FORECASTING UTILITY DEMAND BASED ON HISTORIC DATA APPLYING DEEP LEARNING TECHNIQUES

A PROJECT REPORT

Submitted by

ANSIYA S

TKM20EEPS05

to

the **APJ Abdul Kalam Technological University**

in partial fulfillment of the requirements for the award of the

Degree of

MASTER OF TECHNOLOGY

In

POWER SYSTEMS

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

T.K.M COLLEGE OF ENGINEERING

KOLLAM-5

2020-2022

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

THANGAL KUNJU MUSALIAR COLLEGE OF ENGINEERING

KOLLAM

CERTIFICATE

This is to certify that the Project report entitled **'FORECASTING UTILITY DEMAND BASED ON HISTORIC DATA APPLYING DEEP LEARNING TECHNIQUES '** submitted by **Ms.Ansiya S** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirement for the award of the Degree of Master of Technology in Power Systems,Electrical and Electronics Engineering is a bonafide record of project work carried out by her under our guidance and supervision.This report in any form has not been submitted to any other University or Institute for any purpose.

Dr.BIJUNA KUNJU K
Professor [Internal Supervisor]
Department of Electrical and Electronics

EXTERNAL EXAMINER

Prof. SHANAVAS T N
Associate Professor [PG co-ordinator]
Department of Electrical and Electronics

Dr.SABEENA BEEVI K
Associate Professor[HOD]
Department of Electrical and Electronics

ACKNOWLEDGEMENT

I am obediently thankful to **God Almighty**, praise and glory is to Him, for all His uncountable bounties and guidance, without which, this project would have never been a reality.

It is my privilege and pleasure to express my gratitude and indebtedness to **Dr. T A Shahul Hameed**, Principal of TKM College of Engineering, and **Dr. Sabeena Beevi.K**,Associate Professor, Head of the Department, Dept of Electrical and Electronics Engineering for providing all the necessary facilities and support.

My heartfelt gratitude to **Dr.Bijuna Kunju K**, Professor, and Internal Supervisor, Department of Electrical and Electronics Engineering, for her constructive guidance and suggestions in designing and implementing this project.

I am greatly obliged to **Prof. Shanavas T N**, Associate Professor, PG Coordinator, Department of Electrical and Electronics, for his encouragement and support.

I express my thanks to **Prof. Jibi P Mathew**, Asst. Professor, Project coordinator, Department of Electrical and Electronics, for his constant support and technical guidance provided during this project work.

I show my extreme gratitude to all Faculty member and Technical staffs in Electrical and Electronics Dept. for providing all the help and necessary facilities to present this project and my deep hearted cheers to my parents and all my friends who extended their support and co-operation towards the successful presentation of the project.

ANSIYA S

ABSTRACT

Load Forecasting has been one of the most emerging area of research for the last several years. Power system scheduling, reducing the expense of spot purchase of power, day-to-day operation and efficiency are some of the very interesting outcomes that can be explored by load forecasting. The development of Smart Grid and Energy Management System, aggregates large-size of data adding to the complexity of the system. Big Data Analytics is a modern day technique that can extract information from these complex and large datasets. Typical load profiles exhibit periodicity, allowing to extract patterns from demand time series and available historical recordings. However there are many factors that cause strong variations of the demand patterns from the predicted values.Deep learning models can learn from a considerable volume of big data, insufficient data that contains missing values, heterogenous data. Artificial intelligence (AI) can be combined with big data technology to solve complex problems in demand forecasting. This project is aimed at comparing the load prediction based on Artificial Neural Network(ANN), Long Short Term Memory(LSTM) and Bidirectional Long Short Term Memory(BLSTM).

Contents

List of Figures

List of Tables

ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
BLSTM	Bidirectional Long Short Term Memory
DNN	Deep Neural Network
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Square Error
ReLU	Rectified Linear Units
ADAM	Adaptive Moment Estimation

Chapter 1

INTRODUCTION

1.1 Overview

Load forecast plays a crucial role in all aspects of planning, operation as well as the control of electric power systems.The term load forecasting means estimating the demand in advance It is very important for operating a power system network reliably and economically.The need for load forecasting has now become a much-discussed issue in recent times. Reliable and accurate forecasting tools would enable the electric utilities to plan ahead of time for peak demands and so proper allocation of the resources. Otherwise, this will lead to the underestimation or the overestimation of resources. The underestimation may lead to the poor quality of services and the overestimation may lead to overcapacity and the utility has to overcome this cost of overcapacity. The accurate load forecasting will also reduce the spot purchase of power. The electrical industry is now facing problems in the imbalance between supply and demand.

The innovations in the field of electrical utilities including smart grid technologies and renewable energy resources aggregate a large amount of data.Big data analytics now become the more appropriate solution for the extraction of correlation between these data. The sources of big data are generally social data, machine data, and transactional data.There is an acknowledged understanding that big data analytics can provide a competitive advantage.

Transfer function models were used earlier for demand forecasting. These transfer function models can be bivariate or multivariate depending upon the number of independent parameters in the dataset. The transfer function model for large datasets can be more complex due to the independent variables in the dataset. This complexity has made the utilities switch from the transfer function model to Artificial Intelligence techniques where the relationship between the different parameters is not essential due to the feature extraction technique. Some of the Artificial Intelligence techniques applied to load forecasting, are linear regression,random forest, and support vector machine, among them neural networks(NNs) have gained the most attention due to its ability to handle non-linearity. The fundamental structure of NNs are multilayer perceptron which are used to forecast load on past data. But the most common issue in all of these Neural Networks are vanishing and gradient problem. Recurrent Neural Network (RNN) can be used to analyse time series sequential data. The hidden state of Recurrent neural network has a looping link. This looping requirement ensures that the given data captures sequential information. The long term dependency problem in RNN can be addressed by its variants such as Long Short Term Memory(LSTM) and Gated Recurrent Unit(GRU).LSTM[10] is more accurate when dealing with dataset containing longer sequences.It has the ability to handle temporal features and long term dependencies.LSTM networks are unidirectional models which uses past data as input.BidirectionalLSTM(BLSTM)[8]is a neural network which can propogate both forward and backward, which can extract the information from both past and future data, which can learn the trend of time series pattern from the whole data.

1.1.1 Motivation and objectives

This project aims to build an efficient algorithm by Deep learning using Bidirectional LSTM and compare with other deep learning models for day ahead demand forecasting.

Other objectives also include:

- Minimizing the errors between forecasted demand and actual demand.

Chapter 2

LITERATURE REVIEW

In[2],the author proposed short term load forecasting technique using Artificial neural network.This study anticipate hourly electricity load forecast in the NEPOOL region (ISO New England).The standard Matlab neural network fitting toolkit, which contains three layers and 20 neurons, was employed. The hidden layer's activation function is sigmoid, while the output layer's is linear.Separate study is done to analyse the load profile of weekdays and weekends. M.A.P.E is used to determine the performance of proposed model.

The author Warrior in [1]compares neural networks, decision trees and Conditional Restricted Boltzmann Machines algorithms for forecasting short-term demand with less amount of testing data. Machine learning algorithms can be used to analyse dataset containing less values.Year,month,day of the week,day of the month,power consumption datas are taken as the input features.hour ahead and day ahead power demand forecasting is implemented.Among them neural network outperform decision tree in day ahead prediction.For hour ahead CRBM gives more accurate results.Neural network and decision tree will give more accurate result even the dataset contains inconsistent values.

In[3]the author proposed deep neural network using Long Short Term Memory for the day ahead prediction.Time series analysis is conducted to find out the load characteristics.The input features is selected using correlation analysis.The training and testing set is selected using moving window method.

In[4]a hybrid GRU-CNN model is implemented.The hybrid model is build inorder to combine the benefits of both GRU and CNN.Gated Recurrent Unit network is good at processing sequence data processing.The spatiotemporal matrixes can be easily extracted using Convolutional Neural Network.The proposed hybrid model is compared with other models.The result shows that hybrid GRU-CNN model outperformed other models.

In[5],a stacked Bidirectional LSTM is proposed for the day ahead and week ahead load forecasting based on historical and weather parameters.Unlike by using the unidirectional deep learning models like ANN,LSTM the bidirectional LSTM can propogate in both forward and backward direction, which can extract the hidden pattern of the load effectively.The analysis is done with the residential demand data in Scotland.The SBLSTM is capable of resolving vanishing gradient and vanishing exploding problem in traditional neural netwoks.

In[6]a hybrid Bidirectional Recurrent Neural Network is proposed for the load forecasting.The network is build with multilayer Bidirectional GRU-LSTM model.As it is bidirectional network , it can extract the hidden correlation of the load pattern.This proposed model combines both the benefits of LSTM and GRU network.The proposed model is compared with SVM,LSTM and BP models.The model is also used to forecast the demand for different seasons.

In[7]proposed an attention based Bidirectional LSTM model for the power load forecasting.The attention mechanism in the forecasting model can project the effect of key parameter by increasing the weight of the particular feature.In the study covariance of diffent parameter is analysed.MAPE and RMSE is used as the evaluation metrices.The proposed model have the benefit of the attentian mechanism as well as the forward and backward propogation of the BLSTM model.

In[8]author proposed a deep bidirectional LSTM model to predict day ahead peak load for normal days and special days.The model was trained and tested by using the real time demand data from local utility.The proposed deep Bi-LSTM S2S regression forecasting model proved to be more accurate and flexible for forecasting the non-linear, volatile, and nonstationary day-ahead “peak” electricity demand.The proposed model is compared with other models like shallow Bi-LSTM S2S, deep LSTM S2S and medium Gaussian support vector regression (MG-SVR) forecasting models.rmse and mape are used as the evaluation metrices.

Chapter 3

BIG DATA ANALYTICS

3.1 Introduction

It refers to the use of advanced analytic techniques to very large, heterogeneous data sets, which can contain structured, semi-structured, and unstructured data from a variety of sources with sizes ranging from terabytes to zettabytes.The term Big data is used to describe data sets that are too large or complex for typical relational databases to acquire, maintain, and handle in a timely manner.The analysis enables analysts, academics, and business users to use data that was previously inaccessible or unsuitable to make better and faster decisions. The characteristics of Big data are:

- **volume**
- **velocity**
- **variety**
- **value**
- **veracity**

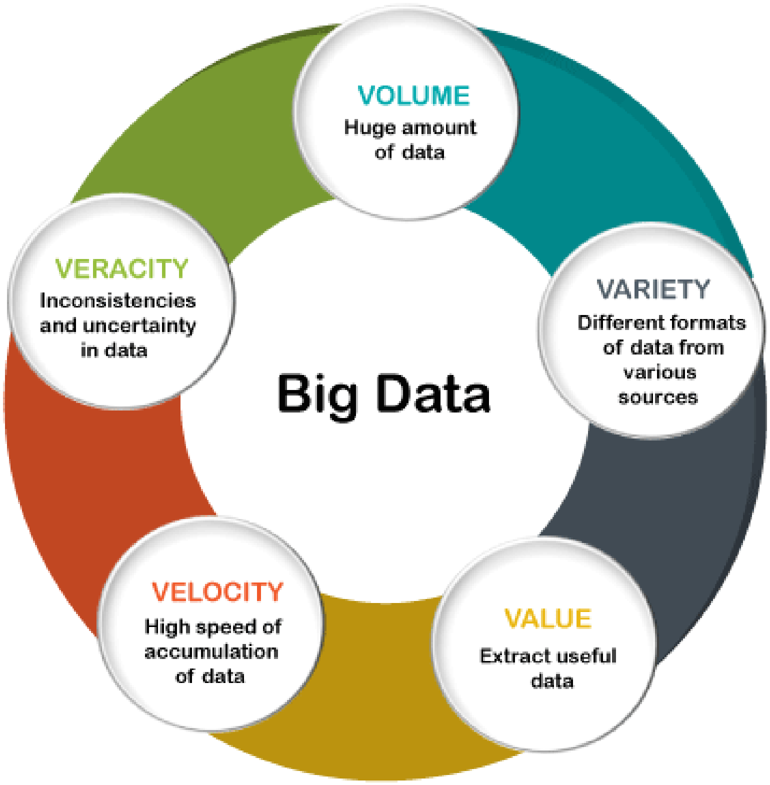

Figure 3.1. Characteristics of Big Data[13]

3.2 Different types of Big Data Analytics

1. Descriptive Analytics-It simplifies the data and summarises the past data into a readable form.

2.Diagonostic Analytics - It gives a detail and depth insight on the root cause of a problem.

3. Predictive Analytics - This type of analytics make use of historical and present data to predict future event.

4. Prescriptive Analytics - It allows business to determine the best possible solution to a problem.

Chapter 4

MACHINE LEARNING

4.1 Introduction

Machine learning (ML) is a subset of Artificial Intelligence,which focuses mostly on creating algorithms that let a machine to independently learn from data and previous experiences.Machine learning algorithms create a mathematical model with the aid of historical sample data, or "training data," that aids in making predictions or judgments without being explicitly programmed. Computer science and statistics are used with machine learning to create prediction models.Machine learning is necessary because it can perform activities that are too complex for a person to carry out directly. As a result of our limitations in being able to access such a vast quantity of data manually, we need computer systems, and here is where machine learning comes in to help us.

4.2 Types of Machine Learning

There are different ways an algorithm can model a problem based on its interaction with the experience or environment or the input data. It is mainly of three types as given below:

4.2.1 Supervised Learning

In this type of learning,machines are training using labelled inputs.This labelled data work as a supervisor that teaches the machine to predict the output.This labelled data will provide input data and output data to the machine so that it can find a relation between the input and output variable.

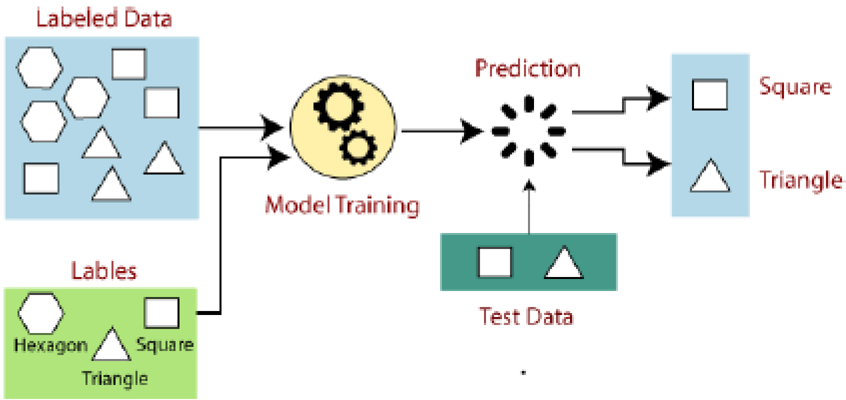

Figure 4.1. Supervised Learning[14]

4.2.2 Unsupervised Learning

In this type of learning,the model will not trained using labelled data,the model itself find the relation from the given data.This cannot be used for a regression or classification problem.Unsupervised learning can used to perform the task of clustering an image dataset into groups by finding the similarity between images.

