VEHICULAR EMISSION TRAJECTORIES IN SRI LANKA: A SCENARIO-BASED ANALYSIS

S. S. JAYAWARDHANA

248013V

Master of Science Dissertation

Department of Civil Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2025

VEHICULAR EMISSION TRAJECTORIES IN SRI LANKA: A SCENARIO-BASED ANALYSIS

Shasthri Sankha Jayawardhana

248013V

A dissertation submitted to the
Department of Civil Engineering, University of Moratuwa
in partial fulfilment of the requirements for the degree of Master of Science

Supervised by: Dr. Varuna Adikariwattage and Dr. Thusitha Sugathapala

Department of Civil Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2025

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Shasthri Sankha Jayawardhana Date:15.10.2025

The above candidate has carried out research for the Masters dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Dr. Varuna Adikariwattage Date: 16/10/2025

Dr. Thusitha Sugathapala Date: 24/10/2025

DEDICATION

To my supervisors, Dr. Varuna Adikariwattage and Dr. Thusitha Sugathapala, your collective expertise in transportation, emissions, and most importantly, in the discipline of research, has guided me not just through this thesis but also in shaping the way I approach inquiry and evidence.

To Prof. Saman Bandara and Dr. Loshaka Perera, for believing in my potential and selecting me as a Research Assistant for this impactful project under the Vehicle Emissions Test Trust Fund.

To the staff of the Transportation Engineering Laboratory, Mrs. Niranjala and Mrs. Melani, whose patience and assistance made the long hours of work possible.

To my close colleague and friend, Kushan Aponsu, my right hand in the project component of the grant, and to my roommates, without whom much of the implementation would not have been possible.

To my parents and my brother, whose faith in me never faltered, and to Sachinthi, for her unwavering love and support through every high and low. I dedicate this thesis to you.

ABSTRACT

This study presents a novel, data-driven framework for estimating and forecasting vehicular emissions in Sri Lanka, addressing a critical gap in disaggregated emissions modelling on a national scale. While current estimations rely on a top-down methodology, this research introduces a bottom-up approach utilising vehicle-level data from the national Vehicle Emissions Test (VET) programme, spanning 2016–2024. The study focuses on internal combustion engine (ICE) passenger cars, and processes raw VET data through cleaning, mileage estimation, and integration of vehicle technology/category classifications and emission factors as per EMEP/EEA Tier 02 guidelines. A Long Short-Term Memory (LSTM) neural network is developed to forecast average annual mileage using time-series data, including fleet composition, economic indicators, and disruption events. The model achieved a MAE of 121.81 km and an R-squared value of 0.85 during validation, showing strong predictive accuracy and a good fit. The trained model is used to predict two scenarios, with Scenario 01 being the continuation of the national vehicle import ban and Scenario 02 being the lifting of the ban, allowing for modernisation with Euro 6-compliant vehicles. Using the predicted annual spanning mileage, annual spanning emissions of CO, NMVOC, NO_x, N_2O, NH_3, and PM are estimated by multiplying by matched emission factors. Results indicate a divergence with Scenario 02 having higher total annual spanning emissions with an increase of 1.34% (CO), 3.68% (NMVOC), 1.62% (NO_x), 10.5% (N_2O), 6.33% (NH_3) and 6.92% (PM), at the end of the prediction horizon, reflecting the impact of increased vehicle activity despite cleaner technologies. This study demonstrates the potential of regulatory datasets for predictive modelling in low-resource environments and supports evidence-based air quality policy planning through a scalable methodology.

Keywords: Vehicular Emissions, Emission Factors, LSTM Prediction, Scenario Analysis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BPTT	Backpropagation Through Time
CEC	Constant Error Carousel
CO	Carbon Monoxide
COPERT	Computer Programme to Calculate Emissions for Road Transport
CO_2	Carbon Dioxide
DMT	Department of Motor Traffic
EMFAC	Emission Factors (Model)
EF	Emission Factor
EMEP/EEA	European Monitoring and Evaluation Program/ European Environment Agency
FFNN	Feed Forward Neural Network
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GRU	Gated Recurrent Unit
ICAP	Indian Clean Air Programme
ICE	Internal Combustion Engine
IPCC	Intergovernmental Panel on Climate Change
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MRV	Monitoring, Reporting and Verification
MSE	Mean Square Error
N_2O	Nitrous Oxide
NH_3	Ammonia
NMVOC	Non-Methane Volatile Organic Compounds
NO_x	Nitrogen Oxides
PEMS	Portable Emission Measurement System
PM	Particulate Matter
RNN	Recurrent Neural Network
RTRL	Real-Time Recurrent Learning
SVR	Support Vector Regression
UNFCCC	United Nations Framework Convention on Climate Change
VET	Vehicle Emissions Test
VKT	Vehicle Kilometres Traveled
VSP	Vehicle-Specific Power
WLTC	World Harmonised Light Duty Vehicle Test Cycle
YOM	Year of Manufacture

CHAPTER 1

INTRODUCTION

A growing concern for air pollution has emerged worldwide in recent decades, with the transportation sector as one of its largest contributors. Yet, a notable research gap exists, particularly in Sri Lanka where air pollutant emissions are not subjected to estimation on a granular scale. This thesis, titled "Vehicular Emission Trajectories in Sri Lanka: A Scenario-Based Analysis" aims to establish a data-driven approach for establishing a baseline and future prediction of vehicular emissions that better represents its actual contribution, aiding regulatory and policy decisions at a national scale.

Chapter 1 outlines the background and motivation of the study, highlighting the requirement for an advanced method of emission estimation in Sri Lanka and the potential of existing, underutilised data sources. It also includes the novelty and contribution of this study to research communities and regulatory organisations in the sector of emissions. The chapter concludes with the thesis structure, providing comprehensibility for the next chapters.

1.1 Background and Motivation

The active vehicle population has seen rapid growth across the world in the past decade due to population growth, urbanisation, and economic development, especially in developing nations [1]. This has also been the case in Sri Lanka, showing 2.5%-13.29% annual growth over the last decade up to 2022 [2], with a slight plateau in the last few years due to economic factors and import restrictions. As a result of the upward trend, vehicular emissions with carbon monoxide (CO), non-methane volatile organic compounds (NMVOC), nitrogen oxides (NO_x), nitrous oxide (N_2O), ammonia (NH_3), and particulate matter (PM) as main pollutants [3] have a significant contribution towards air pollution.

Sri Lanka's Carbon Net-Zero 2050 Roadmap and Strategic Plan [4] note that the transport sector accounts for 21% of total Green House Gas (GHG) emissions, with three-quarters of it coming from road transport emissions. Although this study focuses on air pollutants and not GHGs, this inadvertently showcases the lack of disaggregated measurement, as transport emissions are defined as a percentage of total emissions, on the basis of fuel sales.

1.1.1 Approaches to Estimating Transport Emissions

The bottom-up approach for transportation emissions builds emissions estimates from detailed data on individual travel activities, such as person-trips or freight movements. It considers each trip, including the type of vehicle, fuel efficiency, trip length, and load factors, to calculate emissions more precisely. This method requires disaggregated data inputs such as vehicle kilometres travelled (VKT), fuel consumption by vehicle type, and emission factors (EFs) specific to each vehicle and fuel combination. These models can range from highly detailed disaggregated emissions models, which use extensive local data and sophisticated travel demand models, to simpler spreadsheet tools that incorporate default values and limited local inputs. Three tiers of emission estimation have been defined for the bottom-up approach based on the level of disaggregation, which will be discussed in depth in Section 2.1. This approach allows for tracing the cause and effect of specific mitigation measures, such as improvements in vehicle fuel efficiency, modal shifts, or adoption of alternative fuels [5].

Bottom-up methodologies are often used ex-post to document actual emissions and reductions. They provide the necessary detail to evaluate the effectiveness of interventions and support monitoring, reporting and verification (MRV) systems [6]. Globally, regulatory bodies encourage the adoption of more disaggregated methodologies. However, Sri Lanka has yet to explore or fully incorporate these tiers, even though the necessary vehicle data infrastructure (such as the Vehicle Emissions Test database) is already in place.

1.1.2 Sri Lanka Vehicle Emissions Test Program

The Sri Lanka vehicle emissions test (VET) initiative was created and put into action on a national level. This currently applies to all nine provinces and went into effect in November 2008. The tests used are the snap acceleration test for diesel vehicles and the no-load idle and fast-idle tests for gasoline-powered vehicles. According to the regulations, to get the annual revenue license from the divisional secretariat office, all cars registered under the Department of Motor Traffic (DMT) must obtain an emission test certificate (with a pass rating) by paying the authorised testing price. Vehicles that fail are allowed to retest for free for a period of three months. With online data transfer to a central database, the VET programme was designed as a centralised, test-only system. Two commercial companies, Laugfs Eco Sri Pvt. Ltd. and CleanCO Pvt. Ltd., conduct test centres located throughout the island. These testing facilities are divided into three categories: mobile, semi-mobile, and fixed [7].

The purpose of this test is to assess the eligibility of a particular vehicle to be operated on local roads, considering the environmental impact. Aside from the test result, these databases contain vehicle details at the time of testing for verification or other purposes. The raw data available from the DMT contains vehicular emissions test results from the year range 2016-2024, with individual vehicle records containing vehicle class, make, model, year of manufacture, fuel type, and current mileage alongside volume per volume percentage (v/v%) values for CO and Hydrocarbons at two RPM levels in petrol vehicles, and smoke opacity level (k) in diesel vehicles to check if they comply with the government-mandated thresholds. It is a check for eligibility rather than a calculation of emissions of a particular vehicle over a period of time. Although investigation on the impacts of vehicular emissions is an objective of the VET programme, no further analysis is done with this collected data. The availability of this disaggregated data evokes the possibility of a bottom-up emission estimation at a national scale, aiding policy decisions for mitigation measures that are required to achieve national goals [8].

1.1.3 Vehicle Imports Ban and Related Uncertainties

In the last 5 years, the Sri Lankan government has utilised the temporary suspension of vehicle imports as a measure aimed at conserving foreign exchange reserves. A previous study in Sri Lanka highlights uncertainties that stem from the disruption of the automobile supply chain, which has led to reduced availability of new vehicles and a consequential surge in prices within the used vehicle market. This market contraction has not only altered consumer purchasing behaviour, shifting preferences toward domestic alternatives and aftermarket modifications, but also affected travel behaviour [9]. The government has lifted this temporary suspension via an extraordinary gazette notification effective from February 2025, stating concerns of reduced state income and market stabilisation [9], [10]. Yet an uncertainty persists with regard to the reinstatement of suspensions in the future. Therefore, the estimation of future emissions could fall under multiple scenarios, depending on the growth and proportions of the active vehicle fleet [11], leading to uncertainties in policy implementation. The bottom-up emission estimation required should be able to address these uncertainties.

1.1.4 Incorporation of Data-Driven Techniques for Vehicular Emissions Estimation and Prediction

There has been a significant shift in focus on data-driven approaches for the estimation and prediction of vehicle emissions in the recent research landscape. Traditional emission models, such as average speed models (COmputer Programme to calculate Emissions from Road Transport - COPERT, EMission FACtors - EMFAC) and other modal models are being supplemented by machine learning approaches that utilise real-world data from dynamometer and on-road tests [12]. Traditional models often lack the micro-level detail necessary for tailored decisions, primarily estimating total emissions, particularly when time-dependent data is present. Although higher computational power is required for these techniques, they are able to learn complex patterns and relationships from the available data, leading to more accurate and realistic emission estimations compared to traditional models that often rely on empirical assumptions and parameters that may not be locally applicable.

Artificial Neural Networks (ANN) have become a primary data-driven tool for emissions estimation due to their adaptability and high predictive accuracy [12], [13], [14]. Specific variations of these models are not only capable of time-series prediction [15], but also spatial-temporal mapping through combinations, resulting in improved performance [16], [17]. Beyond ANN, neuro-fuzzy models, genetic algorithms and geospatial information systems (GIS) have also shown successful applications in determining vehicular emission parameters [18]. The availability of a database of historical data from the VET program allows the incorporation of a data-driven approach towards the required bottom-up emission estimation with regard to a time-series prediction that addresses the related uncertainties, and will be discussed in detail under Section 2.2.

1.2 Research Problem

Sri Lanka lacks a disaggregated methodology for estimating road transport-related air pollutant emissions. Current national-level estimations are unable to reflect changes in fleet composition, technology shifts, or policy impacts. Despite operating a centralised VET programme since 2008, the potential of its dataset remains untapped for emissions modelling. Furthermore, no national studies have yet implemented Tier 02 or Tier 03 estimation frameworks, as recommended by global regulatory guidelines [3]. This impedes Sri Lanka's ability to evaluate policy scenarios, track emissions trajectories, and meet air quality goals

1.3 Research Scope

This research focuses on developing a Tier 02-based emissions estimation framework using historical data from the VET programme in Sri Lanka. The scope is limited to ICE passenger cars, which constitute a major share of the national fleet and showcase the widest variability in technology and vintage. It applies data pre-processing, feature engineering, and machine learning (LSTM) to forecast vehicle activity levels and estimate emissions of regulated air pollutants (CO, NMVOC, NO_x, N_2O, NH_3, and PM). Two forward-looking policy scenarios related to the import ban are simulated to evaluate the impact of regulatory change on national emission trajectories.The study does not include CO_2 emissions in the forecasting framework, as Tier 02 emissions estimation is unable to fully encapsulate its contribution. Unlike local pollutants, CO_2 emissions are not affected by after-treatment technologies and are instead directly proportional to the mass of fuel consumed. This would require integration of fuel consumption data on a disaggregated level, which is unavailable under the VET database.

1.4 Research Question and Objectives

The study follows these primary research questions:

- How can vehicle-level test data be used to perform disaggregated emissions estimation in Sri Lanka?
- How to forecast annual vehicle pollutant emissions under changing macroeconomic, technological and fleet conditions?
- What are the differences in emission trajectories under alternative policy futures (continued import ban vs. fleet modernisation)?

The objectives of the research study are as follows:

1. To estimate annual mileage per vehicle using odometer data, enabling disaggregated emissions calculation.
2. To assign vehicle EFs using engine capacity, power output, and year of manufacture.
3. To develop and validate a prediction model for forecasting average annual mileage per vehicle under variable fleet and macroeconomic conditions.
4. To simulate two forward looking policy scenarios (import ban) and estimate their impact on national emissions

1.5 Contribution

This thesis makes several key contributions to the fields of transport emissions analysis, environmental policy planning, and data-driven modelling, with a specific focus on the Sri Lankan context, where such integrated approaches have not been used. The contributions are outlined as follows,

1. **Revamped Methodology:** This research introduces a novel, data-driven bottom-up methodology to estimate and predict emissions in Sri Lanka. It incorporates individual vehicle-level data obtained from the VET program, enabling more granular and accurate emissions estimation that accounts for vehicle characteristics and other external factors, as opposed to the currently used top-down approach that is solely based on aggregate fuel sales.
   The study uses one vehicle class (Motor/Passenger Cars) as a case study, with the notion that the methodology is scalable to all other vehicle classes as required.

