direct impact on future production, imports, exports, and investment. Energy production is a key driver of economic growth and development. It is essential for powering industry and infrastructure, and it also plays a crucial role in shaping a country's trade and investment relationships. The availability of energy affects the cost of production, which in turn affects the competitiveness of a country's products in the global market. Additionally, energy production and availability also affect the cost of living for citizens, as well as the ability for a country to attract and retain businesses and industries. Therefore, forecasting energy demand and planning for energy production and availability is important for a country's long-term economic stability and growth.

Integration of renewable energy sources in the power grid is increasing, but the fluctuating nature of these sources presents challenges for load prediction and energy demand forecasting [1]. Load forecasting is important for regulating the demand and supply of electricity, but it is particularly challenging in the agriculture sector due to factors, such as weather, precipitation, and soil type. According to the 2011 census, a majority of India's population (61.5%) lives in rural areas and is dependent on agriculture. The agriculture sector is also a significant contributor to the Indian economy, accounting for 14.4% of the gross domestic product (GDP). Given the importance of the agriculture sector to the Indian economy and population, forecasting energy demand and ensuring availability of electricity for this sector is crucial for economic growth and development. Figure 1 shows the power supply scenario in India in terms of peak. The country has enough electricity generation capacity to meet the demand, and the difference between the demand and supply is not due to a lack of power generation. The marginal gap may be due to other factors, such as transmission and distribution issues, infrastructure problems, or other non-generation related issues.

Load forecasting is divided into short-term, medium-term, and long-term predictions, with medium-term being the most important for power system operations. Medium-term load forecasting (MTLF) relies on external factors, such as weather, and it is less sensitive than short-term forecasts. Load forecasting is critical for the planning and operation of electrical power systems, such as outages management, significant operations, and schedule

* Department of Computer Science and Engineering, Malaviya National Institute of Technology Jaipur, Rajasthan, India; e-mail: 2018rcp9099@mnit.ac.in; nmittal.cse@mnit.ac.in
** Genus Power Infrastructures Limited, Jaipur, Rajasthan, India; e-mail: anukram.mishra@genus.in; arun.gupta@genus.in
Corresponding author: Megha Sharma

(DOI: 10.2316/J.2023.203-0451)

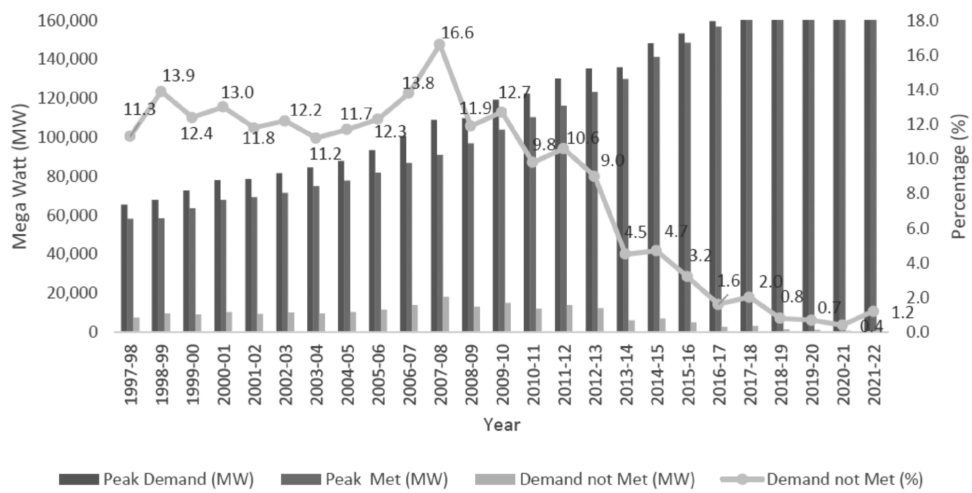

Figure 1. Power supply scenario in terms of Peak [2].

maintenance. There are two main categories of forecasting techniques

1. Statistical approach, which includes time series analysis
2. Intelligent approach, which includes machine learning, ensemble learning, and deep learning.

The proposed work focuses on MTLF for the agricultural sector (AS). Previous work [3] has applied different time-series models to agricultural datasets to predict future consumption, but in this paper, the authors plan to take a different approach by identifying different weather features that affect power consumption in the agricultural sector, and then using that information to predict the load. The proposed work aims to analyse data collected from Jaipur Vidyut Vitran Nigam Limited (JVVNL)[1] in collaboration with Genus Power Infrastructures Limited[2] to identify the peak power consumption seasons in the agricultural sector.

Our contributions

1. The use of seasonal auto-regressive integrated moving average (SARIMA) and exponential smoothing models for time series forecasting in the AS to identify peak electricity consumption of consumers.
2. The application of the machine learning approach to identify significant features that can be used to forecast load.
3. A comparison of the performance of different models and approach for electricity consumption forecasting in the agricultural sector.
4. The identification of peak power consumption season, which can be used to improve the efficiency of demand response management in the agricultural sector.

The organisation of the article allows readers to quickly find the information they need it is as: Section 2 provides an overview on related work done on estimating power consumption. Section 3 describes the proposed work and provides a clear outline of the research methodology and the models that will be used to analyse the data. Section 4 provides detailed information on the findings of the research. Section 5 provides the conclusion with an overall summary of the research.

[1]Dataset collected from Jaipur Vidyut Vitran Nigam Ltd and data accessed on the request basis.

[2]Testing and validation of the work are done with the corporation of the Genus Power Infrastructures Limited, Sitapura, Jaipur by visiting different farms in Rajasthan.

