Prediction of Power Consumption of Electric Vehicles: A Deep Learning Approach

MSc Research Project
Data Analytics

Jatin Rajkumar Singh
Student ID: x20227965

School of Computing
National College of Ireland

Supervisor: Dr.Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Jatin Rajkumar Singh
Student ID:	x20227965
Programme:	Data Analytics
Year:	2022-23
Module:	MSc Research Project
Supervisor:	Dr.Christian Horn
Submission Due Date:	15/12/2022
Project Title:	Prediction of Power Consumption of Electric Vehicles: A Deep Learning Approach
Word Count:	6704
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Jatin Rajkumar Singh
Date:	31st January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Prediction of Power Consumption of Electric Vehicles: A Deep Learning Approach

Jatin Rajkumar Singh
x20227965

Abstract

Electric vehicles are seen as the next big thing in the automobile industry as conventional fossil-fueled vehicles have major drawbacks which are affecting the environment. Due to its relatively new technology, it is yet to touch every segment of the market and get user approvals. There are many hindrances to it, and one of the major reasons is its unpredictable energy requirements for longer trips as users are anxious about its energy requirement. Therefore this research has put forward a method to predict the power consumption of electric vehicles using real-time data. This method has used deep neural networks to build and train the model to estimate power consumption. The deep neural network achieved RMSE and R2 scores of 0.3153 and 0.8896 respectively and did predictions on real-time data and gave several insights into the data. Also, This study found out that there are some major factors which play an important role in power consumption. Factors such as vehicle speed, the distance of the trip, wind speed, road type and climate etc. are the major components in calculating and predicting consumption. This study can help users to reduce their anxiety about power consumption and can help them in planning their journey.

Keywords: Electric Vehicles, Power Consumption, Prediction, Deep Learning

1 Introduction

In recent years, there is an increase in greenhouse gases in the atmosphere which is leading to global warming all over the world. The earth's average temperature is increasing, leading to natural disasters. Hence automobile and transportation segment is looking for alternatives which can solve this major problem and benefit users and industries both. This industry is going through massive changes by trying out multiple options. Electric Vehicles(EV), Hydrogen Cars and Autonomous vehicles are the top options considered by major automobile players. These alternatives have huge potential and are expected to take the automobile industry ahead in the future. EVs are one of the most favourite and affordable options coming into the market and their demand is increasing over time Gong et al. (2020).

Since EVs are relatively new technology, it is undergoing major changes and have some obstacles. There are inadequate charging stations for EVs, higher battery costs and different ranges for different EVs are some of the challenges faced by this industry which is affecting the sales and adoption of EVs all over the world Sankaran et al. (2020). Since there is uncertainty if EVs can be used for long trips as there is a limited charging stations which creates doubt in the driver's minds. Drivers can be anxious that their EV battery will be enough for longer trips. This can be overcome by calculating the energy consumption of the vehicles for a specific distance for a particular trip at a certain time. The total energy consumption of EVs is dependent on multiple factors which can be external as well as internal. Factors such as the climate, temperature of vehicles, road types, the direction of winds, type of terrains, elevation and degree of the slopes are some of the external components which contribute to predicting energy consumption De Cauwer et al. (2017). Also, factors such as type of EV model, average speed, heater and air conditioning consumption and distance of trip are the internal element which plays a major part in determining the consumption Pokharel et al. (2021). All these factors are dynamic and change over time for different profiles which can be challenging while calculating energy consumption. Since EVs require different setups and infrastructures to charge their batteries, there is a shortage of charging stations all over the world. Due to this drivers are not confident about the power consumption of EVs for longer trips due to range anxiety of the EVs Pokharel et al. (2021).

This study tackles the following research question in order to put forward the method for predicting the power consumption of electric vehicles.

To what extent can deep neural networks improve the prediction of the total power consumption of an electric vehicle with real-world data?

The research question stated has helped in examining the effectiveness of neural networks in achieving the results for doing predictions for the power consumption of electric vehicles. The outcome of these results showed that the neural network technique can be a better choice for the predictions also it works well for real-time data. This study can help private as well as public entities to further build and enhance the existing infrastructure for the power charging stations and reduce the driver's anxiety about the long-distance journey. Also, it can be a better environment-friendly substitute for fossil fueled vehicles which can have a longer impact on the automobile industry. The further study covers the following roadmap: Section 2 discusses the related work; section 3 covers the Methodology; Section 4 describes Design Specifications and Implementations; Section 5 covers the Evaluation and Section 6 includes the Conclusion and the Future work.

2 Related Work

This segment of the research analyses the already existing work of the other authors in a similar area in this case field of Electric vehicles. It includes other authors' findings, critical analysis, and shortcomings of the research and draws some motivation to do the research further. This section of the study is divided into subsections to discuss certain points to understand and gain knowledge of existing works.

2.1 Machine Learning Methods for Making the Predictions of Total Power Consumption of Electric Vehicles

In Zhixin et al. (2021), the Naive Bayesian classification algorithm has been used to do prediction of electric vehicle power consumption. In this paper, features such as EV type, total travel distance, and journey time have given good accuracy in predicting EV power consumption. This paper has employed only a single machine-learning algorithm to train the model as other algorithms can also be useful in getting better results. Additionally, data used for training the model is generated using a simulation technique from EV from Nanjing 2021. In future work, this paper can also use more data from different vehicles and from different regions. Also, this paper can also use other factors such as profiles of roads, slopes and temperature etc for predicting power consumption. In Rhode et al. (2020) authors have put forward a method for estimating the total energy consumption of EVs which does not employ the data related to the vehicles. Instead, they have used machine learning algorithms on data produced by the sensors of the vehicles. Also, In order to deal with changing structures, Authors have applied certain kinds of kernel adaptive filtering procedures. This paper did not study the connection between terrains, speed and driver profiles. Produced neural networks, the outcome of a single linear adaptive filter, two state-of-the-art kernel adaptive filters and nine multiple vehicle data have been used to do the assessment of the suggested techniques. In this paper, the model proposed can be trained using a dataset which has more data from several vehicles of different brands. Also, authors can employ different deep-learning algorithms in order to predict the results.

In Liu et al. (2021), Under small and hidden data conditions, authors have put forward a machine-learning technique in order to estimate the energy consumption of EVs. Additionally, The authors have performed data sampling, data augmentation and ensemble methods using several different techniques and methods. Authors have also taken into consideration different drivers' behaviour which is produced by real-time drivers of EVs. As per Liu et al. (2021), Machine Learning Control Variables(MLCV) have better accuracy than traditional machine learning techniques. Additionally, a sample size of data can be enhanced by data augmentation methods which also increases the polarity of the dataset and gives additional training labels for model training. Deep learning algorithms can also help authors to get better results and evaluate the outcomes of multiple techniques.