2. **Scenario-Based Emission Forecasting for Policy Insight:** A major contribution of this research is the scenario-based framework for forecasting future emissions under multiple assumptions about fleet growth, vehicle technology updates, and

import regulations. Modelling of multiple future scenarios provides policymakers with a decision-support tool that maps the emission impact of policy/regulatory decisions.

3. **Application of Underutilised National Datasets:** The research increases the analytical value of the VET database in Sri Lanka, transforming it from a regulatory compliance tool into a rich data set for emissions research and forecasting. This represents a shift in how existing government data infrastructure can be reused for national planning and environmental reporting, including alignment with global climate goals.

4. **Contribution to Local and Regional Knowledge:** With the focus being on the Sri Lankan vehicle fleet, the study provides rare insights into the limited body of research on transport emissions in South Asia. The models and methods allow scalability and transferability to other developing nations with similar data limitations and policy needs.

1.6 Thesis Structure

The thesis is structured under 7 chapters, each building upon the foundation laid in the introduction to develop a clear progression from background context to final conclusions.

Chapter 2: Presents a comprehensive review of existing literature on vehicular emissions estimation, with a focus on bottom-up methodologies, EFs, data-driven modelling approaches, and the importance and application of time-series forecasting for emission prediction. It also discusses previous work relevant to Sri Lanka through comparable developing contexts, identifying gaps that this research aims to address.

Chapter 3: introduces LSTM networks as a machine learning technique suitable for time-dependent data. It outlines the key concepts, mathematical formulations and architectural features of LSTM models, as well as their advantages over traditional statistical and neural network models in handling temporal dependencies in emission data.

Chapter 4: Describes the first section of the methodology, which details the data pre-processing, focusing on the preparation of the VET dataset for modelling. It includes steps such as data integration, feature engineering, and the rationale for selecting specific input variables for model training.

Chapter 5: Includes the second section of the methodology, which describes the development and training of the machine learning model using the pre-processed dataset. It covers the model architecture, hyperparameter tuning, validation and evaluation parameters.

Chapter 6: Presents the results of the selected model, evaluating its performance using appropriate statistical metrics. It includes scenario-based analyses to illustrate the model's applicability to real-world policy contexts under different future fleet compositions and regulatory conditions.

Chapter 7: Summarises the findings of the study, limitations and future research potential.

CHAPTER 2

LITERATURE REVIEW

This chapter includes a comprehensive review of the existing body of research relevant to vehicular emissions estimation and prediction, with a particular focus on methodologies applicable to the Sri Lankan context. It begins by examining the bottom-up approaches used in emissions estimation in terms of the level of disaggregation/tiers. It then explores possible EFs aligned with these tiers, which help localise the emissions to the region of the study based on vehicle characteristics and behaviour. The review then examines the evolution of emissions modelling from traditional deterministic models to data-driven and machine learning-based approaches. Special attention is given to studies that utilise disaggregated vehicle-level data, as well as the role of time-series prediction models in capturing temporal patterns in emissions. This chapter also identifies key research gaps, particularly related to Sri Lanka, where no prior attempts have been made to estimate emissions with an integrated approach.

2.1 Bottom-Up Emissions Estimation

As discussed in Chapter 1, there are three tiers of emission estimation, namely Tier 01, Tier 02, and Tier 03. According to the European Monitoring and Evaluation Programme/European Environment Agency (EMEP/EEA) 2024 guidelines [3], only Tier 01 is fuel-based, while the other two are distance-based with increased parameters for accuracy with disaggregated data. for higher tiers. A governing/regulatory body or a country may select a suitable tier for estimation based on the availability of parameters required for each tier. The parameters and estimation strategies defined under the EMEP/EEA 2024 guidelines are as follows,

2.1.1 Tiers of Emission Estimation (EMEP/EEA Guidelines)

Tier 01 uses the general equation shown by Equation (2.1)

$$E_i = \sum_j \left(\sum_m \left(FC_{j,m} \times EF_{i,j,m} \right) \right) \tag{2.1}$$

E_i = Emissions from pollutant *i (g)*
$FC_{j,m}$ = Fuel efficiency of vehicle category *j* using fuel *m (kg)*
$EF_{i,j,m}$ = Fuel efficiency-specific emission factor of pollutant *i* for vehicle category *j* and fuel *m (g/kg)*

The analysis takes several vehicle categories into account, including passenger cars, light commercial vehicles, heavy-duty vehicles, and those falling under the L-category. It also considers multiple fuel types, such as petrol, diesel, liquefied petroleum gas (LPG), and natural gas. A key aspect of the equation is that it requires disaggregating fuel consumption or sales data by vehicle type, which is a level of detail that is typically not provided in national datasets.

Under Tier 02 estimation, the number of vehicles and the annual mileage per euro vehicle legislation/technology (or the number of vehicle-km per technology) should be provided. These vehicle-km data are multiplied by the Tier 02 EFs as shown by Equations (2.2) and (2.3)

$$E_{i,j} = \sum_{k} (< M_{j,k} > \times EF_{i,j,k}) \tag{2.2}$$

or

$$E_{i,j} = \sum_{k} (N_{j,k} \times M_{j,k} \times EF_{i,j,k}) \tag{2.3}$$

$<M_{j,k}>$ = Total annual distance driven by all vehicles of category *j* and technology *k* *[veh-km]*

$EF_{i,j,k}$ = Technology-specific emission factor of pollutant *i* for vehicle category *j* and technology *k [g/veh-km]*

$M_{j,k}$ = Average annual distance driven per vehicle of category *j* and technology *k* *[km/veh]*

$N_{j,k}$ = Number of vehicles in the nation's fleet of category *j* and technology *k*

The same vehicle categories as Tier 01 apply, while vehicle technology (Pre-ECE to Euro 6) characteristics are also given within the guideline.

Tier 03 emission estimation introduces engine operating conditions, where the total emissions are a sum of hot emissions and emissions during transient thermal engine operation as shown in Equation (2.4)

$$E_{TOTAL} = E_{HOT} + E_{COLD} \tag{2.4}$$

E_{TOTAL} = Total emissions *(g)* of any pollutant for the spatial and temporal resolution of the application

E_{HOT} = Emissions *(g)* during stabilized (hot) engine operation

E_{COLD} = Emissions *(g)* during transient thermal engine operation (cold start)

Additionally, the driving condition is also incorporated, as different driving situations impose different engine operation conditions, and therefore a distinct emission performance as shown in Equation (2.5)

$$E_{TOTAL} = E_{URBAN} + E_{RURAL} + E_{HIGHWAY} \tag{2.5}$$

Total emissions are calculated by combining activity data for each vehicle category with appropriate EFs. The EFs vary according to the input data (driving situations, climatic conditions). The EFs here are dependent on the speed of the vehicles and fit into the equation using a frequency distribution. The proportion of hot and cold emissions are separated using a cold/hot emission quotient for a particular pollutant under a particular vehicle technology.

2.1.2 Selection of a Suitable Tier

The selection of a suitable tier depends on the data available for estimation. The guideline provides a decision tree outlining the reasoning behind each selection as illustrated using Figure 2.1. Since the VET database contains mileage values at the time of testing and vehicle vintage data for each vehicle [7], a Tier 02 estimation is achievable in Sri Lanka. This approach has been established as a practical method of transport emissions estimation in countries such as the Netherlands [19] with regard to road-transport, railways and other mobile machinery.

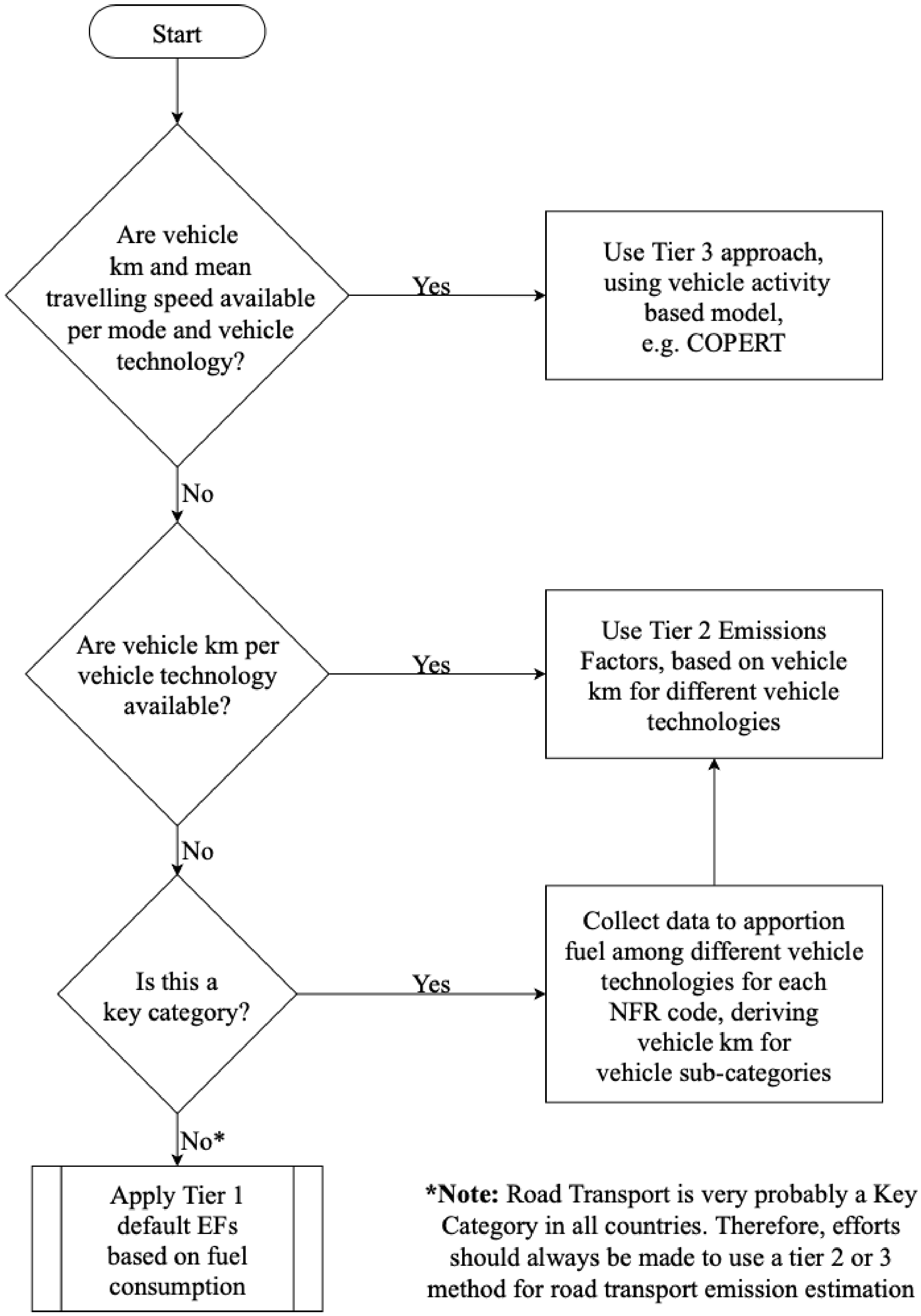


Figure 2.1: Decision tree for exhaust emissions from road transport. Adapted from [3]

2.1.3 Selection of Emission Factors

Vehicular EFs represent the average emission rate of a pollutant per unit of activity, such as per kilometre travelled or per unit of fuel consumed. Four main techniques of calculation and development of EFs can be identified [20] as follows,

1. **Chassis and Engine Dynamometer Testing:** Controlled laboratory methods where vehicles or engines are tested under standardised/localised driving cycles. Emissions are measured directly, and EFs can be developed by plotting aggregated emission data against parameters such as mean speed and acceleration/deceleration [21], [22]. Instantaneous emissions data can also be used to create detailed emission maps as functions of engine speed and power, allowing simulation of emissions for various driving patterns [23].

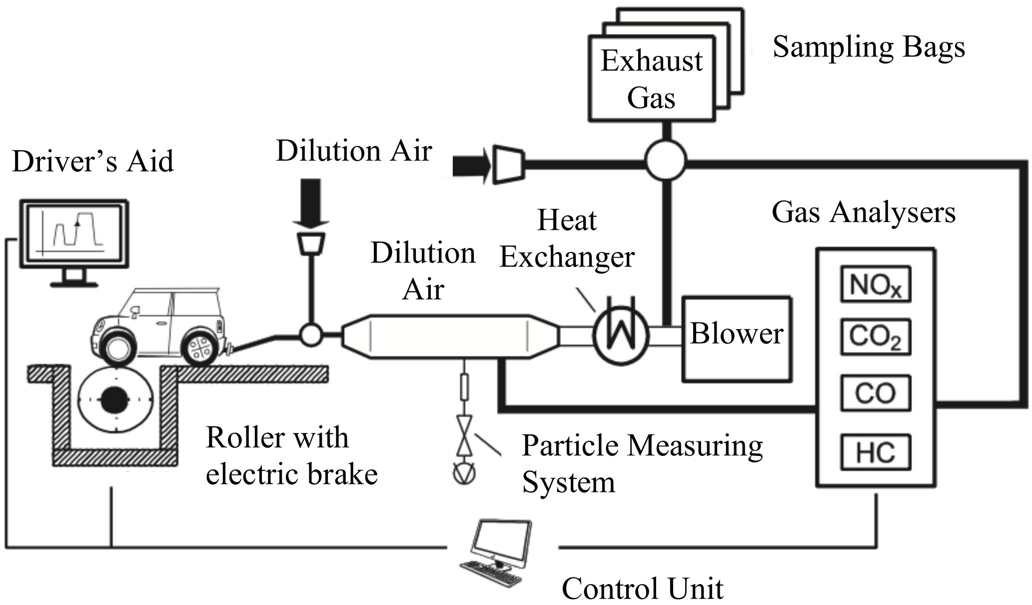


Figure 2.2: Schematic representation of a chassis dynamometer test facility. Adapted from [20]

2. **Remote Sensing:** Measures pollutant concentration ratios in vehicle exhaust plumes on the roadside. Since exhaust is immediately diluted, actual tailpipe concentrations are estimated by assuming turbulent dilution and stoichiometric combustion conditions [24]. EFs are derived using pollutant-to-CO_2 ratios and fuel carbon content, either on a tailpipe concentration basis or a fuel consumption basis. Remote sensing provides snapshots of emissions but has limitations in resolution and uncertainty due to factors like sampling site characteristics and driver behaviour.

