In 2024, Yenagimath et al. developed an advanced electrical load forecasting model by using improved feature selection techniques which comprises of skewness, kurtosis, IPCA and a novel Improved Information Gain approach and hybrid deep learning algorithms that is ANN and LSTM fused together. In feature selection first we extracted statistical characteristics (skewness, Kurtosis, variance). Then, they used Improved PCA to reduce dimensions keeping important patterns in check. Finally, they used Improved Information Gain Technique to select most important feature for load forecasting. The final output was achieved by using hybrid of ANN and LSTM models which resulted in improved forecasting accuracy over traditional methods.

In 2024, Fakhryza et al. studied about the application of artificial neural networks (ANN) for short term peak load forecasting in the 150kV Semarang power system. In their study they used input features such as historical peak load, minimum load, population and energy production which emphasized the role of careful variable selection and model tuning. The final result depicted how ANN models achieved high forecasting accuracy with MAPE values less than 10% which outperformed traditional methods like ARIMA and regression.

In 2024, Gao et al. proposed a load forecasting framework using pre-trained large language models (LLMs). In their approach they transformed load sequence data into natural language which is under stable to LLMs to predict better result. Even they incorporated statistical enhancements to strengthen forecasting accuracy and got rid of the hallucination issues using data enhancement techniques. Overall, the result demonstrated state of art performance across real world datasets which highlighted the potential of LLMs in complex load forecasting task.

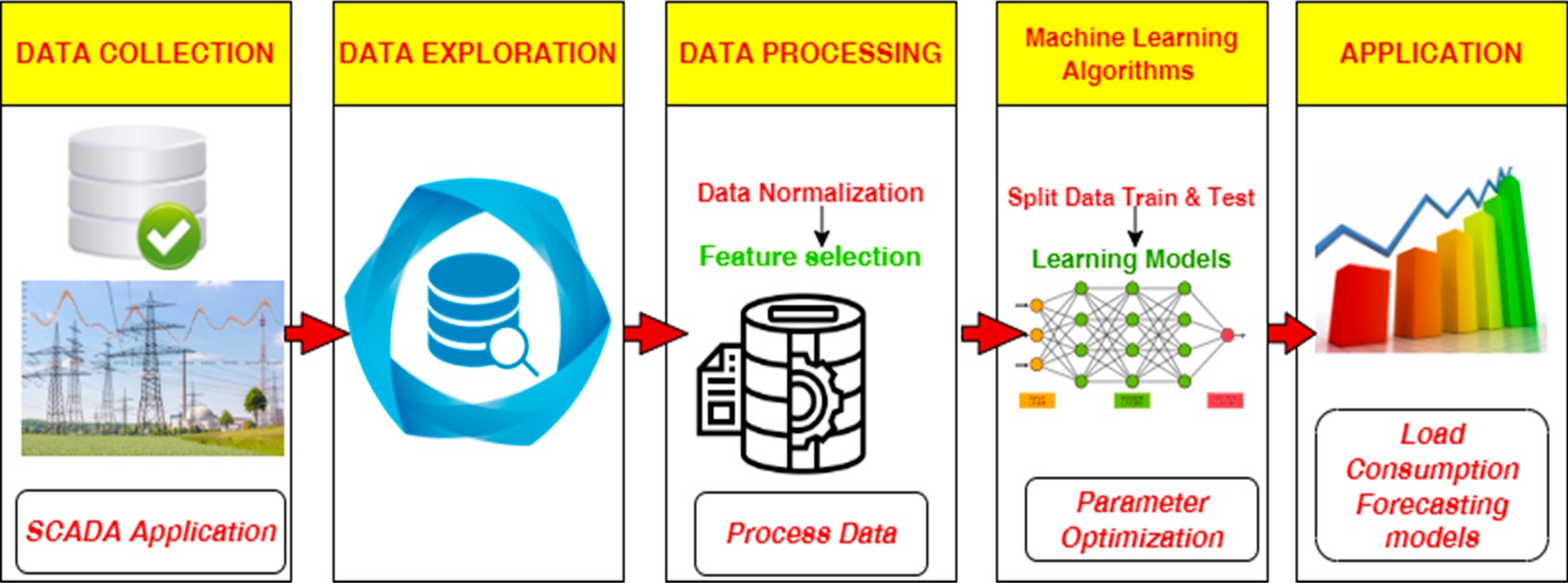
In 2023, Barigye Osbert gave a short- term load forecasting model using ANN, trained with weather, holiday and load data from three towns of Panama. In case of ANN model, it was a multilayer ANN with ReLU activation function used for proper capturing of complex load patterns. Adam optimizer and MSLE loss function was used to avoid overfitting. Evaluation was done using MAPE and correlation between predicted and actual values. This model achieved a high prediction accuracy with a MAPE of 3.27% and it was deployed as a web application vis streamlit for real time forecasting.