2. Related Work

Forecasting electricity demand is an important aspect of managing power systems. By accurately predicting how much electricity will be needed, utilities and grid operators can ensure that there is enough generation capacity to meet demand, and that it is dispatched in the most cost-effective and efficient way. This helps to avoid blackouts, reduce costs, and support the integration of renewable energy sources. Additionally, accurate demand forecasting is also important for long-term planning and investment decisions, such as building new power plants or transmission lines [4]. Load analysis is necessary for power prediction. The main objective of load analysis is to identify viable customers for load management programs, forecast load demand for the short term, and develop rate designs [5]. Time series-based model for monthly power consumption using auto-regressive integrated moving average (ARIMA) model presented in [6] for an institution. The agriculture sector has not seen as much research in the application of time series models for power consumption estimation compared to other sectors. However, there are still a number of models that can be effectively applied to this problem. Some of these models include: SARIMA, exponential smoothing (ES) and intelligent approaches. There are several intelligent methods that can be used to forecast short-term wind and solar power, including machine learning and artificial intelligence (AI) techniques. These methods can be trained on historical data to learn the patterns and relationships between various meteorological and solar conditions, and the corresponding wind and solar power output. Neural network-based wind and solar power

Table 1
A Brief Description of Load Forecasting Models

Forecasting Model	Parameters	Feature Selection	Remark
LSTM-RNN model [15]	Hourly load data and regional weather data	Feature selection: Wrapper method Lag Selection: Genetic algorithm	There is an inversely relation between electricity consumption and temperature
ANN [16]	Population, GDP, crop land and load consumption data	-	Train the model with scaled conjugate gradient (trainscg), traincgf, trainlm
Feature selection [12]		MI, AC, CFS and ReliefF	
Bagging-XGBoost [14]	Weather and time factors	Mutual information	Computational time is high
Linear regression and XGBoost [13]	Load, temperature and calendar rules	Pearson correlation coefficient	Load forecasting for Industrial consumers using SVMD
NN with shark smell optimisation [7]	State of charge (SOC)	Filter method based on MI and interaction gain	STLF of wind and solar power

prediction is proposed in [7]. A wind signal prediction model based on new method for feature selection, empirical mode decomposition (EMD) is presented in [8].

Feature engineering and feature selection are important fields of research in machine learning and AI, and several methods have been proposed to improve the accuracy of models by identifying and selecting a minimal subset of relevant features from the original dataset. In [9], the author gave two-stage forecast engine based on two neural network (NN), Ridgelet NN for pre-processor and Elnam NN for prediction. The Pennsylvania–New Jersey Maryland (PJM) dataset was used for STLF. Feature selection using clustering-based filter method is presented in [10]. Fast correlation-based filter (FCBF), mutual information (MI), and ReliefF (RF) are all feature selection methods that can be used for dimensionality reduction. A residential power prediction using MI, elastic net presented in [11]. RF is an instance-based feature selection method that assigns a weight to each feature based on the differences and similarities across classes. It assigns higher weights to features that can differentiate between different classes, and lower weights to features that are similar across classes [12]. All three methods are commonly used in dimensionality reduction, and they can help to identify the minimal subset of features that retain the most informative and useful characteristics of the original data, while removing irrelevant and redundant features.

A model for STLF using SampEn variational mode decomposition (SVMD) and XGBoost regressor is presented in [13] for industrial consumers. Severe weather can have a significant impact on power consumption and identifying it in advance can help power providers to prepare for the increased demand. Severe weather identification is presented in [14], where Bagging–XGBoost model is used for STLF. Table 1 gives a brief description of the existing load forecasting technique.

Some time series-based models were examined in our earlier work [3], [17] but in this work, time series-based forecasting along with exogenous and endogenous feature-based load forecasting was done to improve the accuracy. The author notes that there is a lack of research in the agriculture sector when it comes to load forecasting, which is crucial for providing farmers with accurate and reliable power. The current research in this area primarily focuses on economic factors, such as GDP, population, and land type, but the impact of weather on agriculture load prediction has not been fully explored. This represents a gap in knowledge that needs to be addressed in order to improve load forecasting for the agriculture sector.

3. Material and Methodology

3.1 Dataset

JVVNL, also known as Jaipur Discom, is responsible for distributing and supplying power to 12 districts in the state of Rajasthan [18]. The data collected from JVVNL includes historical electricity consumption data for residential, industrial, commercial, and agricultural sectors. Weather data consists of mean, max, min value of temperature (Temp) humidity (Hum) pressure (Pre) and collected from meteorological department. The data covers the time period from January 2015 to December 2021. In preprocessing, the data is divided into two parts, one is used for training and another is used for testing. The training data is the past 6 years of monthly data from January 2015 to December 2020, and the testing data is the actual consumption of data from January 2021 to December 2022.

3.2 Proposed Methodology

Time series data is a type of data where a set of observations are collected at regular intervals of time, with each consecutive data point in the series dependent on the previous data point. This type of data is commonly used in fields, such as finance, economics, weather forecasting, and more.

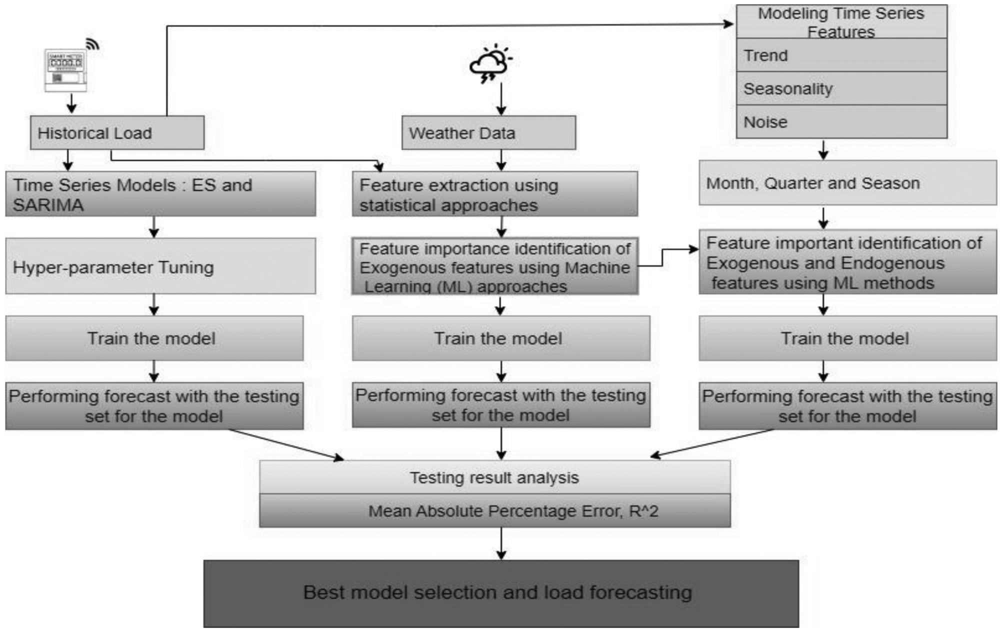

Figure 2. Proposed methodology flowchart.

