ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

COMPARATIVE ANALYSIS OF PREDICTIVE MODELS FOR ENERGY CONSUMPTION IN ELECTRIC VEHICLES

M.Sc. THESIS

Canberk ŞEN

Energy Science and Technology Division

Energy Science and Technology Programme

FEBRUARY 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

COMPARATIVE ANALYSIS OF PREDICTIVE MODELS FOR ENERGY CONSUMPTION IN ELECTRIC VEHICLES

M.Sc. THESIS

Canberk ŞEN
(301211051)

Energy Science and Technology Division

Energy Science and Technology Programme

Thesis Advisor: Asst. Prof. Dr. Mustafa Berker YURTSEVEN

FEBRUARY 2025

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTRİKLİ ARAÇLARDA ENERJİ TÜKETİMİ TAHMİNLEME MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

YÜKSEK LİSANS TEZİ

Canberk ŞEN
(301211051)

Enerji Bilim ve Teknoloji Anabilim Dalı

Enerji Bilim ve Teknoloji Programı

Tez Danışmanı: Dr. Öğr. Ü. Mustafa Berker YURTSEVEN

ŞUBAT 2025

Canberk ŞEN, a M.Sc. student of ITU Graduate School student ID 301211051, successfully defended the thesis entitled “COMPARATIVE ANALYSIS OF PREDICTIVE MODELS FOR ENERGY CONSUMPTION IN ELECTRIC VEHICLES”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Asst. Prof. Dr. M. Berker YURTSEVEN** ....
İstanbul Technical University

Jury Members : **Prof. Dr. Önder GÜLER** ....
Istanbul Technical University

Assoc. Prof. Dr. Bedri KEKEZOĞLU ....
Yıldız Technical University

Date of Submission : 10 January 2025
Date of Defense : 14 February 2025

To my grandfather,

FOREWORD

The study presented here explores energy consumption prediction for EVs, a topic that addresses critical challenges such as range anxiety and energy efficiency in real-world driving scenarios. As global efforts intensify to transition towards greener transportation, the insights gained from this research aim to enhance the usability and acceptance of EVs, contributing to the advancement of this transformative technology.

I would like to express my heartfelt gratitude to my advisor, Asst. Prof. Dr. Mustafa Berker Yurtseven whose guidance and encouragement were invaluable throughout this journey.

I am deeply grateful to my family and friends for their unwavering support, patience and encouragement. Their belief in my abilities and understanding of my commitments were pillars of strength during challenging times.

February 2025

Canberk ŞEN
(Mechanical Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
BEV	: Battery Electric Vehicles
BMS	: Battery Management System
CG	: Center of gravity
CNN	: Convolutional Neural Networks
DL	: Deep Learning
DT	: Decision Tree
ECM	: Equivalent circuit models
EMD	: Empirical Mode Decomposition
ESG	: Ensemble Stacked Generalization
EV	: Electric Vehicle
eVED	: Extended Vehicle Energy Dataset
FL	: Federated Learning
GPS	: Global Positioning System
HEV	: Hybrid Electric Vehicles
HV	: High voltage
HVAC	: Heating, Ventilation and Air Conditioning
ICE	: Internal Combustion Engine
ICEV	: Internal Combustion Engine Vehicle
ID	: Identity document
LSTM	: Long Short Term Memory
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
ML	: Machine Learning
MMC	: Markov Monte Carlo method
MSE	: Mean Squared Error
OAT	: Outside Air Temperature
OBD	: On-board diagnostic
PCC	: Predictive cruise control

PHEV	: Plug-in Hybrid Electric Vehicles
R^2	: Coefficient of Determination
RBF	: Radial Basis Function
ReLu	: Rectified Linear Unit
RF	: Random Forest
RMSE	: Root Mean Squared Error
RNN	: Recurrent Neural Network
SOC	: State of charge
SOH	: State of health
SVM	: Support Vector Machine
SVR	: Suppport Vector Regressor
TPM	: Transition Probability Matrix
TTW	: Tank to wheel
V2I	: Vehicle to infrastructure
V2V	: Vehicle to vehicle
VED	: Vehicle Energy Dataset
VMD	: Variational Mode Decomposition
WTW	: Wheel to wheel
XGBoost	: eXtreme Gradient Boosting

SYMBOLS

α	: Angle of road
°C	: Degree Celcius
A	: Ampere
C	: Regularization parameter
E_{total}	: Total energy consumption
F_{ar}	: Aerodynamic drag
F_{fr}	: Front rolling resistance force
F_g	: Gravitational force
F_i	: Inertial force
F_n	: Normal force
F_{rr}	: Rear rolling resistance force
F_t	: Tractive force
h_{t-1}, h_t, h_{t+1}	: Hidden state
I_b	: Battery current
kWh	: Kilowatthour
m	: Total number of data points
P	: Power
Q_0	: Rated capacity of battery
Q_t	: Remaining battery charge
t	: Time
V	: Voltage
V_b	: Battery voltage
W	: Watt
Wh	: Watthour
X	: Actual value
X_{t-1}, X_t, X_{t+1}	: Input
Y	: Predicted value
γ	: Kernel coefficient
ε	: Epsilon-insensitive loss

LIST OF TABLES

Page

LIST OF FIGURES

Page

COMPARATIVE ANALYSIS OF PREDICTIVE MODELS FOR ENERGY CONSUMPTION IN ELECTRIC VEHICLES

SUMMARY

This study investigates energy consumption prediction for electric vehicles (EVs), a critical area of study to address challenges like range anxiety and improve the efficiency of EVs in real world conditions. Limitations in range estimate and energy management that prevent EVs from being widely adopted must be addressed as EVs become more and more important in lowering greenhouse gas emissions and reducing dependency on fossil fuels. This study evaluates the performance of three advanced machine learning models Support Vector Regression (SVR), eXtreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks, using real world data to develop accurate predictive models for energy consumption.

The analysis is based on the Vehicle Energy Dataset (VED), which provide a comprehensive set of energy and driving statistics from several car models, including hybrid, plug-in hybrid and battery electric vehicles. The dataset, collected over a year in diverse driving conditions, includes key parameters such as GPS trajectories, speed, state of charge (SOC) and ambient temperature factors. Its comprehensive nature ensures the applicability of the models developed in this study to real world scenarios.

A detailed methodology was employed, starting with rigorous data preprocessing to ensure the quality and consistency of the dataset. This involved steps such as cleaning, normalization and feature selection, which are crucial for optimizing the performance of the machine learning models. Each of the three models selected for this study offers unique advantages: SVR effectively captures both linear and nonlinear relationships; XGBoost excels in handling feature interactions and structured data and LSTM networks are well suited for analyzing time series data and identifying sequential dependencies.

The findings indicate the potential of machine learning approaches to estimate EV energy use with high accuracy. These predictions not only mitigate range anxiety but also enable the development of adaptive energy management systems that optimize battery performance and improve overall user experience. These developments also help EV customers make accurate decisions about charging and route planning.

This research highlights the importance of integrating high quality datasets like VED with robust machine learning models to advance EV technology. This research advances the larger objective of developing an efficient and sustainable transportation system by addressing current problems in energy forecasting. Future work could focus on enhancing hybrid model approaches, expanding datasets to include more diverse driving conditions and developing real time adaptable prediction systems, significantly closing the gap between existing constraints and the expanding demands of EV adoption.

ELEKTRİKLİ ARAÇLARDA ENERJİ TÜKETİMİ TAHMİNLEME MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

ÖZET

Bu tez elektrikli araçlar için enerji tüketimi tahminini araştırmakta ve menzil kaygısı zorluklarının üstesinden gelmek ve elektrikli araçların gerçek dünya koşullarındaki verimliliğini artırmak için kritik bir çalışma alanına odaklanmaktadır. Elektrikli araçlar, sera gazı emisyonlarını azaltmak ve fosil yakıtlara bağımlılığı azaltmak için giderek daha önemli bir rol oynamaktadır. Ancak, menzil tahmini ve enerji yönetimindeki sınırlamalar elektrikli araçların daha geniş çapta benimsenmesini engellemektedir. Bu çalışmada, elektrikli araçların enerji tüketimini tahmin etmek için makine öğrenimi modelleri geliştirilmiş ve performansları karşılaştırılmıştır. Çalışmada, Support Vector Regression (SVR), eXtreme Gradient Boosting (XGBoost) ve Long Short-Term Memory (LSTM) olmak üzere üç farklı makine öğrenimi modeli kullanılmış ve enerji tüketimini tahmin etme yetenekleri değerlendirilmiştir.

Model geliştirme sürecinde, gerçek dünya sürüş verilerini içeren Vehicle Energy Dataset (VED) kullanılmıştır. Bir yıl boyunca farklı sürüş koşullarında toplanan bu veri seti GPS rotaları, hız, şarj seviyesi ve ortam sıcaklığı gibi temel parametreleri içermektedir. Veri setinin kapsamlı yapısı, bu çalışmada geliştirilen modellerin gerçek dünya senaryolarına uygulanabilirliğini sağlamaktadır. Araç hızı, ortam sıcaklığı, iklimlendirme sistemlerinin enerji tüketimi, batarya gerilimi ve batarya akımı gibi değişkenler veri setinde yer almakta olup enerji tüketimi tahmin modellerinin eğitimi için temel giriş değişkenleri olarak kullanılmıştır.

Çalışmada, elektrikli araçların enerji tüketimini belirlemek amacıyla batarya akımı ve batarya gerilimi kullanılarak enerji tüketimi hesaplamaları yapılmıştır. Elektrikli araçlarda anlık güç tüketimi, batarya gerilimi ve batarya akımının çarpımı ile hesaplanmaktadır. Anlık güç tüketiminin belirlenmesi, belirli bir süre boyunca enerji tüketimini hesaplamak için temel bir adımdır. Enerji tüketimi, anlık güç değerlerinin zaman ile integrali alınarak hesaplanmıştır. Çalışmada kullanılan VED veri setinde batarya gerilimi ve akım değerleri belirli aralıklarla kaydedildiğinden, enerji tüketimi hesaplaması gerçekleştirilmiştir.

Çalışmada, veri temizleme, normalizasyon ve özellik seçimi gibi adımları içeren kapsamlı bir veri ön işleme süreci gerçekleştirilmiştir. Eksik veya hatalı veriler temizlenmiş, özellik mühendisliği teknikleri ile model performansını artıracak değişkenler belirlenmiştir. Veri temizleme aşamasında, eksik ve hatalı veriler tespit edilerek uygun yöntemlerle işlenmiştir. Veri temizleme işlemi sırasında ayrıca kopya kayıtların kaldırılması ve aykırı değerlerin tespit edilerek model performansını olumsuz etkileyebilecek girdilerin veri setinden çıkarılması sağlanmıştır.

Özellik mühendisliği aşamasında, modelin tahmin doğruluğunu artırmaya yardımcı olacak en önemli değişkenler belirlenmiş. Bu süreçte, değişkenler arasındaki

korelasyon analizleri yapılarak, gereksiz veya yüksek korelasyonlu değişkenler elenmiş, modelde aşırı öğrenmeyi engellemek adına yalnızca anlamlı değişkenler kullanılmıştır.

Veri setinin eğitim ve test verisi olarak bölünmesi aşamasında, modelin genelleme yeteneğinin doğru bir şekilde değerlendirilebilmesi için veri seti %80 eğitim ve %20 test oranında bölünmüştür. Veri bölme işlemi, rastgele seçilen veriler yerine zaman serisi analizine uygun bir şekilde yapılmış, böylece eğitim verilerinin gelecekteki test verileriyle zaman açısından örtüşmemesi sağlanmıştır. Ek olarak, model doğruluğunu daha iyi ölçebilmek için çapraz doğrulama yöntemi uygulanmış ve 5-katlı çapraz doğrulama süreci kullanılarak modelin farklı veri bölümlerindeki performansı test edilmiştir.

Her model, tahmin doğruluğunu artırmak amacıyla hiperparametre optimizasyonu sürecinden geçirilmiş ve en iyi performansa ulaşacak şekilde ayarlanmıştır. Hiperparametre optimizasyonu, modelin öğrenme sürecinde en iyi parametre kombinasyonunu belirleyerek tahmin doğruluğunu artırmayı amaçlamaktadır. Bu süreçte Grid Search ve Random Search gibi yaygın kullanılan teknikler uygulanmıştır.

SVR modeli için çekirdek tipi, düzenleme parametresi ve hata toleransı gibi kritik hiperparametreler optimize edilmiştir. Grid Search yöntemi kullanılarak RBF (Radial Basis Function) ve lineer çekirdek seçenekleri test edilmiş, optimum düzenleme parametresi ve hata toleransı değerleri belirlenmiştir. Modelin aşırı öğrenmemesi için uygun çekirdek fonksiyonu seçilmiş ve çapraz doğrulama ile modelin genelleme yeteneği test edilmiştir.

XGBoost modeli için öğrenme oranı, maksimum derinlik, ağaç sayısı ve alt örnekleme oranı gibi hiperparametreler optimize edilmiştir. Random Search yöntemi kullanılarak geniş bir hiperparametre alanı taranmış ve en iyi parametre kombinasyonu belirlenmiştir. Modelin aşırı uyum göstermesini engellemek için erken durdurm kriteri uygulanmış ve optimum iterasyon sayısı belirlenerek modelin performansı artırılmıştır.

LSTM modeli için gizli katman sayısı, nöron sayısı, öğrenme oranı, optimizasyon fonksiyonu ve toplu işleme boyutu gibi hiperparametreler optimize edilmiştir. LSTM modelinin zaman serisi verilerine uyumunu en iyi hale getirmek için Adam algoritması seçilmiş, en düşük hata değerlerini veren öğrenme oranı seçilmiştir. Modelin daha verimli öğrenebilmesi için geri yayılım sürecinde sönümleme ve L2 düzenleme teknikleri kullanılarak aşırı öğrenme önlenmiştir.

Tüm modellerde hiperparametre optimizasyonu uygulanarak en iyi kombinasyonlar belirlenmiş ve tahmin doğruluğu maksimize edilmiştir. Çapraz doğrulama ile test edilen bu modellerin performansı, R^2, RMSE ve MAE gibi metrikleri ile değerlendirilmiş ve elde edilen sonuçlara göre en başarılı model belirlenmiştir.

Elde edilen sonuçlara göre, LSTM modeli zaman serisi verilerindeki bağımlılıkları başarılı bir şekilde yakalayarak en yüksek doğruluk oranına ulaşmıştır LSTM modelinin en iyi performansı sergilemesinin temel sebebi, zaman serisi verilerinde bağımlılıkları başarılı bir şekilde öğrenebilme yeteneğidir. Elektrikli araçların enerji tüketimi, zamana bağlı değişkenleri içeren bir süreçtir. Araç hızındaki değişimler ve ortam sıcaklığı gibi faktörler geçmiş değerleriyle doğrudan ilişkilidir. LSTM, geçmiş zamandaki girdileri hatırlayarak enerji tüketimi tahmininde daha iyi bir genelleme sağlayabilir. Bu özellik, özellikle araç hızındaki değişimlerin enerji tüketimi üzerindeki etkisini değerlendirme noktasında kritik bir avantaj sunar.

XGBoost modeli ise ağaç tabanlı bir model olup, değişkenler arasındaki ilişkileri öğrenmede oldukça etkilidir, ancak zaman serisi verilerindeki ardışık bağımlılıkları doğrudan modelleyemez. XGBoost, veriler arasında güçlü korelasyonları yakalayarak tahmin doğruluğunu artırabilir, ancak zamanla değişen dinamik ilişkileri doğrudan modelleyemez. Bununla birlikte, değişkenlerin belirli anlık kombinasyonlarına bağlı olarak enerji tüketimi değişimlerini iyi bir şekilde yakalayabilir, ancak geçmiş değerleri hesaba katmada LSTM kadar etkili değildir.

SVR modeli ise doğrusal ve doğrusal olmayan ilişkileri belirli bir düzeye kadar öğrenebilse de, yüksek boyutlu ve karmaşık zaman serisi verilerinde yeterince esnek değildir. SVR, belirli çekirdek fonksiyonları ile doğrusal olmayan yapıları modelleyebilir, ancak verilerdeki uzun süreli bağımlılıkları ve zamana bağlı değişimleri doğrudan öğrenme kapasitesi sınırlıdır. Ayrıca, SVR veri ölçeğine oldukça duyarlı olduğu için geniş veri setlerinde ve değişkenler arası etkileşimlerin yoğun olduğu durumlarda, genelleme konusunda yetersiz kalmaktadır.