3. **Road Tunnel Studies:** Measures pollutant concentrations at tunnel entrances and exits to calculate aggregate real-world EFs using mass balance techniques. Tunnel studies can provide distance or fuel-specific EFs and cover emissions from various

sources, including brake and tire wear. However, they mainly reflect steady-speed, free-flow traffic conditions and have difficulty distinguishing emissions by vehicle class [25]. Tunnel studies have also been used to assess seasonal and long-term trends of EFs [26].

4. **Portable Emission Measurement Systems (PEMS):** PEMS allow on-road, real-world emission measurements over a wide range of driving conditions. Although earlier systems had lower accuracy than laboratory methods, modern PEMS have improved significantly. EFs from PEMS data can be developed similarly to dynamometer data by binning emissions by mean speed or using instantaneous engine map models. Limitations include reduced repeatability and a narrower range of measurable pollutants [27].

While dynamometer tests on driving cycles relevant to a specific country are used to develop national-scale EFs, remote sensing, tunnel studies and PEMS are largely used for validation studies with on-road conditions [20]. Therefore, an analysis of suitable foreign EFs that allow a tier 02 emission estimation (VKM-based) is required since Sri Lanka has no defined set of EFs pertaining to the local conditions.

Although Indian EFs can be used as an alternative with similar traffic behaviour and density, previously developed Indian EFs under the Indian Clean Air Programme (ICAP) [28] have not been updated for over a decade. Other researches are products of individual/project research instead of a guideline from a regulatory body [29], [30], and do not possess adequate separation characteristics based on vehicle technology and power/size as shown in Table 2.1

Table 2.1: EFs for Indian road vehicles [29].

	CO	**NO_X**	**SO_2**	**PM**	**HC**
Bus	3.6	12	1.42	0.56	0.87
Omni buses	3.6	12	1.42	0.56	0.87
Two wheelers	2.2	0.19	0.013	0.05	1.42
Light motor vehicles (passenger)	5.1	1.28	0.029	0.2	0.14
Cars and jeeps	1.98	0.2	0.053	0.03	0.25
Taxi	0.9	0.5	10.3	0.07	0.13
Trucks and lorries	3.6	6.3	1.42	0.28	0.87
Light motor vehicles (goods)	5.1	1.28	1.42	0.2	0.14
Trailers and tractors	5.1	1.28	1.42	0.2	0.14
Others (Averaged)	3.86	3.89	1.94	0.24	0.54
Reference	CPCB, 2007	CPCB, 2007	Kandikar and Ramachandran, 2000	CPCB, 2007	CPCB, 2007

One study indicates that Europe could serve as first-order default values for on-road emission modelling work where more specific regional data are not available [31]. [32] affirms that the current emission inventory in China relies mostly on European emission databases, although it might not reflect local conditions or technological performances. Therefore, it is more suitable to use European defaults. The EMEP/EEA guideline [3], provides a comprehensive set of EFs for Tier 01 and 02, and derivation formulae for Tier 03. Since it was established that Tier 02 is approachable with VET data, this review focuses on EFs under Tier 02. The guideline provides a table of EFs categorized based on, (i) vehicle categories ranging from passenger cars, light commercial vehicles, heavy-duty trucks, buses, and L-category (motorcycles and mopeds), (ii) vehicle technologies based on vintage and fuel type which is a breakdown of euro standards with year-ranges of implementation alongside fuel type, and (iii) Pollutants, including CO, NMVOC, NO_x, N_2O, NH_3, and PM with separate EFs matched to the above two categories. Previous studies have highlighted the influence of deterioration and vehicle technology on the EF for a particular pollutant, requiring either correction factors for previously defined EFs or recently updated guidelines containing EFs [33]. The EMEP/EEA guidelines which were updated in late 2024, provide the latter.

As mentioned in Chapter 1, since this is a case study on motor/passenger cars with the notion that the methodology is scalable to all other vehicle classes as required, Table 2.2 highlights the vehicle type and legislation combinations for which EFs are defined for passenger cars under this guideline. The complete list of EFs is detailed in Appendix.

Table 2.2: Passenger car type and legislation combinations covered by the tier 02 methodology[3].

Type	**Legislation/Technology (k)**
Petrol Mini	Euro 4, Euro 5, Euro 6
Petrol Small, Medium, Large-SUV-Executive	PRE ECE, ECE 15/00-01, ECE 15/02, ECE 15/03, ECE15/04, Improved Conventional, Open-Loop, Euro 1 – Euro 6
Diesel Mini	Euro 4 – Euro 6
Diesel Small, Medium, Large-SUV-Executive	Conventional, Euro 1 – Euro 6
LPG Mini	Euro 4 – Euro 6
LPG Small, Medium, Large-SUV-Executive	Conventional, Euro 1 – Euro 6
Petrol Hybrid	Euro 4 – Euro 6
Petrol PHEV	Euro 6 up to 2016 – Euro 6
Diesel PHEV Large	Euro 6
CNG	Euro 4, Euro 5, Euro 6

The development of localised EFs that reflect the performance of a country's vehicle fleet is vital [34]. Although outdated, the EFs developed under the ICAP provide the most comprehensive separation characteristics for comparison with the European guideline, to observe the regional variation characteristics once Sri Lankan EFs are eventually developed. Table 2.3 showcases this variation using three pollutant EFs available in both sources, under similar fuel-vintage-size combinations.

Table 2.3: Comparison of the regional variation of EFs (ICE passenger cars)

Vehicle Type (EMEP/EEA)	**Equivalent Vehicle Type (ICAP)**	**EFs (g/km)**		
		CO	**NO_x**	**PM**
Petrol-Euro 3-Small		3.15	0.11	0.001
	Petrol-BS 1-(1000-1400cc)	3.01	0.12	0.006
Petrol-Euro 3-Medium		2.83	0.11	0.001
	Petrol-BS 1-(1400cc+)	2.74	0.21	0.006
Petrol-Euro 4-Medium		0.95	0.06	0.001
	Petrol-Post 05-(1400cc+)	0.84	0.09	0.002
Diesel-Euro 3-Small		0.09	0.79	0.046
	Diesel-BS 1-(Below 1600cc)	0.72	0.84	0.190
	Diesel-BS 2-(Below 1600cc)	0.30	0.49	0.060
Diesel-Euro 4-Small		0.09	0.59	0.038
	Diesel-Post 05-(Below 1600cc)	0.06	0.28	0.015
Diesel-Euro 2-Medium		0.29	0.73	0.059
	Diesel-96 to 00-(1600cc+)	0.66	0.61	0.180

With the exception of PM, for which the European guideline only considers PM2.5 while ICAP does not specify that information, CO and NO_x EFs show slightly similar ranges, with European values being higher in some instances and ICAP values being higher in others. This further justifies the use of European values as the first-order default until local EFs are developed for Sri Lanka. The resultant emissions would have a proportional sensitivity to the EFs, since the mileage values remain the same (Eg. CO emissions with ICAP EFs for an annual span would be 96% of the CO emissions with EMEP/EEA EFs for the same time period for a particular vehicle fleet).

2.2 Integration of Machine Learning for Emissions Estimation and Prediction

2.2.1 Traditional Models and Static Machine Learning Models

Prior to the popularisation of data-driven models, emission modelling was primarily done through mechanistic and empirical models, and prediction was governed by statistical principles. These traditional models consist of average-speed models and modal models.

1. **Average-speed models**: EMFAC (developed by the U.S. Environmental Protection Agency) and COPERT (developed by the European Environment Agency) rely on regression functions between average vehicle speed and EFs. These models aggregate emission estimates over time and space, making them suitable for long-term and large-scale inventory development. For example, EMFAC estimates emissions by considering factors like fuel type, ambient conditions, and inspection/maintenance programs, while COPERT incorporates deterioration rates and correction parameters such as road slope, fuel composition, and vehicle load, allowing for broader pollutant coverage [12].

2. **Modal models:** Allow microscale emission estimation through division of driving behaviour into operation modes (idling, cruising, acceleration, etc.). Examples include MOVES, CMEM, and IVE, which use instantaneous driving inputs and vehicle-specific power (VSP) metrics to predict emissions on a second-by-second basis. These are useful in urban air quality assessments [12].

There is a clear shift towards data-driven models, with 30.33% of the recent literature on emissions prediction utilising data-driven methods, primarily in the form of ANNs [12]. While traditional models are still used for large-scale emission inventories, data-driven models are being popularised, especially in scenarios demanding fine spatial and temporal evaluations. ANN models have been implemented using techniques such as back-propagation and correlation-based feature selection, showing high accuracy (up to 98% coefficient of determination, R^2) and adaptability to different engine types and conditions [18]. One study [35] has shown that ANNs trained on chassis dynamometer data were able to predict transient NOx and CO_2 emissions with a mean error under 5%. Moreover, since ML models directly use data for pattern recognition, they can be retrained for new regions, vehicles, or policies without completely reengineering the model.

2.2.2 Limitation of Static Models and the Need for Temporal Forecasting

Both empirical and most data-driven models, even though capable of estimating emissions ex-post, have a primary limitation in forecasting future emissions under scenario-based, real-world conditions. Many of the machine learning models have been designed to predict emissions at a specific moment [12], using historical sensor inputs like speed, torque, or acceleration [35]. These pointwise regressors cannot convert to forecasting emissions under changing policy, vehicle technology, or fleet composition scenarios.

Earlier forecasting approaches like ARIMA have been applied to emission time series [36], [37], but they are known to be sensitive to stationarity assumptions and incapable of modelling nonlinear relationships or long memory [38]. Neural networks, including feedforward ANNs, are more flexible in handling nonlinear patterns but lack any memory structure, making them poorly suited for time-series tasks where the future is dependent on the past [39]. However, recurrent neural network (RNN) models have shown better performance in learning time-dependent emission patterns. Connor and Atlas [40] were early to show that RNNs could outperform feedforward networks in emissions prediction. However, standard RNNs suffer from the gradient vanishing problem and short memory span, limiting their forecasting accuracy. LSTM-based models explicitly address such issues through their gated architecture and are suited for predicting long-term dependencies such as policy decisions [41]. The availability of disaggregated emissions datasets, such as the VET database, further justifies the need for time-aware models, as they can be enhanced with growing data annually.

2.2.3 LSTM Time-Series Forecasting Applications

The LSTM model was designed specifically to overcome the vanishing or exploding gradients of standard RNNs in learning long-term dependencies in sequential data [42]. This is achieved through the constant error carousel (CEC) and gated architecture, which enables it to preserve gradients during training and avoid vanishing or exploding errors [43], which will be discussed in more detail under Chapter 3. In the context of the VET database, individual vehicle emissions aggregated monthly/yearly can be used to predict future emission values using time-series analysis, with the presence of long-term dependencies. Hochreiter and Schmidhuber's [44] foundational work introduced memory cells and gating mechanisms, namely input, forget, and output gates, that enable LSTMs to retain useful temporal features and discard irrelevant ones. This capability has been used for modelling emissions as temporal phenomena. This also allows multivariate temporal inputs, including traffic, meteorological, and fleet composition variables [38].

Multiple studies have been conducted with target outputs such as CO, CO_2, NO_x, and Hydrocarbons (HC), where LSTM-based models outperform baseline methods like SVR, ANN, and ARIMA. For instance, one study [45] has developed LSTM models that consider the time dependence of emissions response to the vehicle operation situation of buses. Another [46] have proposed a parallel attention-based LSTM (PA-LSTM) that reduces the effect of outliers and improves accuracy. Similarly, LSTM showed significantly higher accuracy than ARIMA in forecasting lane-level concentrations of CO, HC, and NO in urban Beijing.[38] The model introduces an important innovation, which is wavelet decomposition of the original emissions time series. The highly variable input data is decomposed into smoother sub-series, each modelled by a dedicated LSTM, before summing the forecasts. This technique has improved performance, especially under noisy or non-stationary data conditions. A similar reduction in data variability can be achieved through introducing moving averages [47].

Another study [48] has proposed a Residual Autoencoder-LSTM (ResAE-LSTM) model, where emissions in time series are decomposed into:

- A linear component, modelled by autoregression,
- A nonlinear residual, modelled via LSTM,
- A compressed feature space, learned through an autoencoder.

This separation improves the model's ability to capture both routine emissions trends and outlier events (spikes from freight traffic or weather-driven anomalies). The model achieved a Root Mean Square Error (RMSE) of just 0.050 for NO forecasting, significantly outperforming ARIMA (0.327) and ANN (0.603). These hybrid models are especially suited to scenario-based forecasting, where emissions inputs may be altered to reflect new fleet mixes, policy mandates, or climate factors. Large-scale disaggregated databases containing vehicle characteristics are still primarily analysed through traditional models such as COPERT and IVE through the application of EFs. There is a clear gap in implementing empirical calculations in conjunction with time-series estimation using LSTM models. Therefore, this thesis intends to predict emissions using a novel framework that combines aggregated averages of emissions on a month-wise basis, with the prediction strategies under LSTM models.

CHAPTER 3

LONG SHORT-TERM MEMORY (LSTM) MODELS

In machine learning applications that involve sequential or time-dependent data, models must capture long-range dependencies across time steps. Traditional feedforward neural networks (FFNNs) do not possess this capability since they are designed for static classification tasks and do not account for temporal relationships between inputs. To overcome this, RNNs have been introduced, enabling information to be retained across time through recurrent connections [49]. However one of the limitations of RNNs is their inability to effectively learn dependencies longer than a few timesteps due to the vanishing or exploding gradient problem during training [43]. LSTM networks, which were introduced by Hochreiter and Schmidhuber [44] in the 1990s, were designed to mitigate this issue. LSTM models feature a unique memory cell and gating mechanisms that regulate the flow of information, enabling the model to preserve gradients and learn from distant past inputs.