There are several types of features that can be present in a time series data set, such as Date and Time features, lag features, window features (based on a fixed-size window of consecutive observations) and domain specific features. In this paper, weather data, such as mean_Temp, max_Temp, min_Temp, *etc.*, are considered as domain-specific features.

The goal of time series load forecasting in the agriculture sector is to predict future demand patterns, typically peak demand during specific seasons, so that a DR program can be designed and implemented to manage and optimise energy consumption. This is typically done by analysing historical consumption data and using various statistical and machine-learning techniques to make accurate forecasts for future time intervals. The proposed work steps include data collection, preprocessing, model selection, model training and evaluation, and implementation of the forecast. Incorporating exogenous features, such as weather data, into the forecasting model can improve the accuracy of agricultural electrical load predictions. Compared to traditional time series-based methods for agricultural electrical load forecasting, the use of exogenous features in the ML approach allows for more accurate predictions by incorporating additional information about external factors that may affect the load. The results of this paper suggest that ML-based approaches are a promising approach for agricultural electrical load forecasting and can improve the accuracy of predictions compared to traditional methods, the proposed methodology is depicted in Fig. 2.

The process for analysing and predicting agricultural sector's power consumption using the data collected from JVVNL and weather data can be broken down into the following steps:

Step 1 (Data Collection and Pre-Processing): Clean, format and process the data to make it ready for analysis and modeling.

Step 2 (Feature Engineering for Time Series Data): Extract relevant endogenous features from data, such as month, quarter, *etc.*

Step 3 (Applying Different Types of Statistical Approaches): Use techniques like ARIMA, SARIMA, and ES models to analyse and model the time series data.

Step 4 (Applying Different Types of Regression and Machine Learning Algorithms): Use algorithms like multiple linear regression (MLR), AdaBoost to model the data and make predictions.

Step 5 (Identify the Best Prediction Model Using Error Matrices): By comparing the accuracy of different models using metrics, such as MAPE and R^2, to identify the best model with the most accurate predictions.

Overall, the main characteristics that will be considered for this research work are the time series features, weather data. These features will be used to train the model and make predictions of the agricultural sector's power consumption.

3.3 Time Series Load Forecasting

The predicting of a variable in time series analysis depends on its historical trends and data gathered throughout time [19]. The previous pattern of the variable is used to forecast future values. (1) is used to calculate the previous pattern and make predictions for the next period.

$$X\mathrm{t} = Xt - 1 + \varepsilon \tag{1}$$

It is often necessary to transform the data into a stationary time series. Stationarity in a time series refers to the stability of the statistical properties of the series over time. A stationary time series has a constant mean, constant variance, and no trend or seasonality. This is important because many time series analysis techniques, such as ARIMA models, assume that the data is stationary.

If the data is not stationary, it must be transformed to make it stationary before applying these techniques. The most common ways of achieving stationarity are using differencing or using a decomposition technique [20]. Auto-correlation function (ACF) is a tool used to find the relation between time series data at different time lags. The ACF can be used to determine the order of a moving average (MA) model, which is represented as MA(q) and reaches its maximum level at q lags. The partial ACF (PACF) is used to determine the order of an Auto-Regressive (AR) model, which is represented as AR(p) and reaches its maximum level at p lags [6].

3.3.1 SARIMA Model

The SARIMA model is a variation of the ARIMA model that includes a seasonal component in addition to the non-seasonal components of AR, MA, and differencing. The SARIMA model is used to forecast time series data with a clear seasonal pattern [21]. The SARIMA model is represented as SARIMA(p,d,q)(P,D,Q)s, where (p,d,q) are the non-seasonal parameters, and (P,D,Q)s are the seasonal parameters. The non-seasonal parameters represent the order of the Autoregressive (p), the order of differencing (d), and the order of the moving average (q) components of the model. The seasonal parameters represent the order of the seasonal autoregressive (P), the order of seasonal differencing (D), and the order of the seasonal moving average (Q) components of the model, and 's' represents the number of time steps per seasonal period. The parameters of the SARIMA model are identified using the ACF and PACF plots of the stationary data, and then the model is fitted to the data. The model can then be used for forecasting future values of the time series.

$$\begin{gathered}\varphi_P(B^s)\,\varnothing_p(B)\,(1-B^s)^D(1-B)^d \\ X_t = \vartheta_Q(B^s)\theta_q(B)Z_t\end{gathered} \tag{2}$$

where

$$\varphi_P(B^s) = 1 - \varphi_1 B^s - \varphi_2 B^{2s} - \cdots\cdots - \varphi_P B^{\mathrm{Ps}}, \tag{3}$$

is the seasonal AR term,

$$\varnothing_p(B) = 1 - \varnothing_1 B - \varnothing_2 B^2 - \cdots\cdots - \varnothing_p B^p, \tag{4}$$

is the AR term,

$$\vartheta_Q(B^s) = 1 + \vartheta_1 B^s + \vartheta_2 B^{2s} + \cdots\cdots + \vartheta_Q B^{\mathrm{Qs}}, \tag{5}$$

is the seasonal MA term

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots\cdots + \theta_q B^q, \text{ is MA term} \tag{6}$$

3.3.2 ES

ES is a time series forecasting method that uses a combination of three parameters (level, trend, and seasonality) represented by α, β, and γ. These parameters are used to assign weights to the current and previous observations to improve the accuracy of the forecast. ES is not based on exogenous variables and is commonly used for time series forecasting. The seasonality component of the model can be either additive or multiplicative [22].

new level l_n represented as:

$$l_n = \propto .(X_n - s_{n-m}) + (1-\propto)(l_{n-1} + t_{n-1}), \tag{7}$$

new trend t_n represented as:

$$t_n = \beta(l_n - l_{n-1}) + (1-\beta)\,t_{n-1}, \tag{8}$$

new seasonal s_n represented as:

$$s_n = \gamma.(X_n - l_n) + (1-\gamma)\,s_{n-m} \tag{9}$$

where h = future forecast step, m = season, length = 12.