Bu çalışma, makine öğrenimi modellerinin elektrikli araçların enerji tüketimini tahmin etme konusundaki etkinliğini ortaya koymuştur. Özellikle LSTM modelinin zaman bağımlı verileri başarılı bir şekilde işleyerek en iyi sonuçları verdiği gösterilmiştir. Geliştirilen modellerin doğru enerji tüketimi tahminleri yapabilmesi, elektrikli araç kullanıcılarının daha güvenilir menzil tahminleri elde etmelerine ve rota planlamalarını daha bilinçli yapmalarına olanak sağlamaktadır. Ayrıca, bu tahminlerin batarya yönetim sistemleri ile entegre edilerek, elektrikli araçların şarj sürelerinin optimize edilmesine ve şebeke yükünün daha iyi yönetilmesine katkı sağlayacaktır.

Sonuç olarak, tez kapsamında geliştirilen makine öğrenimi modelleri, elektrikli araçların enerji tüketimini tahmin etme sürecinde güçlü ve etkili araçlar olarak değerlendirilmiştir. Özellikle LSTM modeli, zaman serisi verileriyle çalışmada üstün performans sergileyerek enerji tüketimi tahmininde en doğru sonuçları elde etmiştir. Çalışmada geliştirilen modeller, EV teknolojisinin daha geniş bir kullanıcı kitlesine yayılmasını sağlayacak doğru menzil tahmini ve enerji yönetimi sistemleri için önemli bir temel oluşturmaktadır. Gelecekte yapılacak çalışmalar, daha büyük ve çeşitli veri kümeleri kullanarak tahmin doğruluğunu artırmaya, hibrit modelleme tekniklerini uygulamaya ve gerçek zamanlı enerji tahmin sistemlerini daha geniş çapta uygulamaya odaklanmalıdır. Bu doğrultuda yapılan çalışmalar, elektrikli araçların kullanım verimliliğini artırarak sürdürülebilir ulaşım sistemlerinin yaygınlaşmasına katkı sağlamaktadır.

1. INTRODUCTION

The growing adoption of electric vehicles (EVs) is a critical element in global strategies aimed at reducing greenhouse gas emissions and transitioning to sustainable transportation system (Ullah et al., 2022). Despite their environmental benefits significant barrier to widespread EV acceptance remains range anxiety drivers’ fear of insufficient battery charge to complete their journey (J. Wang, 2016). This challenge is rooted in the limited driving range of current EV models and the variability of real world driving conditions, which can complicate range estimation.

While extensive research has been conducted on EV energy consumption and prediction, there is a pressing need for experimental comparisons of model based and data driven prediction techniques (Shen, Zhou, Yu, et al., 2023). To address this gap, this thesis focuses on a comparative analysis of different methodologies for predicting EV energy consumption using real world data. By evaluating the performance and accuracy of these approaches, the study aims to identify the most effective methods for energy consumption prediction.

1.1 Electric Vehicles

EVs are increasingly being recognized as a sustainable alternative to conventional internal combustion engine (ICE) vehicles. Their adoption is driven by a combination of environmental, economic and policy related factors. EVs have the potential to significantly reduce greenhouse gas emissions and dependence on fossil fuels, aligning with global efforts to combat climate change (Krogh et al., 2015). The switch to EVs has been further accelerated by government incentives, emissions reduction laws, rising fuel prices and increased environmental awareness (Adnane et al., 2023; Skuza & Jurecki, 2022).

Key advantages of EVs include their lower carbon footprint, increased energy efficiency and operational benefits. EVs produce zero tailpipe emissions and offer advanced features such as energy recuperation and a near ideal speed-torque profile, which contribute to their overall efficiency (Yuan et al., 2024). However, despite these benefits several challenges impede their widespread adoption.

A significant limitation is the restricted driving range of EVs, which is constrained by current battery capacities when compared to the ranges of traditional ICE vehicles (J. Wang, 2016). Infrastructure and charging time are other problems. EVs require considerably longer charging times than the refueling times for conventional vehicles and the number of available charging stations is still limited compared to gas stations (J. Wang, 2016). These challenges collectively impact drivers' acceptance of EVs and hinder their widespread adoption.

To address these challenges, researchs have focused on several key areas. Improving energy efficiency through optimized vehicle components and driving strategies is a critical area of development. Accurate energy consumption prediction models are being developed to provide real time estimations of energy usage (X. Xu et al., 2019; J. Zhang et al., 2020). Route planning algorithms aim to identify the most energy efficient paths for EVs, while advancements in battery technology seek to increase driving range and reduce costs (J. Wang, 2016).

The transition to EVs represents a vital step toward achieving a sustainable transportation system. While EVs offer significant environmental and operational advantages, continued research and development are necessary to address persistent challenges related to range anxiety, energy efficiency and infrastructure limitations. By overcoming these barriers, EVs can become a viable and widely accepted alternative to traditional ICE vehicles (J. Wang, 2016).

1.2 Range Estimation

Range estimation is a complex and critical component of EV technology. Accurate range estimation enables drivers to plan their trips effectively, reduces range anxiety and promotes confidence in EV technology (Albuquerque, 2022; Feng et al., 2024). Analyzing a variety of dynamic variables, such as battery SOC, driving conditions, environmental influences and vehicle parameters are necessary for range estimation. This section explores the challenges of range anxiety, the various estimation models used for range prediction and the methodologies employed to improve their accuracy and reliability.

1.2.1 Range anxiety

Range anxiety is one of the most significant psychological barriers to the widespread adoption of EVs (J. Wang, 2016; Z. Xu et al., 2024). It refers to the fear that an EV's driving range may be insufficient to reach a destination or charging point, causing stress and inconvenience for drivers. This concern is amplified by the relatively shorter range of EVs compared to traditional ICE vehicles, longer charging times and the uneven availability of charging infrastructure (Liu et al., 2021; Mediouni et al., 2022; Ullah et al., 2022; J. Wang, 2016). These factors create uncertainty and hesitation among potential EV users, impacting their willingness to adopt this sustainable technology.

A significant consequence of range anxiety is the reduced usable battery capacity. Many EV drivers maintain a 20–30% battery buffer to avoid running out of charge, even though their vehicles are designed to use the entire battery capacity (Shen, Zhou, Yu, et al., 2023; Ullah et al., 2022). While this precaution minimizes the risk of being stranded, it also restricts the vehicle's effective range, limiting the benefits of EV technology. Additionally, range anxiety often influences drivers' route planning behaviors. Drivers often select less convenient routes with frequent stops for charging because they are worried about running out of battery power.

Hesitancy toward long trips is another notable impact of range anxiety (Z. Xu et al., 2024). EV owners may avoid long distance travel, particularly in areas with meager charging infrastructure. This reluctance limits the utility of EVs for intercity travel and reinforces concerns about their practicality for everyday use. Moreover, the fear of being stranded adds a layer of mental stress to the driving experience, diminishing user confidence and satisfaction with EV technology (Z. Xu et al., 2024).

Several factors contribute to the prevalence of range anxiety. Inaccurate range estimation is a primary concern, as many EVs display predicted ranges based on ideal conditions without accounting for variables such as weather, terrain and individual driving styles (Ullah et al., 2022). This discrepancy between the predicted and actual range reduces trust to the technology. The degree of range anxiety varies among drivers, a phenomenon known as heterogeneous range anxiety (Z. Xu et al., 2024). For instance, new EV drivers, who may lack familiarity with their vehicle's capabilities, are often more prone to anxiety compared to experienced users. The limited availability of charging infrastructure, particularly in rural or underdeveloped areas, amplifies range anxiety and restricts the mobility and appeal of EVs.

Efforts to mitigate range anxiety focus on technological advancements and infrastructure improvements. Accurate energy consumption prediction, which incorporates real time data on driving conditions, road types, weather and traffic, can significantly enhance the reliability of range estimates (Ullah et al., 2022; J. Wang, 2016; Z. Xu et al., 2024) . Personalized range estimation models that consider individual driving behaviors also reduce prediction errors and build driver confidence. Probabilistic models that present predictions with confidence intervals can help drivers better understand and manage uncertainties (Petkevicius et al., 2021). Expanding charging networks and improving charging speeds are equally critical for addressing concerns about range limitations. Furthermore, tools that optimize routes for energy efficiency and charging availability can help drivers plan trips more effectively, reducing the stress associated with range anxiety.

Range anxiety is a complex problem due to the limitations in current EV technology and infrastructure. Addressing this challenge requires a combination of accurate range estimation methods, improved charging infrastructure and strategies to enhance driver trust and confidence in EV technology. By mitigating range anxiety, EVs can become a more viable and attractive alternative to ICE vehicles, accelerating the transition to sustainable transportation (Shen, Zhou, Yu, et al., 2023; Ullah et al., 2022).

1.2.2 Estimation models

Estimation models are essential for predicting the range of EVs and enhancing the overall driving experience. These models analyze various dynamic factors such as driving behavior, environmental conditions, vehicle specifications and battery parameters to estimate energy consumption and the remaining driving range. Estimation models can be categorized into analytical, statistical and machine learning (ML) based approaches, each offering unique advantages and limitations (Liu et al., 2021).

Analytical models rely on the principles of vehicle dynamics and powertrain efficiency to estimate energy consumption. These models calculate the energy required to overcome forces like rolling resistance, aerodynamic drag and gravitational pull during uphill driving while accounting for regenerative braking and motor efficiency (Shen, Zhou, Yu, et al., 2023). This is clearly indicated in Figure 1.1, different forces acting on a vehicle moving uphill on an inclined road. It includes forces such as gravity, normal force, traction force, aerodynamic resistance, rolling resistance and inertia (Albuquerque, 2022). These forces influence the vehicle's motion, energy consumption and overall performance, particularly in varying terrain conditions. Understanding these interactions is crucial for optimizing vehicle efficiency and energy management. Analytical models are computationally efficient and provide a solid theoretical framework for understanding energy consumption. However, their simplicity can be a limitation, as they often fail to capture the complex interactions of real world driving conditions, such as variations in traffic, road surfaces and weather.

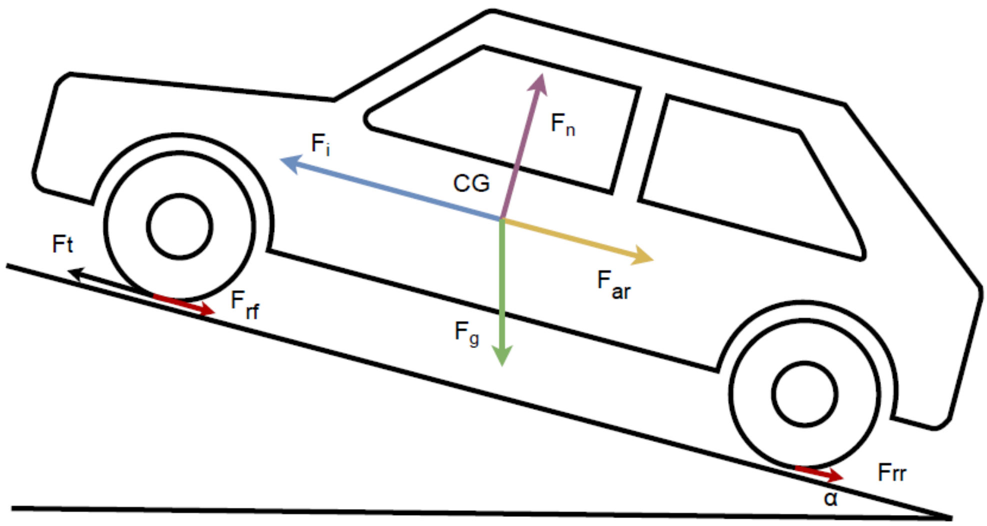

Figure 1.1 : Main influencing forces on a moving vehicle (F_i , inertial force; F_t,tractive force; F_g , gravitational force; F_{rr} , rear rolling resistance force; F_{fr} ,front rolling resistance force; F_{ar} , aerodynamic (air) drag; F_n, normal force; CG, center of gravity; α,the road slope) (Albuquerque, 2022).

Statistical models use historical data to identify patterns and relationships between key parameters and energy consumption. For example, linear regression models can establish dependencies between variables like speed, acceleration and battery consumption. While statistical models are relatively simple and interpretable, they rely on assumptions about linearity and error distributions. These assumptions may not hold in dynamic, real world conditions with nonlinear relationships, leading to reduced accuracy and potential biases in predictions.

Machine learning models have emerged as a robust solution for EV range prediction, offering flexibility and adaptability to complex datasets. These models can capture intricate, nonlinear relationships between input features and energy consumption, making them particularly effective for real world applications. Machine learning approaches can be further divided into single models, ensemble models, hybrid models and deep learning (DL) models. The instantaneous power consumption of an electric vehicle battery is shown in equation 1.1.

$$P(t) = V_b(t) \; x \; I_b(t) \tag{1.1}$$

where power P(t) at any given time (t) is determined by the product of the battery voltage $V_b(t)$ and the battery current $I_b(t)$. Equation 1.2 calculates the total energy consumption over a given time period.

$$E_{total} = \int_{t=0}^{t_0} V_b(t) \; x \; I_b(t) \, dt \tag{1.2}$$

The total energy consumption E_{total} over a time period from t=0 to t=t_0 is determined by integrating P(t), given as the product of $V_b(t)$ and $I_b(t)$ over time. This integral represents the accumulation of power over the specified duration, providing a measure of the total energy consumed, expressed in Watthours.

Single models, such as Support Vector Regression (SVR), Artificial Neural Networks (ANNs) and Decision Trees (DTs), use a single algorithm for prediction tasks. While these models are effective in handling specific use cases, their standalone performance may be limited in highly dynamic scenarios. Ensemble models, such as Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), combine the outputs of multiple models to improve accuracy and robustness. By leveraging the strengths of many algorithms and mitigating individual weaknesses, ensemble methods provide enhanced predictive capabilities.

Hybrid models combine machine learning techniques with other methods, such as optimization algorithms, to enhance performance. For instance, hybrid approaches may integrate linear regression with nonlinear neural networks or use metaheuristic algorithms to optimize model parameters. This combination allows hybrid models to balance simplicity and complexity, improving prediction accuracy across diverse conditions (Chou & Tran, 2018).

DL models, including Long Short-Term Memory (LSTM) networks and Transformer architectures, excel at capturing temporal dependencies in time series data. These models are particularly well suited for EV range prediction, as they can analyze sequential patterns in driving data, such as changes in speed and energy consumption over a trip. Their ability to handle large datasets and learn complex relationships enables DL models to achieve high accuracy, particularly when trained on real world driving data (Sulaiman & Mustaffa, 2024). As shown in Figure 1.2, energy consumption on EVs are highly depending on vehicle speed profile (Yan et al., 2024).

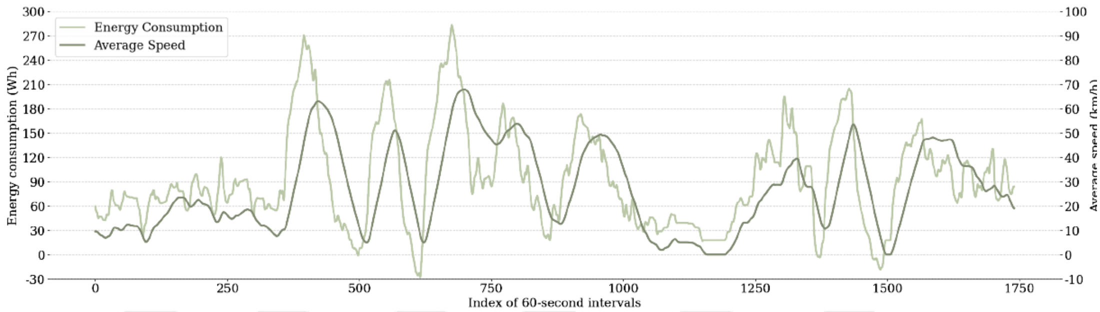

Figure 1.2 : Effect of average speed on energy consumption (Yan et al., 2024).

To further enhance the performance of machine learning models, advanced techniques such as feature selection, data augmentation and transfer learning are often employed. Feature selection identifies the most relevant parameters for modeling, improving prediction accuracy while reducing computational complexity (Chou & Tran, 2018). Data augmentation expands training datasets by creating synthetic samples, enhancing the diversity and robustness of the models. Transfer learning allows models trained on existing EV data to be adapted to new vehicle types or scenarios, ensuring their applicability across diverse conditions.

Evaluating the performance of these estimation models is critical for understanding their effectiveness. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and R-squared (R^2) are commonly used to assess prediction accuracy and model fit. Each metric provides unique insights, helping researchers and developers refine their models for optimal performance (Feng et al., 2024).

Despite their advancements, estimation models face challenges such as data quality, model generalization and real time adaptability. High quality datasets with detailed features like SOC, driving conditions and environmental factors are essential for accurate predictions. Models must also generalize effectively to different topographies, driving styles and EV models, ensuring broad applicability. Furthermore, real time adaptability is crucial for responding to changes in traffic, road conditions and driver behavior.