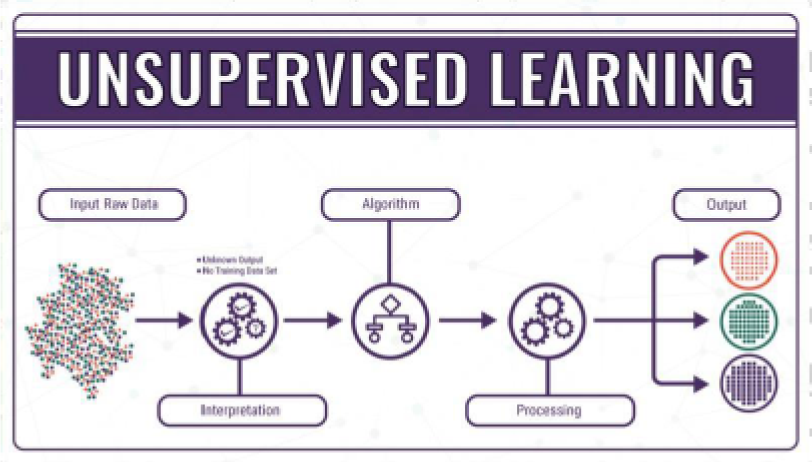

Figure 4.2. Unsupervised Learning[15]

4.2.3 Semi-Supervised Learning

Input data is a mixture of labeled and unlabeled examples. There is a desired prediction problem but the model must learn the structures to organize the data as well as make predictions.

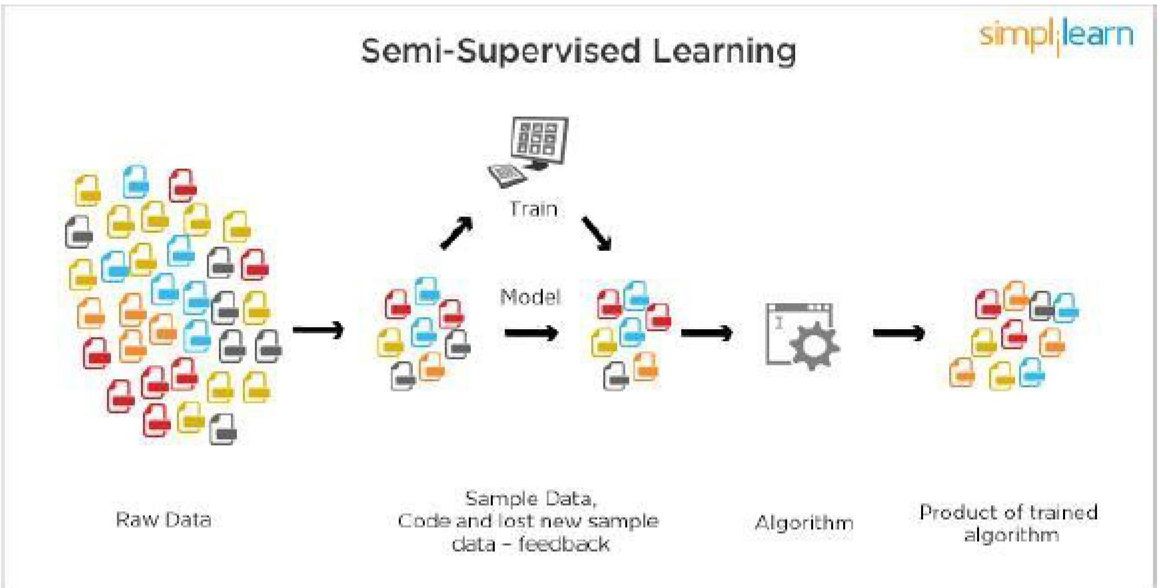

Figure 4.3. Semi-Supervised Learning[16]

Chapter 5

DEEP LEARNING

5.1 Introduction

Deep learning can be considered as a subset of machine learning. It is a field that is built on self-improvement through analysis of computer algorithms.. While machine learning uses simpler concepts,It uses artificial neural networks, which are created to mimic how humans think and learn, whereas machine learning uses simpler concepts. Until recently, neural networks were limited by computing power and thus were limited in complexity. However, advancements in Big Data analytics have permitted larger, sophisticated neural networks, allowing computers to observe, learn, and react to complex situations faster than humans. Deep learning has aided image classification, language translation, speech recognition.

5.2 Artificial Neural Network

Neural Networks are a type of computational learning system that employs a network of functions to recognize and translate a data input in one form into a desired output in another. Human biology and the way neurons in the human brain work together to understand inputs from human senses inspired the artificial neural network concept.In simple terms, they are a set of algorithms that attempt to recognise correlations and information in data using a process that is inspired by and functions in the same way as the human brain.

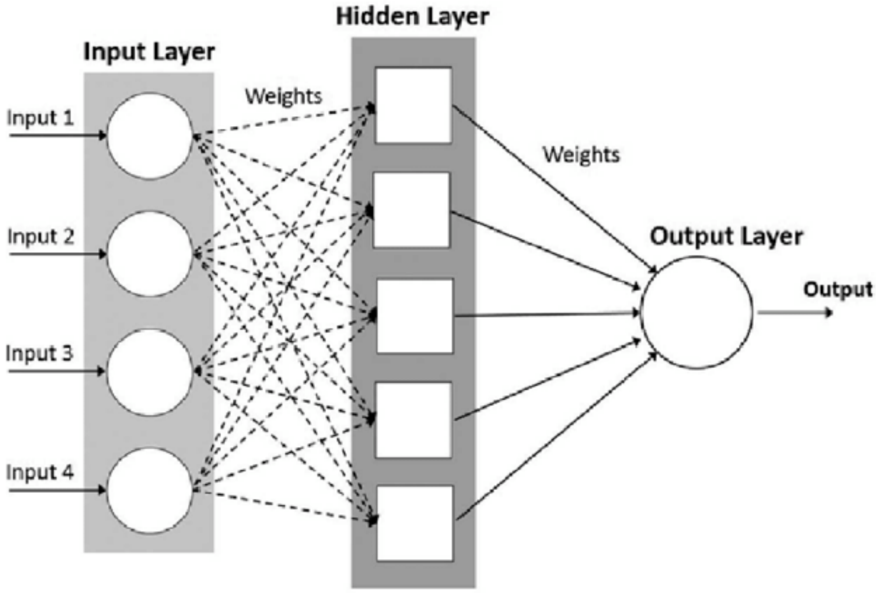

Figure 5.1. Artificial Neural Network[17]

An Artificial Neural network consist of three components.

- **Input Layer**
- **Hidden Layer**
- **Output Layer**
- **Input Layer**-Inputs/information from the outside world that the model uses to learn and draw conclusions are provided by the input nodes.
- **Hidden Layer**-All the computation in the input data is performed in this layer.There may be any number of hidden layer in the neural network.
- **Output Layer** -The model's output/conclusions, which are derived from all of the calculations made, are found in the output layer. The output layer may contain one

or more nodes. In a binary classification problem, the output node is 1, however in a multi-class classification problem, there may be more than one output node.

The working of Artificial Neural Network as follows:

- Input are passed to the hidden layer with some weights applied.The gradient or coefficient of each variable is called weight. It demonstrates the potency of the specific input. A bias variable is included after the weights have been assigned. Bias is a constant that aids in the model's best fit.
- An activation function is then used in the second step. Before the input is passed on to the next layer of neurons, it undergoes a nonlinear alteration called the activation function.
- The last output is given by the hidden layer.After getting predicted output from the hidden layer.The loss is calculated by the differnce in the actual and predicted value.The whole process is called the forward propogation.
- If the loss is large ,backpropogation is done to reduce the error.The process of updating and determining the weights or coefficients that allow the model to minimise error is known as back propagation.Optimizers are used to update the weights. Optimizers are techniques or mathematical formulas that modify the characteristics of neural networks, such as weights, to reduce error.

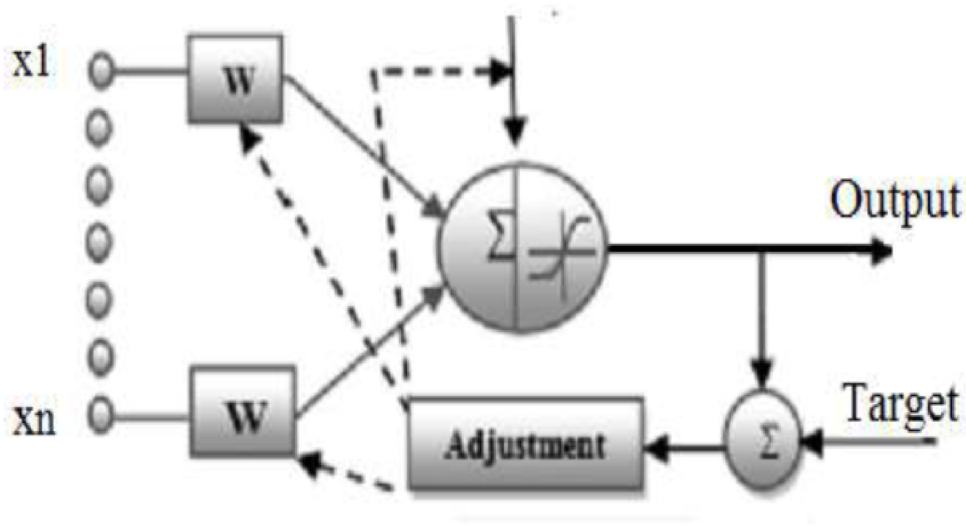

Figure 5.2. Structure of back propagation of ANN[18]

5.3 Recurrent Neural Network

The recurrent neural network (RNN), can operate effectively on sequences of data with variable input length.The input of these network depends upon the output of the previous state.But in case of traditional network these states are independent to each other.The main feature of hidden state of RNN network is it have a memory that can remember the information about a sequence but it cannot handle long sequences.GRU and LSTM are two variants.These variants use a form of memory to help make predictions in sequences over time.

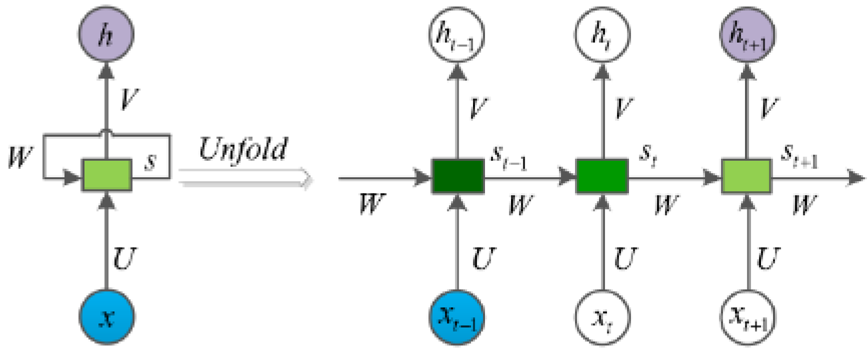

Figure 5.3. RNN and its unfolded network[19]

5.4 Long Short Term Memory

Long Short Term Memory (LSTM) networks are a type of recurrent neural network that may learn order dependence in sequence prediction problems.It was developed by Hochreiter Schmidhuber. It addressed the problem of RNN long-term dependency. As the gap length increases, RNN cannot function effectively. By default, the LSTM may hold data for a very long time. It is employed in the processing, prediction, and classification of time-series data.

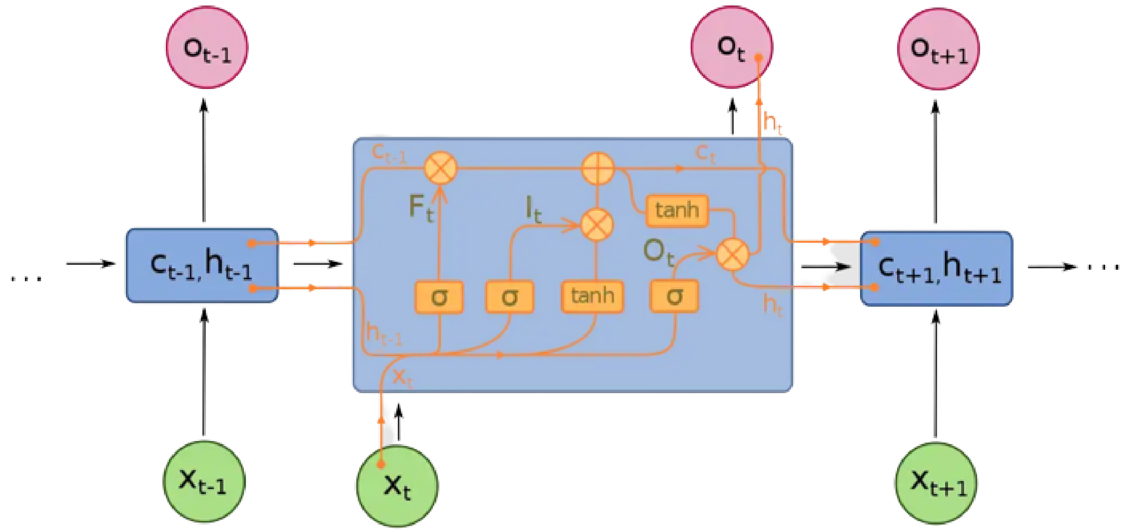

Figure 5.4. Structure of LSTM Cell[20].

LSTM block is made up of three gates and a cell state.

– **Cell State** - The cell state functions as a highway for the transportation of relative information throughout the entire sequence chain.During the processing of the sequence, the cell state carry the relevant meaningful data include . As a result, even knowledge from earlier time steps might reach later time steps, diminishing the impact of short-term memory.

– **Forget Gate** -The forget gate expels the data that is no longer relevant in the cell state. The gate receives two inputs, x t (input at the current time) and h t-1 (prior cell output), which are multiplied with weight matrices before bias is added. The output of the activation function, which receives the outcome, is binary. If a cell state's output is 0, the piece of information is lost, however if it is 1, the information is saved for use in the future.

– **Input Gate** - The input gate updates the cell state with relevant info. The inputs are regulated using the sigmoid function and filter the values that need to be remembered similarly to the forget gate. Then, a vector containing every possible value between h t-1 and x t is produced using the tanh function, which produces an output ranging from -1 to +1. The useful information is finally obtained by multiplying the vector's values .