In Basso et al. (2021), authors have studied electric vehicles time-related routing issues. It includes machine learning models which employ route variances, links of the roads and paths. As per the authors, if the number of vehicle rounds are increased over a period of time, it will increase the model's precision value which indicates that the model is learning from the data it is being trained on. Also, if the driver is driving the vehicle on a continuous path, it will save energy and will not require frequent charging over a period of time. In Singh and Vestlund (2021), authors have used navigational data on machine learning to calculate the energy consumption of electric vehicles. As per the authors, the results of the model may not change as factors such as vehicle speed and acceleration might be changing as drivers may be driving the EVs as per the traffic and road conditions. Also, this study has been carried out in cooler places, this study may not get the same results in places where the temperature is on the higher side. As per this study, there should not be a single absolute value of the drivers driving styles and there should be a different range of values as drivers may change their driving behaviour as per different conditions.

In Ullah et al. (2022), authors have used several machine learning algorithms for the predicting power consumption of EVs. As per the authors, XGBoost and LightGBM perform better and give higher R2 and lower values of RMSE and MAE. Authors have also found that there are other factors such as the heater, trip distance, and air conditioner consumption which play a major role in calculating the power consumption of EVs. This study can utilise different neural networks technique to compare the results on the same dataset as neural networks can perform better and can give good results.

In Croce et al. (2020), authors have used an amalgam of steady-state and quasi-dynamic models to provide the methods for predicting the EV's power consumption. Authors have employed filtering, fusion and transformation techniques for Floating Cara Data(FCD) to traffic predictors such as speed of the vehicle, density and vehicle acceleration. Also, by calculating the product of energy consumption every hour and vehicle's flow in the same hour authors have predicted the power consumption of the EVs. The flow of the vehicle is produced by the product of hourly average values of densities and speeds. In order to do the advancements in Croce et al. (2020), authors can include several others influencers such as terrain, climate, slopes of the roads, and drivers profile to make the calculation of energy consumption. In Pokharel et al. (2021), In order to the prediction of power consumption of EVs, authors have employed several machine learning algorithms. Major energy consumption contributors such as trip distance, variety of tires and other factors can influence the overall power consumption of the vehicles.

In the study Pokharel et al. (2021), trip distance has a 0.87 correlation value which is the highest for energy consumption. Additionally, Extreme Gradient Boosting(XGBoost) provides a maximum accuracy of 92% in all the machine learning techniques used in the study. Major contributors such as the distance of the trip, heating power and readings of the odometer also play important roles in the prediction. This paper has not considered the gradient of the roads. Also, deep learning techniques can also be used to provide better results. Li et al. (2021) have strongly examined the vital component which influences the power consumption of EVs. With the help of big data techniques, authors have estimated the relational coefficient among major factors and energy consumption of the vehicles. As per the study, speed and outside temperature are the most crucial factors which affect energy consumption if the driver's profile is not taken into the consideration. Additionally, a speed between 30 and 40 kilometres per hour and a temperature between 25 °C and 30°C facilitates the lowest range of energy consumption. They suggest that temperature is the most crucial factor when compared to speed in order to predict power consumption.

2.2 External Factors Affecting the Power Consumption of EVs

In Ayman et al. (2021), authors have proposed models for estimating the power consumption of different fleets of transport vehicles. Data from the bus fleet of CARTA has been collected in order to perform the operations which is Chattanooga's public firm. Authors in this paper have proposed the techniques for the collection of the data, data storage and processing and cleaning of domain-connected system-level transport data. In order to get the results, authors have used 21 independent features, 14 different types of roads, overall distance, 3 types of climates and 2 different types of traffic. Authors have found that machine learning does not have better accuracy than deep learning techniques when predicting the power consumption of EVs. Also, it was found that extended trips have lesser energy utilisation. Roads which have different terrains may not have equivalent results. Additionally, the data used in the paper does not have variations as they are taken from the same region. In future work, this research can also employ different driver profiles as they may have different results on power consumption.

In Schmidt and Cheein (2019), authors have performed the analytical predictions of EVs as per the nature of the terrains on which EVs are moving on their journey. In this paper, authors have used artificial vision to create and segregated different terrains such as grass, clay, gravel and pavement terrains etc. This terrain type is one of the major influencers in power consumption operations. The authors have trained 2 different models for 2 different scenarios that are when the vehicle is accelerating and when the vehicle is moving at a constant rate. In order to build and train the models, authors have used a polynomial regression algorithm and have achieved accuracies of around 65% and 90% as details are given by the original manufacturers of the EVs. Authors have employed the polynomial regression method for this study as these methods are sensitive to given input data. Therefore authors can use other regression methods and deep learning methods on datasets to overcome these issues. Since neural networks can give better results, authors may achieve better results and make predictions better.

In Al-Wreikat et al. (2022), authors have employed 4 years of real-time data on the EV from metro politian regions of the UK. This study has studied the effects of temperature and trip distance on the energy requirements of EVs. As per the authors, if the surrounding temperature of the EVs decreases, the energy consumption increases. If the temperature drops then the EV battery needs additional energy to produce heat which takes some power from the battery. Also, as per this study, EVs batter has better performances in urban regions in contrast to rural regions. In Liu et al. (2018), authors have considered the outcomes of outer temperature and surplus load on power utilisation of an EV. As per the study, the energy consumption of EVs is affected by the temperature of the outside weather. The driver of an EV can use the air conditioner if the weather is warm and the temperature is high which will suck up the power of the vehicle. Also, if the temperature is less than 17 degrees, energy utilisation will be lower as drivers may not use the air conditioner and heater as well which means the energy consumed will be lesser. Additionally, the most efficient temperature range for an EV is 21.8°C -25.2°C as vehicles can save up to 10% of electricity every kilometre.

In De Cauwer et al. (2017), authors have put forward methods in order to predict the power consumption of EVs. Authors have employed data which is real-time and have features such as geolocation, weather and type of the roads on which EVs are moving. Authors have used a multiple regression algorithm which combines features such as acceleration, and speed with other external features such as the type of the road and temperature. Neural network techniques are used in order to train the model for temperature and weather conditions. There was a mean absolute error of 12-14% for the trained model. Therefore after the study in the paper, it was found that a data-driven approach in the prediction of power consumption can help the model to adapt to the changing conditions of the climate.

Zhang et al. (2021) Have used the digital twin model technique to calculate and predict the energy consumption of EVs. Real-time data in the study helps in getting more accurate results. As per the paper, the digital twin model is influenced by temperature rather to other small factors such as the speed of the vehicle, acceleration and resistance. Other factors such as slopes, wind direction and air conditioning can not be simulated for the prediction. In Jonas et al. (2022), authors have studied the effects of different types of roads on the calculation of energy consumption. The authors have used real-world driving data which have been generated by the drivers. As per this study, local roads makes vehicles consume more energy as their slopes and terrains are different and soak ups more power from the battery while consumption on roads where terrains slopes are good enough to drive consumer lesser energy. Driving on such roads are easier hence driver makes little changes in their driving and energy needs are lower as well. Authors can also consider the temperature, use of heater and air conditioner, and excess load while making the assumptions and predictions.