Estimation models are a cornerstone of EV range prediction, each contributing unique strengths to the task. By combining advancements in analytical, statistical and machine learning approaches, alongside innovative techniques like hybrid and DL models, researchers and manufacturers can develop more accurate, reliable and adaptive range prediction systems. These systems are essential for addressing range anxiety and supporting the widespread adoption of EVs.

1.3 Literature Review

The literature on EVs covers a wide range of topics, including energy consumption modeling, prediction techniques, influencing factors and the application of ML in improving these processes. This section synthesizes key findings and emerging trends from the reviewed sources, highlighting the complexities and advancements in EV energy consumption prediction and related areas.

Real world driving data is a critical component for accurate EV energy consumption analysis and prediction. Numerous studies emphasize the limitations of laboratory based tests, which often fail to reflect actual driving conditions. Real world datasets provide insights into how various factors affect energy consumption, offering a more comprehensive understanding of EV performance (Achariyaviriya et al., 2024; Shen, Zhou, Yu, et al., 2023).

Several key factors significantly influence EV energy consumption. Ambient temperature plays a major role, with studies noting a substantial increase in energy usage at lower temperatures. For instance, a study reported a 100% rise in energy consumption when the ambient temperature dropped from 20°C to 0°C (Achariyaviriya et al., 2024). Similarly, trip length impacts energy efficiency, with shorter trips consuming up to 10% more energy per unit distance for distances below

16 kilometers (Achariyaviriya et al., 2024). Traffic conditions also have a notable effect, with congestion increasing energy consumption by as much as 40%. Driving behavior, including aggressive acceleration and braking, can raise energy usage by up to 16%. Road grade greatly also affects energy consumption, with a 3% incline increasing usage by 50%, while descending grades can reduce consumption by 80% (Holden et al., 2020). Vehicle auxiliary loads, such as climate control systems, interact with ambient temperature to further influence energy demands (Ullah et al., 2021). Table 1.1 summarizes key factors influencing energy consumption in EVs, highlighting the impact of environmental conditions, driving behavior and vehicle related characteristics. Temperature extremes significantly affect energy consumption with both cold and hot conditions increasing energy demands, especially due to HVAC usage. External conditions such as wind resistance, road inclination and traffic congestion further contribute to variations in energy consumption. Additionally, driving behavior and vehicle load play a role, where aggressive driving and added weight lead to higher consumption. These insights emphasize the importance of optimizing energy efficiency through smart driving strategies, route planning and adaptive vehicle control systems.

Table 1.1 : Summary of energy consumption by the factors analysed (Skuza & Jurecki, 2022).

Type of factor	Energy consumption	Energy consumption
Negative temperatures	-20°C to 0	~6% -14%
High temperature	45°C	~33% - 58%
HVAC heating (in winter)	-	~52% - 94%
HVAC cooling (in summer)	-	~11% - 17%
Wind	15 km/h	~5% - 14%
Inclination	%3	~50%
Route length	4-16 km	~15-29%
Traffic conditions (proportion of stopping time)	%12-18 stop to %24-34	~20%
Load	+250 kg	~7%
Driving style	Agressive	~17%

Lastly, driving speed significantly impacts energy consumption patterns, with hybrid EVs emitting less CO_2 at lower speeds. As Figure 1.3 shows, there is significant gap for CO_2 emisions between BEV, HEV and ICEV for both urban and highway roads.

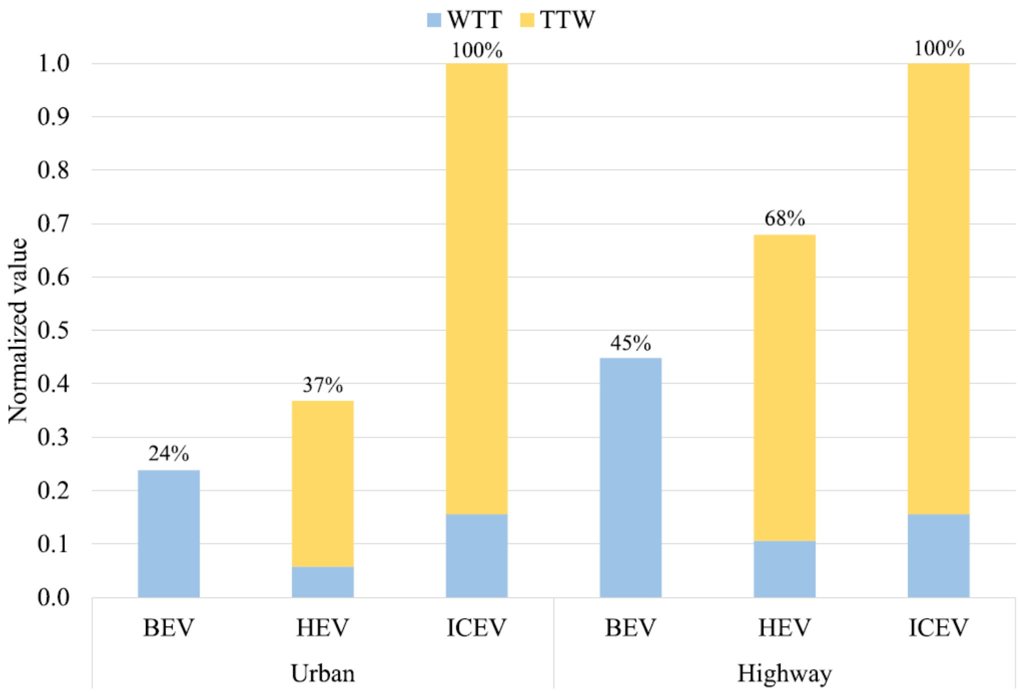

Figure 1.3 : The normalized wheel to wheel (WTW) and tank to wheel (TTW) CO_2 emissions of the test vehicles for each route mode (Achariyaviriya et al., 2024).

Road grade and driving behavior remain critical factors, with numerous studies quantifying their impacts on energy consumption. Regenerative braking has also been a focus of research, as this technology improves energy efficiency by recovering energy during braking events.

The application of ML in predicting EV energy consumption and range has gained significant attention in recent years. Artificial intelligence (AI) based forecasting methods, including ML and DL techniques, have been extensively reviewed and compared for their effectiveness in energy prediction (Adnane et al., 2023; Jui et al., 2024). Commonly used ML algorithms include Artificial Neural Networks (ANN), Support Vector Machines (SVM), RF and LSTM networks (Chou & Tran, 2018; Shin & Woo, 2022; Ullah et al., 2021; Yang et al., 2022). These models are capable of capturing complex relationships in large datasets, making them ideal for energy consumption prediction.

Hybrid models, which combine multiple ML techniques or integrate ML with other methods, have emerged as a powerful approach. Ensemble models, such as RF and Gradient Boosting, improve prediction accuracy by aggregating the outputs of multiple algorithms. Some studies have explored the use of direct function ANN to simultaneously predict multiple energy parameters, comparing these models to inverse function approaches (Bouktif et al., 2018; Chou & Tran, 2018).

ANNs have been employed to develop engine fuel consumption maps for conventional vehicles, which were then compared to parallel hybrid vehicle models. These comparisons demonstrated a notable reduction in fuel consumption and carbon dioxide emissions for hybrid vehicles, showcasing the potential of ANNs to enhance energy efficiency and reduce environmental impact (Adedeji, 2023). Similarly, ANNs have been applied in the design and simulation of pure electric vehicle parameters, where a single model has been shown to accurately predict multiple outputs simultaneously, further emphasizing their versatility.

Innovations in ANN applications have also led to the development of inverse and direct function models. Inverse function models use outputs for prediction and estimation tasks, while direct function models utilize virtual functions as inputs to generate multiple outputs. A notable example is the direct function ANN model, which integrates calculated virtual functions to predict various energy parameters, including city, highway and combined electric consumption, within a single framework (Achariyaviriya et al., 2023). This approach has been shown to achieve higher accuracy compared to traditional inverse function models, making it a promising tool for comprehensive energy consumption analysis.

Time series forecasting methods, particularly for short term load prediction, have been applied to energy consumption analysis in both buildings and microgrids (Wazirali et al., 2023). Data driven models emphasize the importance of preprocessing and feature selection to improve model performance. Predictive cruise control (PCC) systems, which use intelligent driving assistance to optimize energy usage, have also been highlighted as a promising area for reducing energy consumption (Gao et al., 2023).

Model optimization techniques, such as hyperparameter tuning, are commonly used to improve the performance of these predictive models. Studies often evaluate these models using metrics such as MAE, MAPE and RMSE to ensure consistent and comparable performance (Bouktif et al., 2018; Ribeiro et al., 2020).

The reviewed studies utilize a variety of datasets for training and testing predictive models, including real world EV tracking data, traffic flow information, weather data and geographic parameters (Adnane et al., 2023).

Emerging trends in EV energy consumption prediction highlight the growing importance of ML and data driven methods. These techniques enable the processing of large and complex datasets allowing for quick and adaptive decision making [5]. However, challenges remain, particularly regarding the standardization of forecasting models (Chou & Tran, 2018). A common database for comparing models is needed to facilitate more accurate evaluations.

The quality and quantity of training data are critical for ML model accuracy. Datasets must include detailed features such as driving conditions, vehicle parameters, weather and driver behavior to ensure reliable predictions (Chou & Tran, 2018). Real time adaptability is another challenge, as models must dynamically adjust to changes in conditions and individual driving habits (Ribeiro et al., 2020). Longitudinal studies are needed to assess the long term impacts of integrating advanced technologies, such as renewable energy sources, into smart grids (Kiasari et al., 2024).

Interpretability remains a key issue in ML based models. While these models often provide highly accurate predictions, their "black box" nature makes it difficult to understand the underlying factors influencing their outputs (Ribeiro et al., 2020). Techniques that improve model transparency are essential for building trust and enabling practical applications.

Research has explored energy consumption and prediction across various types of EVs. Hybrid Electric Vehicles (HEVs) have been a focus, with studies examining energy management strategies, fuel consumption modeling and performance comparisons with conventional vehicles (Jui et al., 2024). Battery Electric Vehicles (BEVs) have also received attention, particularly regarding their energy consumption and driving range under different conditions (De Cauwer et al., 2015).

1.4 Purpose of Thesis

The purpose of this study is to evaluate and compare the performance of three advanced machine learning models SVR, XGBoost and LSTM networks in predicting energy consumption for EVs using real world data. Accurate energy consumption prediction is essential for addressing range limitations, a critical concern among EV users and enhancing EV efficiency in real world driving scenarios.

This study uses detailed driving data collected from a Nissan Leaf 2013 model, one of the widely studied electric vehicles due to accessibility of data. By leveraging this real world dataset, the study aims to provide a robust comparison of SVR, XGBoost and LSTM models examining their effectiveness in capturing complex patterns and interactions that influence EV energy consumption. Each model represents a unique approach to machine learning, with SVR being a powerful regression method suitable for capturing linear and nonlinear relationships, XGBoost as a highly optimized boosting algorithm known for its performance in tabular data and handling feature interactions and LSTM as a recurrent neural network (RNN) specifically designed to capture temporal dependencies and sequential patterns in time series data.

The performance of these models is evaluated on several metrics to determine their accuracy in estimating energy consumption, which is essential for predicting the remaining driving range of an EV. Reliable range predictions help decrease range anxiety among EV users, enhancing their driving experience and trust in EV technology. Accurate range prediction based on real world data further supports the development of adaptive energy management systems that can optimize battery usage, inform drivers of energy saving practices and improve overall EV performance.

2. THEORETICAL BACKGROUND

2.1 Fundamentals of Electric Vehicles

EVs are a transformative innovation in the transportation industry, offering a sustainable alternative to conventional ICE vehicles. EVs capacity to drastically cut greenhouse gas emissions and reliance on fossil fuels are becoming more popular. This shift is further supported by government incentives, stricter emissions regulations, rising fuel costs and growing awareness of environmental sustainability (Ullah et al., 2021). Compared to ICE vehicles, EVs offer enhanced energy efficiency, lower noise pollution and zero tailpipe emissions, positioning them as a critical component of the global effort to mitigate climate change and reduce urban pollution (J. Wang, 2016).

EVs come in various forms, each suited to specific use cases. BEVs rely entirely on electricity stored in high capacity batteries and produce no emissions during operation. HEVs combine an electric motor and battery system with an internal combustion engine, allowing them to switch seamlessly between power sources for greater fuel efficiency. Plug-in Hybrid Electric Vehicles (PHEVs) represent a hybrid category that can be charged externally while also utilizing an ICE, offering extended range flexibility (Z. Wang et al., 2024). These variations demonstrate the versatility of EV technology in meeting diverse transportation needs.

EVs differentiate with their advanced powertrain systems. The battery system is the primary energy source, with lithium-ion batteries being the most commonly used due to their high energy density, long cycle life and relatively low self-discharge rates (Yang et al., 2022). The SOC of the battery is a crucial parameter that determines the remaining driving range and impacts the overall efficiency of the vehicle (Nabi et al., 2023). Integrated with the motor is a power electronics control system, which regulates the flow of energy between the battery and motor, optimizing performance under varying driving conditions. Also, EVs are equipped with regenerative braking systems, which capture and convert kinetic energy during braking into electrical energy stored in the battery, enhancing overall energy efficiency (Skuza & Jurecki, 2022).

Energy consumption in EVs is influenced by a complex interplay of factors. Vehicle characteristics, such as weight, aerodynamics and rolling resistance, directly affect energy efficiency. Environmental conditions, temperature, wind resistance and road gradients play a significant role in determining energy use during a trip. Driver behavior such as acceleration patterns, average speed and braking intensity further adds variability to consumption. Driving conditions, including traffic congestion and trip length, also contribute to fluctuating energy demands (Feng et al., 2024; Zhu et al., 2024). These dynamic factors underscore the importance of accurate energy consumption prediction for optimizing vehicle performance and mitigating range anxiety. A framework illustrates in Figure 2.1 for predicting energy consumption in EVs by integrating historical driving data, environmental factors and ML techniques. It involves segmenting driving data, predicting driving conditions, extracting relevant features and optimizing an XGBoost model for accurate energy consumption forecasts. The final model uses real time vehicle data to provide future energy consumption predictions, improving efficiency and range estimation.

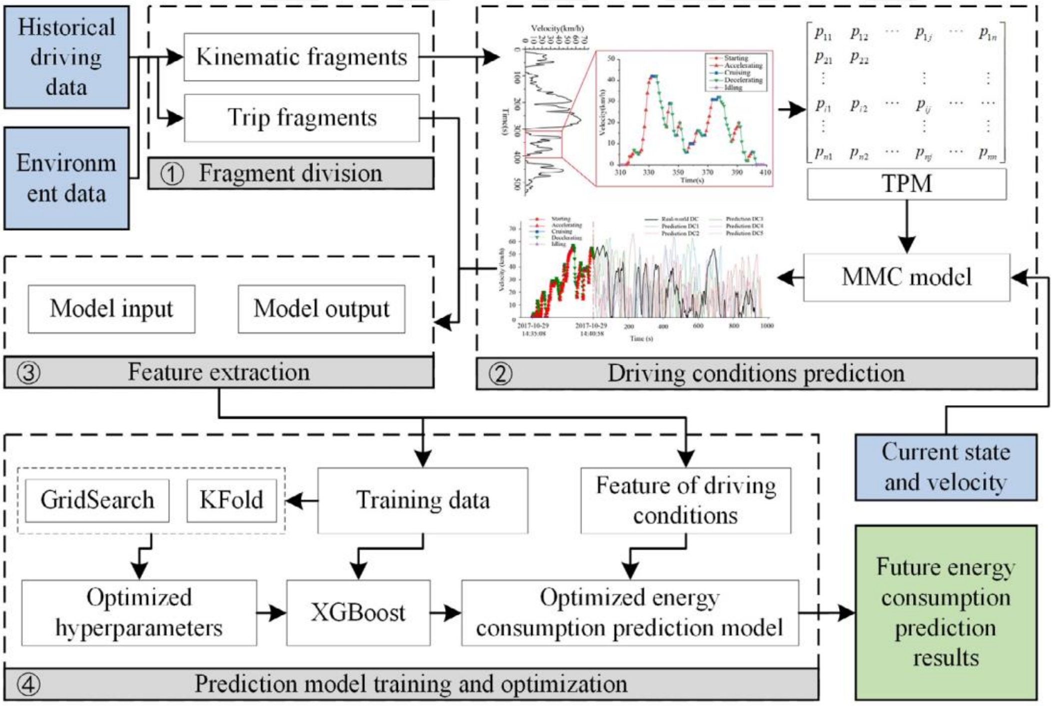

Figure 2.1 : Machine learning based energy consumption prediction framework (J. Zhang et al., 2020).