- **Output Gate**-It will extract the needfull information out of the cell state . By using the tanh function on the cell, a vector is first created. The data is then filtered by the values to be remembered using the inputs h t-1 and x t, and the information is then controlled using the sigmoid function. The vector's values and the controlled values are finally multiplied and supplied as input and output to the following cell, respectively.

$$i_t = (W_{xi} * x_t + W_{hi} * h_t 1 + W_{ci} * c_t - 1 + b_i) \tag{5.4.1}$$

$$f_t = \sigma(W_{\mathrm{xf}} * x_t + W_{\mathrm{hf}} * h_{\mathrm{t\text{-}1}} + W_{\mathrm{cf}} * c_{\mathrm{t\text{-}1}} + bf) \tag{5.4.2}$$

$$o_{\mathrm{t}} = \sigma(W_{\mathrm{xo}} * x_{\mathrm{t}} + W_{\mathrm{ho}} * h_{\mathrm{t\text{-}1}} + W_{\mathrm{co}} * c_{\mathrm{t\text{-}1}} + bo) \tag{5.4.3}$$

$$c_{\mathrm{t}} = f_{\mathrm{t}} c_{\mathrm{t\text{-}1}} + i_{\mathrm{t}} tanh(W_{\mathrm{xc}} x_{\mathrm{t}} + W_{\mathrm{hc}} h_{\mathrm{t\text{-}1}} + bc) \tag{5.4.4}$$

$$h\mathbf{t} = o_{\mathrm{t}} tanh(c_{\mathrm{t}}) \tag{5.4.5}$$

5.5 BIDIRECTIONAL LSTM (BLSTM)

Bidirectional is a variation of RNN, which is developed to train the network with past and future input data sequence. The structure of a unidirectional LSTM depends on learning from data inputs on which its hidden states have passed through. It only sees information from the past. It has no mechanism to enable it to consider the information in the future while predicting the present. The stacked architecture can model more sophisticated data patterns. It can perform a deeper analysis of the training data. On the other hand, BLSTM can process and learn from data in both directions; from future to past and from past to future. BLSTM while going through data combines forward and backward contextual information and uses it to make predictions or classifications.

Bidirectional LSTM is the extension of traditional LSTM to train a network with

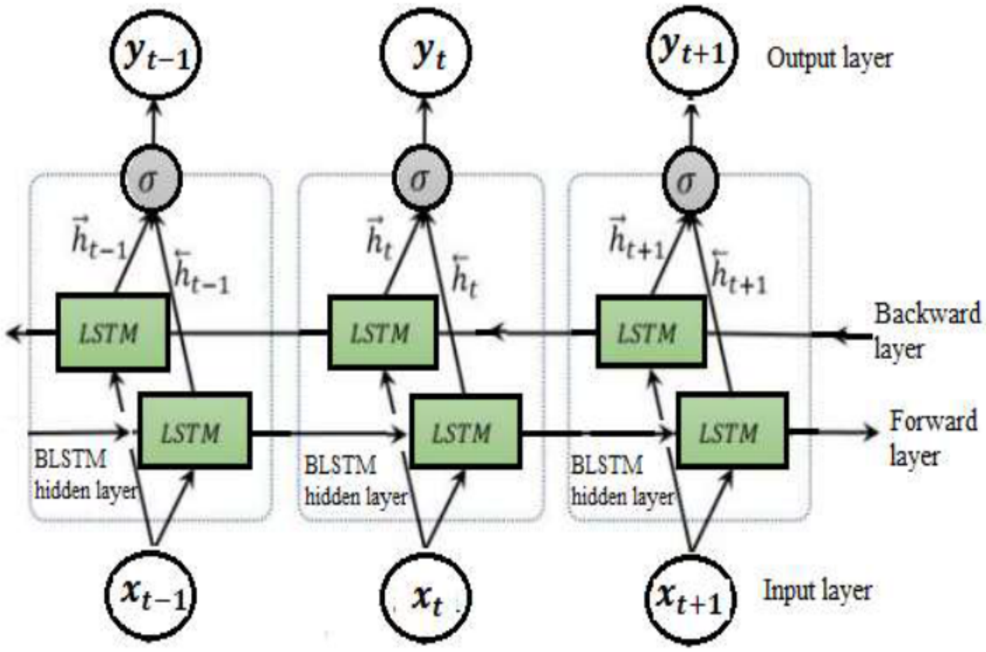

Figure 5.5. Architecture of BLSTM[21]

two LSTMs on the input data sequence. The second LSTM is a reversed copy of the first one, so that we can take full advantage of both past and future input features for a specific time step. The idea is to split the state neurons of a regular RNN in a part that is responsible for the positive time direction (forward states) and another for the negative time direction (backward states).Outputs from forwarding states are not connected to inputs of backward states, and vice versa.

Chapter 6

METHODOLOGY

6.1 Introduction

This project presents a day- ahead load forecasting methodology based on deep learning. Deep learning architectures allow the learning structure to learn complex relationships between inputs and outputs and complex patterns in data. Due to exibility of deep learning networks in python language, TensorFlow is selected as the platform.

TensorFlow is an end-to-end platform that makes it easy to build and deploy ML models. TensorFlow offers multiple levels of abstraction thereby helping to choose the right one according to the need.

6.2 Introduction to Datasets

For the day-ahead prediction, hourly demand data from KSEBL has been taken for the year 2017-2019. The accuracy of the prediction model can be improved by using historical consumption data.

The dataset parameters used for training and testing includes:

- Demand in MW output
- Calendar Holidays (weekdays and off days)
- Day of the week
- Average demand of previous day
- Average demand of previous week
- Hourly average of previous week

	A	B	C	D	E	F	G
1	Date_Time	Kerala Holidays	Day of the week	Average Demand of Prev. D	Average Demand of Prev. w	Hourly Avg. of Prev.	Demand in MW
2	01-01-2017 00:00	TRUE	1	2342.19125	2342.19125	2398.66	2398.66
3	01-01-2017 01:00	TRUE	1	2342.19125	2342.19125	2238.83	2238.83
4	01-01-2017 02:00	TRUE	1	2342.19125	2342.19125	2161.29	2161.29
5	01-01-2017 03:00	TRUE	1	2342.19125	2342.19125	2094.65	2094.65
6	01-01-2017 04:00	TRUE	1	2342.19125	2342.19125	2064.19	2064.19
7	01-01-2017 05:00	TRUE	1	2342.19125	2342.19125	2090.14	2090.14
8	01-01-2017 06:00	TRUE	1	2342.19125	2342.19125	2242.9	2242.9
9	01-01-2017 07:00	TRUE	1	2342.19125	2342.19125	2330.15	2330.15
10	01-01-2017 08:00	TRUE	1	2342.19125	2342.19125	2285.23	2285.23
11	01-01-2017 09:00	TRUE	1	2342.19125	2342.19125	2209.72	2209.72
12	01-01-2017 10:00	TRUE	1	2342.19125	2342.19125	2181.71	2181.71
13	01-01-2017 11:00	TRUE	1	2342.19125	2342.19125	2158.5	2158.5
14	01-01-2017 12:00	TRUE	1	2342.19125	2342.19125	2171.38	2171.38
15	01-01-2017 13:00	TRUE	1	2342.19125	2342.19125	2142.67	2142.67
16	01-01-2017 14:00	TRUE	1	2342.19125	2342.19125	2165.62	2165.62
17	01-01-2017 15:00	TRUE	1	2342.19125	2342.19125	2177.44	2177.44
18	01-01-2017 16:00	TRUE	1	2342.19125	2342.19125	2215.9	2215.9
19	01-01-2017 17:00	TRUE	1	2342.19125	2342.19125	2244.89	2244.89
20	01-01-2017 18:00	TRUE	1	2342.19125	2342.19125	2358.22	2358.22
21	01-01-2017 19:00	TRUE	1	2342.19125	2342.19125	3028.78	3028.78
22	01-01-2017 20:00	TRUE	1	2342.19125	2342.19125	2987.56	2987.56
23	01-01-2017 21:00	TRUE	1	2342.19125	2342.19125	2970.44	2970.44
24	01-01-2017 22:00	TRUE	1	2342.19125	2342.19125	2827.9	2827.9
25	01-01-2017 23:00	TRUE	1	2342.19125	2342.19125	2465.82	2465.82

Figure 6.1. Dataset

- The platform used in this project for the model implementation is Jupyter Notebook platform in Anaconda IDE. A simple representation of prediction process is shown below:

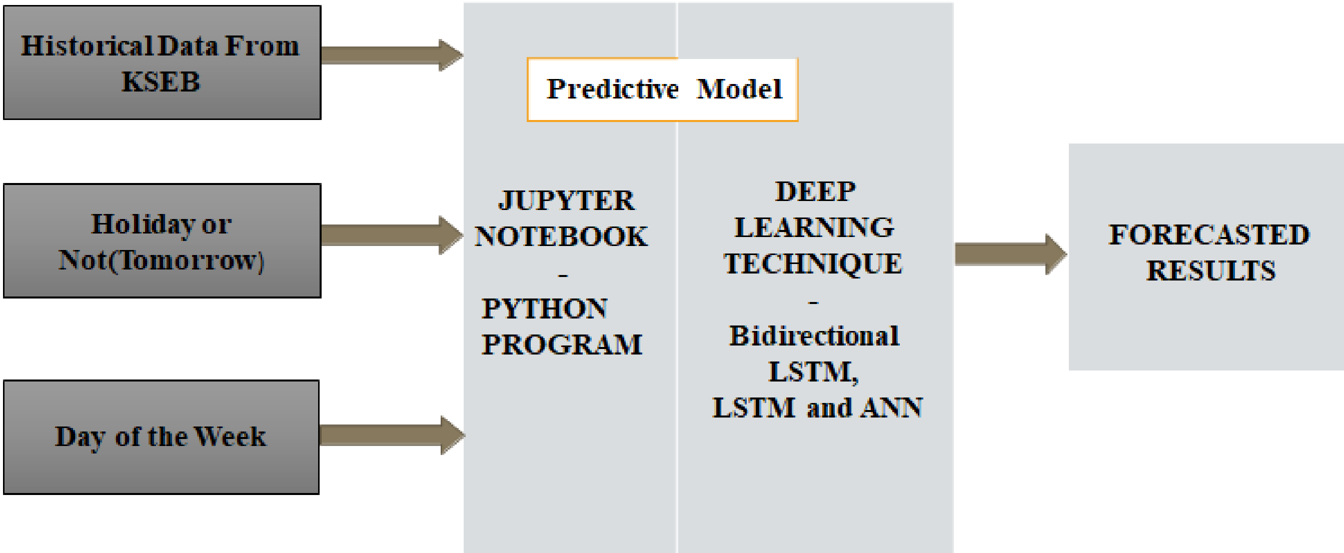

Figure 6.2. Representation of predictive model

6.3 Flowchart of ANN model

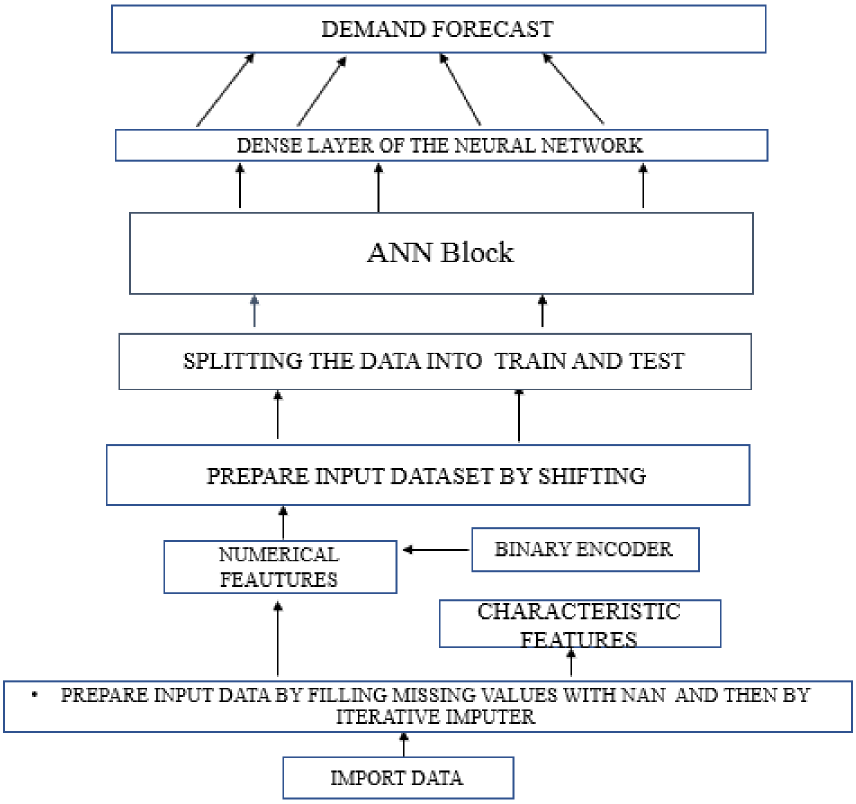

Figure 6.3. Flow chart of ANN model

6.4 Algorithm of ANN model

Step 1: Import data

Step 2: Filling the missing values using the iterative imputer function.

Step 3: Feature scaling. Encoding characteristic parameters with binary encoding.

Step 4: Converting the dataset so that supervised learning can be incorporated.

Step 5: Splitting up the dataset into inputs and targets.

Step 6: Splitting the data into test and train sets.

Step 7: Feature scaling

Step 8: Artificial Neural Network is build

Step 9: Training and testing the model, Error calculation, demand prediction.