In 2024, Nabavi et al. proposed a hybrid model which combined Discrete Wavelet Transformation and Long Short-Term Memory (LSTM) networks to for improved electrical load forecasting accuracy. DWT decomposes load and weather signals into multiple frequency bands to capture patterns. LSTM is used to learn dependencies in load and weather data over time. In case of data preprocessing min-max scaling, missing values filling and separate handling of holidays and weekends is done. Due to the decomposition and enhanced feature extraction model performed superior outperforming LSTM, NARX and SVM models. Even sentiment analysis is also covered that is the role of input features such as solar radiation, holidays etc. in this model.

In 2023, Abumohsen et al. developed deep learning models using LSTM, GRU and RNN algorithms for short term electrical load forecasting in Palestine. These were trained on real time operational data from SCADA systems. Optimized batch size, learning rate, hidden layers, dropout and activation was used for better hyperparameter tuning. To evaluate and compare performance R-squared, MAE, RMSE and MSE were used as a performance metrics to compare all the forecasting models. The results showed that GRU model achieved the best performance with the highest R-squared and the lowest error rates which leads to enhanced load distribution and reduced outages.

1. **Methodology**

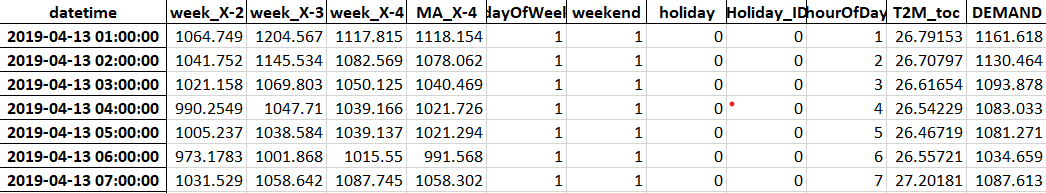
The basic and the initial step of the methodology starts with data collection and preparation. Further steps combine of data exploration, preprocessing of data for the purpose of machine learning. After all the steps mentioned above the next step is to use different machine learning, deep learning and statical algorithms i.e., (ANN, Random Forest, XGBoost, CNN, ARIMA, DWT etc.) for forecasting the electrical load forecasting and using performance metrics to compare different algorithms performance and to select out the best one. In the final step all trained models were integrated in a Streamlit-based interface where user can put their data sets and effectively compare the performance on new datasets. Fig.1 depicts about the methodology used as shown below:



**Figure 1.** Workflow used for the electrical load forecasting models.

* 1. **Data Collection and its description**

The data used here is Electrical Load Forecasting data from Kaggle.com which is a useful data to train and test forecasting models and compare results with the official forecast results. The dataset contains diverse data sources to enhance machine learning learning-based forecasting. It includes hourly electricity load data from Panama’s grid operator which contains both historical post-dispatch as well as weekly pre-dispatch forecasts. Other information such as school calendars, public holidays and weather variables (temperature, humidity, wind speed) enriches the dataset. The dataset contains 121,273 rows and 16 columns which includes features like date, hour, demand, temperature, humidity, precipitation, wind speed etc. Table shows the basic structure of the dataset.