In Hamwi et al. (2022), authors have studied the effects of high environmental temperature on the efficiency of EV when they are exposed to extreme weather conditions. The data used in this study is real world data from the EVs in Kuwait country. This study discovers that EVs in country like Kuwait need more power as vehicles soak up more energy in moving and drivers may use air conditioners and other equipment which can utilise the battery power. Also, this study states that there is a high correlation between total distance travelled and external environment temperature. This paper can also consider different EV model which have different tyres as road type and terrains are similar in entire Kuwait. Also, different profiles of the drivers can also be useful in finding the insights and make the predictions.

2.3 Internal Factors Affecting the Power Consumption of EVs

In Amine and Mokhiamar (2019), authors have applied the latest control methods to study the power consumption of electric vehicles. Authors have put forward neural network controllers to estimate the yaw moment required in stabilising the four-wheel motor EVs. Once simulation operations are executed, it was found that classical controllers have lower stability in contrast to modern controllers. Additionally, modern controllers have lower energy utilisation. Hence it can be said that power consumption is affected by the yaw moment as energy consumption in the modern controller is lesser. Authors can use more real-time data on deep learning algorithms to perform the study. Real-time data can help in determining the actual factors which can consume energy.

In Modi et al. (2020), authors have predicted the overall power consumption of electric vehicles using deep convolution neural networks in order to reduce the anxiety of a driver. The authors have used 3 major parameters which are slope, speed of the vehicle and tractive attempts. In order to train the model, simulation data had been used which was then transformed into images. The model is not trained on any sensor-generated data. As per the study, the gradients influences the power consumption of EVs and affects the battery performance of the vehicle. The authors here have not considered the major components such as the driver's driving styles, climate of the trip, excess loads and different terrains of the roads which can have major say in the prediction of the vehicles.

In Phyo et al. (2022), authors have put forward a methodology to predict the power consumption of electric buses based on driver's behaviour. As per the paper, the speed of the bus plays the important role in the performance evaluation of the electric consumption. If the bus is moving at a fast speed, it will consume more energy. If the bus is accelerating and decelerating at a certain time, it will consume more energy. Additionally, the authors have found that if the number of stops are higher, it will impact power utilisation. This paper can include different types of roads, climate and humidity etc in order to make the prediction better and improve the outcomes.

Kremzow-Tennie et al. (2020) studied the effect of driving resistance or driving force on the energy consumption of EVs. As per the study, increasing the weight inside the vehicle can exert extra pressure on the tire's surface area. Hence if the vehicle is carrying an extra load then it will consume extra electricity. Since acceleration and deceleration need more torque power, it will soak more power from the battery. Since driver's driving profile plays an important role, this study needs to include the type of the driver's driving style so that energy calculation can be more precise. In Li et al. (2022), authors have provided physical and data-driven fusion models in order to predict the energy consumption of electric buses. This study also takes outcomes of rolling drags, breaks of the vehicle and air conditioning temperature in order to train the CatBoost decision tree model. The fusion model achieved one of the best results but the errors are correlated to bad weather. Hence in future work, This study can consider changing weather conditions in order to get better results. Also, other factors such as internal as well as external factors can also be considered for the prediction.

In Desreveaux et al. (2019), authors have studied the effects of velocity on power consumption. In the first stage, The authors validate the energy profile of the real-time vehicles and then speed profiles are produced to reach the last destination. As per the paper here, the highest velocity has the maximum influence on the prediction. If maximum velocity is increased by 20% energy consumed will be increased by 15% to 20%. Also, high acceleration has a minor influence on consumption.

3 Methodology

This research has used a modified approach of Knowledge Discovery in Databases(KDD) methodology in order to conduct the study. KDD is one of the most popular methodologies in order to perform regression prediction experiments. The steps involved in the KDD are recursive, which allows going through every step and getting the best out of the datasets and the model trained. Hence this study has performed the following steps in order to get the results.

3.1 Data Collection

Dataset '0_VED_orig_data.csv' used in this study is taken from the public source Github by Mr Linas P. He is an associate professor Vilnius University, Lithuania. This data source is available on the Google Drive [1]. It involves GPS routes of vehicles as well as time-series data of their fuels and speeds, energy and additional power utilisation. The dataset involves 50 columns and 408057 rows in the CSV format. Table 1 represents breif description of the columns used in the study. DateTime features of the dataset include 'timestamp_ms', 'date', 'datetime', 'daynum' and which represent the date and time of the vehicle at a particular location. 'vehid' represents the id of the vehicle used in the dataset. 'trip' indicates the id of the trip taken by the vehicle. 'lat' and 'lon' values show latitude and longitude values of the vehicles which can help in determining the location of the vehicle at a specific time which can be read by DateTime features. 'speed_kmh' represents speed of the vehicle in km per hour. 'hvbattery_soc_per' represents high voltage battery readings of the vehicle. 'mm_edge_clazzs' represents the type of road on which the vehicle is moving.

3.2 Data Preprocessing

In this step of the methodology, the dataset undergoes different processes in order to improve the data quality. It involves taking out unnecessary elements such as noise, null values and redundant features which do not contribute to the model training process. In the experiment, data is loaded into the notebook using pandas. The total number of unique values of the 'trip' column was calculated so that the total distinct trip value is known for further operations. This dataset had 482 unique trips which needed to be grouped together in order to calculate certain new features which would have contributed

[1] https://drive.google.com/drive/folders/1NxGQzGXARK7qCSMHlsuL-Ovrl-0itisL

Table 1: Column Description

Column Name	**Description**
vehid	Vehicle ID
trip	Trip ID
timestamp_ms	Readings at a particular time in milliseconds
lat	Latitude
lon	Longitude
speed_kmh	Speed of the vehicle in Kilometer per hour
absoluteload	Load in the vehicle
oat_degc	Air temperature in Degrees
Airconditioning_w	Energy consumption of Air Conditioner in watts
heaterpower_w	Heater power in watts
hvbattery_soc_per	EV Battery readings at a timestamp
mm_edge_id	Id of Map matched path or route.
mm_edge_clazzs	Class or type of Map matched path or route
hourlydewpointtemperature	temperature the air needs to be cooled to
hourlydrybulbtemperature	Ambient air temperature in a hour
hourlyprecipitation	Precipitation in an hour
hourlypresentweathertype	Weather type in an current hour