6.5 Flowchart of LSTM model

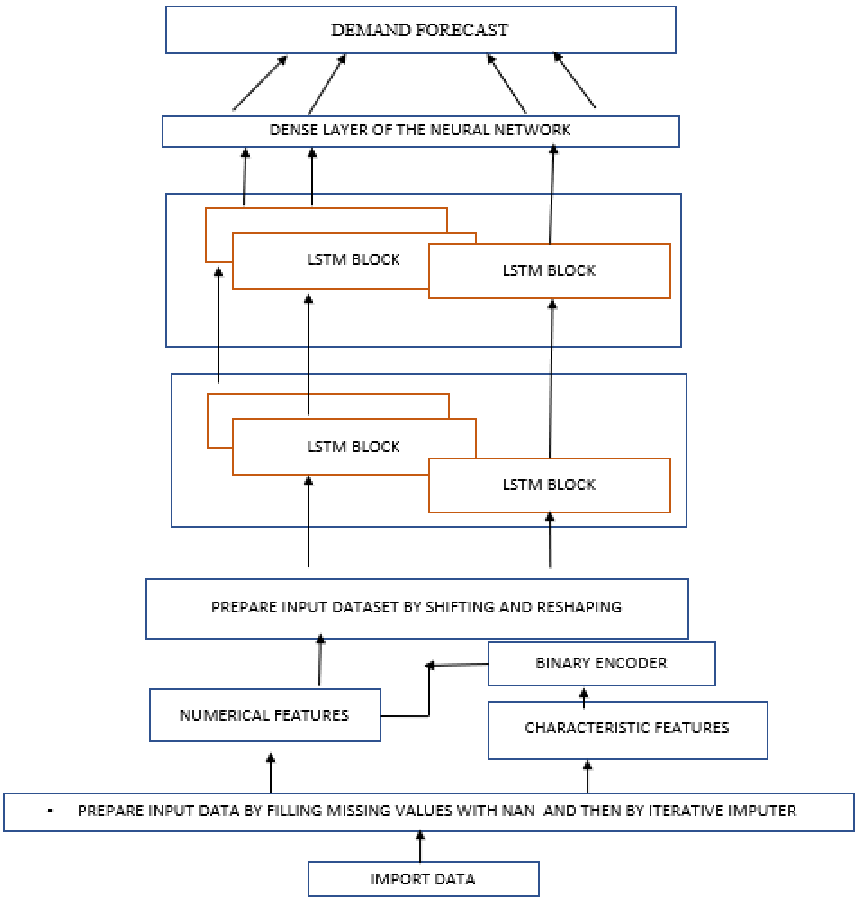

Figure 6.4. Flow chart of LSTM model

6.6 Algorithm of LSTM model

Step 1: Import data

Step 2: Filling the missing values using the iterative imputer function.

Step 3: Feature scaling. Encoding characteristic parameters with binary encoding.

Step 4: Converting the dataset so that supervised learning can be incorporated.

Step 5: Splitting up the dataset into inputs and targets.

Step 6: Splitting the data into test and train sets.

Step 7: Feature scaling

Step 8: Reshaping 2D dataset to 3D dataset.

Step 9: LSTM, Sequence -sequence prediction model is build.

Step 10:Training and testing the model, Error calculation, demand prediction.

6.7 Flowchart of BLSTM model

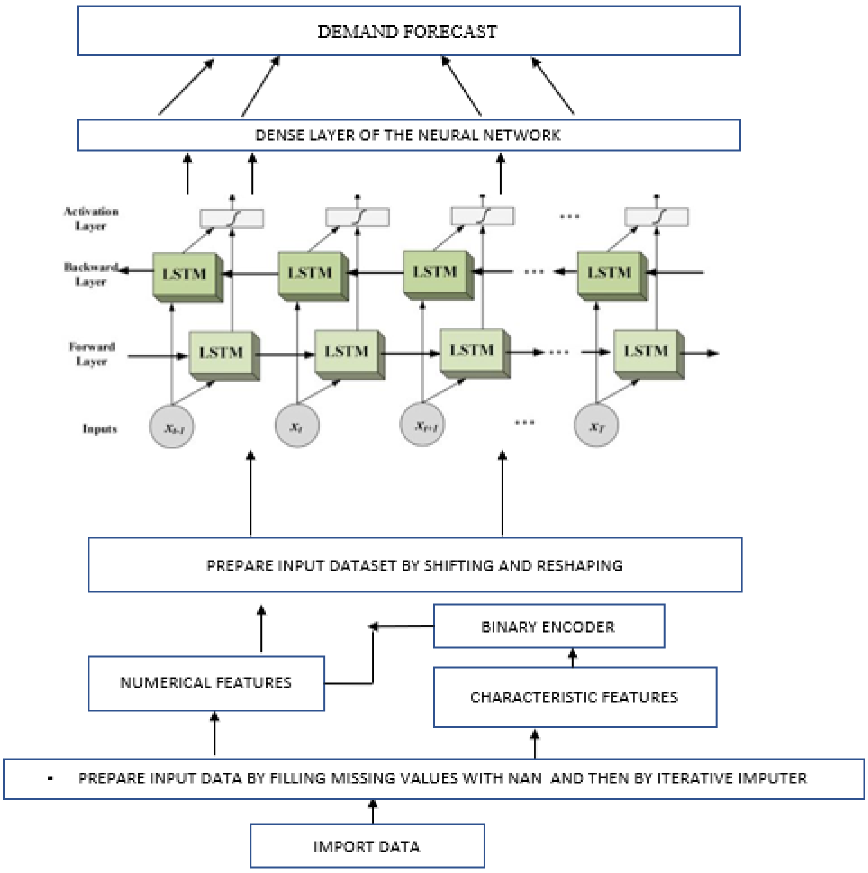

Figure 6.5. Flow chart of BLSTM model

6.8 Algorithm of BLSTM model

Step 1: Import data

Step 2: Filling the missing values using the iterative imputer function.

Step 3: Feature scaling. Encoding characteristic parameters with binary encoding.

Step 4: Converting the dataset so that supervised learning can be incorporated.

Step 5: Splitting up the dataset into inputs and targets.

Step 6: Splitting the data into test and train sets.

Step 7: Feature scaling

Step 8: Reshaping 2D dataset to 3D dataset.

Step 9: BLSTM, Sequence -sequence prediction model is build.

Step 10:Training and testing the model, Error calculation, demand prediction.

In Jupyter notebook platform certain libraries like pandas, tensorflow keras, matplotlib, sklearn, numpy, etc. were used. After importing all these libraries, the dataset was imported.This project uses BidirectionalLSTM model and is compared with other deep learning models like LSTM and ANN.

After importing necessary libraries, the input dataset was prepared by pre-processing. The dataset in csv format is loaded,which contains missing values and were initially replaced with NaN and store it in a csv file.Iterative imputer is used to replace NaN with imputer predicted values.The variation of parameters by graphical library was plotted, this helps to study the variations in the data across dates.The parameters contain numeric as well as characteristic parameters.The characteristic parameters were encoded using encoders. In the dataset only one type of characteristic parameter Kerala Holiday/not is present. For Kerala Holiday/not parameter, the project uses a Binary Encoder to develop a Boolean system which returns True if it is a Holiday and False if it is not. The dataset is converted in to supervised learning type so that the model can learn it easily.In this project we are training the model such that when todays historical parameter is given,the model will predict tommorows demand.Then the input dataset is split into testing and training dataset.This test-train ratio is one of the key factors deciding the accuracy of the model. Normally X is used to denote input and Y for output.

The dataset is then divided into four i.e. input training dataset (Xtrain), input testing dataset (Xtest), output train dataset (Ytrain), output test dataset (Ytest). Reshape function is used to reshape the data from 2D to 3D. This is done because

LSTM and BLSTM only takes 3D inputs, for example if the test data contains (30643,216) datasets it is converted to (30643,24,9). In this way all the datasets are reshaped.The ANN model doesnt need reshaping.The model will accept 1D dataset. The preparation of input data is done here, only after completing these steps the dataset to train the model is given.

The prediction model of BLSTM mainly consist of input layer, hidden layer, fully connected layer and the output layer. The input layer receives and analyses the neural network's input before passing it on to the hidden layer. All computations on the features received through the input layer are performed by the hidden layer, it consist of two bidirectional layers, first layer is the combo of forward propagating and backward propagating LSTM and the second layer is the inverse copy of first layer.The input shape is 3D tensor (number of samples, timesteps, input dimension) to satisfy the characteristics of the hidden layer input. In the experiment, the number of samples is set to zero, the time step is set to 24. The time step is chosen differently based on the type of load. The input feature determines the input dimension . The output layer provides the predicted result.The differnce in LSTM model is that the hidden layer only propogate in forward direction.The hidden layer of BLSTM and LSTM model contains 100 neurons each.The ANN model contains hidden layer having 10,20,5 neurons.

To introduce non-linear components, the rectified linear unit (ReLU) layer is employed as the activation function, which improves the model's expression ability. It does not involve any complicated arithmetic and doesn't have the vanishing gradient problem as suffered by other activation functions like sigmoid or tanh. The proposed model therefore takes less time to converge.The Adam optimisation approach with adaptive learning rate is used in network training. This optimizer can handle sparse matrices and non-stationary goals, as well as calculate multiple adaptive learning rates for different parameters, making it suited for high-dimensional data .The forecasting accuracy can be obtained by using proper evaluation. metrics.The training is done on an intel core i5 laptop clocked at 2.5-3.1 GHz with 8 Gb RAM and the training time took was 11-14 seconds for an epoch. After training the dataset the desired output by using the keyword 'model.predict()' can be predicted.