**Table 1.** Basic overview of the dataset.

* 1. **Exploratory data Analysis**Exploratory data analysis is the processin which features are examined, correlations are found out and hidden pattern is unfurled from the dataset. In this step several methods are utilized. Here firstly all the data was collected from multiple Excel Sheets and concatenated into a single Data frame. Further datetime column was converted to a proper datetime format and then it was to time-index the dataset for temporal analysis. Further missing values were analyzed to ensure data quality before modelling.
     1. Correlation Analysis

Correlation is the measure which is statistical in nature which describes the strength and direction of the linear relationship between two variables or features. Here, a correlation heatmap was generated using Python’s Seaborn and Matplotlib libraries to evaluate the relationships among the numerical features present in the dataset. The specific correlation can be calculated using the formula:

r = (∑ (- ) (- )) / ()

where:

* r = correlation coefficient
* , = individual sample points
* , = mean of x and y respectively

Heatmap provides the visual insight about the positive as well as negative correlations among variables. Here demand shows positive correlation with the features such as hour of the day and previous week demand. Correlation plays a significant role in feature selection as high correlation to the target are most probably the strong predictors which helps in building more effective models. Figure 2 shows feature correlation heatmap without lag features and rolling statistics:

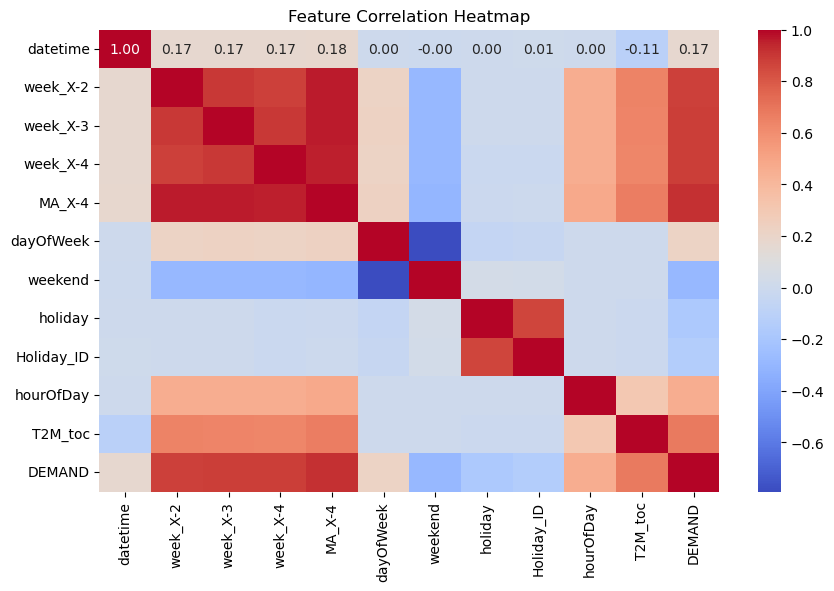


Figure 2: Feature correlation heatmap without lag features and rolling statistics

* + 1. Demand Analysis

To get a better understanding of electricity consumption peaks and drops, boxplots were used to find out the demand distributions across time related dimensions or features. This visualization helps in forecasting the time, months and hours of overloading to take necessary measures to avoid overloading problems. It also helps us to find out the minimum load durations so as we can reduce the losses also.

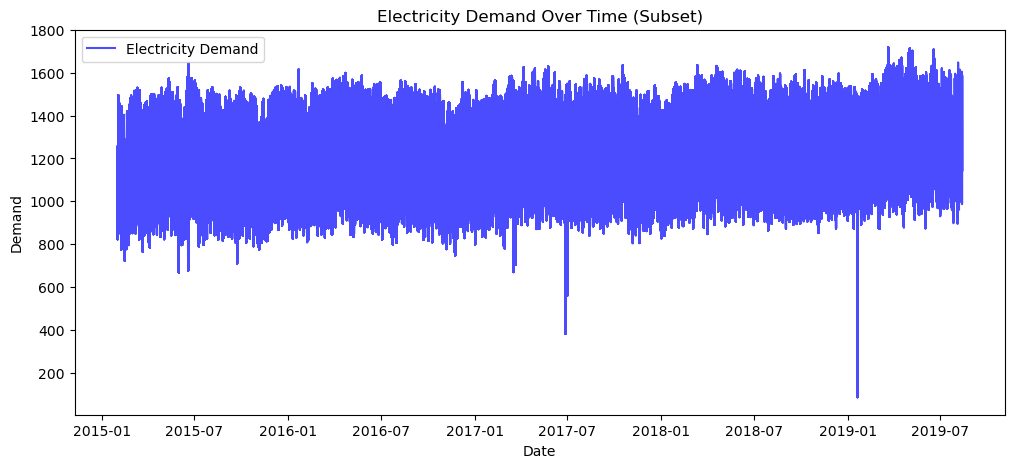


Figure 3: Electricity demand over time.