to the prediction. Each trip had its own values of time, speed and battery consumption etc which were not useful if they were used individually. Hence in order to make data more meaningful, new dataframe 'df' was grouped together as per 'trip' and 'mm_edge_id' which represent trip id and the path id respectively on which they were running. Once grouping was done, a new dataframe 'vals' was created which had sorted grouped rows as per trip and mm_edge id. new dataframe 'vals' had a shape of 19348 rows and 49 columns. Since there were timestamp values in the columns of the dataset, a function 'find_time' was created in order to calculate the time(in seconds) single path and trip. This function helped in calculating the time taken by the vehicle in each trip which then could help in calculating the total driven time of the vehicle. This function subtracted the maximum values from the minimum values of 'timestamp_ms' for 'trip' and 'mm_edge_id' which was then divided by 1000 to keep them in the same unit. 'find_energy' function was also created to find the difference between each reading of 'hvbattery_soc_per' column so that energy consumed between or in the trip can be calculated. 'hvbattery_soc_per' showed the readings and the function could let us know the energy consumed after a certain time. This function also subtracted the maximum values from the minimum values of 'hvbattery_soc_per' for 'trip' and 'mm_edge_id'. Since in order to move ahead in the experiment, it was necessary to calculate the time travelled on each trip. Hence dataframe was grouped as per 'trip' and 'mm_edge_id' and it was passed through a 'find_time' function. It helped in determining the time consumed in a single trip on a particular path that is 'mm_edge_id'. A new feature 'speed' was created using numpy library's mean(). This feature was average speed of the vehicle in a trip. This feature helped determining the average speed of the vehicle in a trip when grouped together in the 'vals' dataframe. 'energy' feature helped in determining the energy consumed by the vehicle in a single trip and path. This feature was the actual column we were trying to predict in the experiment. This was calculated by applying the 'find_energy' to the grouping of trips and paths. Since ‘vals’ was the new dataframe created after grouping the rows on the basis of ‘trip’ and ‘mm_edge’, Vehicles were moving on changing speed hence to calculate the energy consumed in a trip, average speed was calculated. Hence the original dataframe ‘df’ shape changed to 19348 rows and 52 columns which is the new dataframe ‘vals’. A new function ‘find_distance’ was created in order to calculate the distance travelled by the vehicle on each trip. This function used the longitude and latitude value available in the dataset and calculated the distance. Once created this function was applied on the ‘vals’ dataframe which then created a new feature called ‘dist’. The dataframe ‘vals’ contained many columns with null values. Some of the columns had all the null values in their rows. Hence to make data cleaner to transform and then model training, columns without any value or readings were dropped and the rest of the columns which had some null values were treated and transformed into the new ones. Additionally, the columns which were unnecessary or did not add contribute to the predictions were dropped which included DateTime features and longitude and latitude columns. Others were also irrelevant which were dropped and a new dataframe ‘vals2’ was created. Later some columns also had 5-10% of null values, these null values were dropped to make data without any null values. ‘vals3’ dataframe had columns with 0 null values.

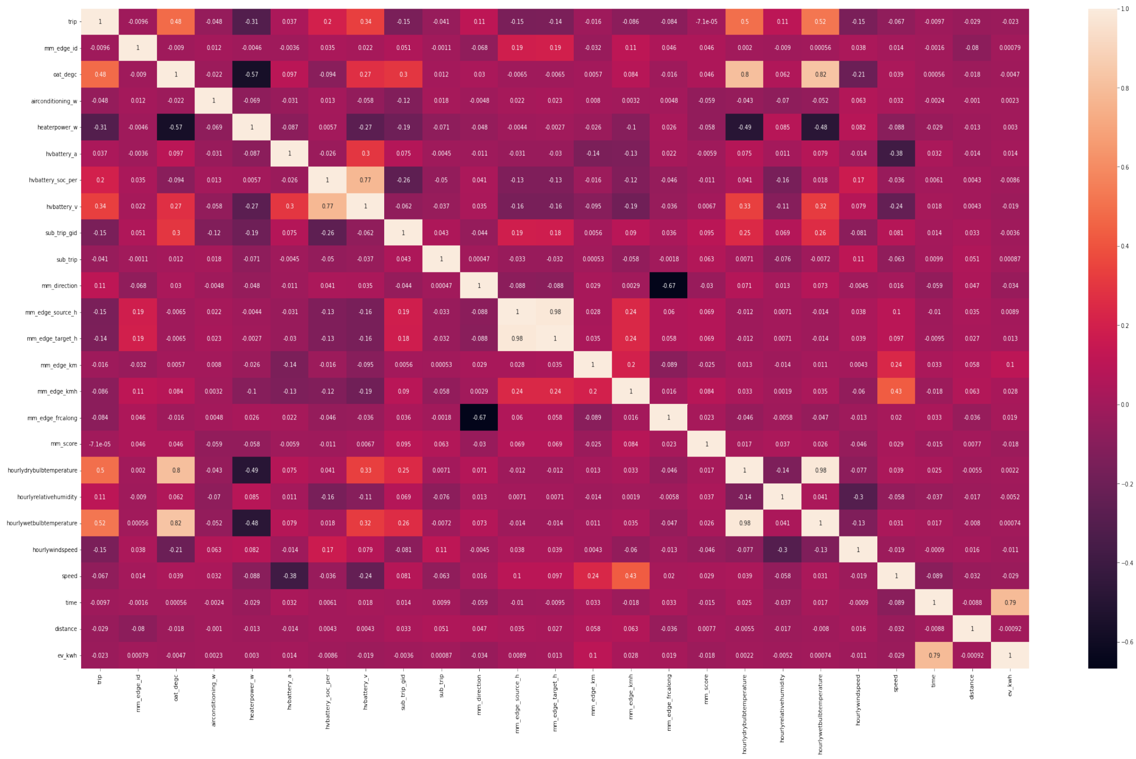

Figure 1: Heatmap for Vals4

The figure 1 represents the heatmap for the dataframe ‘vals4’. It shows that the ‘time’ and ‘mm_edge_km’ features are highly correlated to the target variable ‘ev_kwh’. Other columns whihc have very low correlational values does not seem to be helping in the prediction of the energy consumption, Hence the remaining irrelevant columns were also dropped from the dataframe which provided the ‘vals4’ dataframe. The shape of the ‘vals4’ after dropping and performing the pre-processing steps was 17067 rows and 32 columns.

3.3 Data Transformation

This stage of the methodology includes applying several transformation methods to the data for model training and evaluation operations. It makes researchers choose columns which have major contributions in the prediction stage of the methodology. In dataframe 'vals4', columns such as 'airconditioning_w' and 'heaterpower_w' were in watts which were not in the standard unit. Hence these columns were converted into kilowatts similar to the energy consumption column 'ev_kwh'. 'mm_edge_clazzs' column was the type of road or path on which the vehicle was travelling. This column was a categorical column which needed to be converted into the standard format for training the model. Hence this column was converted using sklearn.preprocessing library onehotencoder(). Once the transformation was done, it was joined with the dataframe and the rest of the irrelevant columns and used columns were dropped. Column 'hourlyprecipitation' had some rows with value 'T' which meant trace, in this experiment it was considered to be 0. Also, some of the columns were in object data type which were then converted into float data type so that they can be processed further for the model training purpose. Once all the irrelevant columns were dropped new dataframe 'dataset' was created in order to make the model training process easier to understand. Since the 'time' column contained some 0 values, those rows were ignored because the model will be trained and tested only on non-zero time values. If the time value was 0 then it did not make any sense to calculate the energy consumption in 0 times. Hence 0 values were dropped from the dataset. In order to perform the data mining and model training process, 'dataset' was divided into independent variable 'X' and dependent variable 'y' where 'y' is the target variable 'ev_kwh' which is the energy consumed in a trip calculated using 'find_energy()' in the data pre-processing stage in a given time. Further, X and y were divided into train, validation and test sets in order to train, validate and test the performance of the model. Therefore dataset has 6 new dataframe to perform the operation. These datasets were not in the standard scale hence standardscaler() from sklearn.preprocessing library was used to transform the data for data mining and training purposes.