Chapter 7

RESULTS

7.1 Parameter Plotting

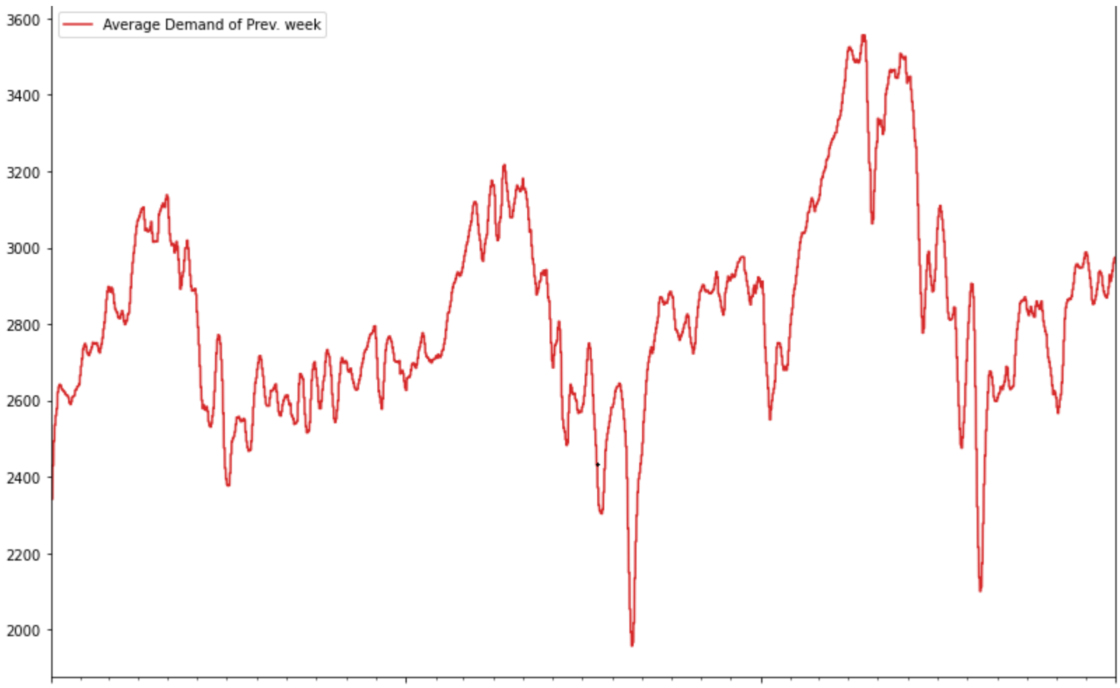

Figure 7.1. Average of Previous Week

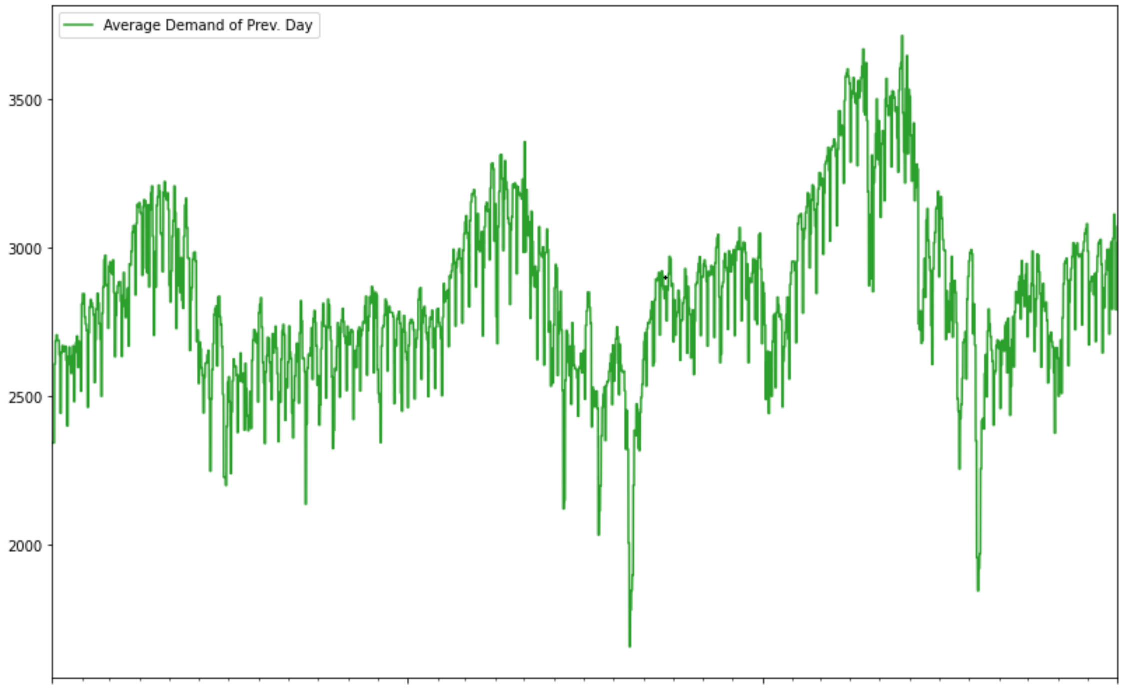

Figure 7.2. Average Demand of Previous Day

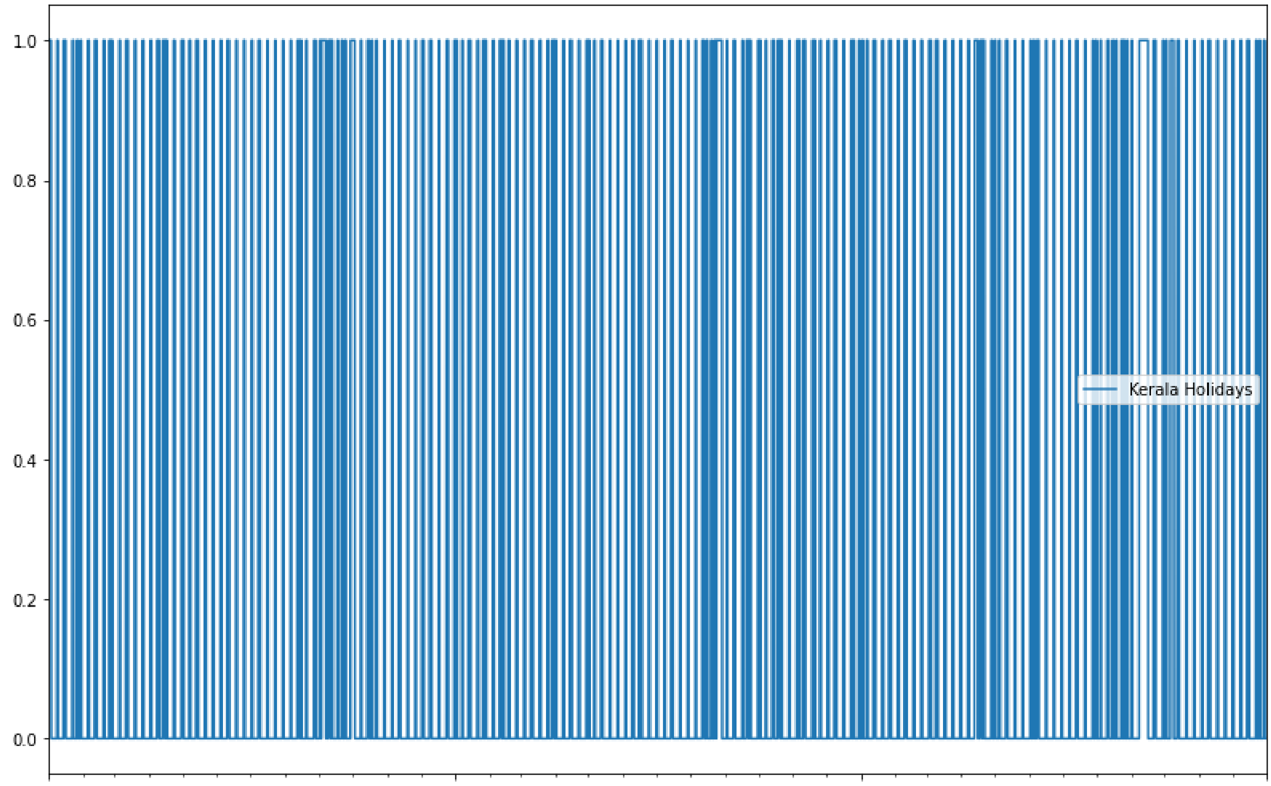

Figure 7.3. Holiday

Figure 7.4. Day of the week

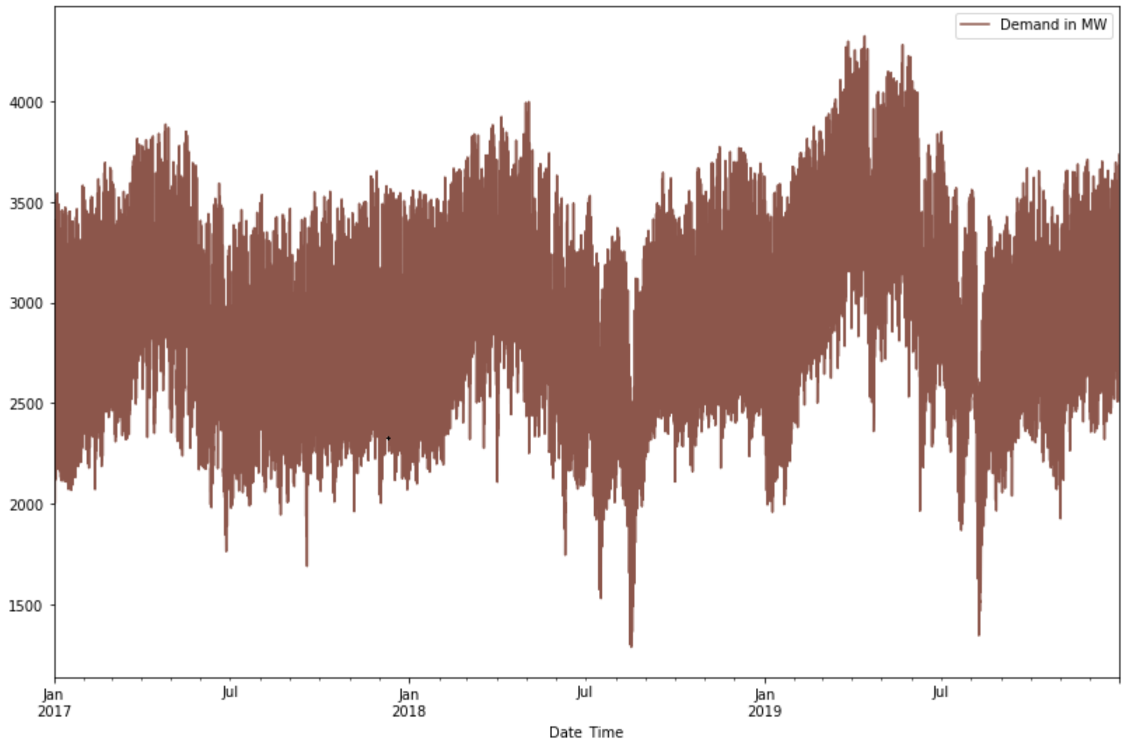

Figure 7.5. Demand in MW

7.2 Hyperparameter Tuning

Hyperparameter tuning is an important part of controlling the performance of a machine learning model.Our predicted model parameters will not minimize the loss function,if hyperparameters are not correctly tuned.which lead to inaccurate results.In this following deeplearning models, the hyper parameters are Batch size,Train-Test ratio,epoch and analysed the loss function values.

- **Batch size** - It determines the amount of samples that will be transmitted via the network.
- **Train-test ratio** -The train-test ratio is the proportion by which the total dataset is divided into training and testing data.
- **Epoch/Iteration**-the number of times that the learning algorithm will work through the entire training dataset.

  The Following figure shows the change in errors withrespect to different hyperparameters.For the analysis, batchsize choosen are 125,250 and 500.Train to test ratio is taken as 70:30,80:20 and 90:10.The selected iteration values are 130,200 and 300.

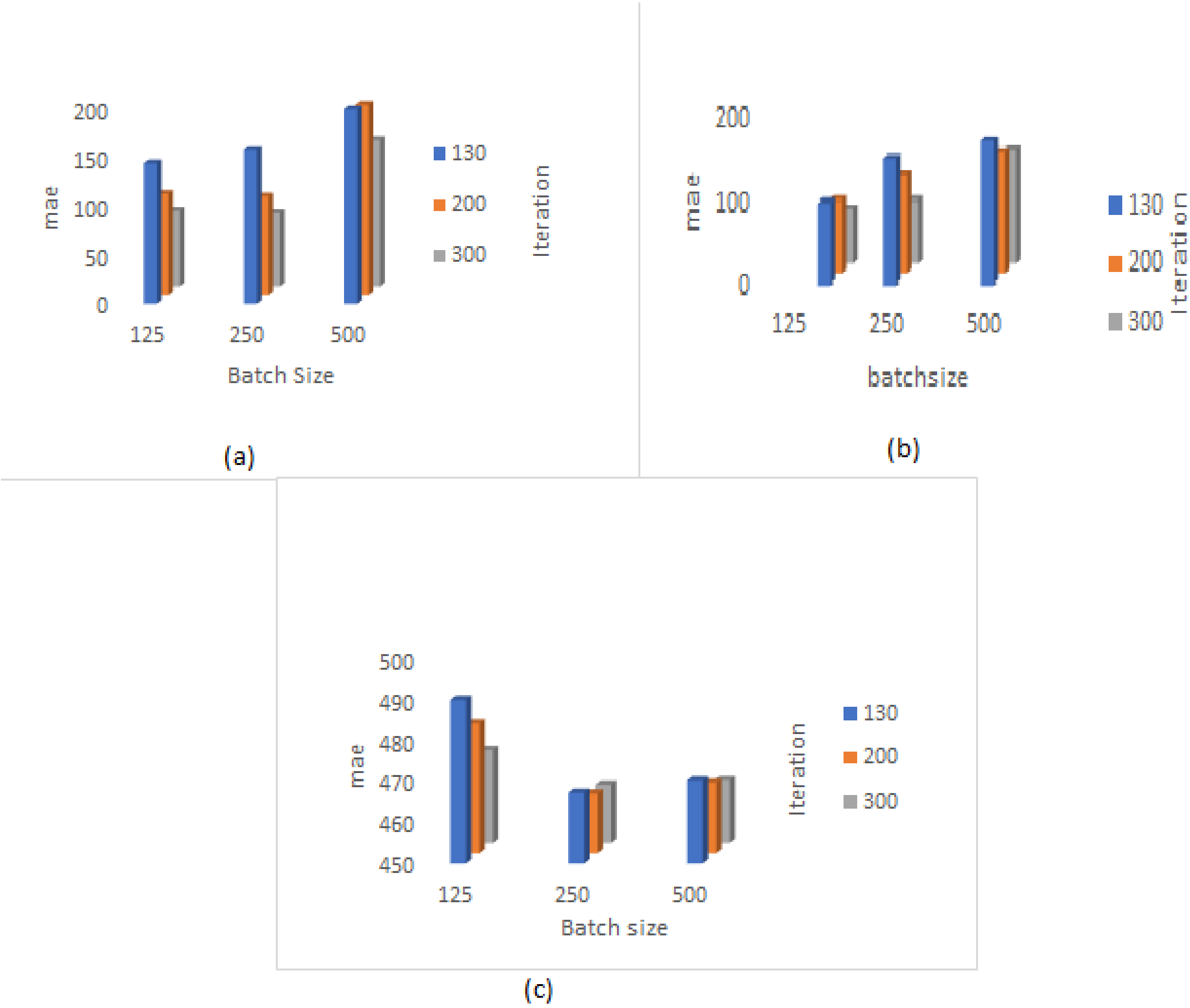

Figure 7.6. mae of BLSTM,LSTM and ANN respectively for the train:test ratio(70:30)

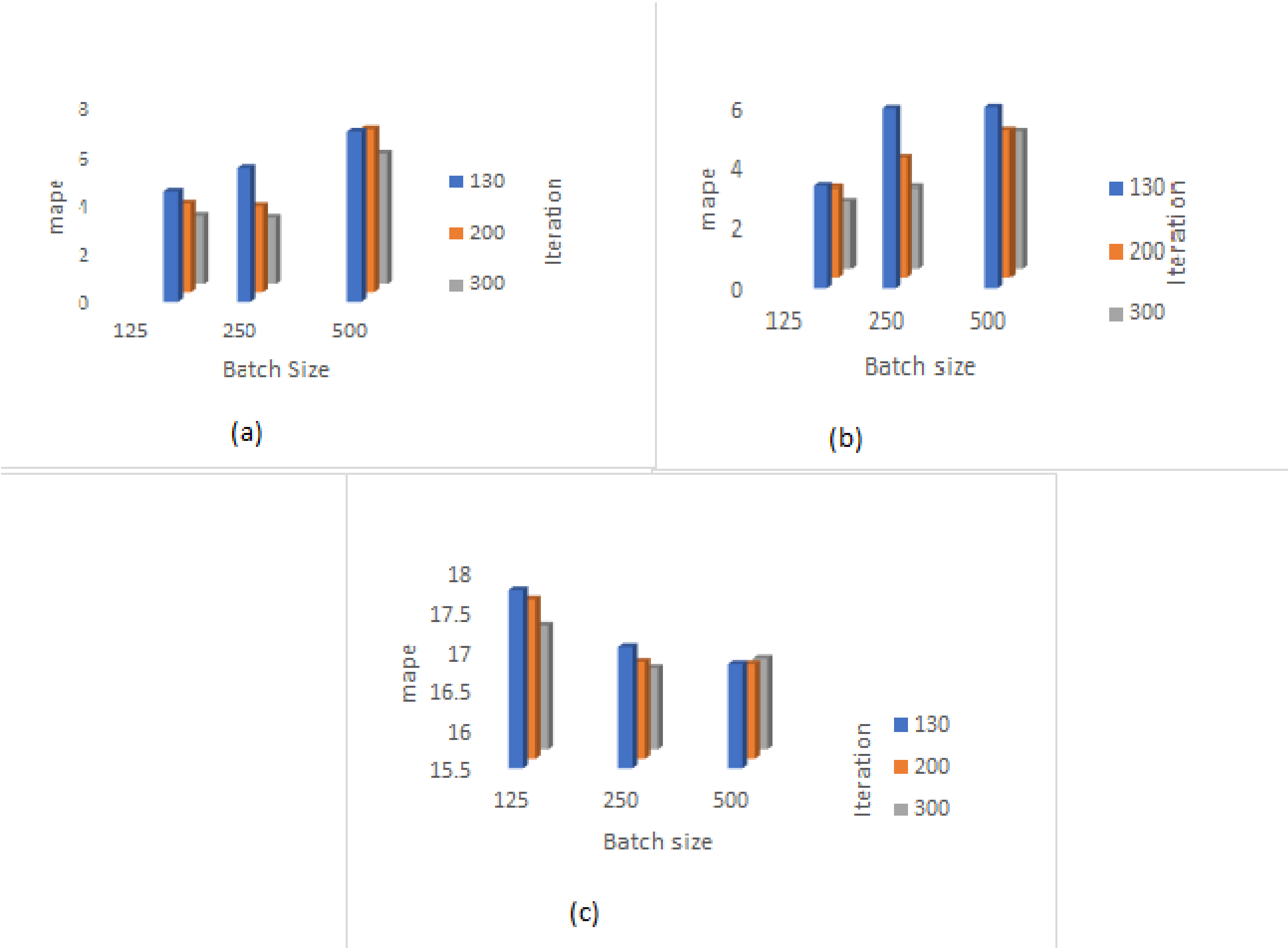

Figure 7.7. mape of BLSTM,LSTM and ANN respectively for the train:test ratio(70:30