1. Additive seasonality represented as:

$$X_{n+h} = l_n + h{\cdot}t_n + s_{n+h-m} \tag{10}$$

2. Multiplicative seasonality defined as:

$$X_{n+h} = (l_n + h{\cdot}t_n){\cdot}s_{n+h-m} \tag{11}$$

3.4 Feature Analysis and Extraction

In the Agricultural (AG) Sector, electricity consumption is closely linked to weather data as the growth and development of crops are heavily influenced by the different seasons and climate conditions. Weather variables, such as temperature, humidity, and pressure, are considered as exogenous inputs for the proposed model as they directly affect electricity consumption for farmers. In addition, the input data is categorised into two parts: continuous data (*e.g.*, temperature, humidity, and pressure) and categorical data (*e.g.*, months, seasons, and quarters) where the latter is derived from time series data.

Climate Season: In India, the climate is characterised by different seasons, such as winter, spring, summer, monsoon, and autumn (Table 2). After analysing the data, it has been observed that the agriculture sector has the highest electricity consumption during the winter and spring seasons. This is likely due to the fact that farmers use electricity for irrigation during these seasons to ensure optimal growth and development of their crops. The consumption in monsoon and autumn may be lower due to the natural water availability during these seasons, as shown in Fig. 3.

Month effect: The residential sector has peak electricity consumption in the summer months, while the commercial sector has peak consumption in the winter months. The industrial sector has more consistent electricity consumption throughout the year. Overall, there are variations in the monthly power consumption across different sectors, with peak consumption occurring at different times of the year. The average electricity consumption in AG sector according to month is shown in Fig. 4. The consumption patterns in the agriculture sector are different from the overall consumption patterns, with peak consumption occurring in February and off-peak consumption in May.

Categorical feature analysis using a statistical approach can be used to understand the relationship between different seasons, months and quarter of the year

Table 2
Classification of Months According to Climate of India

Season	Month	Climate	Season	Month	Climate
Winter	December-January	Very Cool	Monsoon	July-Sept	Wet, hot and humid
Spring	Feb-March	Sunny and pleasant	Autumn	Sept-Nov	Pleasant
Summer	April-June	Hot			

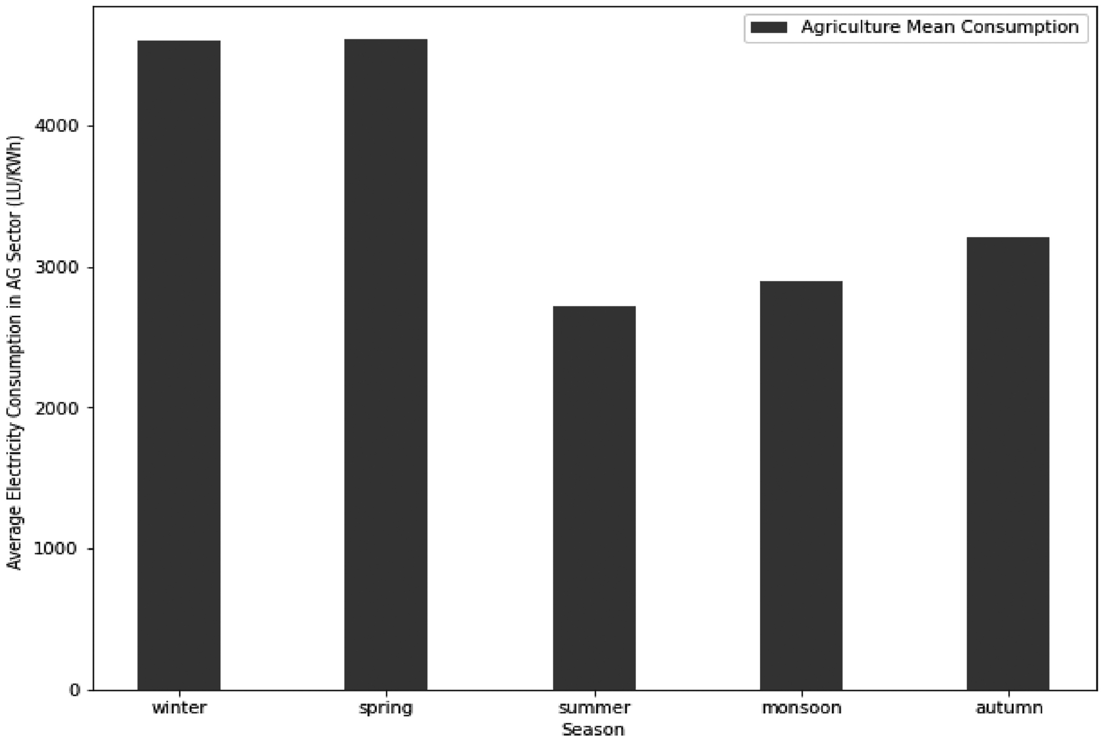

Figure 3. Season effect on AG sector.