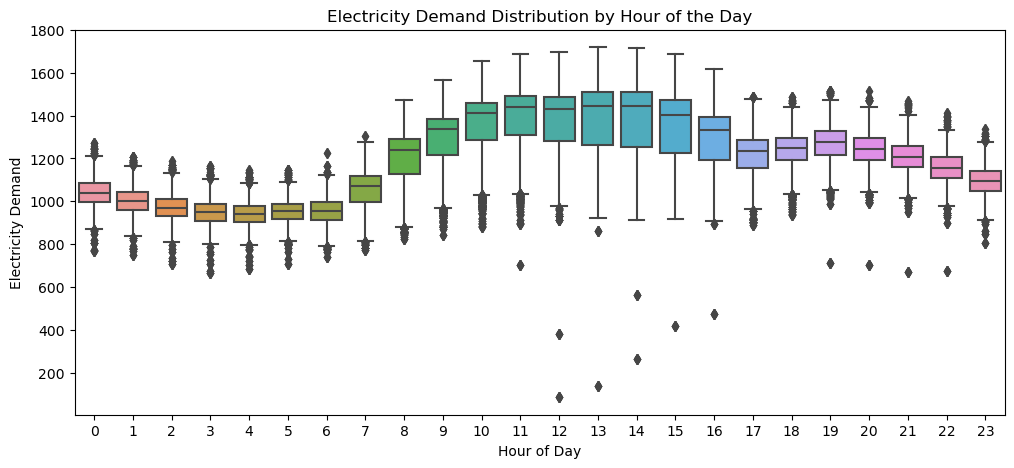


Figure 4: Electricity demand distribution by hour of the day.

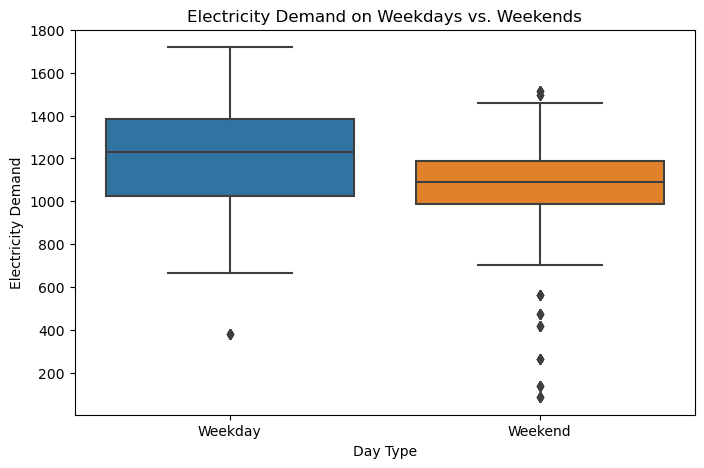


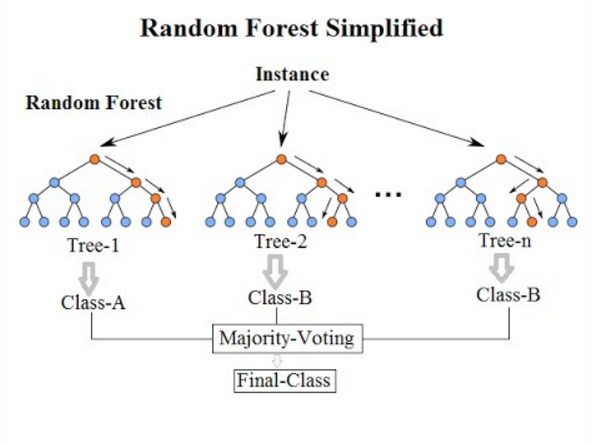
Figure 5: Electricity demand on weekdays vs weekends.

Exploratory data analysis helps us to recognize the patterns and on the basis of that it helps us in making decision to use specific machine learning algorithms. Through visualization we got to know that demand is much greater on weekends also in a day maximum demand is during 10th hour to 16th hour of the day. Overall demand has also increased over the year on a scale with the passing time.