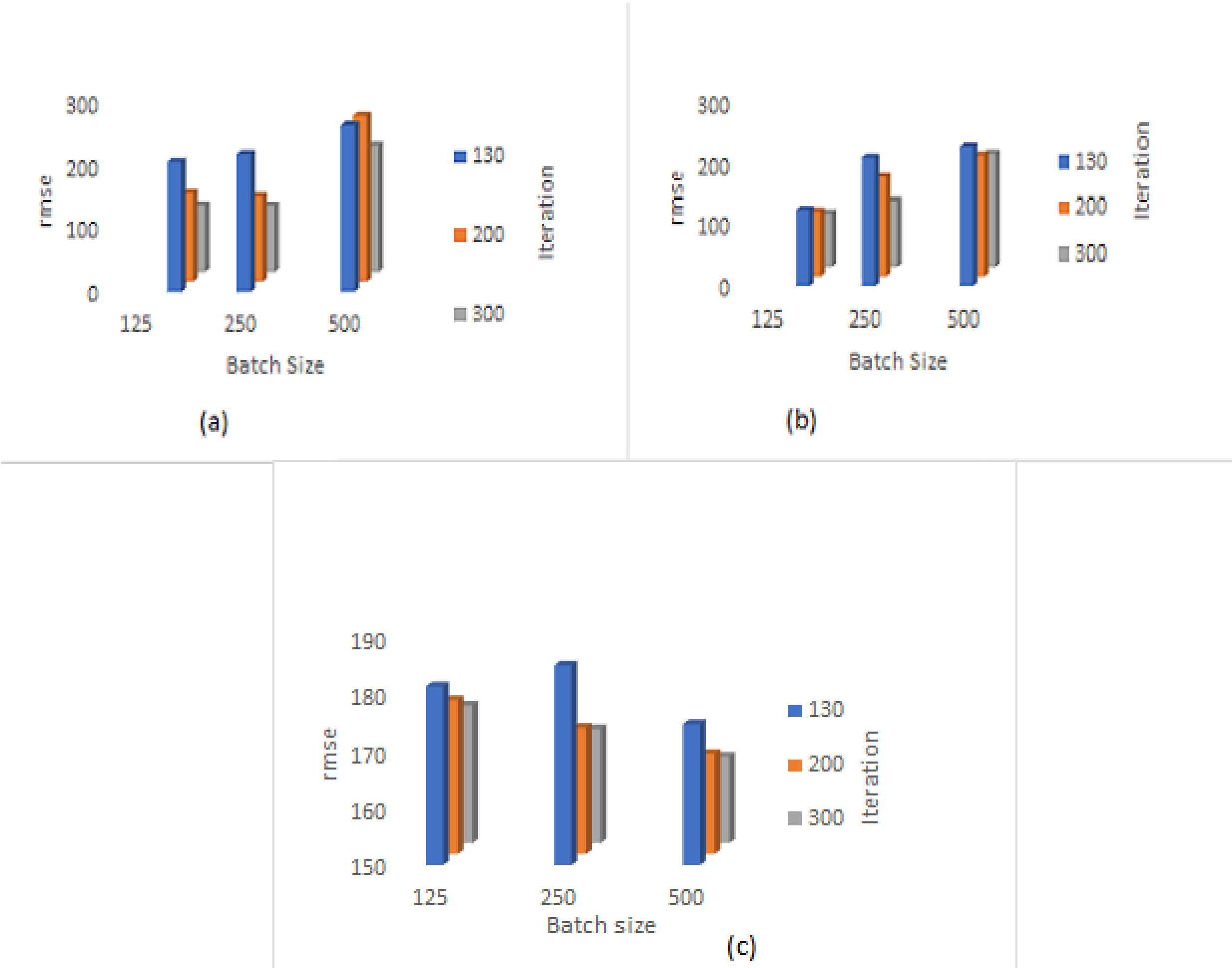

Figure 7.8. rmse of BLSTM,LSTM and ANN respectively for the train:test ratio(70:30

The above figure depicts the mean absolute error value,mean absolute percentage error value and the root mean squared error value of BLSTM,LSTM and ANN respectively for the train to test ratio 70:30.In the case of BLSTM the less error values are achieved for the batchsize 125 and for the epoch value 300.For LSTM the less mae is attained for batchsize 125 and epoch value 300.In case of ANN the error values are found less for the batchsize 250 and epoch value 300.

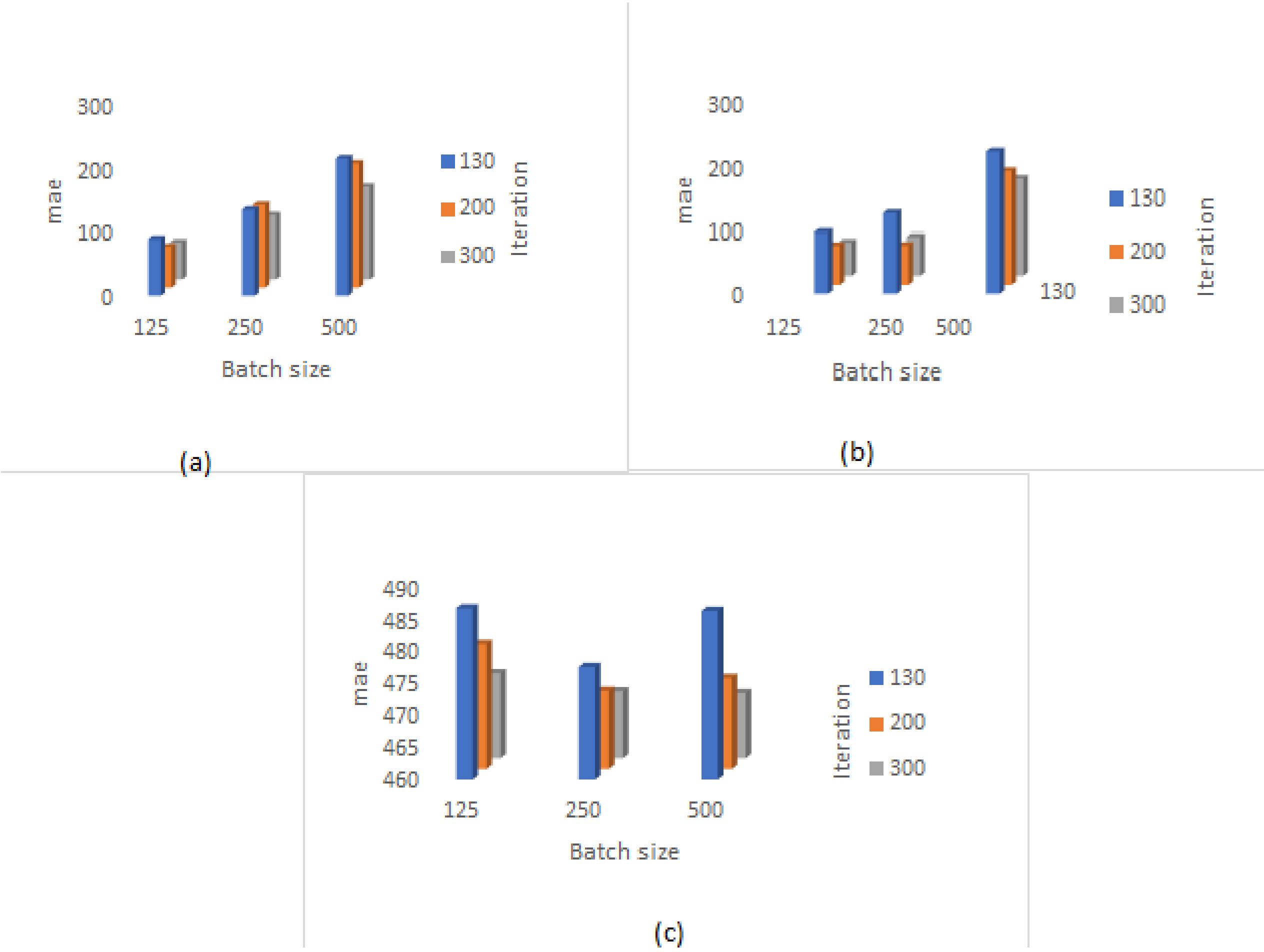

Figure 7.9. mae of BLSTM,LSTM and ANN respectively for the train:test ratio(80:20)

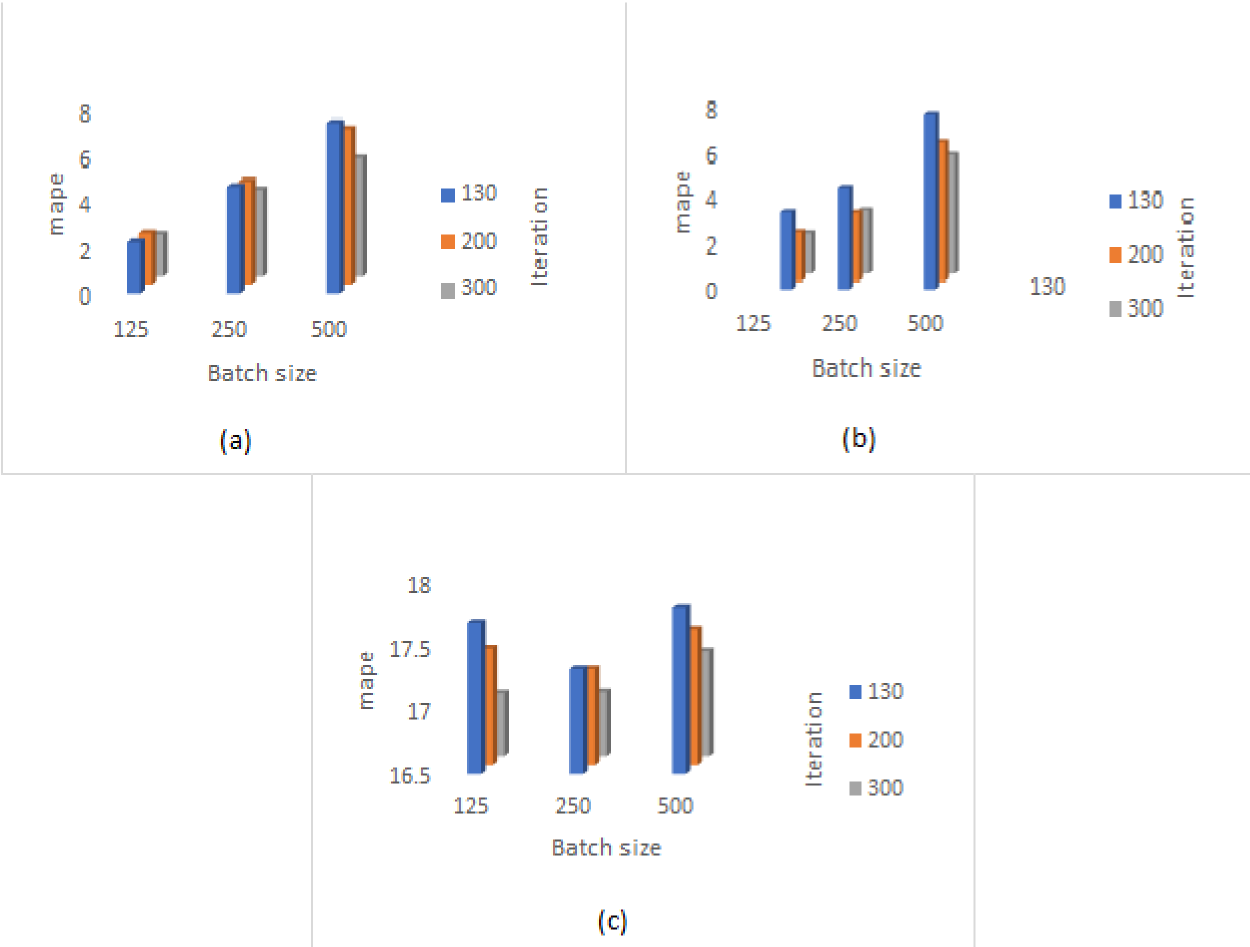

Figure 7.10. mape of BLSTM,LSTM and ANN respectively for the train:test ratio(80:20)

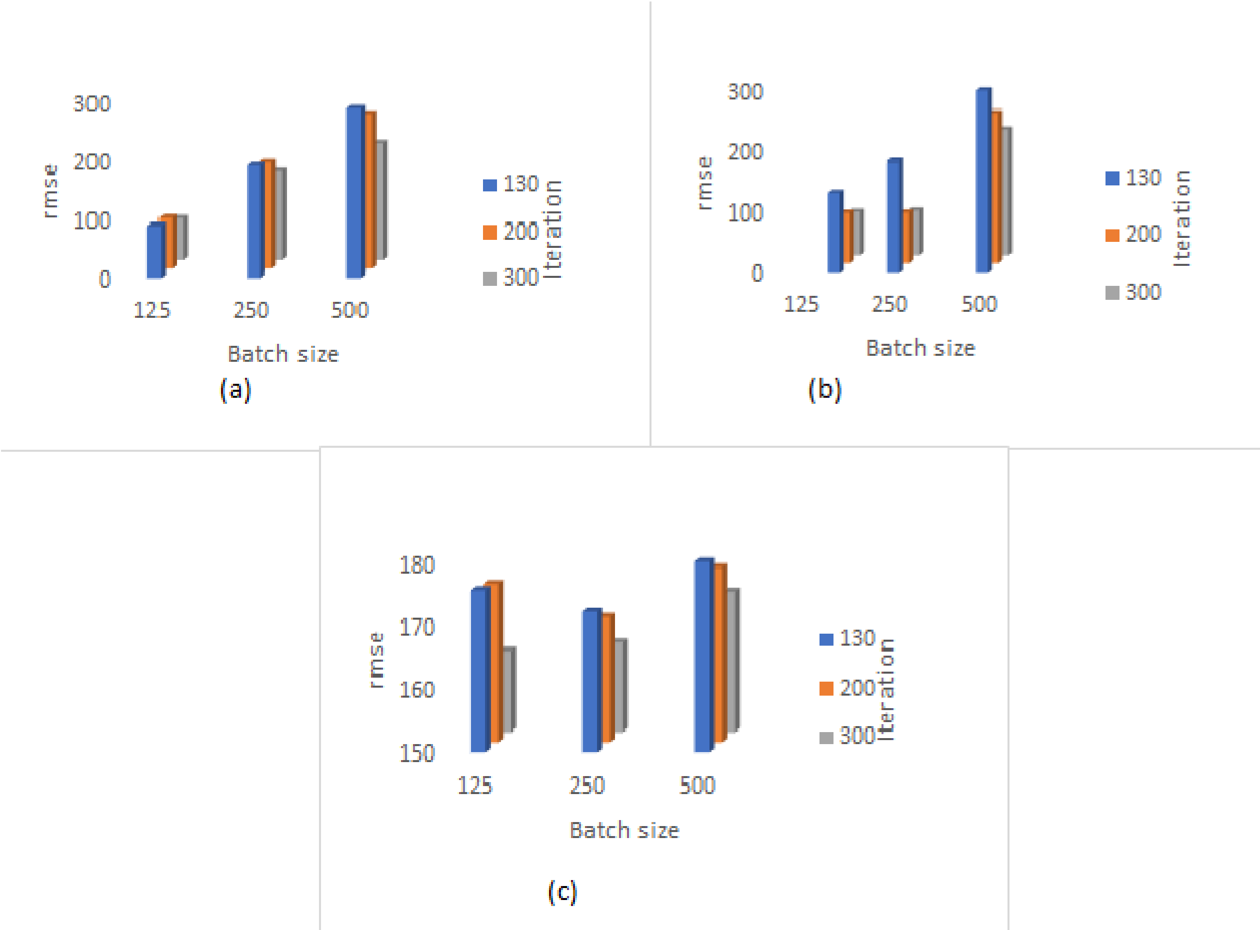

Figure 7.11. rmse of BLSTM,LSTM and ANN respectively for the train:test ratio(80:20)

The above figure shows the mae,mape and rmse for BLSTM,LSTM and ANN model for the train to test ratio 0.8. BLSTM,LSTM and ANN shows better evaluation metrics value than 70:30 train:test ratio.All of three models have less error value for the batchsize 125 with epoch value 300.