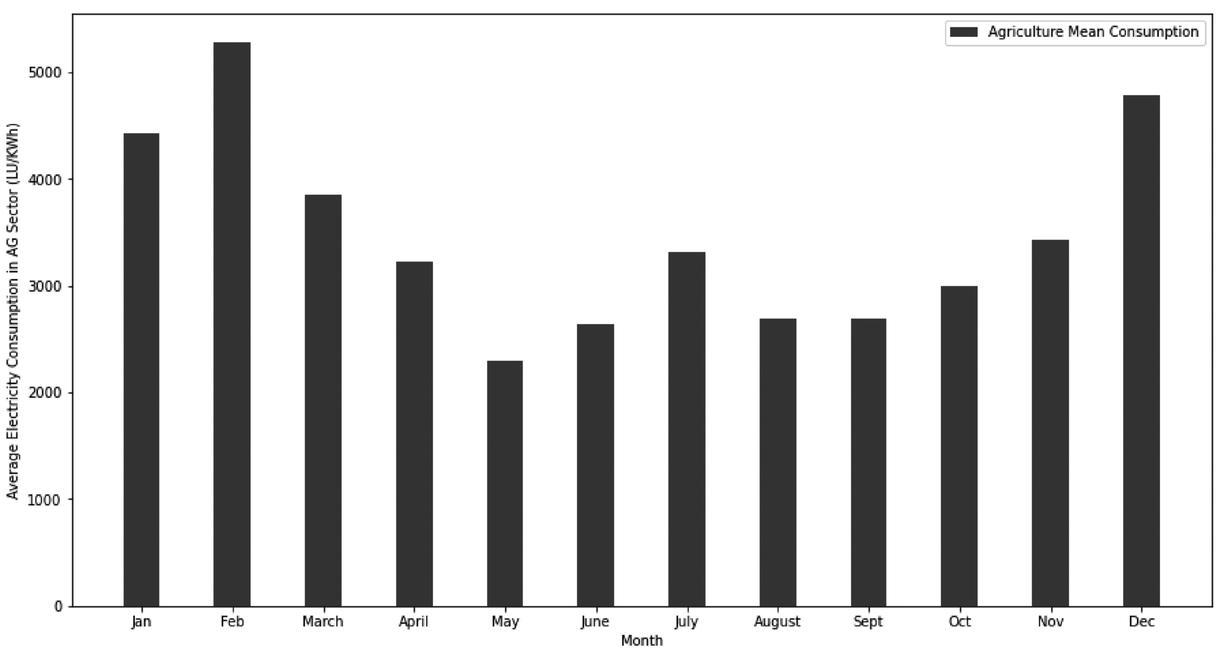

Figure 4. Month effect on AG sector's electricity consumption.

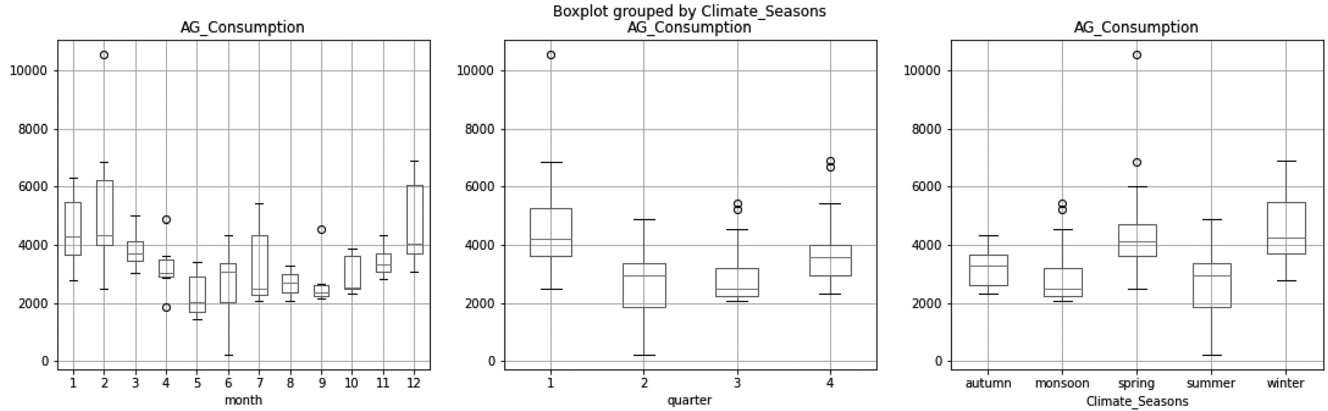

Figure 5. Boxplot diagram of categorical variables (month, quarter and season) and electricity consumption.

Table 3
Selection of Weather Variable based on Correlation Value

Continuous Variables	Correlation Value	Continuous Variables	Correlation Value
Min_Temp (°C)	−0.76	Mean_Temp (°C)	−0.78
Max_Temp (°C)	−0.68	Mean_Pre (mbar)	0.71

and power consumption. Since seasons, month and quarter are the categorical data, the effect these variables on power consumption can be analysed using the box plot method. Boxplots can be used to visualise the relationship between the categorical variable and the continuous variable (power consumption). The strength of the relationship can then be measured using the ANOVA test. Fig. 5 represents the box-plot diagram of the categorical data.

Continuous features analysis: Weather features (temperature, humidity, and air pressure) would come under the continuous data category as these are numerical values. These features can be analysed using scatter plots, which show the relationship between the target variable (power consumption) and the predictor variables (weather features). A correlation matrix is used to measure the strength of the relationship between the different weather features and power consumption. The correlation matrix provides a numeric value (ranging from −1 to 1) that represents the strength of the relationship. Positive values indicate a positive correlation, negative values indicate a negative correlation, and values near 0 indicate a weak correlation. Table 3 shows a correlation between the target variable and input variables.

3.5 Feature Importance Identification using XGBoost

XGBoost is a powerful machine-learning algorithm that is an improved version of gradient boosting decision tree (GBDT). It has several advantages over GBDT, such as faster optimisation and regularisation to avoid over-fitting. Additionally, XGBoost supports parallel processing, which enables it to be more efficient than GBDT. It is a popular method in machine learning competitions and has been used in a variety of classification problems with excellent results. However, its application in regression tasks for load forecasting is relatively limited [13]. Therefore, using XGBoost for load forecasting in the AS has the potential to improve forecasting accuracy.