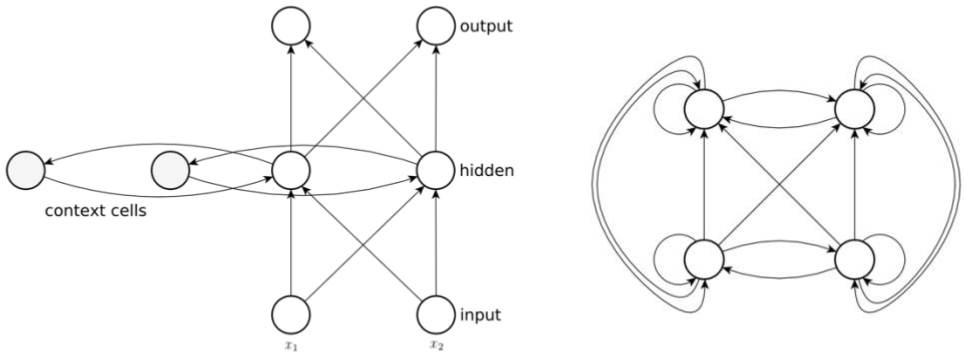
As highlighted by Staudemeyer and Morris [43], LSTM-RNNs are among the most powerful dynamic classifiers currently available. They are capable of modelling sequences over more than 1,000 timesteps, depending on network complexity. Chapter 3 aims to discuss the key concepts, mathematical formulations, and architectural features of LSTM networks. It will also explore training methods, architectural variants, and the limitations and strengths of LSTM networks in modern machine learning applications.

3.1 RNNs and the Vanishing/Exploding Gradient Problem

3.1.1 Basic Architecture

Understanding the evolution of LSTM networks requires exploring the architecture and limitations of RNNs. A basic RNN consists of input, hidden, and output layers, where each hidden unit receives input not only from the previous layer but also from its own previous state at the prior timestep. This self-feedback loop creates temporal dynamics in the network, allowing it to model sequence-dependent tasks.

Elman networks [50] and Jordan networks [43] are classical forms of RNNs that maintain a context layer to store previous hidden states. While simple RNNs can theoretically learn temporal dependencies, in practice their memory depth is severely constrained. Figure 3.1 shows an Elman neural network alongside a fully RNN with self-feedback connections.

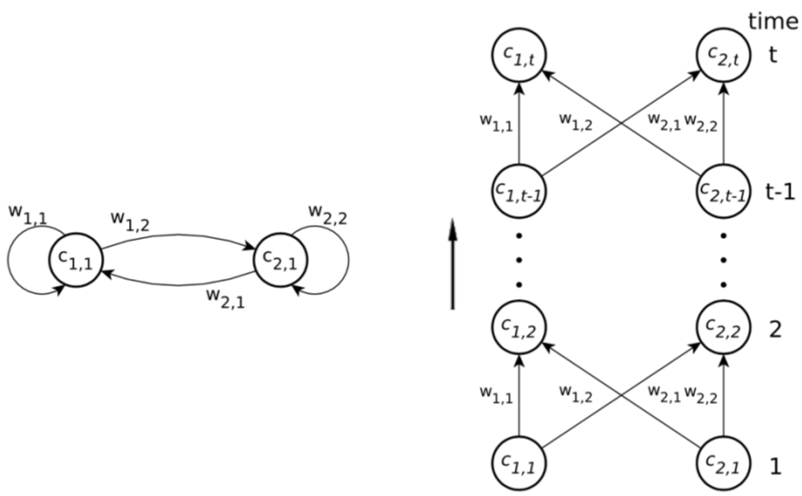


(a) (b)

Figure 3.1: Simple RNN Structures – (a) Elman neural network, (b) fully recurrent neural network. Adapted from [49], [50]

3.1.2 Backpropagation Through Time (BPTT)

As the primary method of training for RNNs, the BPTT algorithm utilises the fact that for a finite period of time, there is a FFNN which behaves identical to an RNN. BPTT unrolls over time, effectively transforming it into a very deep feed-forward network with shared weights across time steps [51]. Figure 3.2 illustrates the unrolling process.



(a) (b)

Figure 3.2: RNN unrolling process – (a) A simple RNN with a two-neuron layer, (b) The same network unfolded over time with a separate layer for each time step (in the form of a FFNN. Adapted from [43]

In this process, he total error over an epoch from time t' to t is defined by equation (3.1)

$$E_{total}(t',t) = \sum_{\tau=t'}^{t} E(\tau) \tag{3.1}$$

Where the instantaneous error E(τ) at time τ is as shown in Equation (3.2)

$$E(\tau) = \frac{1}{2}\sum_{u\in U}(e_u(\tau))^2 \tag{3.2}$$

Here, the unit-wise error $e_u(\tau)$ at a time τ is defined as per Equation (3.3) where d is the target output and u is the observable output.

$$e_u(\tau) = \begin{cases} d_u(\tau) - y_u(\tau) & if\ u \in T(\tau) \\ 0 & otherwise \end{cases} \tag{3.3}$$

The error signal $\vartheta_u(\tau)$ of a non-input unit *u* at a time τ, which is used to adjust the weights, is then backpropagated through time using Equation (3.4), where W is the weight that connects two units, z is the weighted output of a unit, and f' is the derivative of the activation function.

$$\vartheta_u(\tau) = \begin{cases} f'_u(z_u(\tau))e_u(\tau) & if\ \tau \in t \\ f'_u(z_u(\tau))\sum_{k\in U} W_{k,u}\vartheta_k(\tau+1) & otherwise \end{cases} \tag{3.4}$$

This recursive definition illustrates how gradients accumulate backwards in time, making the process vulnerable to numerical instability.

3.1.3 The Vanishing/Exploding Gradient Problem

During the backpropagation process of gradients over timesteps, they are repeatedly multiplied by weights and derivatives of activation functions, such as sigmoid, which are typically less than 1. This can result in exponentially vanishing/exploding gradients, making the model incapable of learning long-range dependencies [44]. When an RNN is backpropagated through time for t - t' $(t' > t)$ time steps, the error signal $\vartheta_u(\tau)$ is scaled by a factor. When this factor is unrolled over time, if $t' \leq \tau \leq t$, and u_r is a non-input layer neuron in one of the replicas in the unrolled network at the time τ, setting $u_t = v$ and $u_{t'} = 0$, Equation (3.5) is obtained

$$\frac{\vartheta_u(t')}{\vartheta_u(t)} = \sum_{u_{t'}\in U} \cdots \sum_{u_{t-1}\in U} \left(\prod_{\tau=t'+11}^{t} f'_{u_\tau}(z_{u_\tau u_\tau}(t-\tau+t'))W_{[u_\tau,u_{\tau-1}]} \right) \tag{3.5}$$

If the terms inside the product are:

< 1 → gradients shrink exponentially (vanishing)

> **1** → gradients grow exponentially (exploding)

This is the issue mitigated by the architecture of LSTM networks [52].

3.2 LSTM Architecture and Operation

3.2.1 Basic Architecture

The LSTM cell operates with a Constant Error Carousel (CEC), a structure that enables the error signal to flow unimpeded across long sequences of timesteps. In an error signal in the form of equation (3.4) when $\tau \in t$, to preserve this error over time, the network enforces the conditions highlighted in Equation (3.6)

$$f'_u(z_u(\tau))W_{u,u} = 1.0 \tag{3.6}$$

Implying that f_u u is the identity function and $W_{u,u} = 1.0$, so that, $y_u(\tau+1) = y_u(\tau)$, and forms the foundation of the CEC and enables stable memory retention over extended timesteps [44].

3.2.2 Memory Block Structure

An LSTM unit is called a memory block and consists of one or more memory cells, an input gate, an output gate and (Optionally) a forget gate. These gates are sigmoid-activated units that modulate the flow of information. The input and output gates determine whether the cell receives or transmits information, while the forget gate decides whether old memory is retained or discarded. Figure 3.3 highlights this structure.

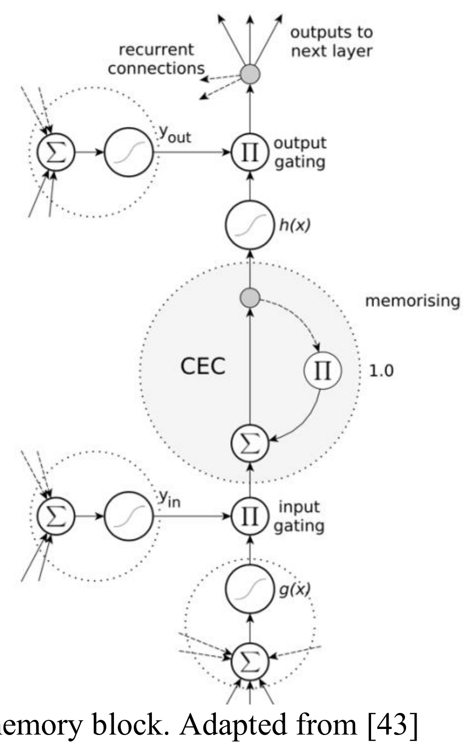


Figure 3.3: An LSTM memory block. Adapted from [43]

The block contains (at least) one cell with a CEC and weight of '1.0'. The state of the cell is denoted as s_c. Read and write access is regulated by the input gate, y_{in}, and the output gate, y_{out}. The internal s_c is calculated by multiplying the result of the squashed input, g, by the result of y_{in}, and then adding the state of the last time step, $s_c(t-1)$. Finally, the cell output is calculated by multiplying s_c by the activation of y_{out}.

3.2.3 Forward Pass Computation

Let $m \in M$ be a memory block and $mc \in m$ a memory cell. Let y_{in_m}, y_{out_m}, and y_ϕ be the activations of the input, output, and forget gates, respectively. Equations (3.7) to (3.13) indicate the functioning of cell structures in a forward pass.

Input and output gate activations.

$$y_{in_m}(\tau+1) = f(z_{in_m}(\tau+1)) \tag{3.7}$$

$$y_{out_m}(\tau+1) = f(z_{out_m}(\tau+1)) \tag{3.8}$$

Weighted input to cell, W and X respectively, are weights and input for a particular unit.

$$z_{mc}(\tau+1) = \sum_u W_{mc,u} \,.\, X_{u,mc}(\tau+1) \tag{3.9}$$

Cell state update

$$s_{mc}(\tau+1) = s_{mc}(\tau).\, y_\phi(\tau+1) + y_{in_m}(\tau+1).\, g(z_{mc}(\tau+1)) \tag{3.10}$$

Where, the squashing function *g(z)* is,

$$g(z) = \frac{4}{1+e^{-z}} - 2 \tag{3.11}$$

Cell output computation,

$$y_{mc}(\tau+1) = y_{out_m}(\tau+1).\, h(sh_{mc}(\tau+1)) \tag{3.12}$$

Where, the output squashing function *h(z)* is,

$$h(z) = \frac{2}{1+e^{-z}} - 1 \tag{3.13}$$

Here, the functionality of the forget gate is as indicated in equations (3.14) and (3.15), where *b* is the bias.

$$y_\phi(\tau+1) = f(z_\phi(\tau+1) + b_\phi) \tag{3.14}$$

Where the input for the forget gate is,

$$z_{\phi}(\tau+1) = \sum_{u} W_{\phi,u} \,.\, X_{u,\phi}(\tau+1) \tag{3.15}$$

The forget gate has been proposed in [53], and its bias is initialised to 1.0 following the recommendation in [54], to encourage initial memory retention and improve training stability.

3.3 LSTM Training

The training process involves calculating gradients of the network's output with respect to its weights over time and updating those weights to minimise prediction error. Due to the internal memory and feedback loops in LSTM networks, training is more complex than in feedforward neural networks and even traditional RNNs. Staudemeyer and Morris [43] explain that a hybrid learning approach combining BPTT and Real-Time Recurrent Learning (RTRL) is typically used [53]. The units after the memory cells are trained using BPTT, while units before and within memory cells are trained using RTRL. This separation ensures preservation of cell states across time, even when no training target is provided at every timestep, which is an inherent capability of RTRL.

3.3.1 Backward Pass using BPTT

Similar to the BPTT procedure in RNNs, the LSTM is unfolded over a time window, and error signals are propagated backwards through time to calculate weight updates. Equations (3.16) to (3.20) explain this process.

For an output unit o, the error at time τ is given by,

$$E(\tau) = \frac{1}{2} \sum_{o \in O} (d_o(\tau) - y_o(\tau))^2 \tag{3.16}$$

The error signal is,

$$\vartheta_o(\tau) = f'_o\big(z_o(\tau)\big) \,.\, (d_o(\tau) - y_o(\tau)) \tag{3.17}$$

The weight update for the output layer is then,

$$\Delta W_{o,v}(\tau) = \eta . \vartheta_o(\tau) . X_{v,o}(\tau) \tag{3.18}$$

For hidden units between the cells and output units, a similar error signal and weight update is obtained.

$$\vartheta_h(\tau) = f'_h\big(z_h(\tau)\big) \,.\, \sum_{o \in O} \Delta W_{o,h} \,.\, \vartheta_o(\tau) \tag{3.19}$$

$$\Delta W_{h,v}(\tau) = \eta . \vartheta_h(\tau) . X_{v,h}(\tau) \tag{3.20}$$

3.3.2 Training via RTRL

Memory cells and gates cannot rely solely on BPTT because their influence propagates indirectly across time, even if no immediate loss is observed. Therefore, RTRL is used to handle such components. The error signal for a memory cell $mc \in m$ is defined as per Equation (3.21), and only holds if the error is truncated [53].

$$\vartheta_{mc}(\tau) = y_{out_m}(\tau) \, . \, h'\big(s_{mc}(\tau)\big) \, . \left(\sum_{o \in O} W_{o,mc} \, . \, \vartheta_o(\tau) \right) + \vartheta_{mc}(\tau+1) \, . \, y_{\phi}(\tau+1) \tag{3.21}$$

This recursive form allows the cell's state to accumulate error contributions across timesteps. The updated weight for the memory cell is given by Equation (3.22)

$$\Delta W_{mc,u}(\tau) = \eta . \vartheta_{mc}(\tau) \, . \frac{\partial s_{mc}(\tau)}{\partial W_{mc,v}} \tag{3.22}$$

Similarly, for the input and forget gates, where $u \in \{in_m, \phi_m)$, the updated weights are given by Equation (3.23)

$$\Delta W_{u,v}(\tau) = \eta . \sum_{mc \in m} \vartheta_{mc}(\tau) \, . \frac{\partial s_{mc}(\tau)}{\partial W_{u,v}} \tag{3.23}$$

Depending on which gate or cell the weight connects to, the corresponding gradient of cell state s_{mc} changes. As an example, the memory cell weight is given in Equation (3.24), where it only holds if truncated.

$$\frac{\partial s_{mc}(\tau+1)}{\partial W_{mc,v}} = \frac{\partial s_{mc}(\tau)}{\partial W_{mc,v}} \, . \, y_{\phi}(\tau+1) + g'\big(z_{mc}(\tau)\big) \, . \, f\left(z_{in_m}(\tau+1)\right) . \, y_v(\tau) \tag{3.23}$$

Similarly, output and forget gates are updated.