Figure 2 provided a description of the dataset after the pre-processing stage. In this figure, it is evident that the average energy consumption in a trip is 0.197474 kWh with a standard deviation of 1.02123. The average travelling time of the vehicles is 26.286 seconds while the mean speed for all the trips is 46.0471 Kmh.

3.4 Data Mining

This stage of the KDD methodology involves finding insights into the data. Once critical insights are drawn out, models can be trained with the data and then can be used to do the prediction tasks. In this study, models are built using deep neural networks and trained to predict the power consumption of electric vehicles. Therefore model was built and its parameters were set. Keras library was used and the sequential class was initialised. Dense and Activation functions were imported from Keras.layers library. Rectified Linear Unit(ReLU) was used as an activation function. The Table 2 represents the number of summary of the neural network.

Once layers were set and parameters were initialised, the model was compiled using compile(). This study used adam optimiser and Mean Squared Error(MSE) as loss parameters. The learning rate of the model was set to 0.01. For training the model, fit() was used with X_train and y_train for the training of the model and X_val and y_val as the validation data. The batch size was selected as 64 while there were 1500 epochs for the

	speed	time	distance	ev_kwh	oat_degc	airconditioning_w	heaterpower_w	mm_direction	mm_edge_km	hourlydrybulbtemperature
count	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000	15337.000000
mean	46.047164	26.286477	0.904910	0.197474	10.178294	0.353179	0.404626	0.210928	0.104006	48.529634
std	18.417525	127.185682	1.323194	1.021234	11.370769	0.363775	0.846311	0.977534	0.111399	19.457025
min	0.000000	0.200000	0.009514	0.000000	-15.500000	0.000000	0.000000	-1.000000	0.003448	-12.000000
25%	33.640908	2.900000	0.094774	0.000000	2.000000	0.000000	0.000000	-1.000000	0.048378	34.000000
50%	47.792855	5.800000	0.318273	0.000000	8.000000	0.300000	0.000000	1.000000	0.072998	45.000000
75%	59.893748	11.900000	1.157310	0.000000	20.500000	0.500000	0.500000	1.000000	0.116420	65.000000
max	122.076663	1876.200000	10.508275	26.463417	34.000000	2.300000	4.250000	1.000000	1.951085	92.000000

Figure 2: Cleaned Dataset Snippet

Table 2: Model Summary

Layers	Output Shape	Parameters
dense	24	600
dense_1	64	1600
dense_2	64	4160
dense_3	128	8320
dense_4	1	129

training of the model.

3.5 Result Evaluation

After training the model, it is important to test the model with data for evaluating the performance. In other words, models need to be exposed to data it has not seen before. This stage is called Result evaluation where the model is tested and its performance is evaluated with the actual results. In this study, once the model was trained, it was tested with the X_test dataset. predict() used for predicting the value of X variables and then results are stored in y_pred. In order to evaluate the results, certain evaluation metrics were used such as Mean Absolute Error(MAE), Mean Squared Error(MSE), Root Mean Square Error(RMSE), and R-Squared (R2). print_evaluate() was created in order to evaluate and print the evaluation results for training, validation and testing datasets. These evaluation metrics were imported using sklearn's metrics library and applied to predicted and actual values of y datasets. All the evaluation metrics had nearly similar values which meant that the model was performing well on training, validation and test datasets.

4 Design Specification and Implementation

In order to implement this project, Python and its related libraries such as pandas, NumPy and seaborn and Keras library were used. Also, this study was conducted on google colab, a cloud-based platform for machine learning and deep learning projects. This study can also be recreated on a standard machine with an i5 or later version of the processor and 8 GB or more of RAM. To run the project, the system needs to have an anaconda or any other platform which supports python or notebook on the system. This study followed a certain flow which helped in the implementation of the study successfully. Figure 3 below shows the project architecture from start to end. It explains and makes the readers understand the overall flow and implementation of the project.

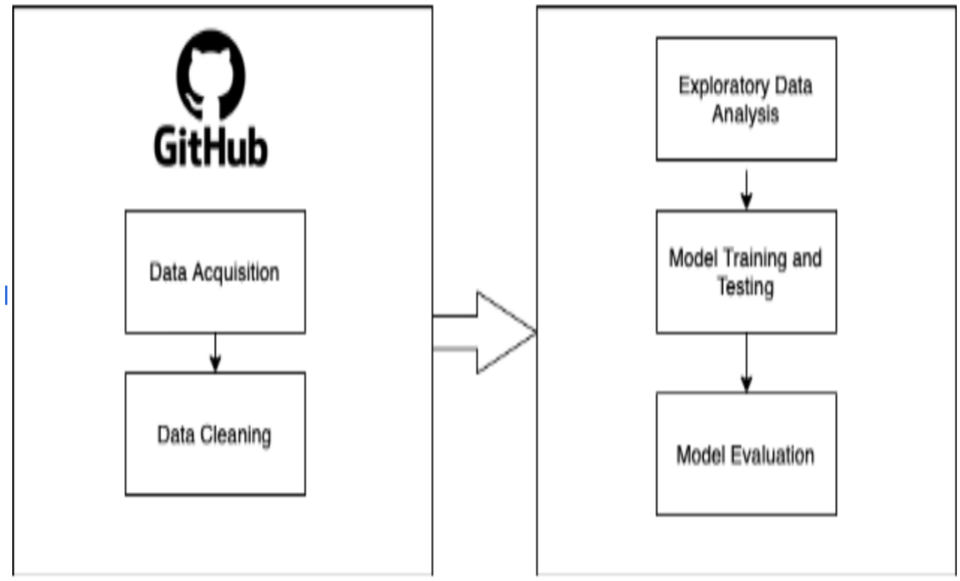

Figure 3: Project Architecture

4.1 Data Acquisition

The original dataset '0_VED_orig_data.csv' is available in the public directory on GitHub. Dimensions of this dataset is 405087 X 50. This dataset is hosted on both GitHub and google drive which can be used from the respective sources. Though the original data is a combination of multiple sources, this study has used the data available on the repository for conducting the study. This dataset is real-time data in which latitude and longitude values are geolocation data which helped in determining the real distance travelled by the vehicle. This dataset also contains map-matches(mm) trips which discuss the route or the path taken by the vehicle during their movement. there were 482 unique 'trip' while 2131 unique 'mm_edge_id' which map matched path id. After grouping the rows the originally used dimensions were reduced to 19348 X 49.