EV energy consumption prediction has evolved through two primary approaches. Traditional model based methods rely on principles of physics and vehicle dynamics to estimate energy use, providing a foundational understanding of how forces like drag, rolling resistance and gradient affect performance. In contrast, data driven techniques, particularly those utilizing ML, analyze historical and real time data to uncover patterns and relationships that influence energy consumption (Shen, Zhou, Yu, et al., 2023). ML based models, such as those employing LSTM networks or ensemble methods, have shown remarkable promise in capturing the nonlinear, multifaceted nature of EV energy consumption (Feng et al., 2024; Ullah et al., 2021). LSTM structural unit shown in Figure 2.2.

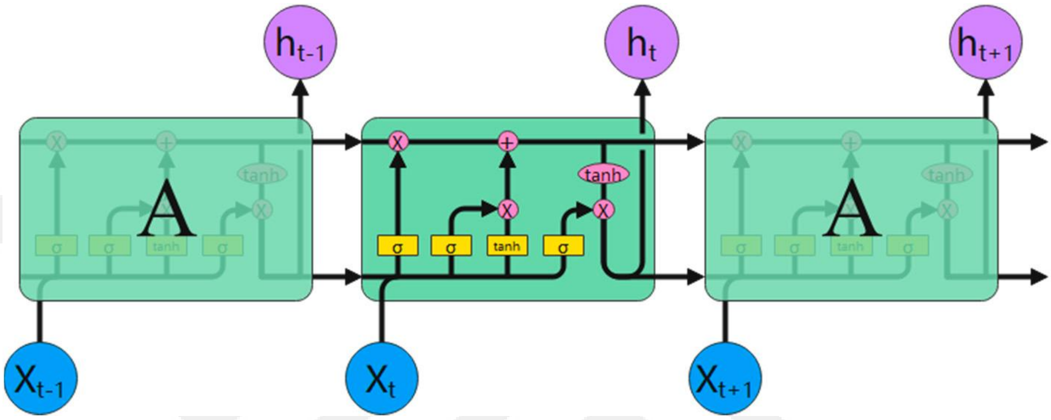

Figure 2.2 : Structural unit of LSTM (X. Zhang et al., 2022).

Battery performance is a key challenge, as factors like energy density, state of health (SOH) and capacity degrade over time, reducing overall efficiency and range. Real world energy consumption variability due to differing driving conditions further complicates the development of standardized energy models (Nabi et al., 2023; Yang et al., 2022).

Ongoing research is directed toward overcoming these challenges and advancing EV technology. Efforts are focused on improving battery materials and designs to enhance energy density and durability, developing sophisticated energy management systems and refining energy consumption prediction models with the help of ML (Ullah et al., 2022). As the transportation sector continues to evolve, EVs are poised to play a pivotal role in achieving sustainable mobility, underscoring the importance of continued innovation and infrastructure development (J. Wang, 2016).

2.2 Factors Affecting Energy Consumption

Energy consumption in EVs is influenced by a wide array of interconnected factors, which can be broadly categorized into vehicle related, environmental, driver related and driving condition factors. Understanding and optimizing these factors is crucial for improving energy efficiency and extending the driving range (Mediouni et al., 2022; Ullah et al., 2022).

Vehicle specific characteristics play a critical role in determining energy consumption. Key parameters include the vehicle's weight, the efficiency of its components and the use of auxiliary systems. Vehicle weight significantly impacts energy demand, as heavier vehicles require more energy for acceleration and maintaining speed. Reducing vehicle mass through lightweight materials and design innovations is an effective strategy to lower overall energy consumption (Liu et al., 2021). Auxiliary systems, such as heating, ventilation and air conditioning (HVAC), also contribute to energy usage. These systems, especially under extreme temperatures, can significantly reduce the vehicle's range (Ullah et al., 2022). The type of tire used may have a minor impact on energy consumption. While differences in tire design, such as rolling resistance, tread pattern and material composition, can influence how efficiently a vehicle moves, their effect on overall energy consumption is generally small compared to other factors like vehicle speed, weight and driving behavior (Pokharel et al., 2021).

Battery system specifications, including energy density, capacity and SOH are fundamental determinants of energy consumption. The SOC of the battery directly influences energy usage and driving range. In addition, motor and drivetrain efficiency, rolling resistance determined by tire pressure and design and the vehicle’s aerodynamic profile all contribute to the energy efficiency of the EV (Mediouni et al., 2022). SOC of a battery is a key parameter that represents the remaining capacity of the battery relative to its full charge as seen in equation 2.1.

$$SOC = \frac{Q_t}{Q_0} x100\% \tag{2.1}$$

where Q_t is the charge remaining in the battery at a given time and Q_0 is the total charge capacity of the battery when fully charged. This equation expresses SOC as a percentage, indicating how much usable energy is left before the battery needs to be recharged.

There is a direct linear relationship between battery current and energy consumption in electric vehicles. As the battery current increases, energy consumption also rises, reflecting the proportional demand for power during higher current usage. Conversely, when the battery current decreases, energy consumption correspondingly reduces (Achariyaviriya et al., 2023).

Energy use is significantly impacted by external environmental factors. Ambient temperature is one of the most significant factors, affecting both battery performance and auxiliary system usage (J. Zhang et al., 2020). In colder climates, energy consumption rises due to increased HVAC usage and reduced battery efficiency, while in excessively hot conditions, air conditioning demands may lower energy efficiency (Skuza & Jurecki, 2022).

Road conditions, the surface quality and slope, further influence energy consumption. Uphill driving and uneven road surfaces increase energy requirements, while downhill slopes can reduce consumption and facilitate regenerative braking (Ullah et al., 2021). Traffic congestion with its stop and go patterns increases energy usage by making the trip take longer.

Route planning and charging habits also play a role. Choosing energy efficient routes with fewer inclines and less congestion can help conserve energy. Charging strategies, including the timing and frequency of charges, can affect battery longevity and energy efficiency (Feng et al., 2024).

Trip characteristics and operational factors have a direct effect on energy consumption. Average speed, speed variability and average acceleration are key indicators of energy usage (Feng et al., 2024). Driving cycles' standardized patterns used to simulate typical driving conditions may differ significantly from real world conditions, affecting energy predictions. Longer trips generally require more energy but efficient driving during extended journeys can offset consumption to some extent (Skuza & Jurecki, 2022).

It is important to note that many of these factors interact in nonlinear ways, creating complex dependencies that can amplify or mitigate energy consumption under varying conditions. For instance, environmental factors such as temperature can compound vehicle related inefficiencies, while regenerative breaking can counterbalance the impact of steep inclines.

2.2.1 Driving behavior

Driving behavior has a profound effect on the energy consumption of EVs. The way a driver handles acceleration and braking in response to real time driving conditions directly impacts the vehicle's kinematic changes. Behaviors such as rapid acceleration, abrupt braking and inconsistent speeds are characteristic of aggressive driving styles, which can lead to a significant reduction in driving range, sometimes by as much as 35%. On the other hand, adopting energy efficient driving habits like smooth acceleration, steady cruising and strategic use of regenerative braking can extend the range by up to 27% (Feng et al., 2024).

The energy demand during driving often varies depending on specific vehicle states, such as starting, accelerating, cruising, decelerating and idling. For example, starting and accelerating typically require the most energy, while effective braking techniques can facilitate energy recovery through regenerative systems. Drivers who frequently engage in hard braking generally achieve lower levels of energy regeneration compared to those with more controlled braking habits (J. Zhang et al., 2020). These personal driving patterns affect how energy is consumed and managed during trips. In Figure 2.3 shows, sample of 5 main profiles distribution with velocity changes.

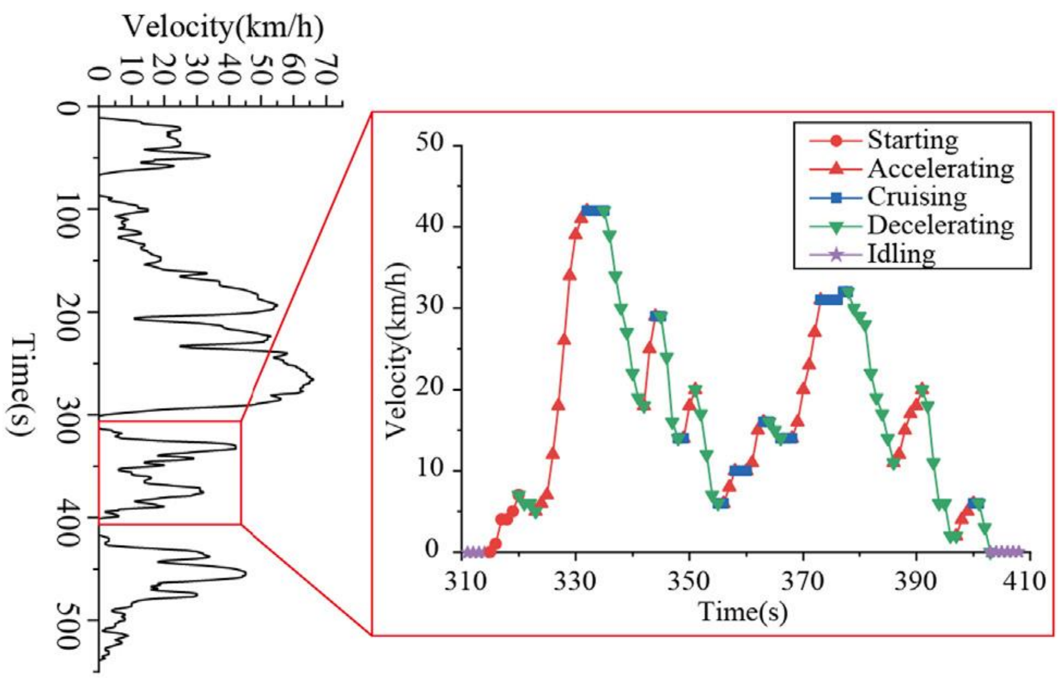

Figure 2.3 : Example of trip fragment, micro-fragments and the corresponding kinematic fragments (J. Zhang et al., 2020).

Quantifying and analyzing driving behavior often involves assessing parameters such as acceleration and deceleration rates, which can then be used to classify driving styles. Clustering methods categorize driving patterns into distinct groups, such as high or low acceleration and deceleration combinations, providing insights into the relationship between driving styles and energy consumption (J. Zhang et al., 2020).

2.2.2 Environmental conditions

Environmental factors play a pivotal role in shaping the energy consumption of EVs. Ambient temperature is particularly influential, as it directly affects battery performance and the operation of auxiliary systems. Energy consumption generally increases in colder temperatures, where battery efficiency declines and HVAC demands rise. Conversely, extremely high temperatures can also reduce range due to increased cooling requirements (Mediouni et al., 2022). Studies suggest that the optimal temperature range for minimizing energy consumption is between 15°C and 20°C (J. Zhang et al., 2020). Figure 2.4 illustrates the relationship between SOC and distance traveled under varying ambient temperature conditions. The distribution suggests that SOC depletes more rapidly in extreme temperatures, particularly at lower temperatures where energy consumption is higher due to reduced battery efficiency.

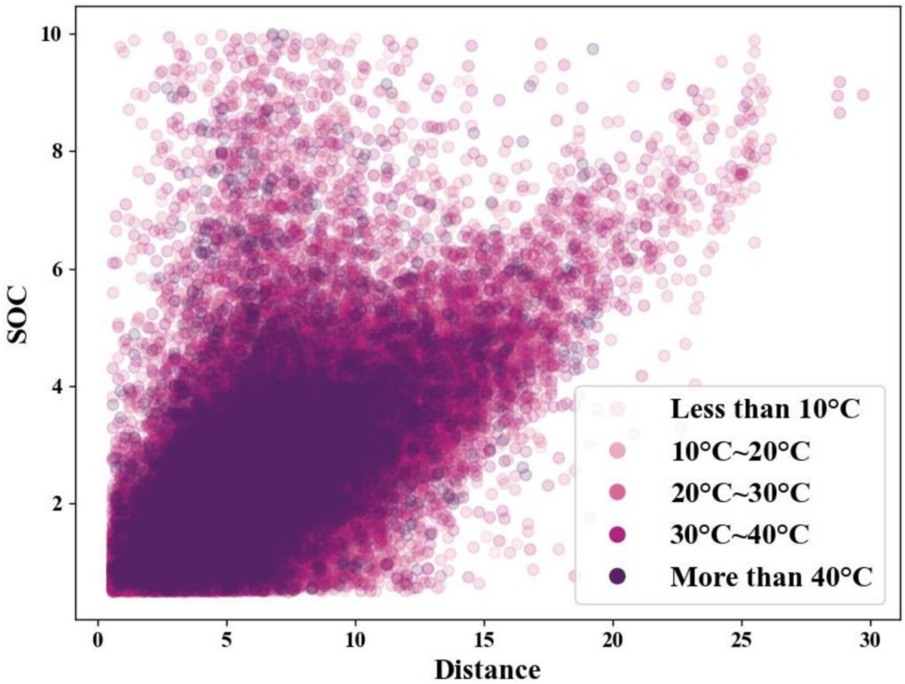

Figure 2.4 : Comparison of energy consumption per unit distance traveled in different temperature zones (Feng et al., 2024).

Road conditions also significantly affect energy efficiency. Uphill driving demands greater energy expenditure, whereas downhill driving offers opportunities for energy recovery through regenerative braking. Additionally, the road surface influences rolling resistance, with smoother surfaces requiring less energy (Mediouni et al., 2022; J. Wang, 2016; X. Xu et al., 2019).

Weather conditions, such as rain or snow, create additional challenges by altering road traction and driving dynamics, which can increase energy demand (Petkevicius et al., 2021). Traffic patterns, ranging from free flowing conditions to severe congestion, also have a cascading impact on energy use. These factors often interact in nonlinear ways, making it challenging to predict energy consumption accurately across varying environmental conditions.

2.3 Range Estimation Techniques

Range estimation techniques for EVs are critical for predicting energy consumption and providing accurate estimations of driving range. These techniques play a pivotal role in mitigating range anxiety, optimizing energy use and enhancing the overall driving experience. Range estimation methods can be broadly classified into three categories: data driven models, physical models and hybrid models. Each approach leverages unique methodologies to address the challenges associated with energy consumption prediction, with strengths and limitations that make them suitable for different applications.

2.3.1 Data driven models

Data driven models, often referred to as "black box" approaches, rely on ML and DL ,which is a subset of ML, techniques to predict energy consumption based on historical and real time data (Heinrich & Pruckner, 2022). These models learn complex patterns from data without requiring explicit knowledge of vehicle dynamics or physical principles.

Machine learning algorithms such as SVR, RF, Gaussian Process Regression and XGBoost have been widely applied in energy consumption estimation (Achariyaviriya et al., 2024). Advanced neural network architectures, including LSTM networks (Feng et al., 2024) and Convolutional Neural Networks (CNN) (Chen et al., 2023) are frequently used for modeling sequential data and predicting battery SOH or energy consumption (Wazirali et al., 2023).

Figure 2.5 outlines a process for estimating the SOH of a battery through data preprocessing, model training and evaluation. Raw battery data undergoes preprocessing, including voltage normalization and dataset structuring, before being split into training, validation and test sets. Machine learning models, including CNN, LSTM and a hybrid CNN-LSTM, are trained to estimate voltage and predict battery capacity, which are key indicators of SOH. The final stage involves evaluating model performance to ensure accurate and reliable SOH estimation, essential for effective Battery Management Systems (BMS) in EVs.

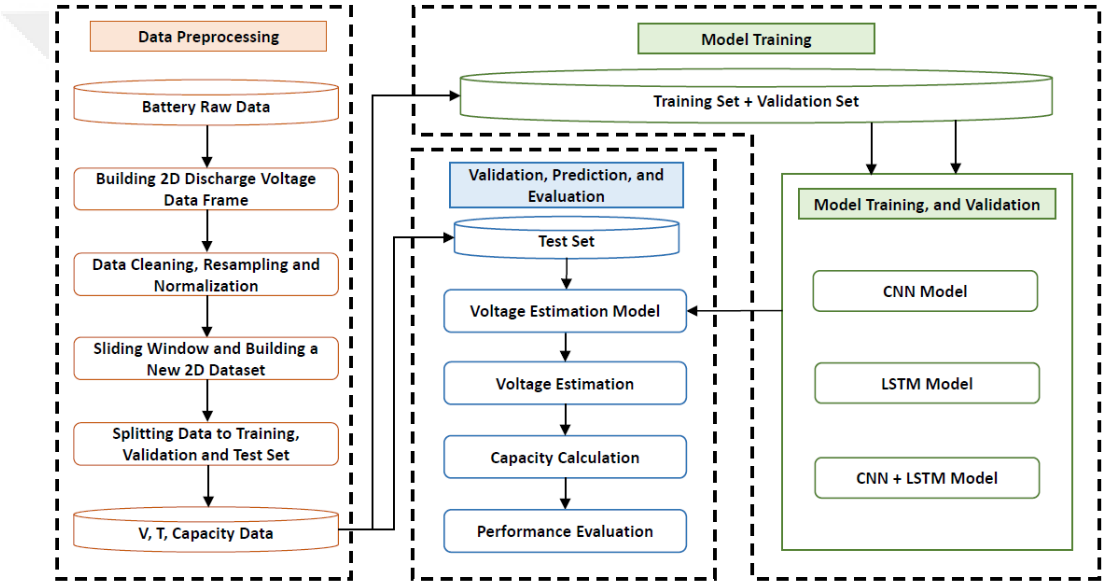

Figure 2.5 : Battery SOH prediction framework (Safavi et al., 2024).