XGBoost algorithm was first proposed by Dr. Tong He in 2016 and is an implementation of the Gradient Boosting framework. It combines multiple weak learners into a strong learner using a technique called boosting. The weak learner used in XGBoost is the classification and regression tree (CART) algorithm. The algorithm iteratively adds new weak learners to the model and updates the model by adjusting the residuals of the previous weak learners. This process can be thought of as an ensemble of CART models, where each iteration builds upon the previous one to create a more accurate and robust model. The end result is a strong learner that is able to make accurate predictions.

$$y_{\mathrm{xi}} = \sum_{n=1}^{N} f_n\left(x_i\right), \ f_n \in F \tag{12}$$

Where y_{xi} is the forecasted value of the ith sample, N is the total number of CART tree, $f_n\left(x_i\right)$ is the forecasted value of the ith sample of the nth tree, and F is the function.

4. Experimental Result

This section presents the main results obtained by different models. The result analysis considered 3 cases. *Case 1:* Time series load forecasting for peak season identification, *Case 2:* Load forecasting based on exogenous variables (weather data), *Case 3:* Load forecasting using exogenous variable and endogenous variables. Mean absolute percentage error (MAPE) and R^2 have been used for measuring the performance of load forecasting.

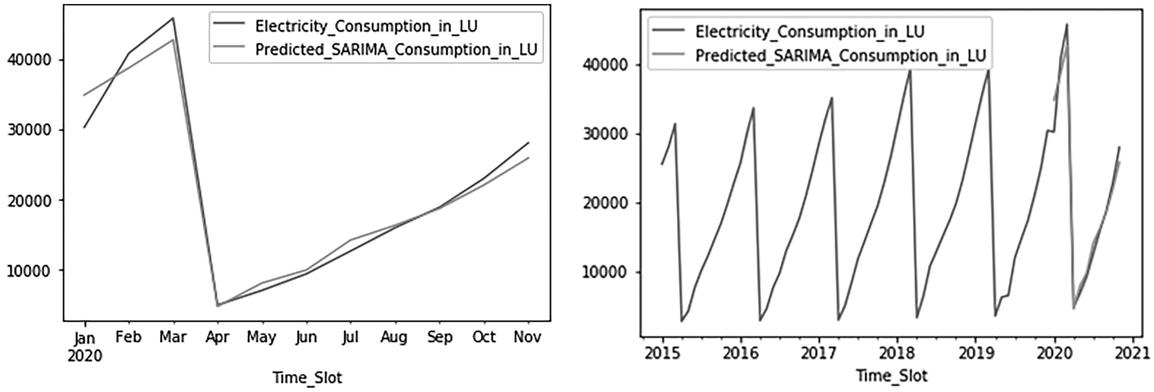

Figure 6. Seasonal ARIMA model-based (a) Month-wise load prediction and (b) Year-wise load prediction.

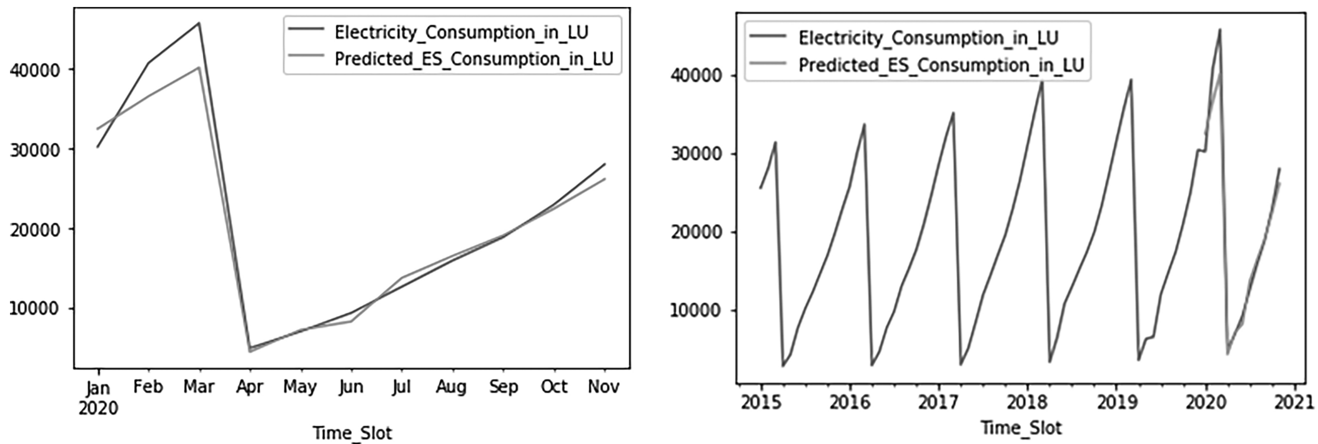

Figure 7. Exponential smoothing: (a) Month-wise load prediction and (b) Year-wise load prediction.

1. *MAPE*: An MAPE error can be calculated using (13), where y_{actual} denotes actual values and y_{pre} denotes predicted load.

$$\text{MAPE} = \frac{\sum_{i=1}^{n} \frac{|y_{\text{pre}} - y_{\text{actual}}|}{y_{\text{actual}}}}{n} \tag{13}$$

(2) R^2: R-squared value varies from 0 to 1. If value near about 1 then it attends more accuracy.

$$R^2 = 1 - \frac{\sum \left(y_{\text{pre}} - y_{\text{actual}}\right)^2}{\left(y_{\text{actual}} - y_{\text{average}}\right)^2} \tag{14}$$

Case 1—Result Analysis Using SARIMA Model and ES Model for Peak Identification:

Agricultural electricity consumption is affected by the seasons. The SARIMA model is a time series forecasting model that can be used to predict future power consumption patterns based on historical data. The one-year-ahead predicted power consumption pattern, as depicted in Fig. 6, shows that the forecasted load pattern is similar to the actual load. This suggests that the SARIMA model is able to accurately capture the seasonal and monthly variations in power consumption. The electricity consumption is measured in lakh units/Kilo Watt hour (LU/KWh) which is a unit of measurement commonly used in India.

In Rajasthan, during the monsoon season, there is usually a decrease in electricity consumption due to the fact that crops are being irrigated by rainwater instead of using electricity-powered irrigation systems. This season is known as Kharif season. As monsoon weakens in Mid-September, the Rainfall decreases in October and November which results in an increase in electricity consumption. This corresponds to the Rabi season where electricity-powered irrigation is used for most of the farming. The results show that the Zaid season necessitates minimal electricity, Kharif crops necessitate a modest amount of electricity because most farming is done with rainwater, and the Rabi crops require a significant quantity of electricity, with a peak demand period from December to March as shown in Figs. 6 and 7.