4.2 Data Cleaning

The original data was in a raw format which could not help in understanding and giving insights. Hence data cleaning stage helped in removing noise, null values and irrelevant columns from the dataset. Initially, unnamed columns were removed. later columns which had all the rows as null values were dropped from the original dataset to have only relevant columns. This dataset was grouped as per 'trip' and 'mm_edge_id' which created a new dataframe. Further after dropping the null columns dimensions of the dataframe were reduced from 408058 X 50 to 17067 X 25.

4.3 Exploratory Data Analysis

This stage of the study helped in understanding the insights of the data which could help in building the model. The figure below is the bar plot for the 'mm_edge_clazzs' column which indicates the distribution of the type of roads on which vehicles were moving.

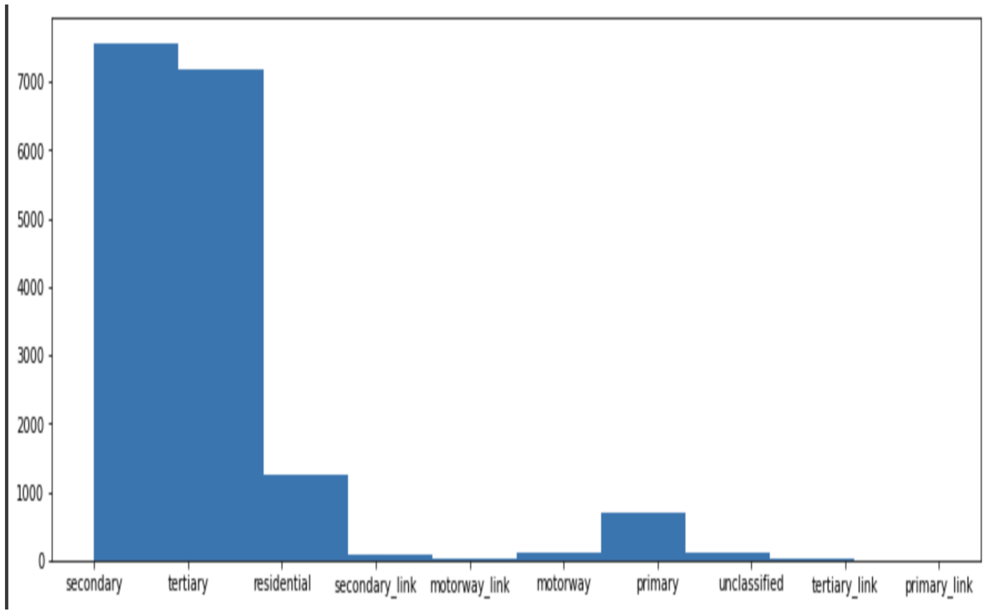

Figure 4: Bar Plot for type of Roads

From the Figure 4 it is evident nearly 90% of the vehicles were running on the secondary, tertiary and residential paths. Other functions such head(), isna(), sum() and etc helped in understanding the data and information related to the dataset.

Figure 5 shows the top 50 trips which have the highest total distance and energy consumed by the vehicles. This graph helps in understanding distances and energy consumed by the vehicles. In other words, how much a vehicle has travelled and how much energy it has consumed in a single trip.

Figure 6 shows the distribution of energy consumption and the total distance travelled by the vehicles.

Figure 7 shows the distribution of energy consumption and the average speed of the vehicles in the trips

4.4 Model Training and Testing

In this stage of the model, the neural network model was built, trained and tested on the dataset. Keras library helped in creating layers and fitting the model on the dataset. 'X' and 'y' datasets had independent and dependent columns which helped in training the model. The figure shows the summary of the model built. This model was then trained on 'X_train' and' X_val' This model then was used for the prediction of the X_test dataset.