Data driven models use a wide range of features, including driving conditions (e.g., speed, acceleration and road grade) (X. Xu et al., 2019), battery parameters (e.g., constant current charging time, charging capacity and voltage curves) (Chen et al., 2023) and environmental factors (e.g., ambient temperature, wind velocity and precipitation). By analyzing these variables, these models can provide accurate predictions of energy consumption under varying conditions.

Ensemble methods, such as Ensemble Stacked Generalization (ESG), further improve prediction accuracy by combining the outputs of multiple base models which shown in Figure 2.6.

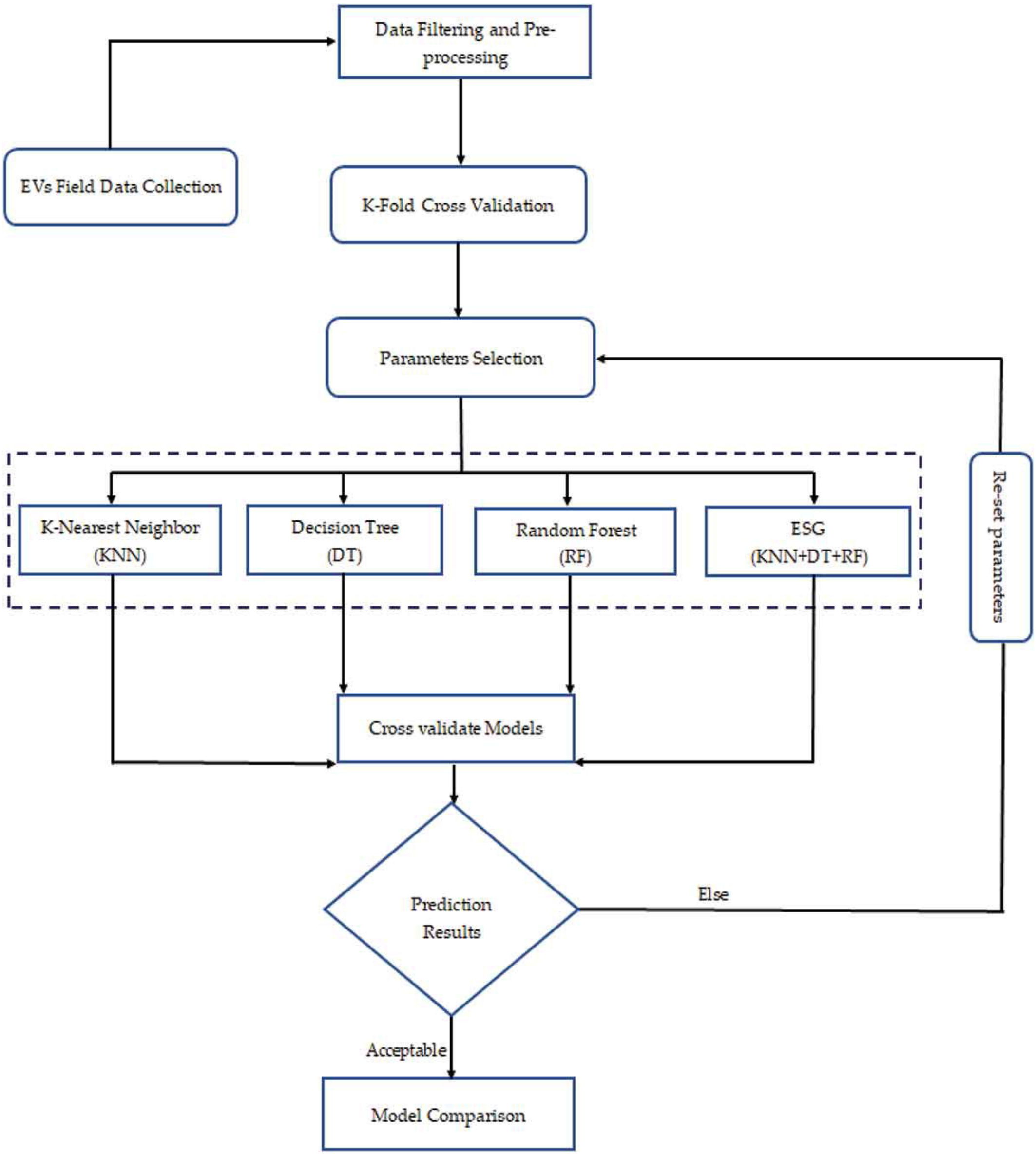

Figure 2.6 : Hybrid ESG method flowchart (Ullah et al., 2021).

Hybrid machine learning techniques, which integrate different ML methods or optimization algorithms, enhance forecasting accuracy by leveraging the strengths of individual techniques (Z. Zhang et al., 2024).

Federated Learning (FL) is a cutting edge approach for developing energy consumption models while preserving data privacy. In FL, local models are trained directly on data from individual EVs without sharing raw data externally (Yan et al., 2024). These models can be implemented in two structures: centralized and decentralized which are both shown in Figure 2.7 and Figure 2.8. In a centralized setup, each vehicle trains its model locally and sends updates to a central server, which aggregates the results to create a global model. Alternatively, in a decentralized structure, vehicles communicate directly with one another, collaboratively refining their models. Transfer learning further enhances energy consumption prediction by applying knowledge from well studied EV models to newer models with limited data, assuming the variables of both models share a similar distribution. Additionally, data decomposition techniques like Variational Mode Decomposition (VMD) are employed to break down complex time series data into high and low frequency components, enabling more accurate analysis and identification of underlying patterns that influence energy consumption (Cheng et al., 2023).

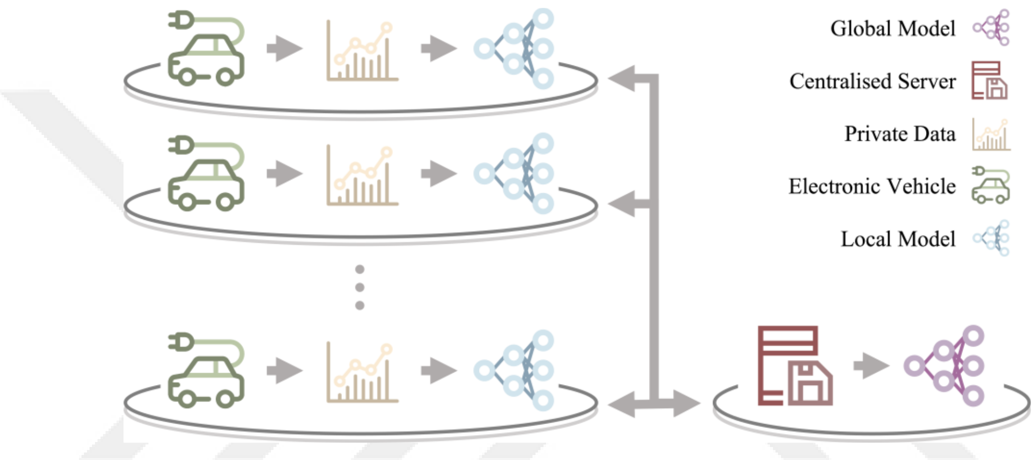

Figure 2.7 : Centralized FL architecture (Cheng et al., 2023).

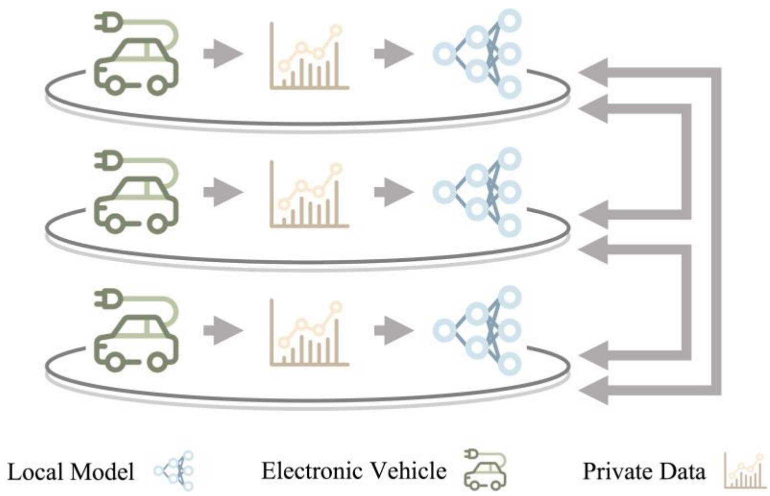

Figure 2.8 : Decentralized FL architecture (Cheng et al., 2023).

2.3.2 Physical models

Physical models, also known as "white box" or model based approaches, are grounded in the principles of physics and engineering. These models use mathematical equations to simulate the dynamics of EV energy consumption, incorporating factors such as rolling resistance, aerodynamic drag, gravitational forces and battery characteristics (Liu et al., 2021).

Physical models are advantageous for their strong theoretical foundations and interpretability. They offer a detailed understanding of the physical processes affecting energy consumption, making them particularly useful for diagnostic and optimization purposes (Shen, Zhou, Yu, et al., 2023). For instance, models based on longitudinal vehicle dynamics calculate the power required at the wheels by considering forces acting on the vehicle. Battery models, such as Lithium-Ion battery and electrochemical mechanism models, simulate energy usage and regenerative processes (Chen et al., 2023).

Despite their strengths, physical models face challenges in calibration. Accurate parameterization requires detailed knowledge of vehicle specific characteristics, which can be labor intensive to obtain. Additionally, physical models may require significant computational resources, especially for real time applications (Shen, Zhou, Yu, et al., 2023).

Examples of physical models include equivalent circuit models (ECM) for battery SOH estimation and mathematical formulations that describe battery power consumption under different conditions (Chen et al., 2023).

2.3.3 Hybrid models

Hybrid models combine the strengths of both data driven and physical approaches to create robust and versatile prediction frameworks. These models often integrate physical insights with the flexibility and adaptability of ML methods. Figure 2.9 illustrates a hybrid modeling approach that integrates both data driven and physical models for improved prediction accuracy. Historical data is used to train multiple models, each capturing different aspects of the system's behavior. These models are optimized through an optimizer to enhance their performance. The trained models are then combined using a model combiner, which integrates insights from multiple sources to generate refined predictions. Additionally, live data is incorporated to update and improve the model in real time, ensuring adaptability to changing conditions (Chou & Tran, 2018).

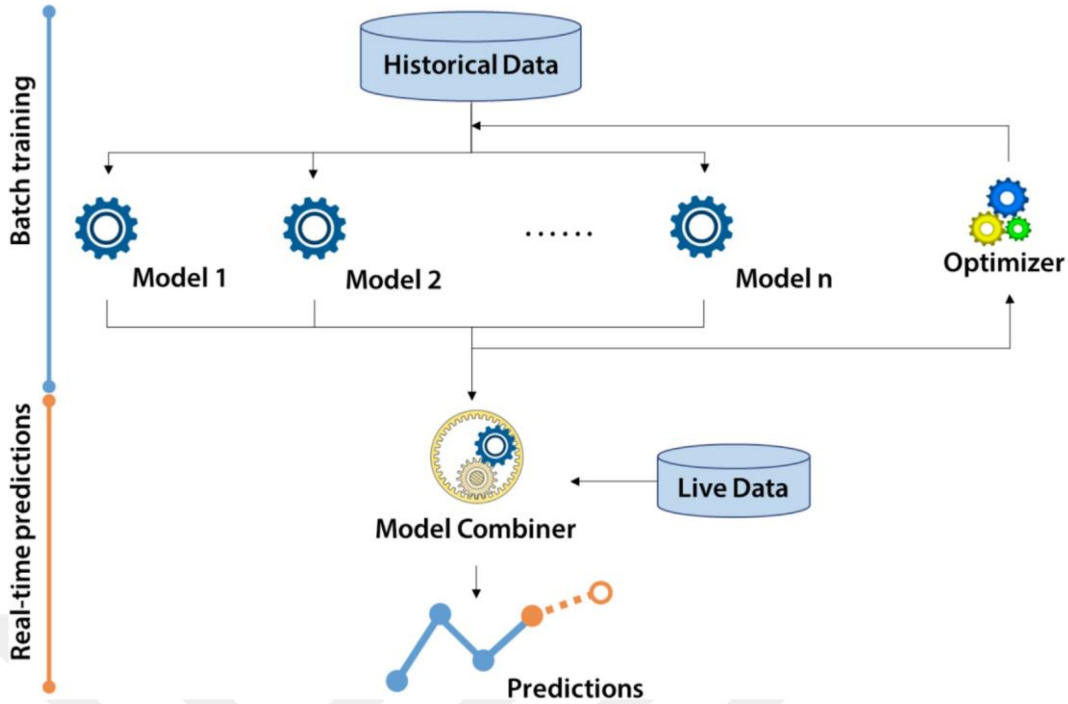

Figure 2.9 : Single phase hybrid model (Chou & Tran, 2018).

For instance, a hybrid model may use a physical model to simulate battery power during short trips and employ machine learning to predict cumulative trip level energy consumption. This approach leverages the interpretability of physical models while addressing their limitations through data driven refinement. Hybrid models also use machine learning to optimize the parameters of physical models, thereby improving their accuracy and adaptability (Zhu et al., 2024).

Some hybrid models incorporate optimization algorithms with machine learning or integrate time series analysis to capture temporal dependencies in driving data. Modal decomposition techniques, such as Empirical Mode Decomposition (EMD), are used to analyze complex signals, enhancing the robustness of hybrid approaches (Wazirali et al., 2023).

Hybrid models are particularly effective when data availability is limited or when the complexity of real world conditions necessitates combining multiple methodologies (Shen, Zhou, Ahn, et al., 2023). They strike a balance between computational efficiency and prediction accuracy, making them ideal for practical applications.

3. METHODOLOGY

3.1 Dataset Description

The Vehicle Energy Dataset (VED) is a comprehensive dataset designed to capture detailed insights into real world driving energy consumption patterns. It was developed through a collaborative effort involving the University of Michigan, Argonne National Lab and Idaho National Lab and spans data collected between November 2017 and November 2018. The dataset includes approximately 374,000 miles of driving data recorded across diverse road types, ranging from highways to dense urban environments, in Ann Arbor, Michigan. The dataset's large scale, covering 383 vehicles, provides a robust foundation for analyzing energy consumption trends in various driving and environmental conditions (Oh et al., 2019). Vehicle ID 10,455 and 541 are EVs in that dataset and also unique number of trips shown in Figure 3.1.

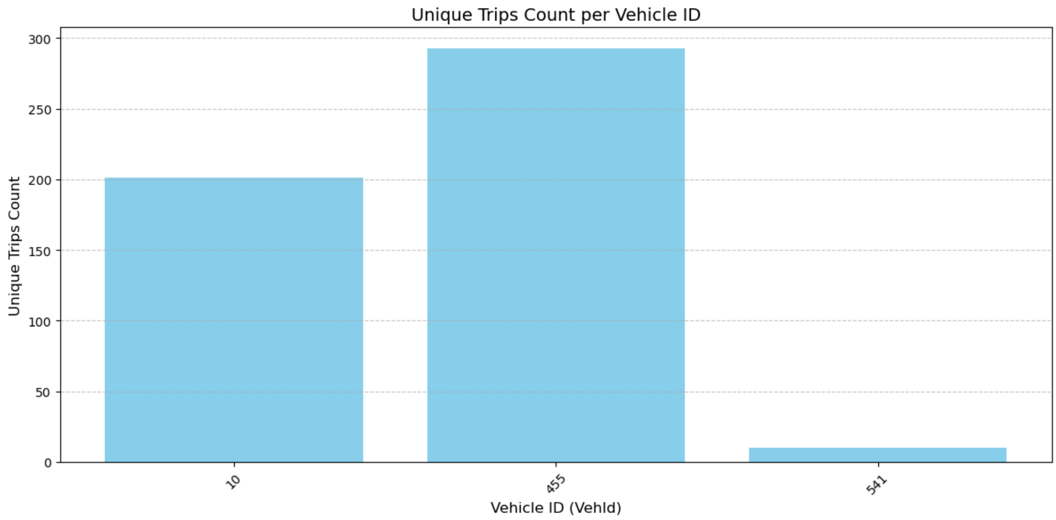

Figure 3.1 : VED unique trips count in 3 EVs.