3.3.3 Training Complexity

LSTM training complexity depends on the number of memory blocks, input/output units, and gating mechanisms. Staudemeyer and Morris [43] have defined this complexity measure as $O(B^2 \, . \, C^2)$, where B is the number of memory blocks and C is the number of memory cells per block.

3.4 Extensions and Variants of LSTM Networks

This section outlines some prominent LSTM variants that have several enhancements and structural modifications to further improve performance required for specific tasks.

3.4.1 Bidirectional LSTM

Usually, RNNs and LSTMs process inputs in a single direction from past to future. But in applications such as speech/handwriting recognition and phoneme classification [55], both forward and backward context is utilised with Bidirectional LSTM. BLSTM utilises two parallel networks, one that processes the input sequence in forward time order and one that processes in reverse time order. Both these outputs are then concatenated and passed to a shared output layer.

3.4.2 Standard Stacked LSTM

In a stacked LSTM, each layer consists of one or more memory blocks. The hidden state $\vec{y}_{hi}$ from the i^{th} LSTM layer is used as the input to the next layer *(i+1)*. This allows the model to learn low-level temporal features, such as short-term fluctuations in lower layers while learning higher-level abstract features, such as trends in upper layers. Stacked LSTMs have been found to outperform single-layer LSTMs in tasks such as speech recognition, machine translation and time-series prediction [56].

3.4.3 Grid LSTM

The Grid LSTM architecture generalises standard LSTM to multi-dimensional spaces, enabling cells to interact not only over time, but also across spatial or hierarchical dimensions such as layers or positions in an image [57]. Unlike stacked LSTM, which passes only hidden states between layers, Grid LSTM includes memory cells across depth and spatial dimensions, allowing it to propagate memory across different axes and avoid instability found in standard multidimensional LSTM [58].

The core mechanism of grid LSTMs are shown using equations (3.24) and 3.25). If $\vec{y}_{h1}, \dots, \vec{y}_{hN}$ are hidden vectors and $\vec{s}_{m1}, \dots, \vec{s}_{mN}$ are memory vectors across N dimensions,

$$X = \begin{bmatrix} \vec{y}_{h1} \\ \dots \\ \vec{y}_{hN} \end{bmatrix} \tag{3.24}$$

Each dimension i then performs an LSTM transformation, with a modular, multi-dimensional structure.

$$(X, \vec{s}_{mi}) \xrightarrow{W_i} (\vec{y}_{hi}, \vec{s}_{mi}) \tag{3.25}$$

3.4.4 Other Notable LSTM Variants

Gated Recurrent Units is a simplified variant of LSTM that merges the memory cell and hidden state and uses only two gates, named update gate and reset gate [59]. This is a faster and more memory-efficient architecture with comparable or even better performance than LSTM on certain tasks, such as machine translation. Sequence-to-sequence models are encoder-decoder frameworks where an LSTM encodes an input sequence into a fixed vector, then another LSTM decodes it into an output sequence [60]. Attention mechanisms are advanced sequence-to-sequence models that allow the decoder to dynamically focus on relevant encoder hidden states [61].

CHAPTER 4

METHODOLOGY I - DATA PRE-PROCESSING

This chapter focuses on the pre-processing of the VET dataset, laying the groundwork for the predictive modelling framework presented in the subsequent chapters. This includes data integration, feature engineering, as well as input variable selection and justification.

As per the scope of the utilised data, as explained in Section 1.2, the study focuses on one vehicle class (Motor/ Passenger Cars) as a case study, with the notion that the methodology is scalable to all other vehicle classes as required. The selection of passenger cars as the vehicle category is motivated by several factors.

1. **Representation of Vehicle Technology:** Passenger cars in Sri Lanka exhibit the widest range of emission control technologies, from Pre-ECE to Euro 6d standards. This spectrum allows the analysis to capture emissions behaviour across different regulatory eras and engine technologies for each fuel type, making the case study more representative of real-world variability.

2. **Policy Influence:** Passenger cars form a significant portion of the national vehicle fleet [2], and the government is most likely to implement policy measures related to vehicle import regulations, emissions thresholds, and taxation based on passenger car trends.

3. **Data Availability:** The comprehensive and consistent coverage of the VET dataset allows passenger car data to provide sufficient temporal depth and sample size for effective prediction.

However, the VET dataset imposes a further limitation due to the lack of standardised and centralised emissions test data for hybrid and electric vehicles in Sri Lanka, as they are not subjected to emissions testing under the existing VET program. Therefore, this research focuses on ICE passenger cars, with a methodology that can be adopted towards hybrid and electric vehicles using relevant EFs (Appendix) when a centralised database is developed.

4.1 Data Integration and Cleaning

The VET program operates under two authorised private agencies (CleanCO Pvt. Ltd. and Laugfs Eco Sri Pvt. Ltd.), where both conduct mandatory emissions tests for motor vehicles and maintain separate digital records of test results on a monthly basis. The dataset provided for this study spans from January 2016 to October 2024, where each record corresponds to a single vehicle tested during a specific month and includes a range of information related to vehicle characteristics, test results, and tailpipe pollutant volume/volume readings (for petrol vehicles) or smoke opacity readings (for diesel vehicles).

4.1.1 Raw Data Structure

The datasets from the two agencies differed slightly in terms of column naming and formatting, though they contained broadly similar information, as shown in Table 4.1

Table 4.1: Raw data column headers from CleanCO and Laugfs Eco Sri.

	CleanCO	**Laugfs Eco Sri**
General	Agent, IsTestRegOnly, IsRefund, MachineID	Agent, REFNo, IsRefund, Reason, Status
Vehicle Data	VehRegNo, VehRegYear, VehProvinceCode, VehDistrictCode, VehInstitute, VehMake, VehModel, VehManuYear, VehEngNo, VehChassisNo, VehCylinders, VehStroke, VehFuelType, VehCurrentMileage, VehClass	FirstRegYear, ProvCode, DistrictCode, Institute, Make, Model, ManuYear, EngNo, ChaNo, Cylinders, Stroke, Fuel, Meter, Class
Test Data	TestDateTime, TestResult, TestCertificateNo, TestLocation, TestMobOrFix, TestType, TestInspector, TestLane, TestOilTemp	TestDate, TestResult, SERIAL, Location, Fixed, TestType, UserID, Lane, Temp
Readings	K1Value, RPM1, K2Value, RPM2, K3Value, RPM3, Kavg, AccHC, AccCO, AccCO2, AccO2, AccLambda, AccRPM, IdleHC, IdleCO, IdleCO2, IdleO2, IdleLambda, IdleRPM	F1Test1, F1RPM1, F1Test2, F1RPM2, F1Test3, F1RPM3, KAvg, F1AccHC, F1AccCO, F1AccCO2, F1AccO2, F1AccLam, F1AccRPM, F1IdleHC, F1IdleCO, F1IdleCO2, F1IdleO2, F1IdleLam, F1IdleRPM

These records include administrative and technical information about the vehicles as well as the odometer reading at the time of the test. For tailpipe pollutant readings, the datasets contain volume per volume percentage (v/v%) values for CO and HC measured at both idle and accelerated RPMs for petrol vehicles, and smoke opacity levels (K value or Lambda value) for diesel vehicles.

4.1.2 Data Harmonisation and Filtering

The following steps are applied to convert the data into a coherent structure required for feature engineering.

1. **Column Mapping and Renaming:** Column headers from the Laugfs dataset are converted to match CleanCO nomenclature to ensure consistency across both data sources. This includes mapping Excel-based columns (Meter → VehCurrentMileage, Make → VehMake) and standardising data types.

2. **Vehicle Type Filtering and CSV conversion:** The VehClass column is used to isolate records corresponding to passenger cars, and the filtered dataset is converted into CSV format for increased computational efficiency. Additionally, outlier removal and quality control are performed, excluding failed tests (TestResult = FAIL), and missing/unrealistic odometer readings (VehCurrentMileage → negative values or values exceeding expected upper thresholds for vehicle age).

   Odometer outliers: The upper threshold is selected based on the average mileage of vehicles under 5-year ranges of the year of manufacture (YOM). The outliers observed are mainly due to input errors, giving 10^6 range mileage difference compared to other vehicles in the YOM range. The total outlier removal procedure removes around 14% of the data points. (5,365,517 data points for passenger cars before filtering)

3. **Dataset Combination:** The monthly datasets from CleanCO and Laugfs are then merged while maintaining their monthly structure, resulting in a unified dataset containing VET records for ICE passenger cars on a monthly basis from January 2016 to October 2024.

4.1.3 Final Dataset Structure

The final extracted dataset includes only the essential columns required for feature generation and trend analysis, as indicated by Table 4.2

Table 4.2: Representative portion of the filtered data

Registration Number	Test Result	Province Code	Make	Model	YOM	Fuel Type	Current Mileage
302-XXXX	P	1	DAEWOO	MATIZ	2000	P	78486.0
64- XXXX	P	2	NISSAN	SUNNY	1992	D	415306.0
JZ- XXXX	P	1	MARUTI	800	2005	P	71418.0
CB- XXXX	P	1	TOYOTA	3BA.NZT260 PREMIO	2020	P	17805.0
KL- XXXX	P	1	VOLKSWAGON	PASSAT	2009	P	175297.0
GE- XXXX	P	9	NISSAN	SUNNY	1997	P	198000.0
HZ- XXXX	P	6	NISSAN	SUNNY	2000	P	0.0
CB- XXXX	P	1	TOYOTA	DBA NGX50CHR	2017	P	29909.0

KY- XXXX	P	1	SUZUKI	ALTO-800-LXI	2014	P	95000.0
KW- XXXX	P	7	MICRO	PANDA LC 1.0	2013	P	13234.0
CA- XXXX	P	1	TOYOTA	FORTUNER	2013	D	83410.0
300- XXXX	F	1	TOYOTA	COROLLA	1997	P	142163.0
300- XXXX	F	1	DAIHATSU	L50IRSGMDS	1998	P	286000.0

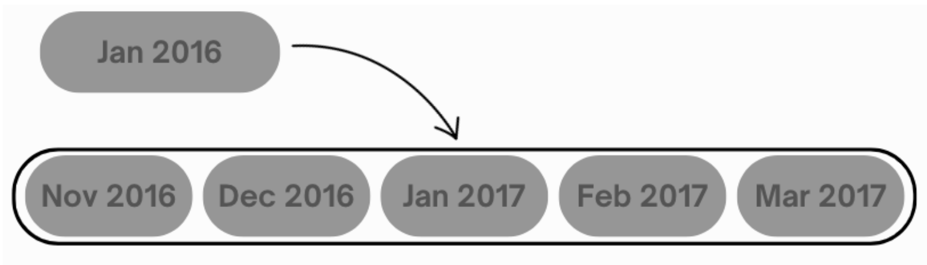
These cleaned records indicate the achievement of the first objective of this study and form the basis for the feature engineering described in the next section.

4.2 Feature Engineering

This section outlines the methods used to manipulate vehicular data to obtain annual mileage values, assign vehicle categories and emission technologies, and transform them into aggregated inputs suitable for time-series modelling, while maintaining the monthly data structure for better temporal depth.

4.2.1 Annual Mileage Estimation

As per Equations (2.2) and (2.3) under Sub-section 2.1.1, annual distance travelled is essential for a Tier 02 emissions estimation. Since the VET dataset contains odometer readings (VehCurrentMileage) but does not directly provide mileage differences or yearly usage, an algorithm is developed to calculate annual mileage for individual vehicles based on consecutive test records. It runs through each record and matches it against a subsequent test result for the same vehicle, using the vehicle registration



number (VehRegNo) as the unique identifier. For a given reference month, the algorithm searches for a matching test within a five-month window around one year later, as indicated in Figure 4.1. This flexible matching range accounts for variations in actual re-test intervals due to user behaviour, registration cycle, or external factors.

Figure 4.1: Example range for cross-matching

When a suitable match is found, the algorithm adjusts the mileage difference to reflect an exact 12-month span, assuming linear mileage accumulation across time since individual vehicle records are being considered. This is supported by Khare et al., who have found that the accumulation of mileage is approximately linear with time in a study related to automotive field failure analysis with a large database of vehicle data [62]. Another study in China, where inspection and maintenance data were used similar to this study, visualises the validity of the assumption of linear mileage accumulation against time, within an emissions scope [63]. Additionally, a study in the US based on the 2017 national household travel survey data provides empirical motivation and pragmatic justification for linear annual accumulation as a working assumption [64]. The calculated difference is then stored under the original month (Mileage differences for a January 2016 test are stored under "January 2016" regardless of whether the match occurs in December 2016 or February 2017). This allows the construction of monthly data frames containing vehicle-level annual mileage estimates.

4.2.2 Integration of Euro Vehicle Category and Vehicle Technology Parameters

As explained under Table 2.1, the EMEP/EEA guidelines have defined four vehicle categories for both petrol and diesel passenger cars: Mini, Small, Medium and Large-SUV-Executive. This classification is based on engine size and power output (horsepower) according to vehicle specification databases and marketplaces based in Europe [65], [66]. Table 4.3 indicates these engine sizes and power output ranges.

Table 4.3: Engine size and power output ranges for Euro vehicle categories. Adapted from [65]

Vehicle Category	**Engine Size (L)**	**Power Output (kW)**
Mini	< 1.0L	22.4-52.2 kW (30-70 Hp)
Small	1.0L ≤ and < 1.4L	52.5-82.0 kW (70-110 Hp)
Medium	1.4L ≤ and < 1.8/2.0L	82.0-111.86 kW (110-150 Hp)
Large-SUV-Executive	Beyond 1.8/2.0L	111.86+ kW (150+ Hp)

In order to incorporate the power output values for individual vehicle records, a web-scraping algorithm is developed to extract horsepower data for specific makes and models from a prominent auto specifications database named 'Auto-data' [66]. It extracts all vehicle brand (make) links available on the base webpage in alphabetical order, then extracts vehicle model links for each model while looping under each brand link. The algorithm uses the BeautifulSoup library to read through each model page and locate the power value or value range, and appends it to a csv file alongside the model and make.

In order to reliably align scraped vehicle specs with VET entries, given inconsistencies and human errors in the VET records (spelling variations, casing, spacing), a nine-step fuzzy matching algorithm is developed where,

- Step 01: Exact Matches on both make and model
- Step 02 – Step 07: Exact match on make and fuzzy match on model with decreasing thresholds for similarity
- Step 08 – Step 09: Fuzzy matches on both make and model with decreasing thresholds for similarity

The resultant dataframe assigns power output values for all unique make and model combinations in the filtered VET dataset containing annual mileage estimations. Table

4.4 indicates a representative proportion of these matched make/model and power output combinations.