* 1. **Forecasting Methodology**

This part is focused to give ways to forecast using different algorithms with the datasets accumulated from above steps. Here we preprocess the data using different techniques such as data normalization specifically min-max scaling, feature selection that is figuring out the dimensions to be used for prediction and the predicted result separately to evaluate the models. Further the data is used to train the machine learning algorithms. The machine learning algorithms used are:

* **Random Forest**: Random Forest is an ensemble learning algorithm whose base is Decision Tree. It builds multiple decision trees during training and merges their output to improve accuracy and control overfitting. Each decision tree is trained on a random subset of data and feature making the model robust to noise and variance. In our case it captures non linear relationships between variables which makes it efficient enough to use in forecasting scenarios.

****

**Figure 6:** Random Forest working structure.

* XGBoost (Extreme Gradient Boosting): It is a powerful boosting algorithm which builds trees sequentially. Here the error is rectified as we go forward in sequence as the tree is increased. Here regularization and tree pruning are used to prevent overfitting and give high accuracy. This algorithm handles the complex relationships between features making it useful enough for forecasting.

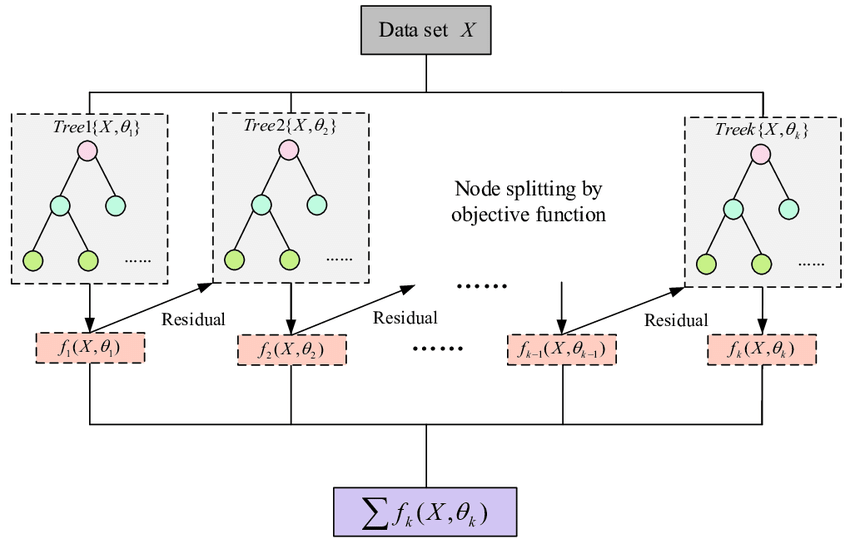


Figure 7: XGBoost working structure.

* **Artificial Neural Network:** ANNs are a similar structure of human brain which consist of interconnected layers of neurons. In case of load forecasting ANNs learn from the historical demand and the features which are influencing like hours of day, temperature etc. to predict the consumption. ANNs have great ability to capture nonlinear dependencies which makes them suitable for this use case.

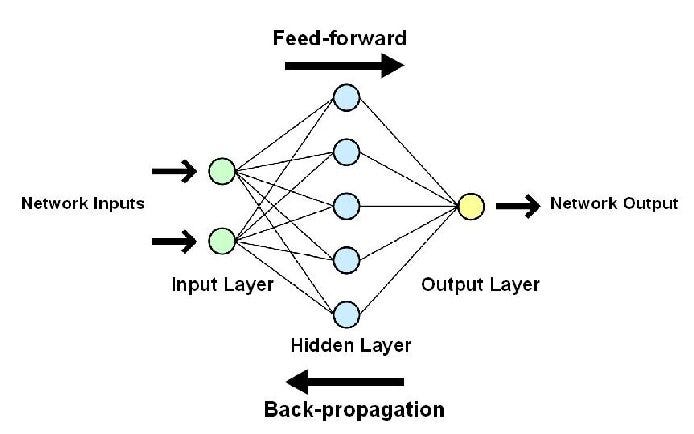
****

Figure 8: ANN working structure.

* **Long Short-Term Memory (LSTM):** It is typeof recurrent neural network (RNN) which is specially designed for sequence and time series data. It most specific feature which makes it different from others is that LSTM retains long-term dependencies through memory cells and gating mechanism. In case of load forecasting which has data, which is in timeseries i.e. dataset has temporal dependencies so LSTM is an ideal and efficient model to be used.