VED contains data for four major categories of vehicles, ICEs, HEVs, PHEVs and EVs. The EV's in VED are identical 2013 Nissan Leaf which have 24 kWh battery. This diversity ensures a comprehensive representation of vehicle types, driving behaviors and energy consumption patterns. The dataset spans all seasons, capturing the influence of weather and environmental factors on vehicle energy consumption.

Data collection was conducted using onboard OBD-II loggers, which recorded time series data of various vehicle parameters and operational metrics. These features enable researchers to explore the interactions between vehicle dynamics, driver behavior and environmental conditions in shaping energy usage.

In addition to the original VED dataset, an extended version known as eVED was developed to enhance the dataset's utility (S. Zhang et al., 2022). eVED includes enriched data features such as road elevation, speed limits, intersections and traffic signal locations. This additional information allows for more precise analyses of the influence of road characteristics and driving conditions on energy consumption. Both the VED and eVED datasets are publicly available and have been widely adopted for energy consumption modeling and prediction research. Table 3.1 provides an overview of the time stamped dynamic data collected for energy consumption analysis in EVs.

Table 3.1 : Contents of time stamped dynamic data.

Data Name	Data Type	Sampling Time
GPS	Latitude / Longitude (deg)	3 sec
Outside Air Temperature (°C)	-	60 sec
Auxiliary Power (HVAC)	AirCon Power (kW)	60 sec
	Heater Power (W)	60 sec
Battery Info	Battery SOC (%)	60 sec
	Battery Voltage (V)	5 sec
	Battery Current (A)	1 sec

The dataset includes key parameters such as GPS coordinates, ambient temperature, HVAC and battery related metrics like SOC, voltage and current, each recorded at different sampling intervals. This structured data allows for a comprehensive analysis of how environmental conditions, vehicle performance and energy usage interact over time (Oh et al., 2019).

The VED dataset includes a rich array of features and parameters that reflect both static vehicle characteristics and dynamic operational metrics. Static data captures the inherent properties of each vehicle, including its make, model and year, as well as key specifications such as weight and engine configuration. These features provide a foundational understanding of how different vehicle types contribute to energy consumption under similar driving conditions. Figure 3.2 illustrates the unique vehicle speed distribution.

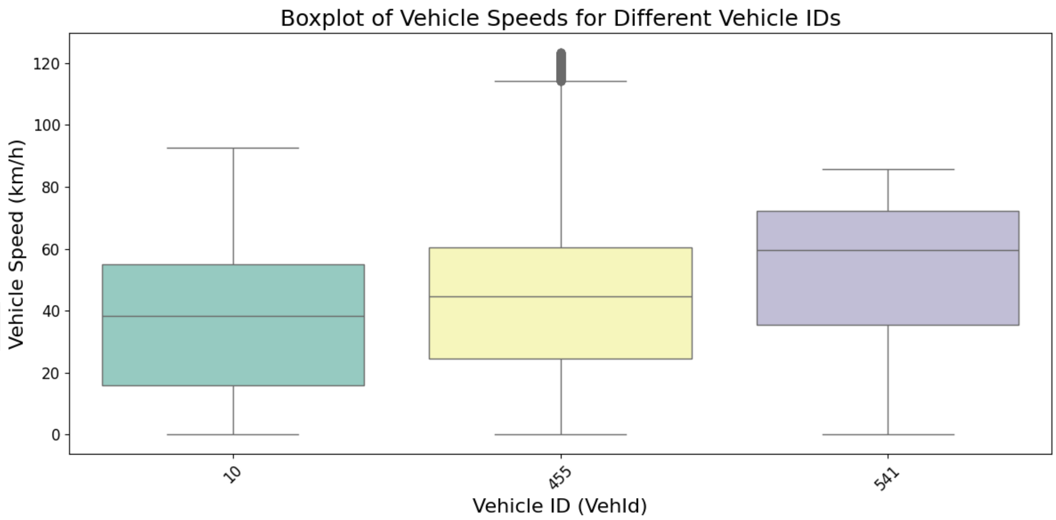

Figure 3.2 : Speed distribution of EVs in VED.

Dynamic data, consists of time series measurements that detail real time vehicle operations. Key parameters include GPS based location data, speed, fuel or energy consumption and auxiliary power usage, such as HVAC systems. Battery related metrics, including the SOC are particularly important for understanding the energy dynamics of EVs shown in Figure 3.3.

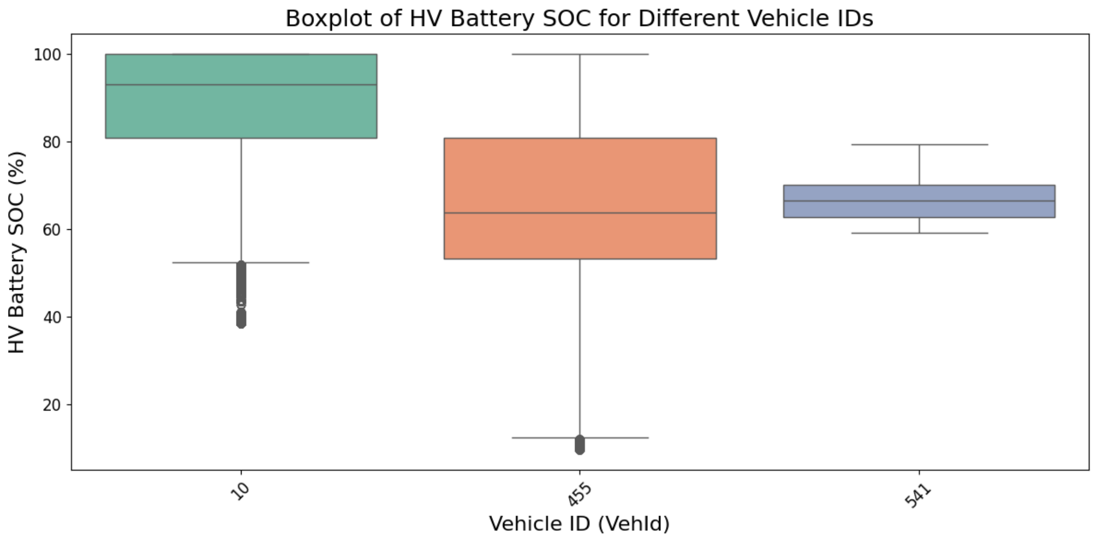

Figure 3.3 : Distribution of SOC on EVs in VED.

The features and parameters in the VED dataset enable the development of robust models for predicting energy consumption and analyzing driving patterns. By integrating diverse data sources and metrics, the dataset supports comprehensive research to improve the efficiency and range of electric vehicles.

3.2 Model Development

The model development process involved implementing three different machine learning approaches XGBoost, SVR and LSTM networks to predict the energy consumption of EV batteries. Each method was chosen for its unique advantages in modeling structured data, including gradient boosting, nonlinear regression and sequential data modeling.

3.2.1 Selection of algorithms

XGBoost was utilized as a gradient boosting algorithm to model the energy consumption using features such as vehicle speed, outside air temperature, air conditioning power, heater power, battery current and battery voltage. The dataset was split into training and testing subsets and the model was trained using the "XGBRegressor" implementation. To optimize the hyperparameters, GridSearchCV was employed, exploring combinations of the number of estimators, maximum tree depth, learning rate, subsample fraction and column sampling fraction. The model was evaluated on the test set and performance metrics, including RMSE and R^2, were computed to assess prediction accuracy. The gradient boosting approach demonstrated its ability to model nonlinear relationships and interactions between features, yielding strong predictive performance. Table 3.2 shows the best parameters of XGBoost Regressor.

Table 3.2 : XGBRegressor the best parameters.

Model	subsample	n estimators	min child weight	max depth	learning rate	gamma	colsample bytree	reg lambda
XGBoost	0.8	500	5	6	0.01	0	1	1

The second approach involved SVR, a nonlinear regression method particularly suited for capturing complex relationships in smaller datasets. A pipeline was created with a "StandardScaler" to normalize feature distributions and the SVR model with a radial basis function (RBF) kernel. Hyperparameters, including the regularization parameter

(C), epsilon-insensitive loss (ε) and kernel coefficient (γ) were optimized using GridSearchCV with three-fold cross validation. This ensured the model was fine tuned to balance bias and variance. Once the model was trained, predictions were made on the test set and metrics such as RMSE and R^2 were calculated. The SVR model showed its efficacy in handling nonlinear patterns in the dataset, providing competitive accuracy compared to XGBoost. The best parameters of SVR shown in Table 3.3.

Table 3.3 : SVR best parameters.

Model	gamma	epsilon	C
SVR	scale	0.1	1

The third approach employed a LSTM network to capture temporal dependencies in the energy consumption data. Unlike XGBoost and SVR, LSTM leverages the sequential nature of the data by using historical energy consumption values to predict future values. The target variable, energy consumption, was scaled to a range of [0, 1] using MinMaxScaler to ensure stability during training. Sequential data structures were constructed by creating overlapping windows of fixed length, with each sequence containing 30 past energy consumption values as input and the next energy consumption value as the target. The LSTM model architecture consisted of two stacked LSTM layers. The first LSTM layer returned sequences, allowing the second layer to process the temporal representation further. The output of the LSTM layers was passed through a dense layer, the final layer producing a single energy consumption prediction. The model was compiled using the Adam optimizer and the loss function was MSE. Early stopping was employed to halt training if the loss did not improve for 20 consecutive epochs, preventing overfitting. The model was evaluated on the test set and predictions were inverse transformed to their original scale for performance metrics computation. Table 3.4 shows the layers and parameters of LSTM.

Table 3.4 : LSTM best parameters.

Model	First Layer		Second Layer		Dense Layer		Optimizer
	Units	Dropout	Units	Dropout	Neurons	Activation	
LSTM	64	0.2	32	0.2	16	ReLu	Adam

Each model was evaluated using the same set of performance metrics to ensure consistency in comparison. The R^2 score quantified the proportion of variance explained by the model, while RMSE, MAE and MSE provided insights into the error magnitude. XGBoost excelled in handling nonlinear interactions and feature importance, SVR demonstrated robustness with smaller datasets and LSTM effectively captured temporal patterns in sequential data. This comprehensive approach enabled a robust comparison of different modeling techniques, paving the way for an integrated predictive system for battery energy consumption estimation.

3.2.2 Input-output mapping

The input-output mapping for predicting energy consumption involved defining the relationship between operational features (inputs) and the target variable (output). The chosen features were selected based on their relevance to the battery's state and their potential to influence energy consumption. These inputs and outputs were consistently structured for all three algorithms to ensure comparability and consistency in modeling.

Input Variables:

- Vehicle Speed [km/h]: Represents the driving behavior and energy consumption rate.
- Outside Air Temperature (OAT) [°C]: Influences battery performance, as temperature fluctuations affect energy efficiency and battery health.
- Air Conditioning Power [kWatt]: Reflects the additional load on the battery due to climate control systems.
- Heater Power [Watt]: Represents another auxiliary load affecting energy consumption.
- HV Battery Current [A]: Directly indicates the flow of electrical current from or to the battery.
- HV Battery Voltage [V]: Provides insights into the battery's charging state and health.

Output Variable:

- Energy Consumption (Wh): Represents the total electrical energy used by the vehicle, influenced by driving behavior, environmental conditions and auxiliary power demands.

For LSTM, energy consumption was normalized to the range [0, 1] using MinMaxScaler to improve numerical stability and convergence during training.

For XGBoost and SVR, the input-output mapping involved a direct regression task where the six input features were fed into the model and the predicted energy consumption was compared against the actual energy consumption for error minimization. The mapping followed a tabular structure where each row represented one observation, with the six features as inputs and the output.

The input-output mapping for LSTM was framed as a time series problem to leverage the sequential nature of energy consumption data. The model was provided with sequences of 30 consecutive values as inputs, representing historical patterns and tasked to predict the energy consumption value at the next time step. This mapping allowed LSTM to capture temporal dependencies, which were not explicitly modeled by XGBoost or SVR.

By establishing a consistent input-output mapping across the three algorithms, the predictive frameworks could be compared on equal footing, offering insights into their respective abilities to capture the relationship between operational parameters and the energy consumption of EVs.

3.2.3 Training and validation

The training and validation process was designed to optimize the predictive performance of the models to achieve this, the dataset was split into training and testing subsets and appropriate validation strategies were employed for each algorithm.

The dataset was divided into training and testing sets, with 80% of the data allocated for training and 20% reserved for testing. This split ensured that the models could learn the underlying patterns from the training data while being evaluated on unseen data during testing. For the LSTM model, overlapping sequences were generated from the training set to create time series inputs for the model, where each sequence comprised

30 historical energy consumption values as inputs and the next energy consumption value as the target.

XGBoost was trained using the XGBRegressor implementation. GridSearchCV was utilized for hyperparameter optimization, exploring combinations of the number of estimators, maximum tree depth, learning rate, subsampling ratio and column sampling fraction. The model was trained to minimize the MSE on the training data. Cross validation with three folds was used to evaluate the model during training, ensuring robust performance across different subsets of the training data.

For SVR, the input features were first normalized using StandardScaler to ensure consistent scaling. The model was encapsulated in a pipeline with the RBF kernel, which is well suited for capturing nonlinear relationships. GridSearchCV was employed to optimize the hyperparameters, including the regularization parameter (C), epsilon-insensitive loss (ε) and kernel coefficient (γ). Three fold cross validation was conducted during the training phase to identify the best combination of hyperparameters, reducing the risk of overfitting.

The model architecture consisted of two stacked LSTM layers, followed by dense layers to produce a single output. Energy consumption values were scaled to the range [0, 1] to improve training stability. The model was compiled with the Adam optimizer and mean squared error as the loss function. Early stopping was employed, monitoring the loss on the training set and halting training if no improvement was observed over ten consecutive epochs. This mechanism prevented overfitting and ensured efficient training.

All models were validated using their respective cross validation methods and subsequently tested on the 20% holdout test set. Predictions from each model were compared against the actual energy consumption values in the test set. For LSTM, predictions were inverse transformed to their original scale using the same scaler applied during preprocessing.

The combination of cross validation during training and evaluation on an independent test set ensured that the models were both optimized and generalizable. This systematic approach to training and validation provided a fair and comprehensive comparison of the performance of XGBoost, SVR and LSTM models in predicting energy consumption for electric vehicle batteries.

3.3 Performance Metrics for Model Evaluation

The evaluation of predictive models for estimating energy consumption requires a systematic assessment of their accuracy and generalization capabilities. To achieve this, a set of well established performance metrics was employed, including MSE and RMSE (Chicco et al., 2021). These metrics collectively provided a comprehensive understanding of the models' ability to replicate actual energy consumption values and minimize errors.

The R^2 is a statistical measure that explains the proportion of variance in the dependent variable that can be attributed to the independent variables. It quantifies the goodness of fit of the model, where an R^2 value closer to 1 indicates that the model can explain most of the variability in the target variable. For regression tasks such as energy consumption prediction, R^2 offers valuable insights into how well the model performs relative to a baseline mean model which shown in equation 3.1.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}\left(\overline{Y} - Y_i\right)^2} \tag{3.1}$$

- X_i represents the predicted values,
- Y_i represents the actual values,
- $\overline{Y}$ is the mean of the actual values,
- m is the total number of observations.

MAE measures the average magnitude of errors between predicted and actual values without considering their direction. It provides an intuitive understanding of the model's performance by directly reflecting the average deviation. A lower MAE signifies a model capable of consistently making accurate predictions. However, it does not penalize larger errors as strongly as squared-error metrics, making it less sensitive to outliers shown in equation 3.2.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \tag{3.2}$$

MSE extends the evaluation by calculating the average squared difference between predicted and actual values. By squaring the errors, MSE amplifies the impact of larger deviations, making it particularly useful when larger errors are more critical to address as shown in equation 3.3. However, the squared nature of the metric means its scale differs from the original data, which may reduce interpretability.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \tag{3.3}$$

RMSE, derived as the square root of MSE, addresses the scale issue by expressing errors in the same units as the target variable. RMSE is widely used as a benchmark metric for regression models because it offers a balanced assessment of error magnitude while maintaining sensitivity to larger deviations which can be seen in equation 3.4. A lower RMSE indicates a model capable of closely approximating actual target values with minimal deviations.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{3.4}$$

These metrics were applied consistently across XGBoost, SVR and LSTM models. Predictions generated by the models were compared against actual target values from the test set, ensuring a fair and transparent evaluation process. For the LSTM model, predictions were inverse transformed to their original scale before metric computation to maintain consistency. By leveraging these metrics, the performance of the three modeling approaches was rigorously assessed, enabling a robust comparison and identification of the most effective algorithm for energy consumption prediction.