Exponential Smoothing Result: Figure 7 shows the forecasted electricity consumption for next year, where forecasted load pattern is similar to the actual load.

Case 2—Load Forecasting Using Historical Load and Weather Data

In this case, weather data (mean_Temp, max_Temp, min_Temp, and mean_Pre) are used as input, as these features are selected from continuous features analysis using correlation value. Figure 8 shows that mean_Temp has the highest impact on load forecasting among weather data. The result of load forecasting using XGBoost of different cases (in reference to input variables) comparing with MLR

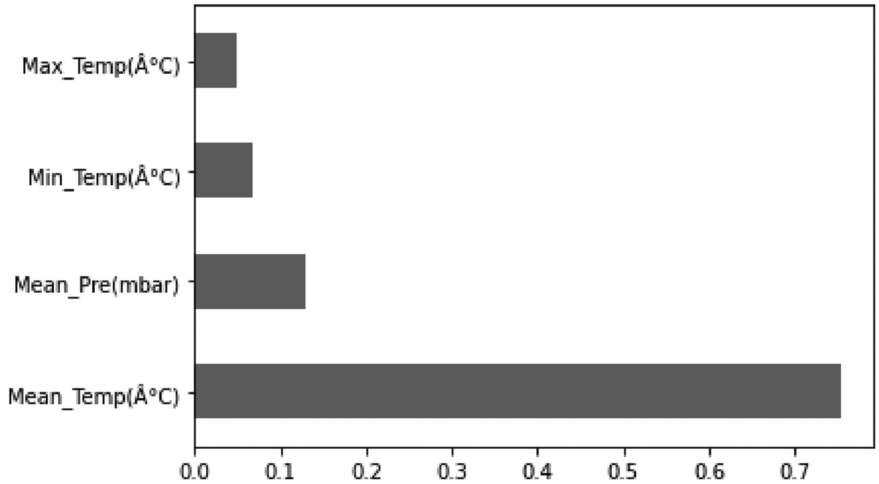

Figure 8. Significance of feature using XGBoost for case 2.

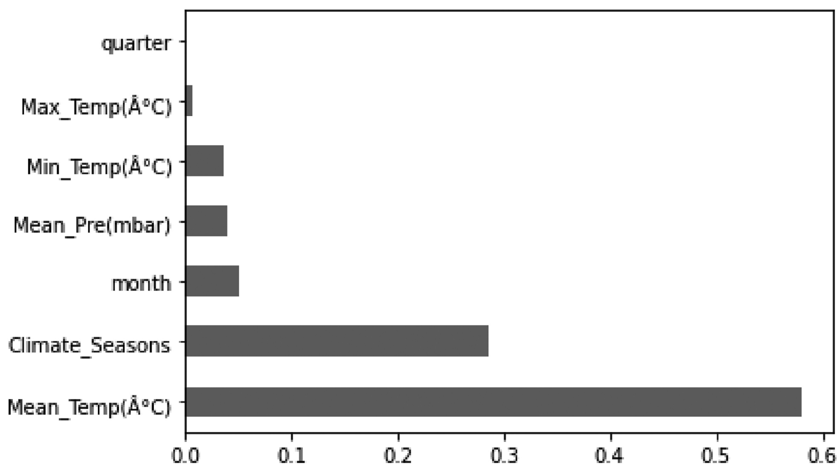

Figure 9. Significance of feature using XGBoost for case 3.

Table 4
Comparative Analysis with Existing Techniques for Case 2

Model	R^2-test	MAPE (%)
MLR-Weather data	0.57	85.2
XGBoost-Weather data	0.82	31.62

Table 5
Comparative Analysis with Existing Techniques for Case 3

Model	R^2-test	MAPE (%)
MLR-weather data	0.77	44.53
XGBoost-weather data	0.85	24.07

model which is the base model. The result analysis of load forecasting is shown in Table 4 using case 2.

Case 3—Load Forecasting Using Historical Load, Weather Data and Endogenous Variables: The endogenous variables (month, climate_season, and quarter) are derived from time series data. Table 5 shows that by adding

Table 6
Comparative Analysis with Existing Techniques

Model	R^2-test	Reference
XGBoost	0.71	[23]
RF-XGBoost	0.81	[23]
Proposed Model	0.85	–

endogenous features improves the accuracy and Fig. 9 depicts the ranking of feature using XGBoost.

The result of load forecasting using XGBoost of different cases (in reference to input variables) comparing with the MLR model which is the base model (Tables 4 and 5). Table 6 depicts the comparison among existing models and the proposed model. The results show that the addition of weather data and time series features in historical load data improves the accuracy of the load forecasting.

5. Conclusion

This article focuses on the importance of forecasting power demand in the agriculture sector. The proposed work uses XGBoost techniques to predict power consumption patterns and compares the results with a base model (MLR). The addition of exogenous and endogenous factors improves the results of the model. The results of the study show that the Rabi season has the highest demand for electricity, followed by the Kharif season with an average demand and the Zaid season with the lowest or off-peak demand. The results also indicate that the inclusion of weather data, seasons and endogenous factors from historical load data improves the accuracy of the forecasting as seen from the improvement in the MAPE and R^2 values. To create a more effective model in the future, we intend to combine the existing model with additional elements like geographic data. In conclusion, the paper emphasises that accurate load forecasting is necessary to ensure reliable and adequate energy services in the agriculture sector.

References

[1] W. Cai, R. Mohammaditab, G. Fathi, K. Wakil, A.G. Ebadi, and N. Ghadimi, Optimal bidding and offering strategies of compressed air energy storage: A hybrid robust-stochastic approach, *Renewable Energy*, *143*, 2019, 1–8.