Table 4.4: Representative portion of matched make/model and power output combinations

Make	Model	Power Output (Hp)
LAND ROVER	FREELANDER	150-240 Hp
MARUTI	800	35-45 Hp
TATA	INDICA XETA	61-85 Hp
MERCEDES BENZ	E 200	310 Hp
NISSAN	SUNNY	79-175 Hp
MITSUBISHI	LANCER	140 Hp
PERODUA	AMIZHR	106 Hp
MARUTI	ALTO VXI	45-63 Hp

Once these values are integrated, another algorithm is developed to include emission legislation vintage (Pre-ECE to Euro 6. The vehicle's year of manufacture (VehManuYear) is compared against defined regulatory intervals in the EMEP/EEA guidelines, as indicated by Table 4.5, to assign each vehicle to its corresponding emission standard. This assignment is direct, as legislation vintages are defined by year bands, allowing for one-to-one mapping using VehManuYear.

Table 4.5: Implementation years of the European emission standards for ICE passenger cars. Adapted from [3]

Fuel Type	Euro Standard	Start Year	End Year
Petrol	PRE ECE		up to 1971
	ECE 15/00-01	1972	1977
	ECE 15/02	1978	1980
	ECE 15/03	1981	1985
	ECE 15/04	1985	1992
	Improved Conventional	1985	1990
	Open Loop	1985	1990
	Euro 1	1992	1996
	Euro 2	1996	1999
	Euro 3	2000	2004
	Euro 4	2005	2009
	Euro 5	2010	2014
	Euro 6 a/b/c	2014	2019
	Euro 6 d-temp	2019	2020
	Euro 6 d/e	2021	2026
	Euro 7	2026 and later	
Diesel	Conventional		up to 1992
	Euro 1	1992	1996

	Euro 2	1996	2000
	Euro 3	2000	2005
	Euro 4	2005	2010
	Euro 5	2010	2014
	Euro 6 a/b/c	2014	2019
	Euro 6 d-temp	2019	2020
	Euro 6 d/e	2021	2026
	Euro 7	2026 and later	

This marks the achievement of the second objective of this study.

4.2.3 Emission Factor Assignment

Based on the two assigned parameters, each vehicle is matched with pollutant-specific EFs extracted from Appendix. These factors are expressed in grams per kilometre (g/km) and are appended to each vehicle record for baseline emissions calculation for the existing months from January 2016 to October 2024.

4.2.4 Monthly Aggregation of Vehicle-Level Data

A monthly aggregation is performed to transform vehicle-level data into time series data suitable for supervised learning and emissions trend analysis.

1. **Average Annual Mileage Per Vehicle:** Rather than using the sum of annual mileage across vehicles in a given month, the study adopts a normalised approach to account for varying sample sizes of VET tests over time. For each reference month (As explained in Sub-section 4.2.1), the average annual mileage per ICE passenger car was computed by dividing the total annual mileage (for the 12 months starting from reference month) of the passenger car category by the number of matched vehicle records for that reference month, as indicated in Equation 4.1

$$AvgMileage_t = \frac{1}{N_t}\sum_{i=1}^{N_t} AnnualMileage_t \qquad (4.1)$$

where,
$AvgMileage_t$ = average annual mileage for reference month t
N_t = number of valid vehicle matches in the reference month t
$AnnualMileage_t$ = Computed mileage for the i^{th} vehicle

This average annual mileage matched to the reference month, expressed in kilometres per vehicle per year, serves as the target variable for the predictive model developed in Section 5. Prediction of this variable for future months allows calculation of emissions, directly through Tier 02 emission estimation equations under Sub-section 2.1.1, eliminating the requirement of development/training of models for each pollutant.

2. **Fleet Composition by Fuel-Technology-Category:** To capture the underlying structure of the passenger car fleet and account for the variability in emissions across vehicle types, the dataset was enhanced with monthly fleet composition proportions for combined groupings of fuel type, euro technology (Pre-ECE, Euro 1 to Euro 6d), and vehicle category (Mini, Small, Medium, Large-SUV-Executive) [67].

   This allows each monthly observation to be represented by a vector of composition shares that is representative of the national fleet for a yearly span starting from the reference month. These compositions are also utilised in conjunction with their respective EFs to calculate emissions in Chapter 6, as per the Tier 02 emission estimation equation given in the EMEP/EEA guideline.

4.2.5 External Covariates

To enhance the predictive capacity of the dataset beyond individual vehicle characteristics, a selected set of external covariates is incorporated. These variables represent factors known to affect aggregate vehicle usage patterns, such as economic conditions, fleet growth, and other external disruptions. The integration of these covariates allows the predictive model to capture effects that would otherwise be latent.

1. **National Fleet Size:** Monthly records of the total number of registered passenger cars are obtained from the National Transport Commission statistics. This variable reflects the evolving baseline size of the vehicle fleet on the road, serving as an indicator of potential total vehicle activity. Although the predictive target is average annual mileage per vehicle, total fleet size accounts for macro-level scaling effects. It provides context for usage intensity trends relative to fleet growth. Although the active national fleet would provide better context in this regard, there are no Sri Lankan statistics or studies conducted to calculate those values in a temporal context.

2. **Gross Domestic Product (GDP):** Economic activity is a known driver of transport demand [68]. To represent this, quarterly GDP data was sourced from the Central Bank of Sri Lanka reports [69] to form a monthly series. GDP represents aggregate demand for mobility, influencing trip frequency, fuel affordability, and overall travel behaviour. Incorporating GDP provides the model with visibility into macroeconomic cycles that may affect vehicle use even when fleet size and fuel prices remain stable.

These two variables are then averaged across one-year spans to be matched with the reference months for which the annual average mileage is represented.

3. **Disruption Flag (Covid-19 and Economic Crisis):** To reflect the effect of exogenous shocks that significantly altered transport behaviour, a binary flag variable is created. This was assigned a value of 1 during mobility restriction periods, and 0 otherwise. This flag captures periods of lockdown, reduced commuting, fuel shortages, and cost spikes that may have negatively affected vehicle usage. It allows the model to adjust its expectations during known non-linear shifts that would otherwise be treated as unexplained variance.

4.2.6 Forward Projection of Covariates for Scenario Analysis

To enable future predictions of emissions and mileage under different policy conditions under the import ban, forward values for all external covariates (except the target variable) are generated for the prediction horizon. Two distinct scenarios are constructed as follows.

1. **Scenario 01 – Continuation of the Vehicle Imports Ban:** Fleet composition values for future reference months are extended using Holt-Winters triple exponential smoothing, which preserves monthly seasonality patterns [70]. This is adopted since composition is also used in the calculation of total emissions through multiplication with EFs. GDP values are extended using national GDP projections [69]. National fleet growth is projected using the growth rate observed during the period of reference months falling under the import ban.

2. **Scenario 02 – Lifting of the Vehicle Imports Ban:** Shares of Euro 06 compliant vehicles are increased progressively using a custom Python algorithm that applies row-wise increase in proportion to specific fuel–technology–size combinations (Eg: petrol-Euro 6-mini). This algorithm ensures proportionality by normalising rows after scaling, preserving the sum of composition proportions at 1.0. GDP projections are maintained consistent with Scenario 01 for comparability. A fleet growth rate similar to the reference months prior to the import ban is projected assuming that the demand burst is limited due to the phased implementation of the lifting of the import ban and the excise duty at clearance, valued closer to the vehicle value itself.

These projected covariates serve as inputs to the model in Chapter 5, allowing for the simulation of future trends in average mileage and resulting emissions under two policy scenarios.

4.3 Model Input

Table 4.6 includes a summary of all features used for the modelling framework in Chapter 5.

Table 4.6: Summary of final model features

Feature Name	Type	Source	Description/Notes
Average Annual Mileage	Target variable	Derived from VET dataset	Average one-year mileage per vehicle within a reference month, computed from matched odometer readings
Composition	Feature (numeric)	Engineered from VET data	Proportional fleet composition by fuel type, emission standard, and vehicle category combinations
Fleet Size	Feature (numeric)	National Transport Commission	Monthly total registered passenger cars, averaged annually per reference month
GDP	Feature (numeric)	Central Bank of Sri Lanka	Quarterly GDP, averaged annually per reference month
Disruption Flag	Feature (binary)	Manually coded	Binary flag for COVID-19 and economic crisis periods
Month	Feature (categorical/time)	Derived	Month indicator used for seasonal pattern recognition in modelling

CHAPTER 5

METHODOLOGY II - MODEL DEVELOPMENT AND EVALUATION

This chapter includes the development of a LSTM neural network model to predict average annual mileage per ICE passenger car, based on reference months. Building upon the structured dataset derived in Chapter 4, the model incorporates both internal fleet composition indicators and external variables to capture temporal trends in VKT.

The chapter presents the structure of the LSTM model, the selected hyperparameters, and the preprocessing pipeline applied to transform the dataset into a suitable input format for training. It also describes the postprocessing steps used to interpret model outputs and the evaluation techniques employed to validate model performance. Scenario-based inputs are used to forecast future trends, which are then carried forward for emissions estimation in Chapter 6.

5.1 LSTM Network Architecture and Hyperparameters

The network architecture is implemented using TensorFlow's Keras API and is designed for recursive multi-step forecasting, where predictions are generated one step at a time, with each prediction used as input for the next time step. The model uses a stacked LSTM configuration, as explained in Sub-section 3.4.2. It is designed for univariate regression based on multivariate time-series input, with two LSTM layers and a dense output layer, implemented sequentially. Each component of the model architecture is described as follows.

Input Layer: The model accepts input as a 3D tensor with the shape (batch_size, sequence_length, num_features) where,

- batch_size = Number of samples processed per training step (dynamically determined during training)
- sequence_length = Number of prior time steps considered (set to 12 months)
- num_features = number of input variables per time (93, including vehicle composition, GDP, fleet size, disruption flags, and seasonal encodings).

This format allows the model to observe temporal patterns across multiple variables in a fixed time window.

LSTM Layer 01: This layer learns temporal dependencies across the input sequence, and returns a sequence of hidden states (one for each time step). It consists of 16 memory cells that are used to capture internal representations of the sequence. Tanh (Hyperbolic Tangent) activation function is used for the output gate, a standard choice in LSTM architectures to allow output values between -1 and 1 [43]. The output retains a sequence format (shape: (batch_size,12,16)), which is required for stacking another LSTM layer.

LSTM Layer 02: This layer processes the output of the previous LSTM and returns a single vector corresponding to the last time step, summarising the entire input sequence. It has 16 memory cell units and uses a tanh activation function similar to the previous layer, but only outputs the final hidden state (shape: (batch_size,16)).

Dropout Layers: Dropouts are typically used to prevent overfitting by randomly deactivating a fraction of neurons during training [71]. But due to the limited temporal depth of the input parameters (monthly inputs from January 2016 to October 2024), the dropout rate is set to zero for each layer.

Dense Output Layer: This fully connected layer maps the final LSTM output vector to a single numeric value, which is the predicted average annual mileage for the upcoming reference month (shape: (batch_size, 1)).

The LSTM model is trained using the hyperparameters detailed in Table 5.1, which are selected through manual tuning and empirical validation to balance model complexity, generalisation ability, and training efficiency.

Table 5.1: LSTM model hyperparameters

Hyperparameter	**Value**	**Description**
Sequence Length	12	The number of past time steps (months) considered in each sequence.
LSTM Units	16 per layer	Number of hidden units in each of the two stacked LSTM laters
Activation Function	tanh	Used in both LSTM layers. Squashes outputs between -1 and 1
Dropout rate	0	No dropout applied
Optimiser	Adam	Adaptive learning rate optimiser which is commonly used
Learning Rate	0.005	Controls the step size during weight updates. Tuned for convergence stability
Loss Function	Mean Absolute Error (MAE)	Measures the average magnitude of prediction errors. Less sensitive to outliers than MSE.
Epochs	50	Maximum number of training iterations over the dataset
Batch Size	12	Number of samples processed before each weight update
Restore Best Weights	True	Ensures the model retains the best-performing parameters.

Each hyperparameter is chosen based on the characteristics of the dataset and the recursive forecasting objective. The sequence length of 12 months is selected to capture seasonal effects and maintain a manageable input size [72]. The 16 LSTM units in each layer allow the model to learn nonlinear patterns without overfitting. The Adam optimiser [73] with a learning rate of 0.005 is chosen after extensive experimentation with different configurations to find optimum performance with a balance between stable convergence and reasonable training speed. Although dropout layers are included for architectural flexibility, they are disabled (rate = 0) to avoid overfitting with the limited temporal depth of the training data [71].

5.2 Data Preparation and Preprocessing

As preparatory steps for model training, the input dataset discussed in Chapter 4 undergoes a final preprocessing pipeline to prepare it for LSTM time series modelling. This process involves target smoothing, feature selection, time-aware scaling, and the construction of supervised learning sequences. Additional features are introduced to account for policy shifts and seasonality to ensure the model has access to real-world contextual signals during training and forecasting.

Target Transformation: The target variable for prediction, which is the average annual mileage per ICE passenger car on a reference month basis, is smoothed using a three-month moving average. This operation reduces the influence of short-term noise while preserving long-term trends [47]. This smoothed series (*Mileage_MA*) is used as the regression target throughout the modelling process.

Feature Selection and Temporal Feature Engineering: From the full input dataset, a set of 93 features are selected by removing irrelevant identifiers. The retained features include,

- Monthly proportions of each fuel–technology–size vehicle category
- National fleet size and GDP (averaged over 1-year windows)
- Seasonal encodings (Month_sin, Month_cos)

  Since the LSTM does not have an inherent awareness of time cycles such as months or seasons, sinusoidal transformations of the month variable are used to capture cyclical seasonality, as indicated by Equations 5.1 and 5.2

$$Month\_sin = sin\left(\frac{2\pi \times month}{12}\right) \tag{5.1}$$

$$Month\_cos = cos\left(\frac{2\pi \times month}{12}\right) \tag{5.2}$$

  This transformation encodes each month on the unit circle, allowing the model to learn smooth seasonal trends across months [74] (Eg. December to January is not a sharp jump as it would be with integer encoding).