4. RESULTS AND DISCUSSION

4.1 Model Evaluation

The evaluation of the predictive models SVR, XGBoost and LSTM was conducted to assess their performance in predicting energy consumption for electric vehicle batteries. Each model was trained and tested using the same dataset, ensuring a fair comparison. The metrics employed for evaluation included the R^2, MAE, MSE and RMSE. These metrics collectively provided insights into the accuracy, precision and reliability of the models' predictions which shown in Table 4.1.

Table 4.1 : Comparison of energy consumption prediction metrics for SVR, XGBoost and LSTM.

Model	R^2	MAE	MSE	RMSE
SVR	0.5149	0.7783	1.6009	1.2653
XGBoost	0.6038	0.8166	1.3796	1.1745
LSTM	0.6571	0.7008	1.1591	1.0766

For SVR, the model demonstrated a moderate predictive capability with an R^2 score of 0.5149, indicating the variance in energy consumption could be explained by the model. The MAE and RMSE values were 0.7783 and 1.2653, highlighting the model's ability to produce relatively consistent predictions with moderate error margins. However, its performance was surpassed by the other models in terms of both accuracy and error reduction.

XGBoost, a gradient boosting framework, achieved an R^2 score of 0.6038, signifying that it explained the variance in the target variable. Despite its higher R^2 compared to SVR, the MAE and RMSE values of 0.8166 and 1.1745 suggested that while the model captured the broader patterns in the data, it struggled with precise predictions, particularly for outliers or extreme values. Also, XGBoost learning curve illustrated in Figure 4.1. In this figure, training error and testing error values getting closer to each other with incrementation of training set size. That type of trends indicates a better generalized model.

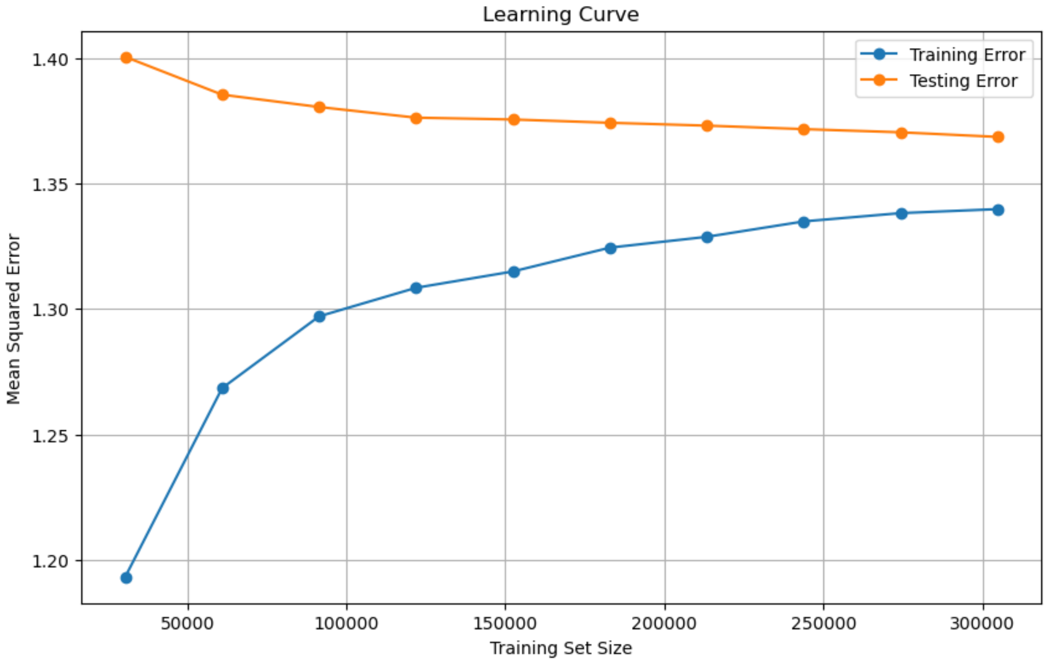

Figure 4.1 : XGBoost learning curve.

Figure 4.2 presents the structure of the decision tree model used in XGBoost for predicting energy consumption. The tree begins with the root node, which represents the most influential feature in this case, the EV battery current as it has the highest impact on energy consumption. As the model progresses downward, it makes recursive splits based on different feature values. Each decision node refines the dataset by directing data points along different branches depending on threshold values.

The tree structure illustrates how XGBoost captures complex, nonlinear relationships between input variables and energy consumption. At each split, the model selects the feature that maximally reduces error, ensuring that the most relevant attributes are prioritized. The process continues until the model reaches the leaf nodes, where final predictions are made. These leaf nodes contain energy consumption values, which are computed based on the statistical properties of the grouped data points.

This visualization highlights the interpretability of XGBoost's decision-making process, demonstrating how multiple factors interact to influence energy consumption. By leveraging such a structured approach, the model effectively identifies patterns and dependencies, enabling accurate energy consumption predictions for EVs in varying driving and environmental conditions.

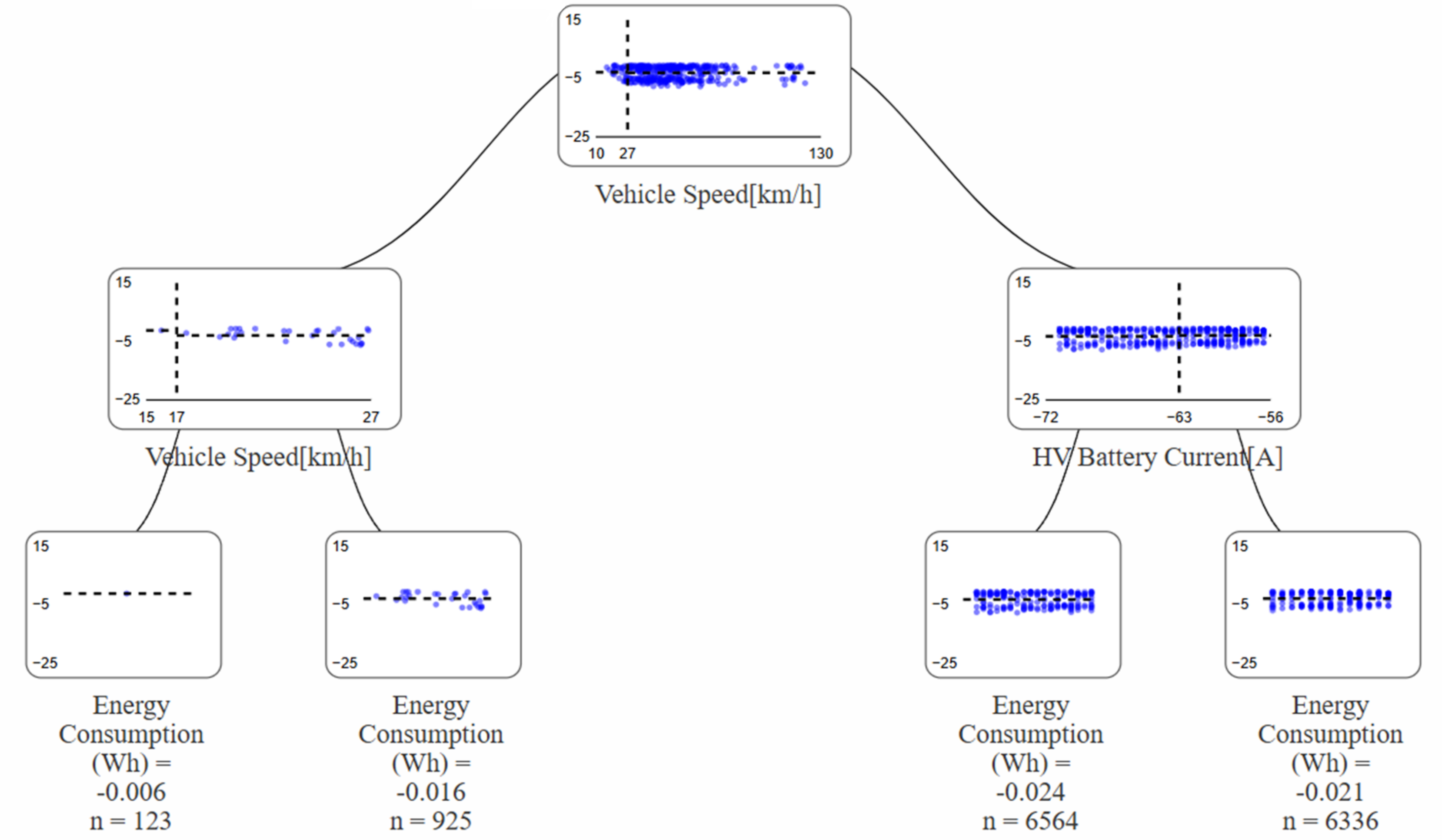

Figure 4.2 : XGBoost tree visualization.

The LSTM model outperformed the other two approaches, achieving an R^2 score of 0.6571, indicating better alignment between predicted and actual energy consumption values. The MAE and RMSE values of 0.7008 and 1.0766 reflected its superior ability to minimize both overall and individual errors. The MSE value of 1.1591 further confirmed the model’s improved accuracy and ability to generalize across the test dataset. Predicted values in LSTM model shown in Figure 4.3.

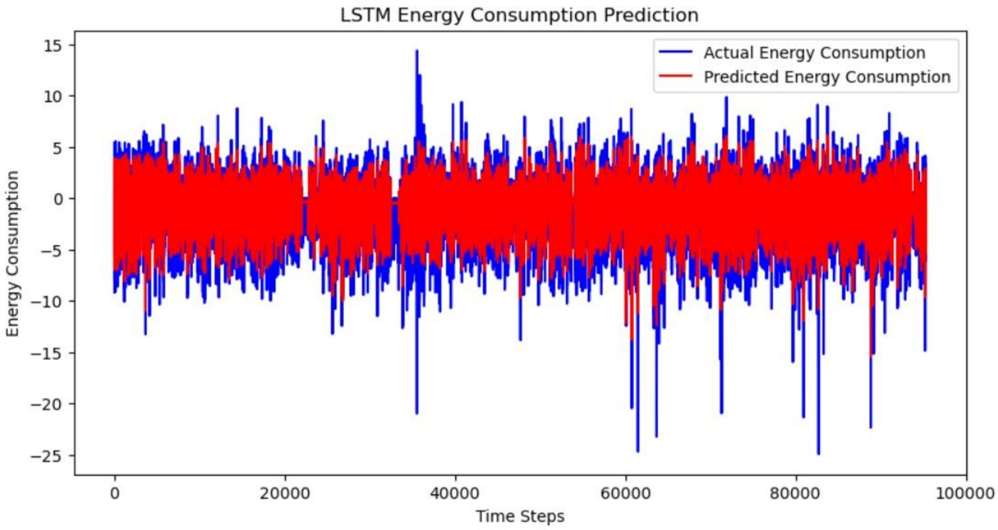

Figure 4.3 : LSTM energy consumption prediction graph.

4.2 Proposed Model

Based on the evaluation and comparative analysis, the LSTM model emerged as the proposed solution for energy consumption prediction in EV batteries. Its performance metrics confirmed its superiority in handling sequential data, minimizing errors and accurately predicting energy consumption values. The model's architecture and training process enabled it to adapt to the complexities of the dataset, making it highly reliable for real world applications.

4.3 Impact of Driving Conditions

Driving conditions, including variations in speed, external temperature and auxiliary power usage, were found to significantly influence energy consumption predictions. The LSTM model demonstrated resilience under varying driving conditions, effectively accounting for the temporal dependencies introduced by these factors. In contrast, SVR and XGBoost exhibited limitations in capturing the dynamic nature of energy consumption under fluctuating driving conditions, leading to higher error rates.

The analysis underscores the importance of considering driving conditions in energy consumption prediction models. The LSTM model's ability to integrate sequential data with external factors makes it an ideal candidate for predictive applications in battery management systems, particularly for electric vehicles operating under diverse and challenging conditions.

Figure 4.4 provides insights into the relationships between key factors influencing EV performance. It highlights how variables such as vehicle speed, external temperature, air conditioning and heater power, battery current and voltage interact with one another. Notably, environmental factors like outside air temperature show a strong inverse correlation with heater power, indicating increased energy consumption in colder conditions.

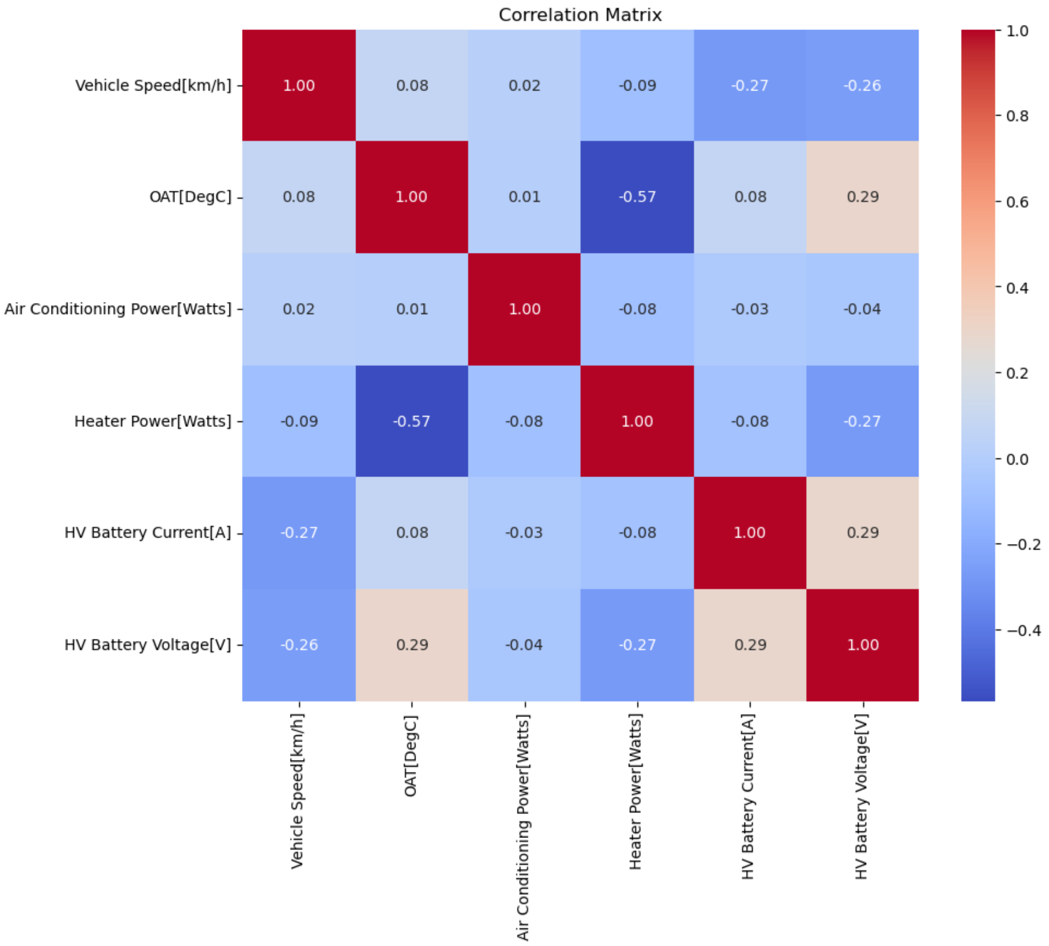

Figure 4.4 : Pearson correlation matrix.

Additionally, battery current and voltage exhibit a moderate positive correlation, reflecting their interdependence in power management. Other relationships, such as the weak correlation between vehicle speed and auxiliary power consumption, suggest that energy usage is influenced by multiple dynamic factors rather than a single variable. This analysis underscores the complexity of EV energy consumption and the importance of comprehensive models for accurate predictions and efficient energy management.

5. CONCLUSION

This study makes significant contributions to the field of EV energy consumption prediction, focusing on machine learning based models for accurate and efficient battery management and range estimation.

One of the major contributions of this research is the comparative analysis of three distinct machine learning models SVR, XGBoost and LSTM for EV energy consumption prediction. The study highlights the strengths and limitations of each model.

LSTM emerged as the most effective model, demonstrating superior accuracy in capturing temporal dependencies and sequential energy consumption patterns.

XGBoost showed strong feature learning capabilities, particularly in handling complex relationships between variables, though it lacked the ability to process sequential data efficiently.

SVR exhibited the lowest performance, proving less effective in dealing with the high dimensional, nonlinear nature of EV energy consumption data.

This study successfully developed an LSTM-based prediction model that achieves high accuracy in estimating EV energy consumption, outperforming traditional statistical and ML approaches. A detailed analysis of real world EV energy consumption, ensuring that the models were trained and validated on realistic, practical datasets rather than simulated data with VED.