[2] Central Electricity Authority, Ministry of Power, Government of India, CEA Annual Report 2021-22, CEA, New Delhi, 2020. https://cea.nic.in/annual-report/?lang=en

[3] M. Sharma, N. Mittal, A. Mishra, and A. Gupta, Machine learning-based electricity load forecast for the agriculture sector, *International Journal of Software Innovation*, *11*(1), 2023, 1–21.

[4] K.M. Powell, A. Sriprasad, W.J. Cole, and T.F. Edgar, Heating, cooling, and electrical load forecasting for a large-scale district energy system, *Energy*, *74*, 2014, 877–885.

[5] A. Domijan, A. Islam, M. Islam, A.M. Bonet, A. Omole, and H.A. Lladro, Price responsive customer screening using load curve with inverted price tier, *International Journal of Power and Energy Systems*, *31*(1), 2011, 67.

[6] H. Kaur and S. Ahuja, “Time series analysis and prediction of electricity consumption of health care institution using ARIMA model,” *Proc. 6th International Conf. on Soft Computing for Problem Solving (SocProS)*, vol. 2. Singapore, 2017, pp. 347–358.

[7] F. Mirzapour, M. Lakzaei, G. Varamini, M. Teimourian, and N. Ghadimi, A new prediction model of battery and wind-solar output in hybrid power system, *Journal of Ambient Intelligence and Humanized Computing*, *10*(1), 2019, 77–87.

[8] M. Mir, M. Shafieezadeh, M.A. Heidari, and N. Ghadimi, Application of hybrid forecast engine based intelligent algorithm and feature selection for wind signal prediction, *Evolving Systems*, *11*(4), 2020, 559–573.

[9] N. Ghadimi, A. Akbarimajd, H. Shayeghi, and O. Abedinia, Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting, *Energy*, *161*, 2018, 130–142.

[10] S.S. Subbiah and J. Chinnappan, An improved short term load forecasting with ranker based feature selection technique, *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *39*(5), 2020, 6783–6800.

[11] A.M. Pirbazari, A. Chakravorty, and C. Rong, Evaluating feature selection methods for short-term load forecasting, *Proc. IEEE Int. Conf. on Big Data and Smart Computing (BigComp)*, Kyoto, 2019, 1–8.

[12] A.R. Gollou and N. Ghadimi, A new feature selection and hybrid forecast engine for day-ahead price forecasting of electricity markets, *Journal of Intelligent & Fuzzy Systems*, *32*(6), 2017, 4031–4045.

[13] Y. Wang, S. Sun, X. Chen, X. Zeng, Y. Kong, J. Chen, *Y.* Guo, and T. Wang, Short-term load forecasting of industrial customers based on SVMD and XGBoost, *International Journal of Electrical Power & Energy Systems*, *129*, 2021, 106830.

[14] X. Deng, A. Ye, J. Zhong, D. Xu, W. Yang, Z. Song, Z. Zhang, J. Guo, T. Wang, Y. Tian, H. Pan, Z. Zhang, H. Wang, C. Wu, J. Shao, and X. Chen, Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model, *Energy Reports*, *8*, 2022, 8661–8674.

[15] S. Bouktif, A. Fiaz, A. Ouni, and M.A. Serhani, Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches, *Energies*, *11*(7), 2018, 1636.

[16] S. Saravanan and K. Karunanithi, Forecasting of electric energy consumption in agriculture sector of India using ANN technique, *International Journal of Pure and Applied Mathematics*, *119*(10), 2018, 261–271.

[17] M. Sharma, N. Mittal, A. Mishra, and A. Gupta, “Analytical machine learning for medium-term load forecasting towards agricultural sector,” *Proc. 2nd Doctoral Symposium on Computational Intelligence (DoSCI)*, Singapore, 2021, pp. 581–592.

[18] JVVNL, Jaipur Vidyut Vitaran Nigam Ltd., Department of Energy, Government of Rajasthan, 2020. https://energy.rajasthan.gov.in/content/raj/energy-department/jaipur-vidyut-vitran-nigam-ltd-/en/home.html#

[19] F. Martínez-Álvarez, A. Troncoso, G. Asencio-Cortés, and J.C. Riquelme, A survey on data mining techniques applied to electricity-related time series forecasting, *Energies*, *8*(11), 2015, 13162–13193.

[20] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, *Time series analysis: Forecasting and control.* (Hoboken, NJ: Wiley, 2015).

[21] M. Cai, M. Pipattanasomporn, and S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, *Applied Energy*, *236*, 2019, 1078–1088.

[22] J.W. Taylor, Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, *54*(8), 2003, 799–805.

[23] R. Banik, P. Das, S. Ray, and A. Biswas, Prediction of electrical energy consumption based on machine learning technique, *Electrical Engineering*, *103*(2), 2021, 909–920.

Biographies

Megha Sharma received the B.Tech. and M.Tech. degrees in computer engineering from Rajasthan Technical University, Kota, India, in 2011 and 2016, respectively. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Malaviya National Institute of Technology Jaipur, India. Her current research areas are load forecasting, demand response, and artificial intelligence.

Namita Mittal received the Ph.D. degree in computer science and engineering from Malaviya National Institute of Technology Jaipur, India, in 2011. She is an Associate Professor with the Department of Computer Science and Engineering, Malaviya National Institute of Technology Jaipur, India. Her research areas are data science, information retrieval, data mining and NLP.

Anukram Mishra received the B.E. degree in electronics and communication from MBM, Jodhpur, India, in 1986, the M.Tech. degree in aircraft production engineering from the Indian Institute of Technology Madras, India, in 1990, and the Ph.D. degree in computer science and engineering from Malaviya National Institute of Technology Jaipur, India, in 2014. He is currently a Chief Technology Officer (CTO, metering solution) with the Genus Power Infrastructures Limited, Jaipur. His research areas are smart grid, advance metering infrastructures and demand response.

Arun Gupta received the B.E. degree in computer science and engineering. He is currently working as the Head of Software Platforms with Genus Power Infrastructures Limited for over two years. He had been working in the Software Industries for over 20 years with 12 years dedicated to advanced metering solutions, delivered across all continents. He enjoys designing scalable secure solutions and has patented his architecture for integrated head end utility metering systems.