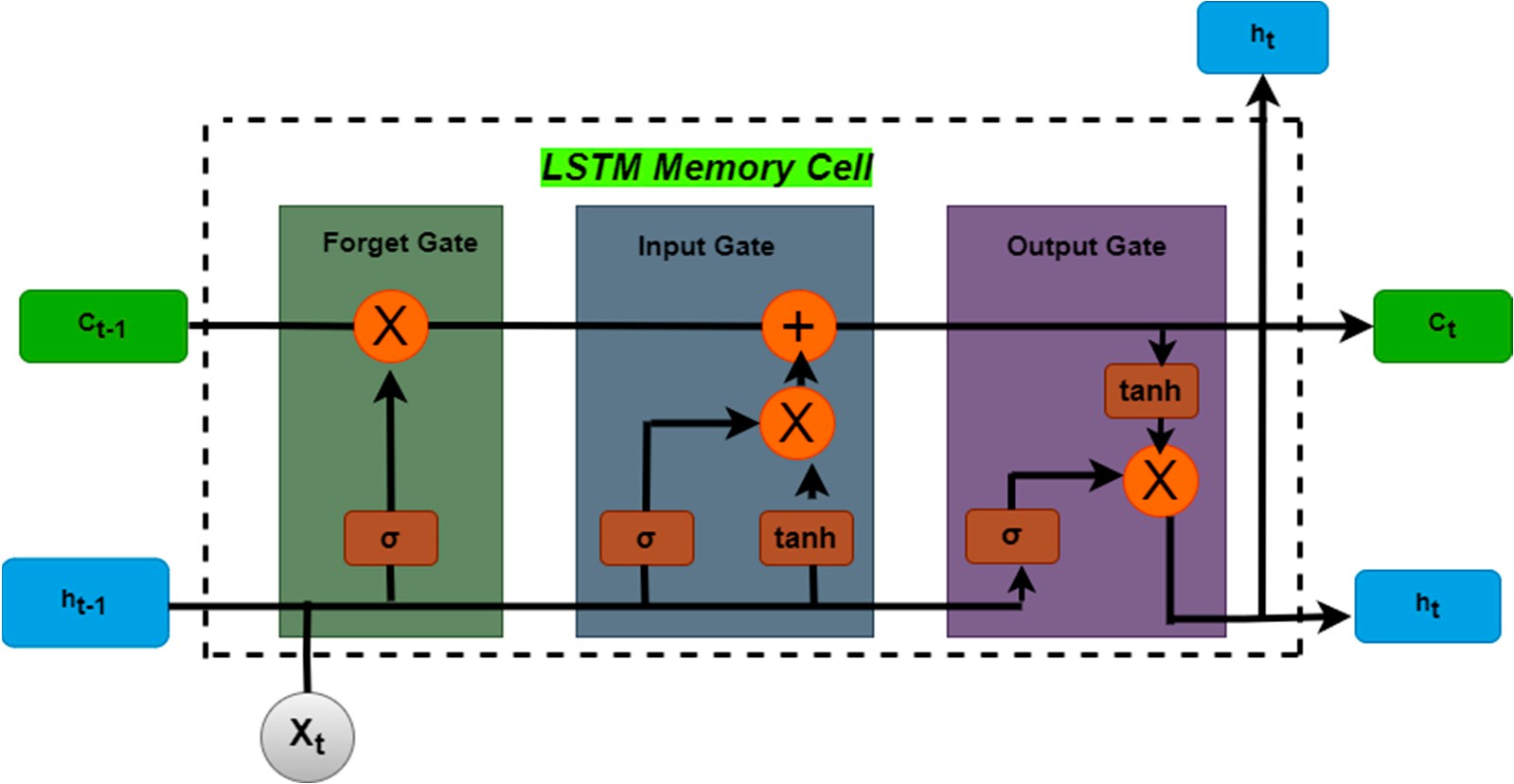


Figure 9: LSTM working structure.

* **ARIMA (AutoRegressive Integrated Moving Average):** It is a classical time series forecasting model which combines autoregression (AR), differencing (I) and moving average (MA) components. It is used much for well- suited univariate stationary time series data which helps to catch linear trends and seasonality. ARIMA struggles with non linear patterns and with data sets having high noise.