Energy consumption calculations based on battery current and voltage, providing an accurate estimation methodology that aligns with actual EV operation. Robust validation through cross validation techniques, ensuring the model’s generalizability across different driving conditions. The use of real world data enhances the practical applicability of the research and ensures that the developed models can be effectively deployed in commercial EV systems.

The study also contributes to the advancement of predictive maintenance techniques for EV batteries. Identify patterns in battery degradation and energy consumption, allowing early detection of performance issues. Support proactive maintenance planning, reducing unexpected failures and increasing battery lifespan. Enhance vehicle reliability, making EVs more attractive to consumers and fleet operators. This approach improves cost efficiency and sustainability by minimizing battery waste and ensuring optimal energy utilization.

5.1 Practical Application of This Study

This study demonstrated the effectiveness of XGBoost, SVR and LSTM models in predicting the energy consumption of EVs using real world driving data. A comparative analysis of these models revealed that the LSTM model outperformed the others. LSTM ability to leverage sequential dependencies and capture temporal patterns in energy consumption. The XGBoost model, while effective at detecting complex relationships, lacked the sequential processing capability required for time dependent predictions. Similarly, SVR struggled with high dimensional, non-linear relationships, making it the least effective model for energy consumption forecasting.

The practical application of this research lies in its potential integration into BMS for real time energy prediction, route optimization and efficiency improvements. Accurate forecasting of energy consumption is crucial for reducing range anxiety, enhancing battery lifespan and optimizing EV performance. By implementing the LSTM model into an onboard BMS, EVs can dynamically adjust their energy management strategies in response to changing driving conditions, ensuring optimal battery utilization and more reliable range estimations.

A key achievement of this study was the successful implementation and validation of ML based models using the VED. The dataset provided a rich source of real world driving behavior, capturing variables such as speed, battery current, voltage, temperature and auxiliary power consumption. This allowed for the development of robust, data driven models that can be directly applied in real world EV applications.

5.2 Future Directions in Research

This study has made significant progress in EV energy consumption prediction, there are still several opportunities for future improvements that can further refine predictive accuracy and real world applicability.

One major direction for future research is dataset enrichment, where additional environmental and contextual variables, such as road elevation, real time traffic congestion, road surface conditions and weather variability, could be integrated into the models. For example, elevation data could enhance accuracy in hilly terrain, while real time traffic flow data could improve route planning and adaptive energy estimation.

Moreover, incorporating Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) communication could enable real time, dynamic updates in energy predictions. By allowing EVs to share anonymized data on driving conditions and energy consumption, ML models could continuously adapt and refine their predictions based on changing external factors. FL could also play a crucial role in privacy preserving model improvements, where EVs collaboratively update their models without exchanging sensitive raw driving data.

Another important research avenue is the development of hybrid modeling approaches that combine physics based models with ML techniques. While ML models provide superiority at pattern recognition, they lack interpretability, making their deployment in safety critical applications challenging. A hybrid approach ,where ML models are guided by first principles physics based models, could improve both accuracy and explainability.

Additionally, real time adaptive learning is an area of growing interest. Most energy consumption models are trained on prerecorded datasets, which may not account for sudden changes in driver behavior or environmental conditions. Implementing online learning techniques that update model parameters in real time, based on incoming sensor data, could significantly improve prediction reliability.

Another challenge in EV energy prediction is generalizability across different vehicle models. Currently, most models are trained on specific vehicle types, making them less adaptable to new EV models. Future research should explore transfer learning techniques, where a model trained on a widely used EV dataset could be fine tuned for new models with minimal additional data collection.

5.3 Implications for Electric Vehicle Technology

The findings of this study have profound implications for the evolution of electric vehicle technology. Accurate energy consumption prediction models are pivotal for improving battery efficiency, extending vehicle range and minimizing energy wastage. By integrating these models into advanced BMS, electric vehicles can dynamically adapt to real time conditions, optimizing energy consumption based on driving patterns and auxiliary loads.

Furthermore, the study highlights the potential for machine learning to enhance user experience in electric vehicles. With precise energy consumption predictions, drivers can make informed decisions regarding trip planning, reducing range anxiety and promoting confidence in electric vehicle adoption. This technology could also support predictive maintenance, identifying potential battery degradation before it impacts vehicle performance.

The study underscores the need for continued research and collaboration between academia, industry and policymakers to unlock the full potential of machine learning in revolutionizing electric vehicle technology.

REFERENCES

Achariyaviriya, W., Wongsapai, W., Janpoom, K., Katongtung, T., Mona, Y., Tippayawong, N., & Suttakul, P. (2023). Estimating Energy Consumption of Battery Electric Vehicles Using Vehicle Sensor Data and Machine Learning Approaches. *Energies*, *16*(17), 6351. https://doi.org/10.3390/en16176351

Achariyaviriya, W., Wongsapai, W., Rinchumphu, D., Tippayawong, N., Tippayawong, K. Y., & Suttakul, P. (2024). A comparative study of vehicle powertrain efficiency: Data-driven analyzing energy consumption and environmental impact. *Transportation Engineering*, *18*, 100286. https://doi.org/10.1016/j.treng.2024.100286

Adedeji, B. P. (2023). Electric vehicles survey and a multifunctional artificial neural network for predicting energy consumption in all-electric vehicles. *Results in Engineering*, *19*, 101283. https://doi.org/10.1016/j.rineng.2023.101283

Adnane, M., Khoumsi, A., & Trovão, J. P. F. (2023). Efficient Management of Energy Consumption of Electric Vehicles Using Machine Learning—A Systematic and Comprehensive Survey. *Energies*, *16*(13), 4897. https://doi.org/10.3390/en16134897

Albuquerque, D. A. S. G. (2022). *Electric Vehicle X Driving Range Prediction – EV X DRP* [MSc Thesis]. INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. (2018). Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †. *Energies*, *11*(7), 1636. https://doi.org/10.3390/en11071636

Chen, S.-Z., Liang, Z., Yuan, H., Yang, L., Xu, F., & Fan, Y. (2023). A novel state of health estimation method for lithium-ion batteries based on constant-voltage charging partial data and convolutional neural network. *Energy*, *283*, 129103. https://doi.org/10.1016/j.energy.2023.129103

Cheng, N., Zheng, P., Ruan, X., & Zhu, Z. (2023). Electric vehicle charging load prediction based on variational mode decomposition and Prophet-LSTM. *Frontiers in Energy Research*, *11*, 1297849. https://doi.org/10.3389/fenrg.2023.1297849

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, *165*, 709–726. https://doi.org/10.1016/j.energy.2018.09.144

De Cauwer, C., Van Mierlo, J., & Coosemans, T. (2015). Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. *Energies*, *8*(8), 8573–8593. https://doi.org/10.3390/en8088573

Feng, Z., Zhang, J., Jiang, H., Yao, X., Qian, Y., & Zhang, H. (2024). Energy consumption prediction strategy for electric vehicle based on LSTM-transformer framework. *Energy*, *302*, 131780. https://doi.org/10.1016/j.energy.2024.131780

Gao, B., Wan, K., Chen, Q., Wang, Z., Li, R., Jiang, Y., Mei, R., Luo, Y., & Li, K. (2023). A Review and Outlook on Predictive Cruise Control of Vehicles and Typical Applications Under Cloud Control System. *Machine Intelligence Research*, *20*(5), 614–639. https://doi.org/10.1007/s11633-022-1395-3

Heinrich, F., & Pruckner, M. (2022). Virtual experiments for battery state of health estimation based on neural networks and in-vehicle data. *Journal of Energy Storage*, *48*, 103856. https://doi.org/10.1016/j.est.2021.103856

Holden, J., Reinicke, N., & Cappellucci, J. (2020). RouteE: A Vehicle Energy Consumption Prediction Engine. *SAE International Journal of Advances and Current Practices in Mobility*, *2*(5), 2760–2767. https://doi.org/10.4271/2020-01-0939

Jui, J. J., Ahmad, M. A., Molla, M. M. I., & Rashid, M. I. M. (2024). Optimal energy management strategies for hybrid electric vehicles: A recent survey of machine learning approaches. *Journal of Engineering Research*, *12*(3), 454–467. https://doi.org/10.1016/j.jer.2024.01.016

Kiasari, M., Ghaffari, M., & Aly, H. (2024). A Comprehensive Review of the Current Status of Smart Grid Technologies for Renewable Energies Integration and Future Trends: The Role of Machine Learning and Energy Storage Systems. *Energies*, *17*(16), 4128. https://doi.org/10.3390/en17164128

Krogh, B., Andersen, O., & Torp, K. (2015). Analyzing Electric Vehicle Energy Consumption Using Very Large Data Sets. In M. Renz, C. Shahabi, X. Zhou, & M. A. Cheema (Eds.), *Database Systems for Advanced Applications* (Vol. 9050, pp. 471–487). Springer International Publishing. https://doi.org/10.1007/978-3-319-18123-3_28

Liu, Y., Zhang, Q., Lyu, C., & 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. https://doi.org/10.1016/j.tra.2021.10.009

Mediouni, H., Ezzouhri, A., Charouh, Z., El Harouri, K., El Hani, S., & Ghogho, M. (2022). Energy Consumption Prediction and Analysis for Electric Vehicles: A Hybrid Approach. *Energies*, *15*(17), 6490. https://doi.org/10.3390/en15176490

Nabi, Md. N., Ray, B., Rashid, F., Al Hussam, W., & Muyeen, S. M. (2023). Parametric analysis and prediction of energy consumption of electric vehicles using machine learning. *Journal of Energy Storage*, *72*, 108226. https://doi.org/10.1016/j.est.2023.108226

Oh, G. S., Leblanc, D. J., & Peng, H. (2019). *Vehicle Energy Dataset (VED), A Large-scale Dataset for Vehicle Energy Consumption Research* (No. arXiv:1905.02081). arXiv. https://doi.org/10.48550/arXiv.1905.02081

Petkevicius, L., Saltenis, S., Civilis, A., & Torp, K. (2021). Probabilistic Deep Learning for Electric-Vehicle Energy-Use Prediction. *17th International Symposium on Spatial and Temporal Databases*, 85–95. https://doi.org/10.1145/3469830.3470915

Pokharel, S., Sah, P., & 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. https://doi.org/10.3390/wevj12030094

Ribeiro, A. M. N. C., Do Carmo, P. R. X., Rodrigues, I. R., Sadok, D., Lynn, T., & Endo, P. T. (2020). Short-Term Firm-Level Energy-Consumption Forecasting for Energy-Intensive Manufacturing: A Comparison of Machine Learning and Deep Learning Models. *Algorithms*, *13*(11), 274. https://doi.org/10.3390/a13110274

Safavi, V., Bazmohammadi, N., Vasquez, J. C., & Guerrero, J. M. (2024). Battery State-of-Health Estimation: A Step towards Battery Digital Twins. *Electronics*, *13*(3), 587. https://doi.org/10.3390/electronics13030587

Shen, H., Zhou, X., Ahn, H., Lamantia, M., Chen, P., & Wang, J. (2023). Personalized Velocity and Energy Prediction for Electric Vehicles With Road Features in Consideration. *IEEE Transactions on Transportation Electrification*, *9*(3), 3958–3969. https://doi.org/10.1109/TTE.2023.3241098

Shen, H., Zhou, X., Yu, A., Lamantia, M., Chen, P., & Wang, J. (2023). An Experimental Comparison of Physics-based and Machine-Learning-based Electric Vehicle Energy Consumption Estimation Methods. *IFAC-PapersOnLine*, *56*(2), 8672–8677. https://doi.org/10.1016/j.ifacol.2023.10.045

Shin, S.-Y., & Woo, H.-G. (2022). Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies*, *15*(13), 4880. https://doi.org/10.3390/en15134880

Skuza, A., & Jurecki, R. S. (2022). Analysis of factors affecting the energy consumption of an EV vehicle—A literature study. *IOP Conference Series: Materials Science and Engineering*, *1247*(1), 012001. https://doi.org/10.1088/1757-899X/1247/1/012001

Sulaiman, M. H., & Mustaffa, Z. (2024). State of charge estimation for electric vehicles using random forest. *Green Energy and Intelligent Transportation*, *3*(5), 100177. https://doi.org/10.1016/j.geits.2024.100177

Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R. E., & 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. https://doi.org/10.1177/0958305X211044998

Ullah, I., Liu, K., Yamamoto, T., Zahid, M., & Jamal, A. (2021). Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach. *International Journal of Green Energy*, *18*(9), 896–909. https://doi.org/10.1080/15435075.2021.1881902

Wang, J. (2016). *Battery electric vehicle energy consumption modelling, testing and prediction: A practical case study* [Phd Thesis]. Technische Universiteit Eindhoven.

Wang, Z., Wei, H., Xi, Y., & Xiao, G. (2024). Data-driven energy utilization for plug-in hybrid electric bus with driving patten application and battery health considerations. *Energy*, *310*, 133041. https://doi.org/10.1016/j.energy.2024.133041

Wazirali, R., Yaghoubi, E., Abujazar, M. S. S., Ahmad, R., & Vakili, A. H. (2023). State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. *Electric Power Systems Research*, *225*, 109792. https://doi.org/10.1016/j.epsr.2023.109792

Xu, X., Aziz, H. M. A., & Guensler, R. (2019). A modal-based approach for estimating electric vehicle energy consumption in transportation networks. *Transportation Research Part D: Transport and Environment*, *75*, 249–264. https://doi.org/10.1016/j.trd.2019.09.001

Xu, Z., Peng, Y., Li, G., Chen, A., & Liu, X. (2024). Range-constrained traffic assignment for electric vehicles under heterogeneous range anxiety. *Transportation Research* **Part** *C: Emerging Technologies*, *158*, 104419. https://doi.org/10.1016/j.trc.2023.104419

Yan, S., Fang, H., Li, J., Ward, T., O'Connor, N., & Liu, M. (2024). Privacy-Aware Energy Consumption Modeling of Connected Battery Electric Vehicles using Federated Learning. *IEEE* **Transactions** *on Transportation Electrification*, *10*(3), 6663–6675. https://doi.org/10.1109/TTE.2023.3343106

Yang, N., Song, Z., Hofmann, H., & Sun, J. (2022). Robust State of Health estimation of lithium-ion batteries using convolutional neural network and random forest. **Journal** *of Energy Storage*, *48*, 103857. https://doi.org/10.1016/j.est.2021.103857

Yuan, X., He, J., Li, Y., Liu, Y., Ma, Y., Bao, B., Gu, L., Li, L., Zhang, H., Jin, Y., & Sun, L. (2024). Data-driven evaluation of electric vehicle energy consumption for generalizing **standard** testing to real-world driving. *Patterns*, *5*(4), 100950. https://doi.org/10.1016/j.patter.2024.100950

Zhang, J., Wang, Z., Liu, P., & Zhang, Z. (2020). Energy consumption analysis and prediction of electric vehicles based on real-world driving data. *Applied Energy*, *275*, 115408. https://doi.org/10.1016/j.apenergy.2020.115408

Zhang, S., Fatih, D., Abdulqadir, F., Schwarz, T., & Ma, X. (2022). *Extended vehicle energy dataset (***eVED***): An enhanced large-scale dataset for* *deep learning on vehicle trip energy consumption* (No. arXiv:2203.08630). arXiv. https://doi.org/10.48550/arXiv.2203.08630

Zhang, X., Hou, J., Wang, Z., & Jiang, Y. (2022). Study of SOC Estimation by the Ampere-Hour Integral Method with Capacity Correction Based on LSTM. *Batteries*, *8*(10), 170. https://doi.org/10.3390/batteries8100170

Zhang, Z., Ye, B., Wang, S., & Ma, Y. (2024). Analysis and estimation of energy consumption of electric buses using real-world data. *Transportation Research Part D: Transport and Environment*, *126*, 104017. https://doi.org/10.1016/j.trd.2023.104017

Zhu, Q., Huang, Y., Lee, C. F., Liu, P., Zhang, J., & Wik, T. (2024). Predicting Electric Vehicle Energy Consumption from Field Data Using Machine Learning. *IEEE Transactions on Transportation Electrification*, 1–1. https://doi.org/10.1109/TTE.2024.3416532

CURRICULUM VITAE

Name Surname : Canberk Şen

EDUCATION :

- **B.Sc.** : 2021, İstanbul Technical University, Faculty of Mechanical Engineering, Department of Mechanical Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- Project Engineer, Arçelik Global, 2021-2022
- Project Engineer, Kordsa, 2022-Ongoing

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Şen C., Yurtseven M.B.** (2024, May 8-10). *Comparative Analysis of Predictive Models for Energy Consumption in Electric Vehicles* [Conference presentation abstract]. 3rd International Graduate Research Symposium, Istanbul, Turkey. https://polen.itu.edu.tr:8443/server/api/core/bitstreams/aa4a063b-5269-401f-80dc-59a6274a78e4/content