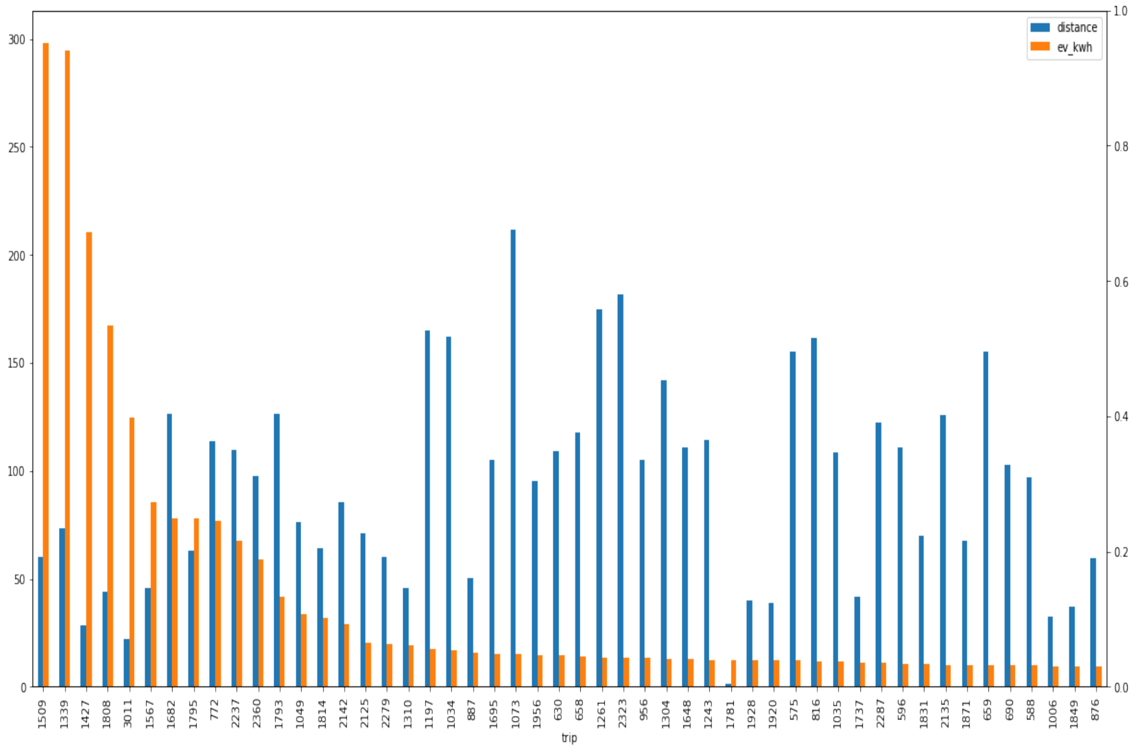

Figure 5: Total Distance and Energy for Top 50 Trips

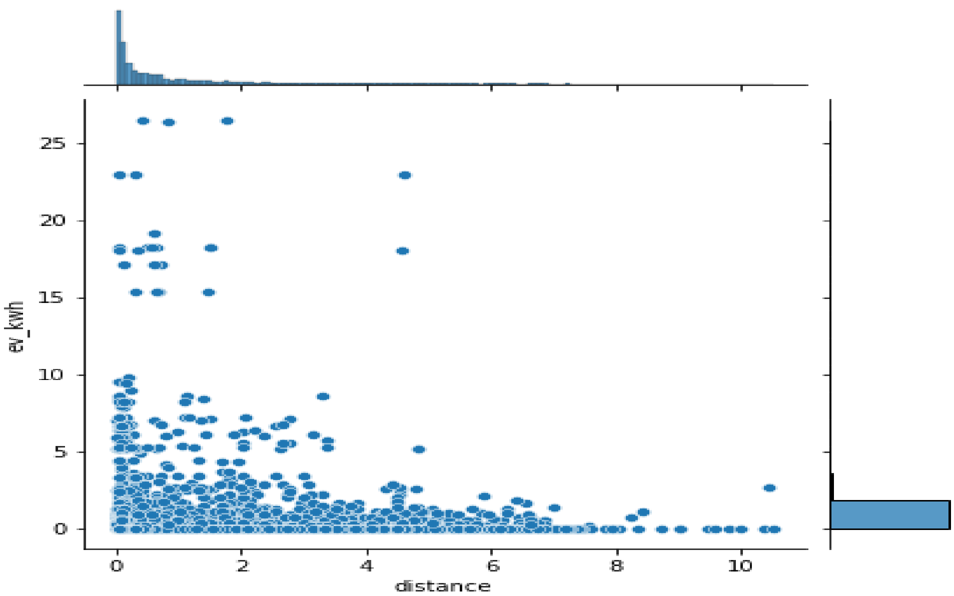

Figure 6: Distance vs ev_kwh

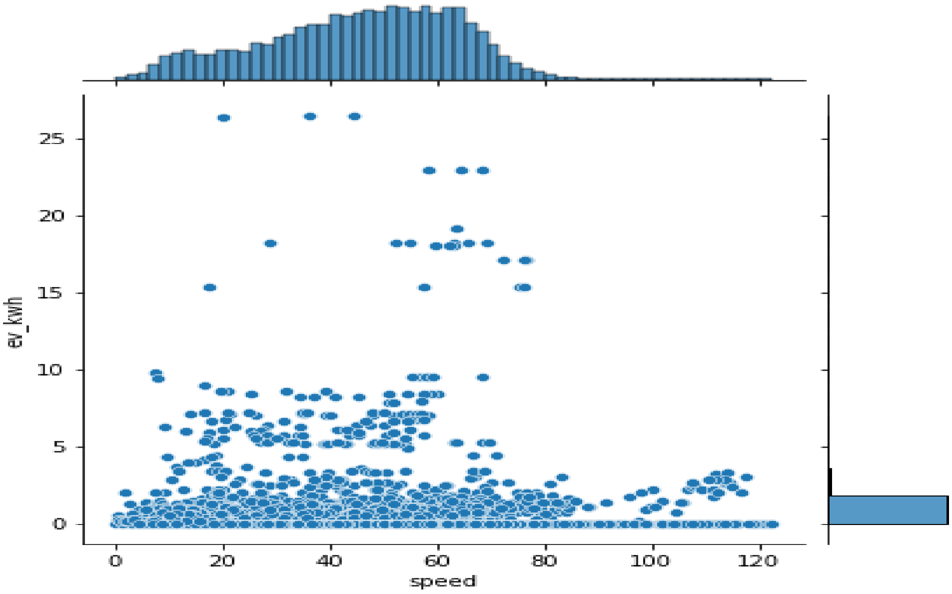

Figure 7: Speed vs ev_kwh

The figure shows the graph between training loss vs validation loss. From the figure, it is evident that the model is performing well on training as well as the validation dataset.

4.5 Model Evaluation

Once the model was trained on the training dataset, it was necessary to make it do the prediction on the testing datasets as well. The performance of the model needed to be evaluated on certain evaluation criteria as discussed in the methodology section. Figure 8 shows the training loss and validation loss on the respective datasets.

5 Evaluation

As mentioned in the result evaluation stage of the methodology, trained model was evaluated on metrics such as MAE, MSE, RMSE and R-Squared. Table shows the results obtained in the study.

Table 3: Results

Parameters	**MAE**	**MSE**	**RMSE**	**R2**
Training	0.1296	0.0589	0.2428	0.9433
Validation	0.1667	0.1252	0.3538	0.8961
Testing	0.1639	0.0994	0.3153	0.8896

From Table 3, it is evident that MAE for all the sets are lower which means that model was able to predict the energy consumption accurately with the average error of around 0.14 kwh. MSE and RMSE values of the model for all subsets are also less which suggests is able to give an accurate prediction of energy consumed by the electric vehicle. A higher value of R-squared also means that the deep neural is performing well and can give good predictions of energy consumption. For all the sets, the R-Squared value is

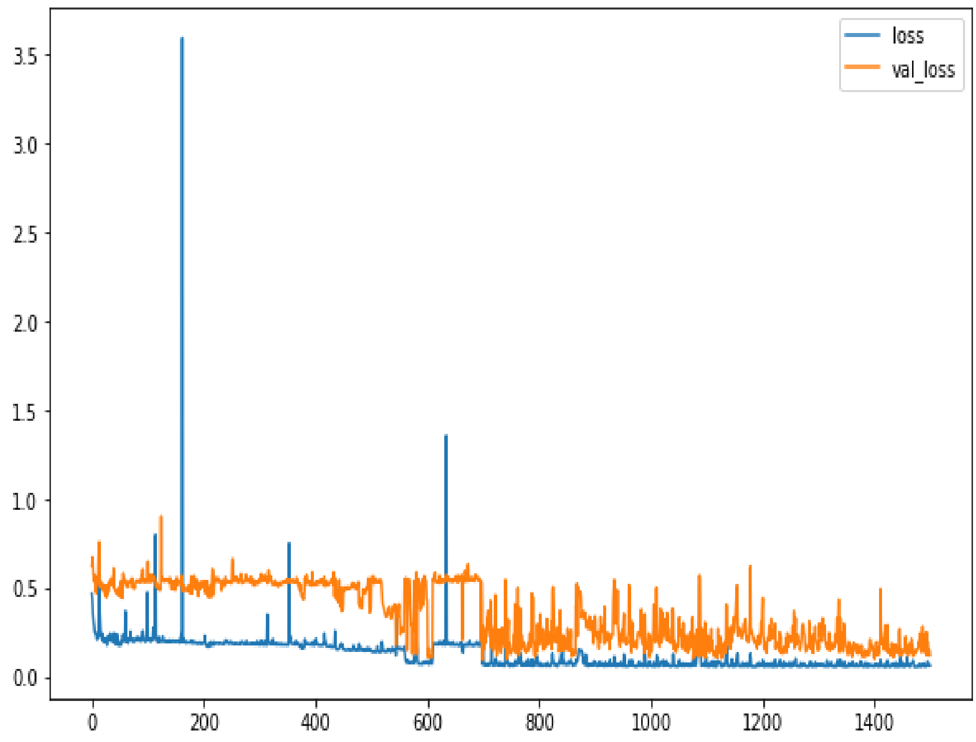

Figure 8: Training and Validation Loss

nearly 90 and above which means that neural network performs well on the real-time dataset used in the experiment.