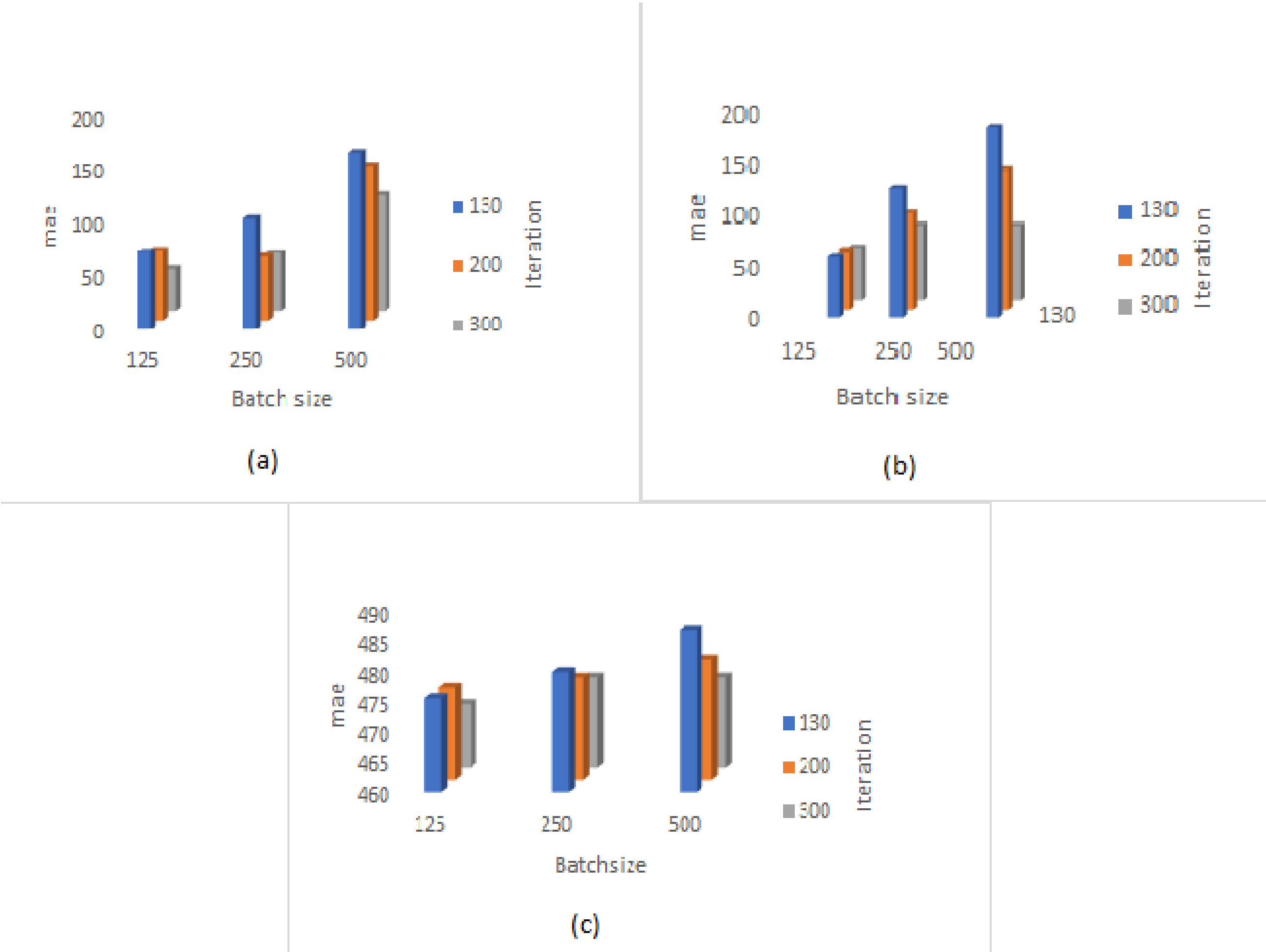

Figure 7.12. mae of BLSTM,LSTM and ANN respectively for the train:test ratio(90:10)

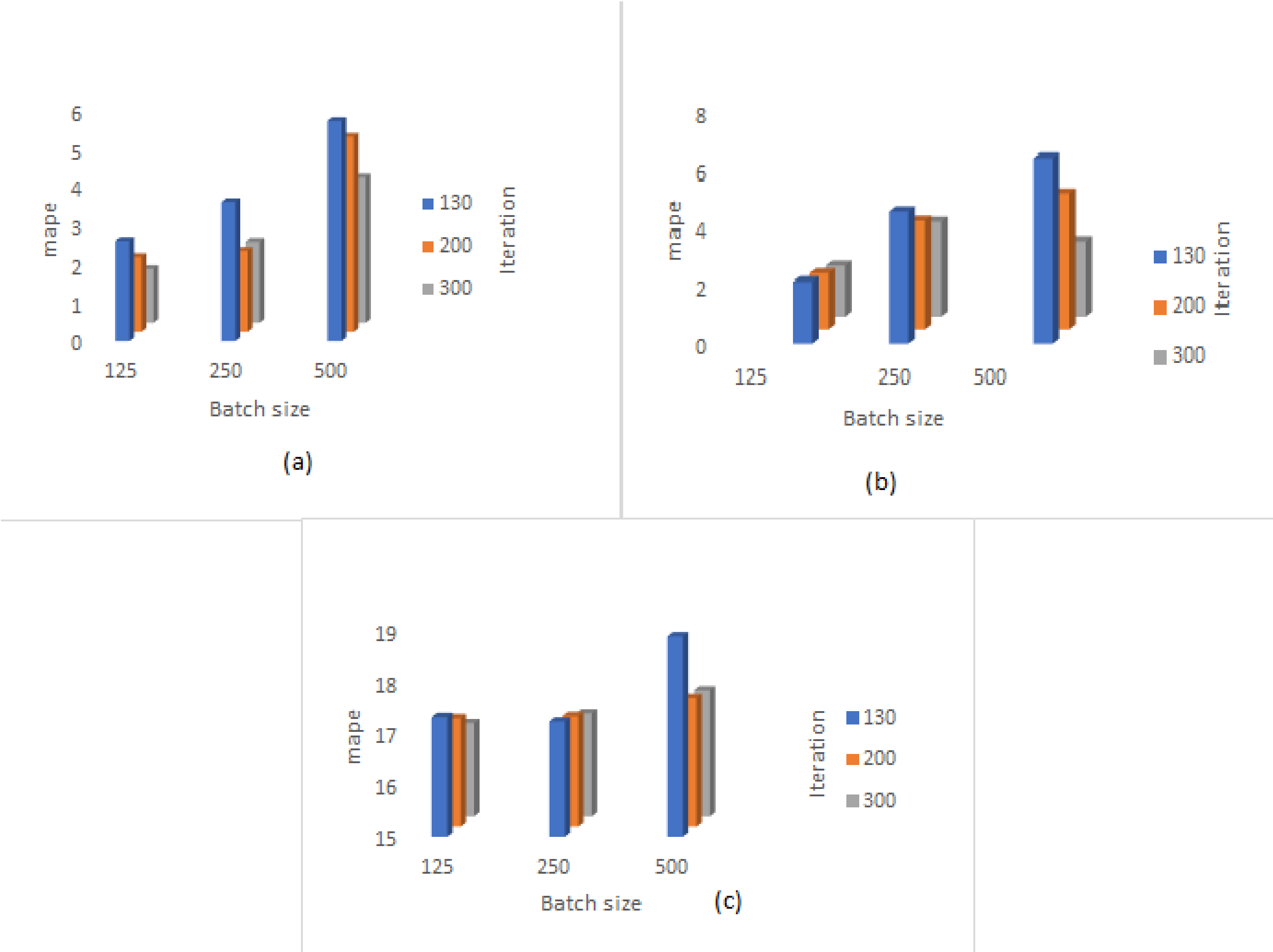

Figure 7.13. mape of BLSTM,LSTM and ANN respectively for the train:test ratio(90:10)

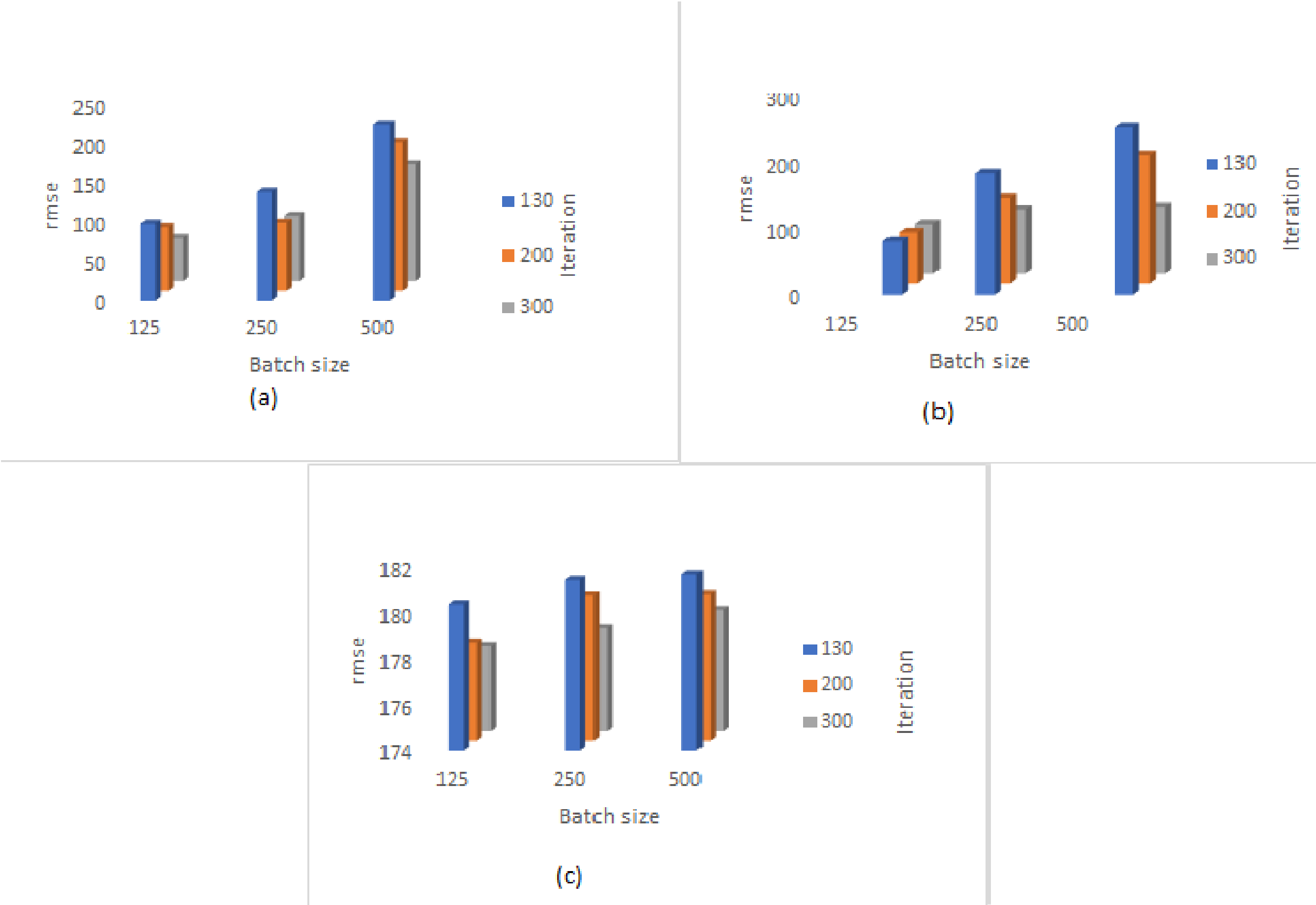

Figure 7.14. rmse of BLSTM,LSTM and ANN respectively for the train:test ratio(90:10)

The above figure shows the mae,mape and rmse value for the BLSTM,LSTM and ANN respectively for the train to test ratio 90:10.BLSTM model have the less error value when compared to 70:30 and 80:20 train:test ratio.For LSTM model error is higher in case of 90:10 train to test ratio.ANN also has less error in case of 80:10.

It can be see that the loss function value will vary according to the epoch,Batch size and the train-test value.From the analysis of this hyperparameter tuning less batchsize value and higher number of epoch value helps to reduce the error value.

For Bidirectional LSTM,less error is found when train to test ratio is 90:10,125 batch-size and epoch value equal to 300.The train:test ratio,batch size and epoch value for

LSTM is found to be 80:20,125 and 300 respectively.For ANN model train:test ratio is 80:20,batch size equal 125 and epoch value equal to 300.

7.3 Learning Curves

Learning curves are plot to show the model performance on the training and testing datasets and also to diagonse an underfit,overfit and goodfit model.