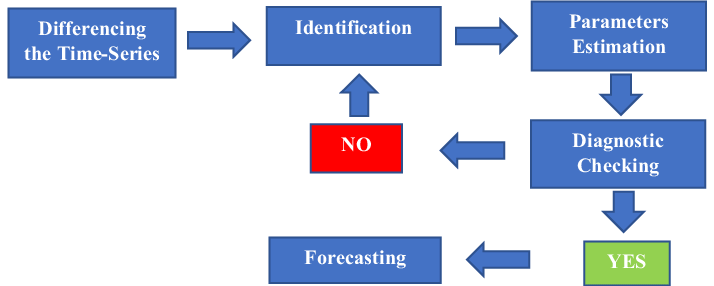
****

Figure 10: Block Diagram of ARIMA model.

* **GRU (Gated Recurrent Unit)**: It is a simplified variant of LSTM which retains memory from gated mechanisms but uses fewer parameters which makes it faster and less prone to overfitting. Temporal dependencies are efficiently handled by this model without generating much complexity and quick training.

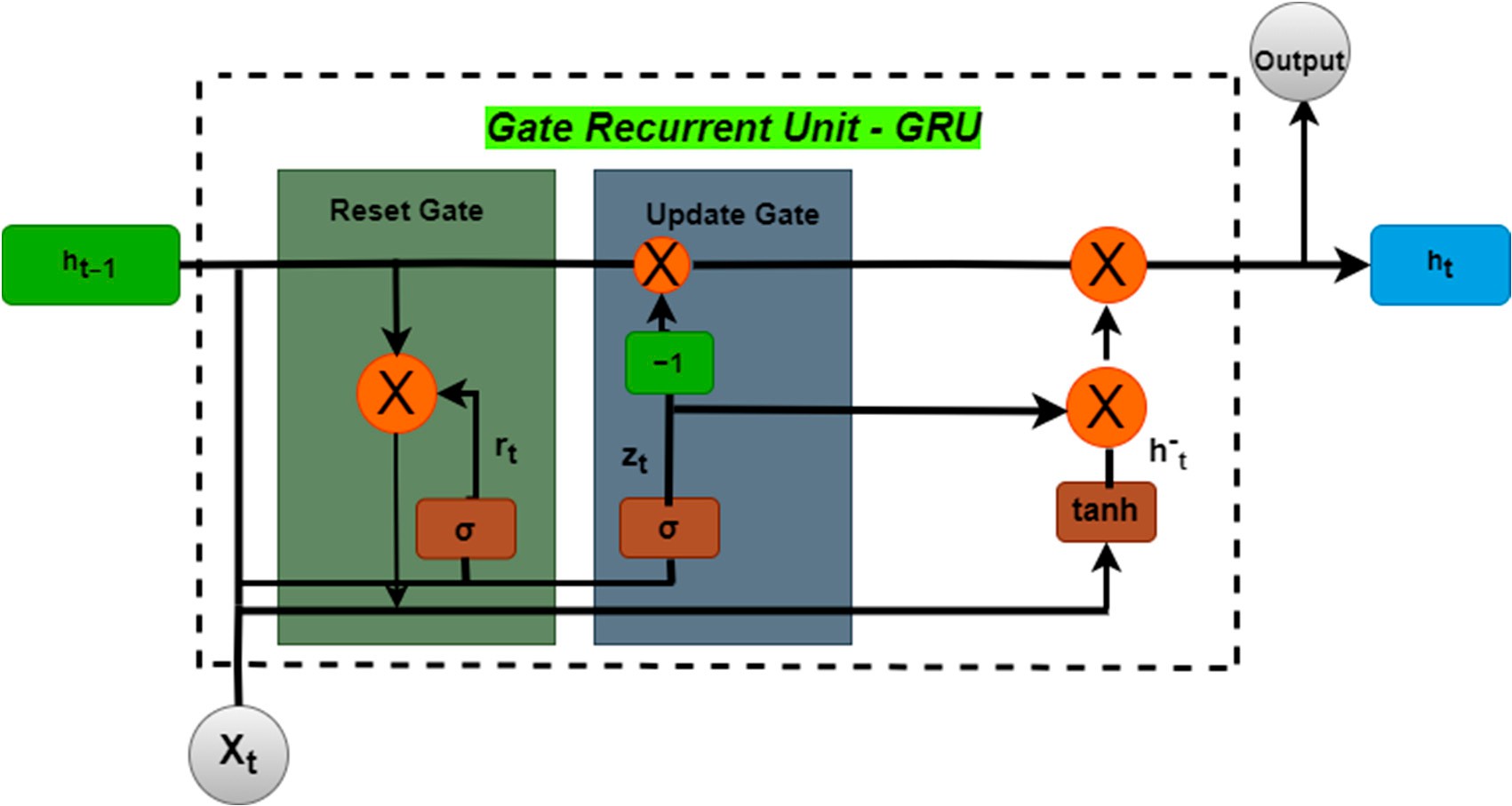


Figure 11: GRU working structure.