- Binary disruption flag as a disruption-aware dummy variable (Scenario_Dummy)

Feature and Target MinMax Scaling: Due to the use of MAE as the loss function, there is an inherent scaling issue due to large differences in feature magnitudes. This results in the error calculation being dominated by features with larger values, diminishing the effect of smaller values. Min-Max scaling effectively standardises data to a specific range [0,1], preserving relative relationships between features and reducing the influence of outliers [75]. A single scaler (scaler_features) is fitted on the combined matrix of historical and projected feature values across both scenarios. This approach ensures consistent normalisation for all inputs during forecasting. A separate scaler (scaler_target) is fitted only on the historical target values to avoid data leakage from future periods.

Sequence Construction for Supervised Learning: To reframe the time series data for supervised learning, the historical data is converted into fixed-length input sequences and associated single-step targets using a custom function, as indicated by Figure 5.1.

```
def create_sequences_recursive(data, sequence_length):
    X, y = [], []
    for i in range(sequence_length, len(data)):
        X.append(data[i-sequence_length:i, :]) # 12-step input
        y.append(data[i, -1]) # 1 step ahead output
    return np.array(X), np.array(y)
```
Figure 5.1: Custom function for sequence creation

Each sample X_i consists of 12 consecutive months of scaled features (shape: 12×93), and the corresponding target y_i is the annual average mileage value of the next reference month. This format enables the LSTM to learn temporal patterns and generate one-step-ahead predictions, which are later repeated recursively during forecasting.

Chronological Train-Test Split: To preserve the temporal structure of the dataset, a chronological split is used to separate training and validation data, with a 90:10 split ratio. The training set includes the first 90% of all constructed sequences, while the validation set contains the final 10% corresponding to the most recent reference months before the forecast horizon. This split makes sure that all training chronologically precedes validation, eliminating any lookahead bias. The resulting arrays are passed directly to the model without shuffling to preserve the sequential characteristics needed for LSTM training.

5.3 Recursive Forecasting and Postprocessing

5.3.1 Forecasting Configuration

After the model training process, the LSTM is utilised in a recursive forecasting configuration to generate predictions of average annual mileage matched onto future reference months, across a predefined forecast horizon.

At the end of the historical dataset, the last 12-month sequence of scaled inputs (both features and the target) is extracted. This functions as the initial input for the model's recursive inference. For each future reference month, he model receives the current 12-month input sequence and generates the next annual average mileage in scaled form. This prediction is appended to the input sequence along with the corresponding scenario-specific feature values (vehicle composition, GDP, fleet size) as defined under Sub-section 4.2.6. The updated sequence is then used as input for the next reference month's prediction. This is repeated for the entire forecast horizon

5.3.2 Postprocessing and Output Transformation

The postprocessing involves converting the LSTM outputs, which are in scaled form, back to their original units (in km) and aligning with the respective reference months. Then the results are prepared for visualisation, evaluation and emissions estimation.

Inverse Transformations for Predicted Mileage: The model's predicted outputs are reshaped and passed through the previously fitted MinMax scaler for the target variable. This reverses the normalisation and restores the values to their original scale. This is done for validation and training outputs, as well as future forecasts under each scenario. These rescaled predictions are stored as flat arrays and matched with the corresponding months of the forecast horizon. The final recursive mileage predictions for both Scenario 1 and Scenario 2 are saved as monthly sequences, beginning from the first forecasted month immediately after the validation set.

5.4 Model Evaluation and Validation

To assess the performance of the trained LSTM model, evaluation is conducted on both the training and validation sets using standard metrics. The validation set, representing the most recent 10% of the dataset prior to the forecast horizon, is used to simulate the model's performance on unseen data. During training, the verbose parameter is set to 1 (verbose = 1), allowing the loss value to be recorded for each epoch, alongside the evaluation loss from the validation set. Convergence and stability of the training and validation error values across epochs indicate that the model has a good fit [76].

The following evaluation metrics are used to evaluate the model's predictive accuracy.

Mean Absolute Error (MAE): Represents the average absolute difference between actual and predicted mileage values. MAE is straightforward and accounts for outliers.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5.3}$$

Mean Squared Error (MSE): Penalises larger errors more severely, indicating volatility and prediction stability.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.4}$$

Coefficient of Determination (R^2): Measures the proportion of variance in the actual values explained by the predictions. A value closer to 1 indicates strong model performance.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \tag{5.5}$$

This indicates the completion of the third objective of the study.

CHAPTER 6

RESULTS

This chapter focuses on the results of the LSTM model developed in the previous chapter and interprets the outputs to establish emission baselines and trajectories for Sri Lanka. It establishes the model performance using the metrics described in Section 5.4, demonstrating its predictive accuracy, while utilising the two defined scenarios to predict pollutant-wise emissions predictions that reflect the effect of the continuation or lifting of the import ban for vehicles.

6.1 Training Results and Model Validation

To monitor convergence and avoid overfitting, the model is trained with 'restore best weights' enabled, tracking MAE on both training and validation sets over 50 epochs. According to the results presented in Figure 6.1, the loss curves show a steady decline in MAE on both sets, with diminishing signs of divergence, indicating stable learning throughout training. The late-stage convergence indicates that the model has captured the underlying patterns effectively without overfitting to the training data.

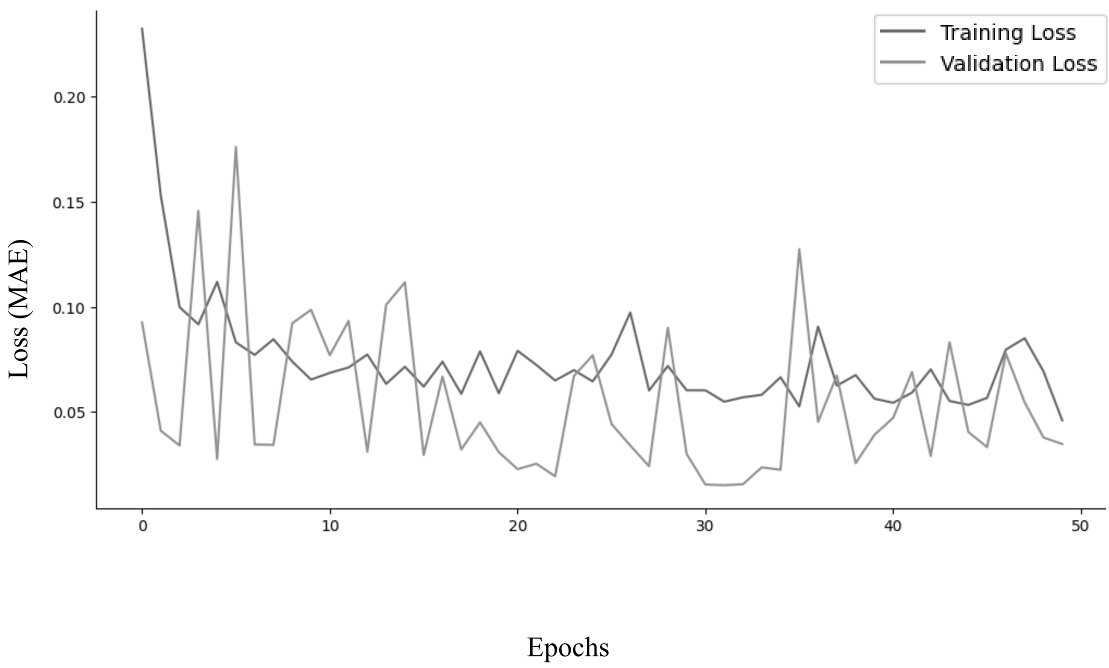


Figure 6.1: Training vs. validation loss (MAE)

6.2 Model Performance Evaluation

With the chronological split that was imposed on the dataset as per Section 5.2, the validation dataset period is used for model validation purposes. Figure 6.2 indicates the alignment between the actual and predicted annual average mileage values on the reference months falling under the validation split.

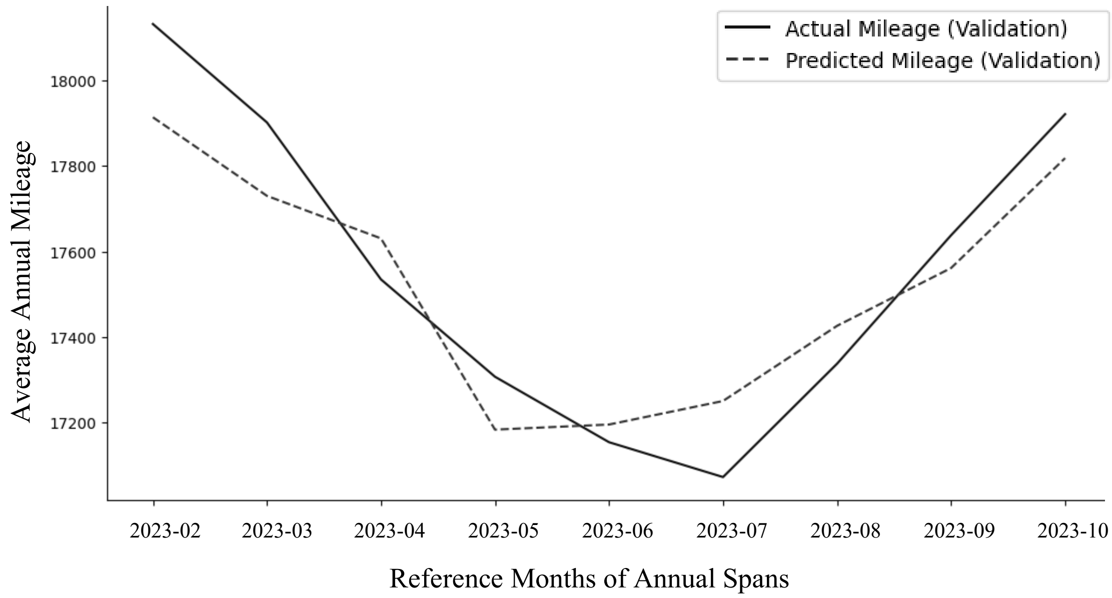


Figure 6.2: Actual vs. predicted mileage (Validation set)

The validation dataset is used to evaluate the prediction capability of the model on unseen data, with the previously defined evaluation metrics giving the following results

MAE = 121.81 MSE = 17675.97 $R^2 = 0.8548$

Collectively, these evaluation metrics suggest that the developed LSTM model demonstrates strong performance in forecasting mileage on unseen data. The MAE value suggests that, on average, the model's predictions for mileage on the validation data are off by approximately 121.81 km (around 0.69% considering the input value range), demonstrating high accuracy in prediction. Although the MSE values are higher than the MAE due to the squaring of errors, it is within a reasonable range considering the scale of the mileage data. The R^2 value obtained signifies that approximately 85.48% of the variance in the actual annual average mileage values on the validation set can be explained by the model's predictions. The obtained value, which is on the closer end to one (1.0), indicates that the model provides a good fit to the validation data and is capable of explaining a substantial portion of the variability in average annual mileage.

6.3 Scenario-Based Forecasting Results

Once validated, recursive prediction is used as described in Sub-section 5.3.1 to generate forecasts under the two distinct scenarios of policy decisions related to the vehicle import ban as indicated in Figure 6.3. Scenario 01 assumes a continued import restriction, reflecting a constrained vehicle composition and fleet size, while scenario 02 assumes the reopening of the import market, resulting in increased proportions of newer vehicles, particularly Euro 6-compliant categories.

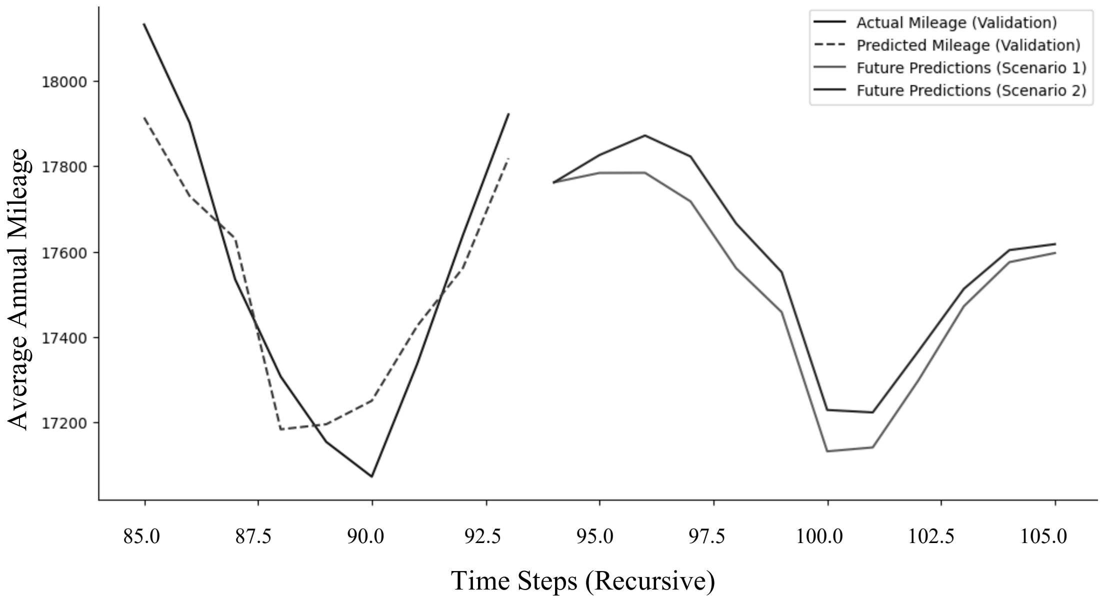


Figure 6.3: Average annual mileage predictions (Validation and recursive prediction)

The observed variation across the prediction horizon can be justified by observing the overall trend of the target variable across the training and validation sets that precede it, as observed in Figure 6.4.