Table 4: Comparison of Results

Study	**MAE**	**MSE**	**RMSE**	**R2**
Current Study	0.1639	0.0994	0.3153	0.8896
De Cauwer et al. (2017)	0.3160	0.0589	0.4710	
Pokharel et al. (2021)	4.551		0.9490	0.9186
Ullah et al. (2022)	24.94		35.12	0.9100

Table 4 shows the comparison of different results of neural networks and machine learning algorithms. from the table it is clear that neural networks perform better and give better results as compared to other algorithms. This study had achieved better MAE, MSE, RMSE and R2 for given dataset as results has surpassed the other given results with better results. Hence neural networks performs better for predicting the power consumption of electric vehicles for given dataset.

6 Conclusion and Future Work

This study proposed a deep neural network method to predict the total power consumption of electric vehicles when trained with real-time data. The deep learning model trained in this study was able to give good results with the lower values of MAE, MSE, RMSE and higher R-Squared which means that a neural network can be a good option when making the predictions for energy consumption. This study also considered the internal, as well as external factors which can affect the utilisation of the power by the vehicle such as speed, temperature, trip distance, trip time, wind speed and humidity etc. these, are the major contributors and responsible for the consumption of the energy. This study also employed the real-time data captured with the geolocation data such as latitude and longitude which eventually helped in calculating the distance. Dataset used in this study was taken from urban driving and considered only a single EV model. It would be fascinating to train the deep neural network on a wider dataset which will have more rows and will cover more regions. Furthermore, drivers' profiles can also be considered since they can have a bigger impact on energy consumption. It would be interesting to see the performance of the model if the dataset contains the type of tire, and different terrains and gives detail about the load in the vehicles as these factors affect the overall consumption of the EVs. Since every region has their own challenges and factors to consider, in the future work of this study, the model can be trained on different brands and hybrid models to see how deep neural networks perform.

References

Al-Wreikat, Y., Serrano, C. and Sodré, J. R. (2022). Effects of ambient temperature and trip characteristics on the energy consumption of an electric vehicle, *Energy* **238**: 122028.

Amine, S. and Mokhiamar, O. (2019). A study of stability and power consumption of electric vehicles using different modern control strategies, *Alexandria Engineering Journal* **58**(4): 1281–1290.

Ayman, A., Sivagnanam, A., Wilbur, M., Pugliese, P., Dubey, A. and Laszka, A. (2021). Data-driven prediction and optimization of energy use for transit fleets of electric and ice vehicles, *ACM Transactions on Internet Technology (TOIT)* **22**(1): 1–29.

Basso, R., Kulcsár, B. and Sanchez-Diaz, I. (2021). Electric vehicle routing problem with machine learning for energy prediction, *Transportation Research Part B: Methodological* **145**: 24–55.

Croce, A. I., Musolino, G., Rindone, C. and Vitetta, A. (2020). Energy consumption of electric vehicles: Models' estimation using big data (fcd), *Transportation Research Procedia* **47**: 211–218.

De Cauwer, C., Verbeke, W., Coosemans, T., Faid, S. and Van Mierlo, J. (2017). A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions, *Energies* **10**(5): 608.

Desreveaux, A., Bouscayrol, A., Trigui, R., Castex, E. and Klein, J. (2019). Impact of the velocity profile on energy consumption of electric vehicles, *IEEE transactions on Vehicular Technology* **68**(12): 11420–11426.

Gong, B., Wang, J. and Cheng, J. (2020). Market demand for electric vehicles under technology improvements and tax relief, *Emerging Markets Finance and Trade* **56**(8): 1715–1729.

Hamwi, H., Rushby, T., Mahdy, M. and Bahaj, A. S. (2022). Effects of high ambient temperature on electric vehicle efficiency and range: Case study of kuwait, *Energies* **15**(9): 3178.

Jonas, T., Wilde, T., Hunter, C. D. and Macht, G. A. (2022). The impact of road types on the energy consumption of electric vehicles, *Journal of Advanced Transportation* **2022**.

Kremzow-Tennie, S., Hellwig, M. and Pautzke, F. (2020). A study on the influencing factors regarding energy consumption of electric vehicles, *2020 21st International Conference on Research and Education in Mechatronics (REM)*, IEEE, pp. 1–6.

Li, X., Gao, R. and Yu, H. (2021). Influence factors and correction coefficient for energy consumption of electric vehicles, *E3S Web of Conferences*, Vol. 252, EDP Sciences, p. 02059.

Li, X., Wang, T., Li, J., Tian, Y. and Tian, J. (2022). Energy consumption estimation for electric buses based on a physical and data-driven fusion model, *Energies* **15**(11): 4160.

Liu, K., Wang, J., Yamamoto, T. and Morikawa, T. (2018). Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption, *Applied Energy* **227**: 324–331.

Liu, Y., Zhang, Q., Lyu, C. and Liu, Z. (2021). Modelling the energy consumption of electric vehicles under uncertain and small data conditions, *Transportation Research Part A: Policy and Practice* **154**: 313–328.

Modi, S., Bhattacharya, J. and Basak, P. (2020). Estimation of energy consumption of electric vehicles using deep convolutional neural network to reduce driver's range anxiety, *ISA transactions* **98**: 454–470.

Phyo, L. Y., Depaiwa, N., Yamakita, M., Kerdsup, B. and Masomtob, M. (2022). Impact of driving behavior on power consumption of electric bus: A case study on rama ix bridge, *2022 International Electrical Engineering Congress (iEECON)*, IEEE, pp. 1–4.

Pokharel, S., Sah, P. and Ganta, D. (2021). Improved prediction of total energy consumption and feature analysis in electric vehicles using machine learning and shapley additive explanations method, *World Electric Vehicle Journal* **12**(3): 94.

Rhode, S., Van Vaerenbergh, S. and Pfriem, M. (2020). Power prediction for electric vehicles using online machine learning, *Engineering Applications of Artificial Intelligence* **87**: 103278.

Sankaran, G., Venkatesan, S. and Prabhahar, M. (2020). Range anxiety on electric vehicles in india-impact on customer and factors influencing range anxiety, *Materials Today: Proceedings* **33**: 895–901.

Schmidt, J. R. and Cheein, F. A. (2019). Prognosis of the energy and instantaneous power consumption in electric vehicles enhanced by visual terrain classification, *Computers & Electrical Engineering* **78**: 120–131.

Singh, A. and Vestlund, N. (2021). Prediction of energy consumption for battery electric vehicles.

Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R. E. and Jamal, A. (2022). A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability, *Energy & Environment* **33**(8): 1583–1612.

Zhang, Z., Zou, Y., Zhou, T., Zhang, X. and Xu, Z. (2021). Energy consumption prediction of electric vehicles based on digital twin technology, *World Electric Vehicle Journal* **12**(4): 160.

Zhixin, F., Hanzhou, B. and Yue, Y. (2021). Calculation method of electric vehicle power consumption based on naive bayes classification, *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*, IEEE, pp. 3562–3566.