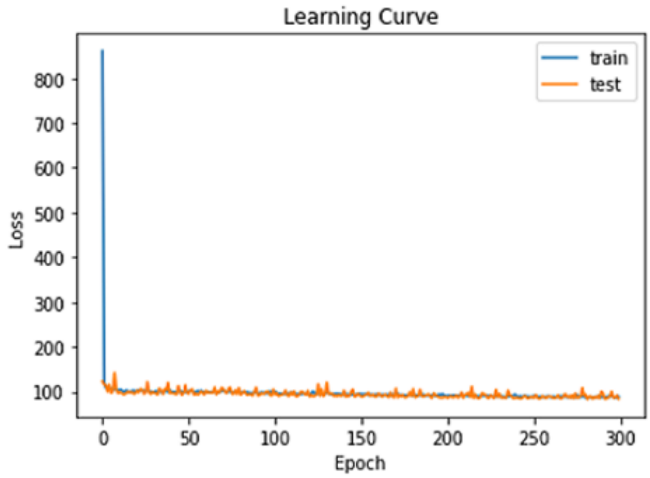

Figure 7.15. Learning curve plot showing training and testing loss with no. of iterations for ANN

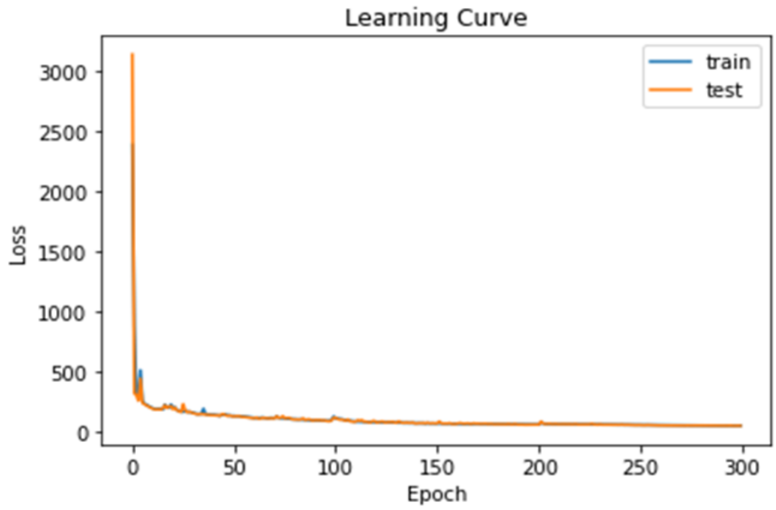

Figure 7.16. Learning curve plot showing training and testing loss with no. of iterations for LSTM

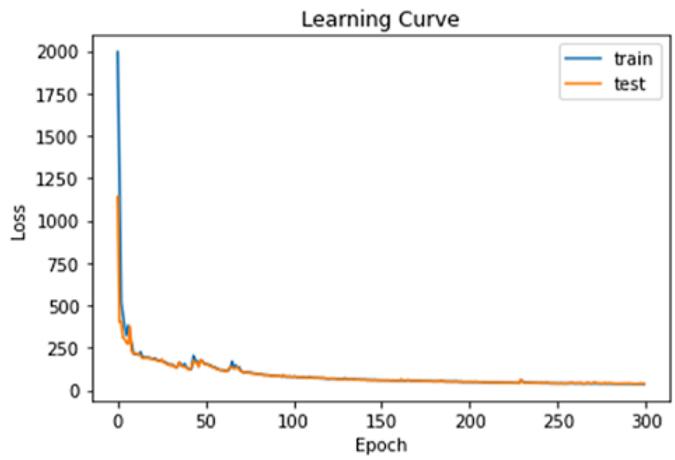

Figure 7.17. Learning curve plot showing training and testing loss with no. of iterations for BLSTM

7.4 Predicted Demand V/s Actual Demand

The figure below shows the predicted vs actual values for Bi-LSTM,LSTM and ANN respectively.The demand (MW) is shown by the vertical axis, while the time(hour)is represented by the horizontal axes over a week.The performance of the corresponding models is represented by the difference between the actual and predicted curves.Among the three techniques the Bi-LSTM has the less difference between the actual and predicted value.The table(7.1) below shows the forecasting error for different techniques

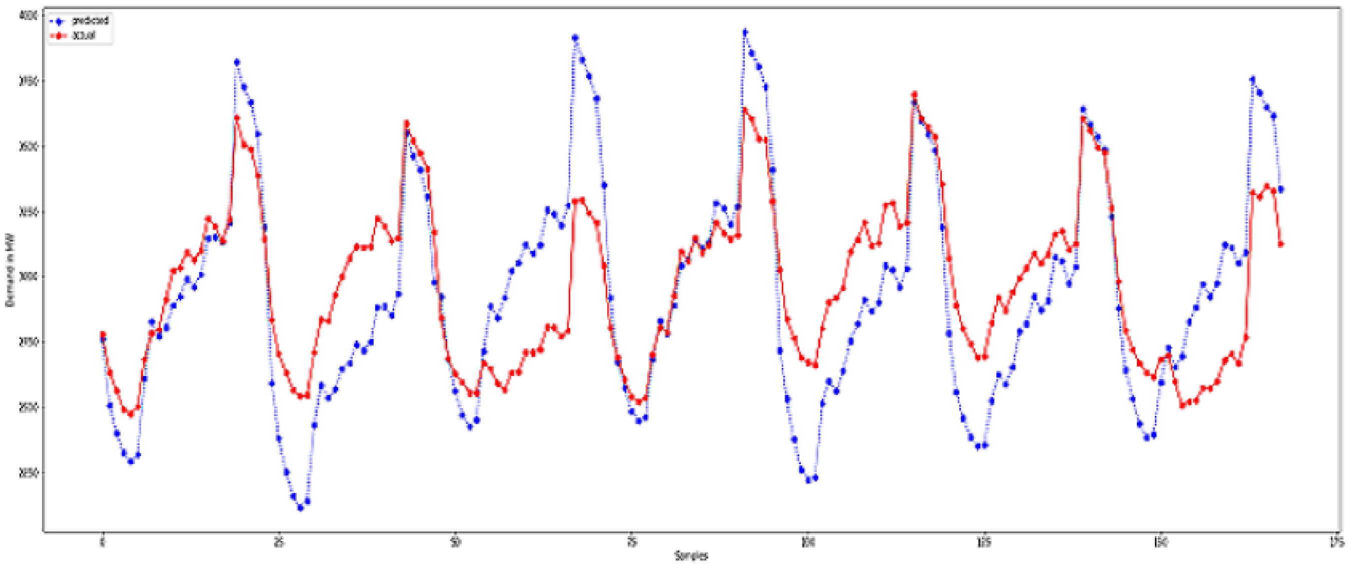

Figure 7.18. Actual vs predicted for ANN.

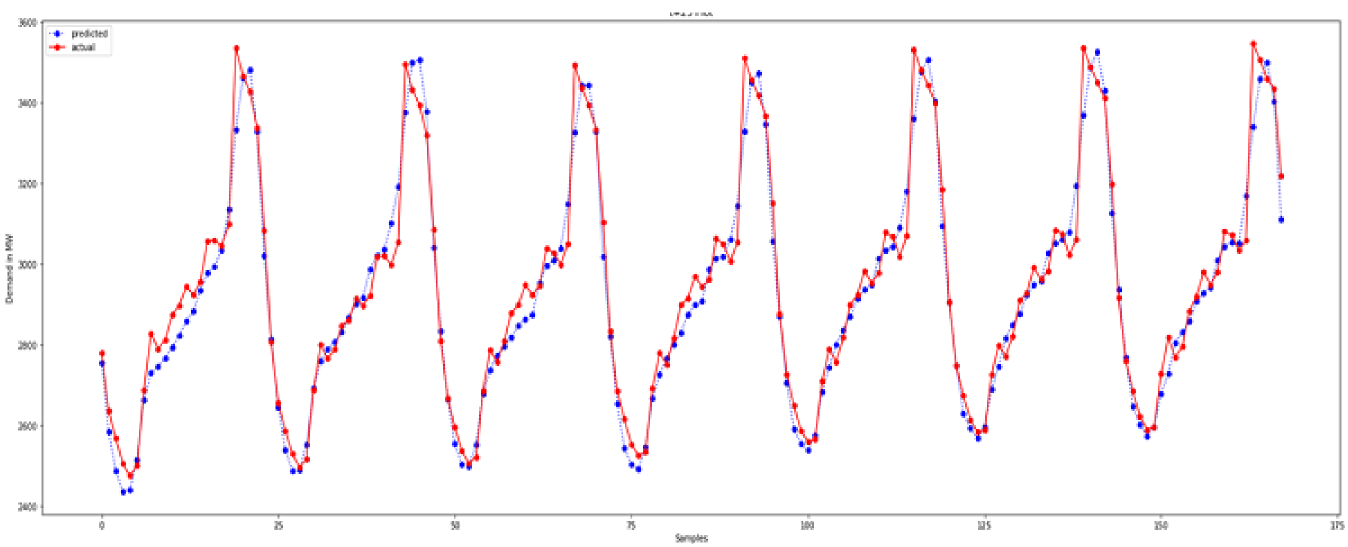

Figure 7.19. Actual vs predicted for LSTM.

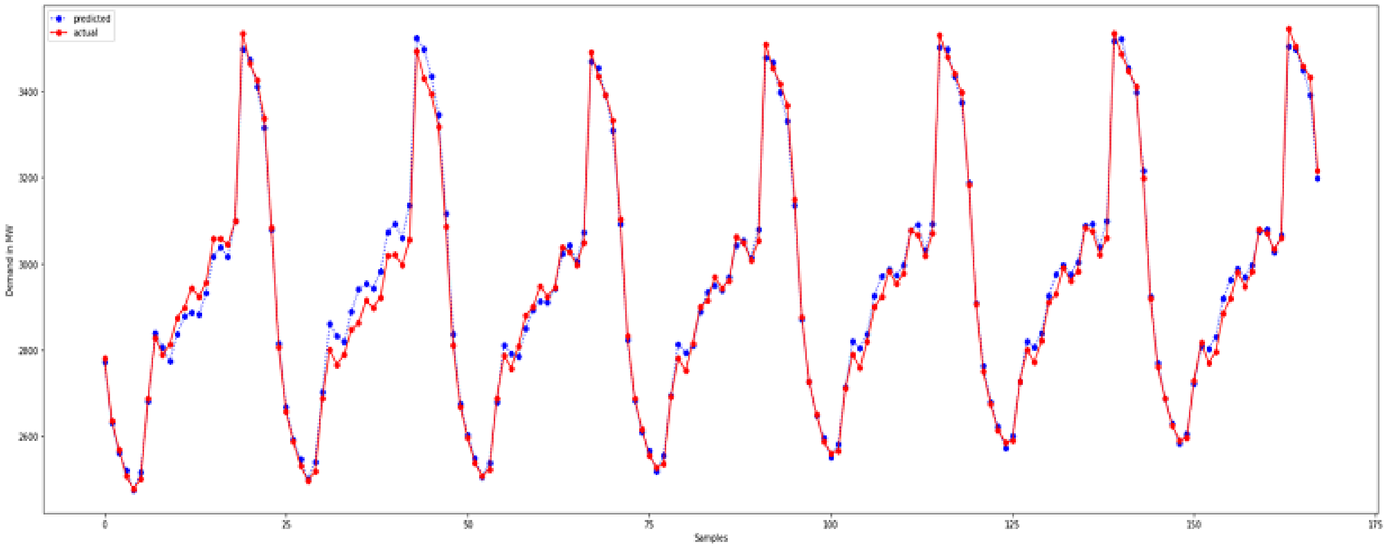

Figure 7.20. Actual vs predicted for BLSTM.

7.5 Evaluation Metrics

Table 7.1. Evaluation metrics value of BLSTM,LSTM and ANN

Name	mae(%)	mape(%)	rmse(%)	R-squared score
Bi-LSTM	0.9314	1.412	1.32	0.9799
LSTM	1.188	1.772	1.672	0.967
ANN	11.105	16.931	3.829	0.8431

The above result shows that Bidirection LSTM shows more accurate performance.It has the less mean absolute error,mean absolute percentage error and root mean squared error value and has greater r squared value.

Chapter 8

CONCLUSION

This project aims to develop a reliable and accurate model which can predict the load demand for the next 24 hours.To perform the day-ahead load forecasting, a forecasting system based on deep neural networks with Bidirectional LSTM technique is proposed.The results demonstrate BLSTM's ability to execute sequential data models and extract information from feature sequences by dealing with both forward and backward dependencies.The proposed BiLSTM is compared with the unidirectional LSTM and Artificial Neural network,it can be seen that BLSTM outperforms the other deep learning models with less error values.

REFERENCES

[1] **Warrior, Karun P., M. Shrenik, and Nimish Soni.**, "Short-term electrical load forecasting using predictive machine learning models", *IEEE Annual India Conference (INDICON)*, pp. 1-6,2016.

[2] **Singh, Saurabh, Shoeb Hussain, and Mohammad Abid Bazaz.**, "Short term load forecasting using artificial neural network.", In 2017*Fourth International Conference on Image Information Processing* (ICIIP), pp. 1-5,2017.

[3] **Kwon, Bo-Sung, Rae-Jun Park, and Kyung-Bin Song.**, "Short-term load forecasting based on deep neural networks using LSTM layer.", *Journal of Electrical Engineering Technology* "15 , no. 4: 1501-1509,2020.

[4] **Wu, Lizhen, Chun Kong, Xiaohong Hao, and Wei Chen.**, "A short-term load forecasting method based on GRU-CNN hybrid neural network model.", *Mathematical Problems in Engineering*,2020.

[5] **Zou, Mingzhe, Duo Fang, Gareth Harrison, and Sasa Djokic.**, "Weather based day-ahead and week-ahead load forecasting using deep recurrent neural network.", In 2019 *IEEE 5th International forum on Research and Technology for Society and Industry* (RTSI), pp. 341-34,6,2019.

[6] **Tang, Xianlun, Yuyan Dai, Ting Wang, and Yingjie Chen.**, "Short-term power load forecasting based on multi-layer bidirectional recurrent neural network.", *IET Generation, Transmission Distribution 13*, no. 17: 3847-3854,2019.

[7] **Wang, Zhangliang, Li Jia, and Chang Ren.**, "Attention-Bidirectional LSTM Based Short Term Power Load Forecasting.", In 2021 *Power System and Green En-*

ergy Conference (PSGEC) , pp. 171-175,2021.

[8] **Mughees, Neelam, Syed Ali Mohsin, Abdullah Mughees, and Anam Mughees.**, "Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting.", *Expert Systems with Applications*, 175 : 114844,2021.

[9] **Wei, Li, and Zhang Zhen-gang.**, "Based on time sequence of ARIMA model in the application of short-term electricity load forecasting.", 2009 *International Conference on Research Challenges in Computer Science*, pp. 11-14,2009.

[10] **Ciechulski, Tomasz, and Stanisław Osowski.**(2021), " High Precision LSTM Model for Short-Time Load Forecasting in Power Systems.",*Energies 14*, no. 11 : 2983,2021.

[11] **Rafi, Shafiul Hasan, Shohana Rahman Deeba, and Eklas Hossain.**, " A short-term load forecasting method using integrated CNN and LSTM network.",*IEEE Access*, 9 : 32436-32448,2021.

[12] **Zang, Haixiang, Ruiqi Xu, Lilin Cheng, Tao Ding, Ling Liu, Zhinong Wei, and Guoqiang Sun.**(2021), " Residential load forecasting based on LSTM fusing self-attention mechanism with pooling.", *Energy*, 229 : 120682,2021.

[13]https://images.app.goo.gl/SU5iK4ckYHCrorbA8

[14]https://www.javatpoint.com/supervised-machine-learning

[15]https://becominghuman.ai/supervised-learning-vs-unsupervised-learning-8af1bc803210

[16]https://empow-her.com/bing/bfdbd37834125.html?p=6.28.4116093.1.18.43 [17]https://cdnsc
2019-0101

[18]https://datathings.com/blog/post/neuralnet/

[19]https://www.techtarget.com/searchenterpriseai/definition/recurrent-neural-networks

[20]https://commons.wikimedia.org/wiki/File:Long-Short-Term-Memory.svg

[21]https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/