* **Temporal Fusion Transformer (TFT)**: TFT is a deep learning architecture specially designed for multi horizon time series forecasting which combines attention mechanism, recurrent layers and interpretable outputs to capture all type of dependencies. It effectively handles multiple inputs and provides high accuracy along with high interpretability.

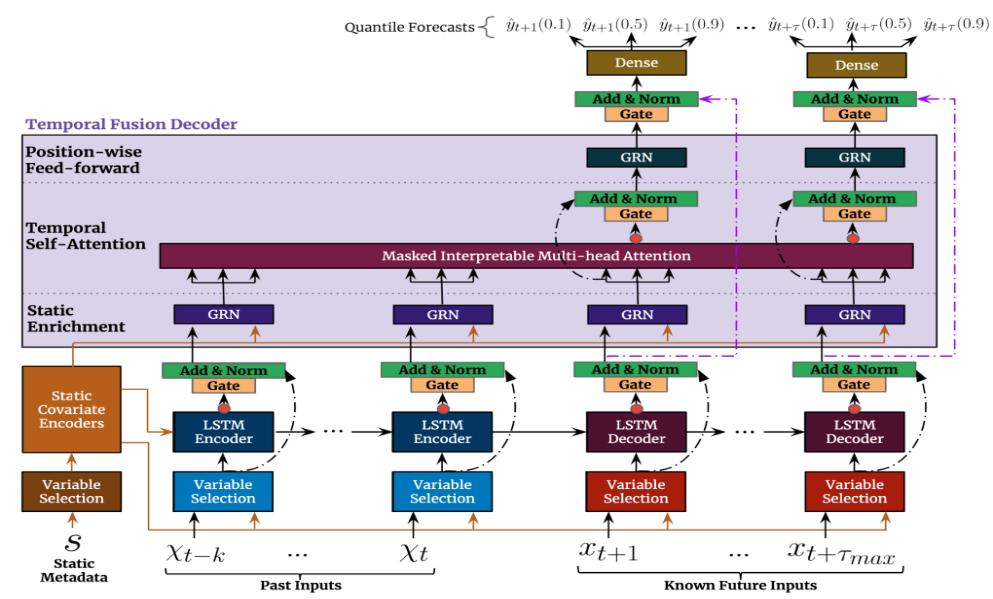


Figure 12: TFT working structure.

Apart from all these models’ hybrid models are also used which is combination of the specified any of the two models from above mentioned one. Discrete Wavelet Transform (DWT), a signal processing technique which decomposes time series data into different frequency components is also used along with LSTM to enhance model performance, reduce noise and extract meaningful features before feeding into neural networks. Few of supervised machine learning algorithms such as SVR which tries to fit the best line or curve with specified margin has also been taken in consideration to evaluate the accuracy of different algorithms and models.

We also focused on hyperparameters tuning for the models. Tuning refers to the process of choosing optimal set of hyperparameters for the learning algorithms which are mentioned as follows:

* Optimizer
* Hidden layers
* Epochs
* Batch size
* Learning rate
* Activation Function

For the purpose of tuning GridSearchCV is used to find out the best combinations to be used to get the best result where we specified the values of each parameter and it ties all the combinations and gives the final result as the one with best accuracy as per the evaluation metrices.

There are several metrics to measure the performance. Following are the ones we have used here:

* **Mean Square Error**: It is the calculation of mean squared deviation between the actual and the predicted values.

Where yt is the original value and ytp  is the predicted value.

* **Root Mean Square Error**: It is mathematically equal to the square root of mean squared error.

* **Coefficient of determination (R2 Score)**: It is a statistical measure of how well the predicted values approximates the actual data. It shows the proportion of variance in the dependent variable that can be predicted from the independent variables.

Where SS regression is the regression sum of squares and SS total is the sum of all squares.

* **Mean Absolute Error**: It is the mean of the absolute value of the errors.