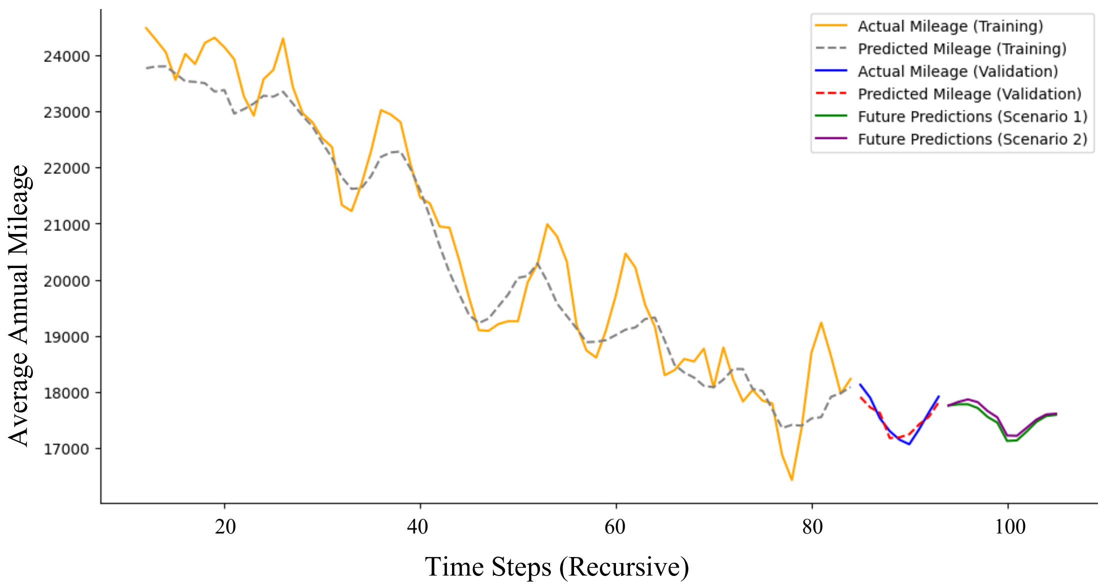


Figure 6.4: Average annual mileage (Training, Validation and Forecast Horizon)

The prediction horizon follows the underlying trend observed in the training stage, maintaining a gradually reducing decline in the average annual mileage. The values revolve around the 17,600 km mark, with the two scenarios diverging as the forecast moves forward. The peak-and-trough nature of the values of the prediction horizon is also reasonable, considering the overall value variation during training. Since the target variable indicates vehicle activity, the results suggest that under scenario 2, the annual mileage could increase up to 21 km per vehicle (while having a higher fleet size) at the end of the prediction horizon. The observed divergence in the forecasts denotes that the LSTM model has identified policy cues with the feature values that have been provided, adjusting the predictions accordingly. The model is able to capture nonlinear responses to feature changes, where fleet upgrades do not simply maintain standard behaviour, but shift the prediction into new dynamics.

6.4 Tier 02 Emissions Estimation

For both scenarios, the average annual emission values fixed onto reference months are converted into pollutant-specific emission values using the Tier 02 emissions estimation methodology, using the EMEP/EEA guidebook. Annual spanning emissions based on reference months for each pollutant are calculated using Equation 6.1,

$$E_{p,t} = \left[M_t \sum_{t=1}^{n} (EF_{i,p} \times Prop_{i,t}) \right] \times N_t \tag{6.1}$$

$E_{p,t}$ = Total annual spanning emissions of pollutant *p* under reference month *t*

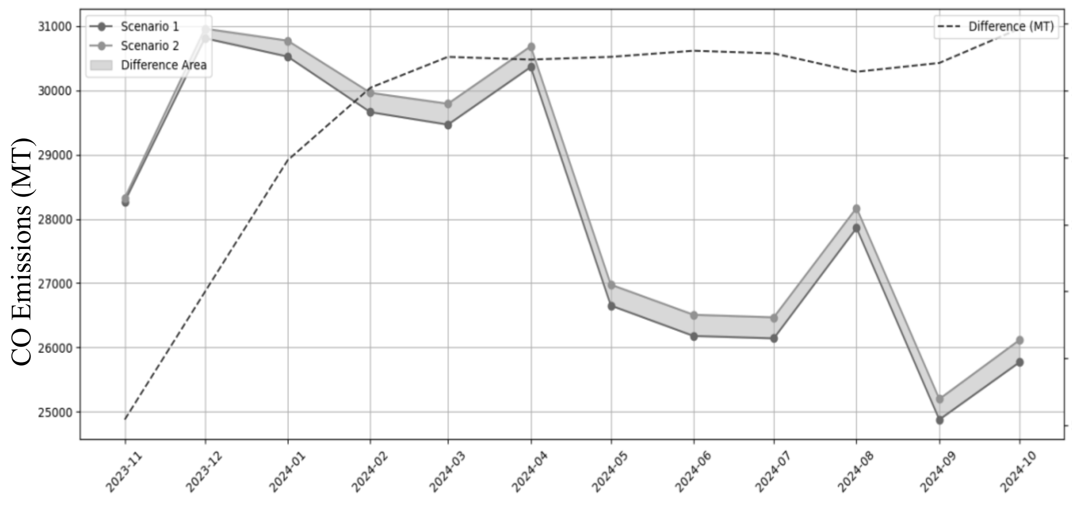
M_t = Annual average mileage for passenger cars for annual span with reference month *t [km/veh]*

$EF_{i,p}$ = Technology-specific emission factor of pollutant *p* for fuel-technology-category combination *i [g/veh-km]*

$Prop_{i,t}$ = Fleet share (proportion) of fuel-technology-category combination *i* for reference month *t*

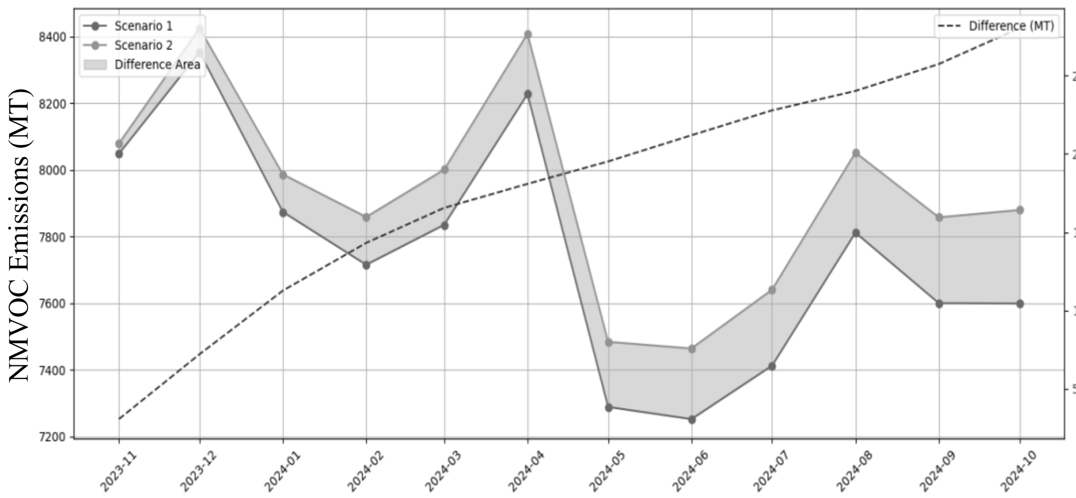
N_t = National fleet for the corresponding annual span with reference month *t*

The following combination of tables and figures illustrates the comparative variation of each pollutant across the annual spanning ranges of the 12-month forecast horizon, with the reference months fixed as the first month of each annual span. (E.g. 2023-08 represents the annual span from 2023-08 to 2024-07). For each figure, a deviation is observed, with scenario 02 producing higher emissions as the annual spans progress forward. The figures provide the magnitude of the deviation for each pollutant as well.



Reference Months of Annual Spans

Figure 6.5: Comparative CO emissions (Metric tons)



Reference Months of Annual Spans

Figure 6.6: Comparative NMVOC emissions (Metric tons)

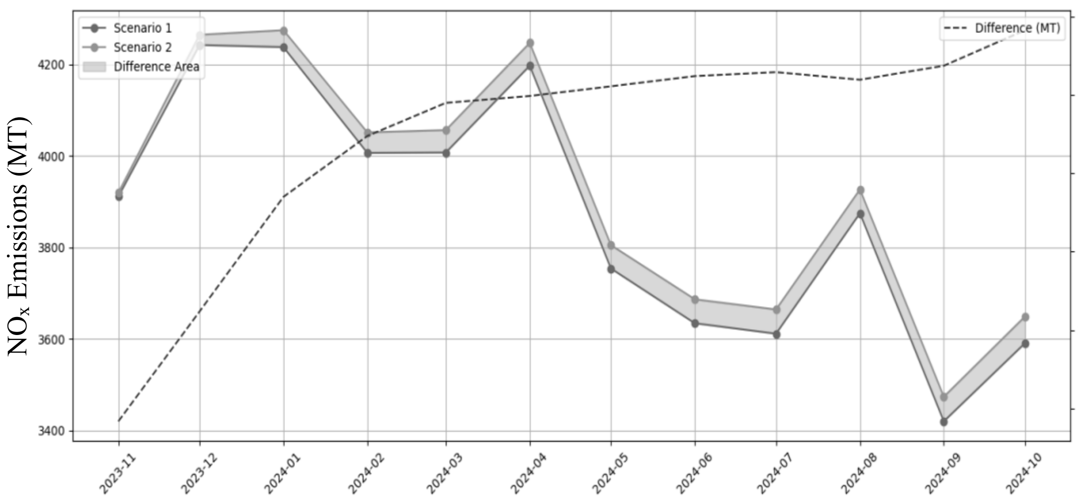


Figure 6.7: Comparative NO_x emissions (Metric tons)

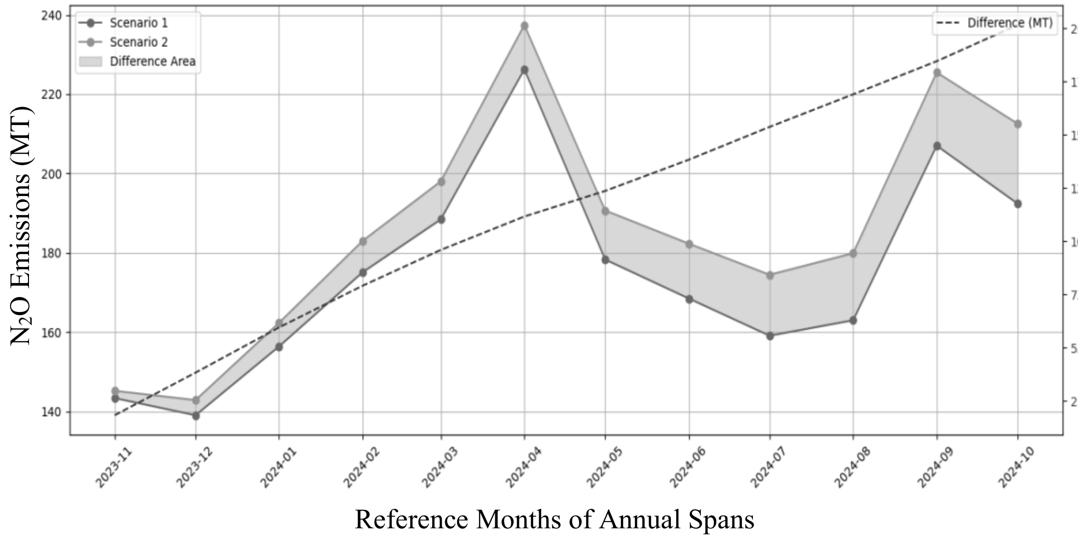


Figure 6.8: Comparative N_2O emissions (Metric tons)

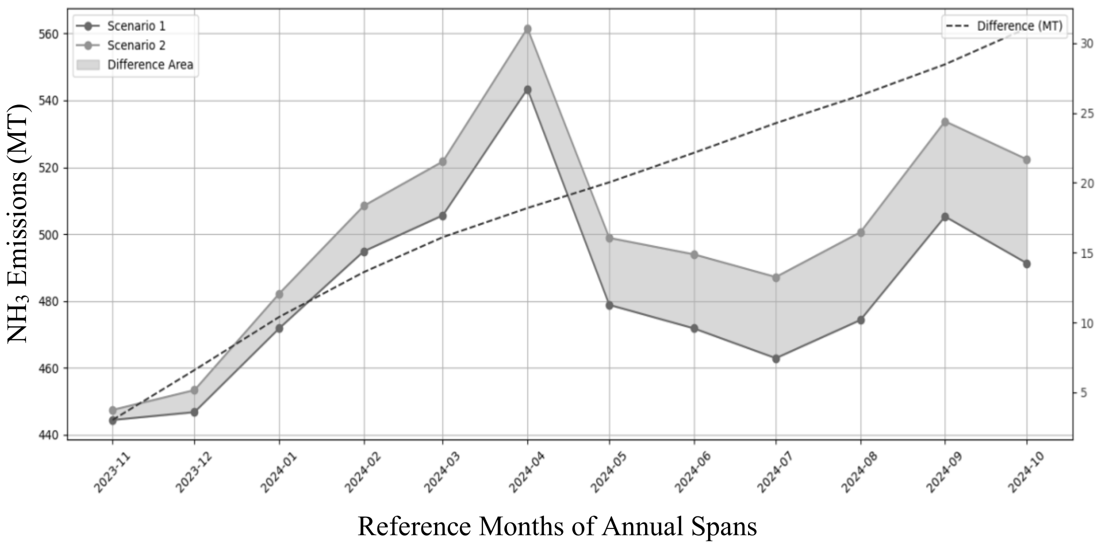


Figure 6.9: Comparative NH_3 emissions (Metric tons)

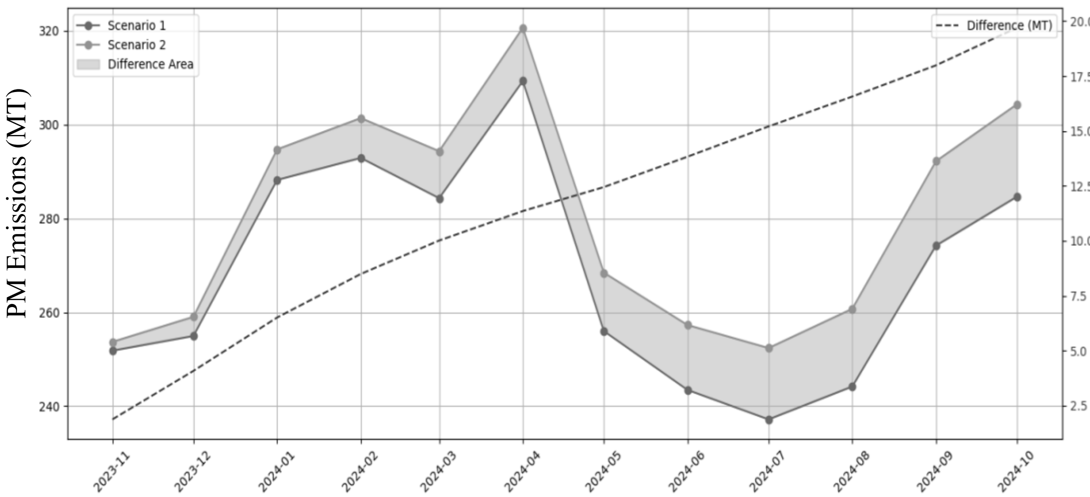


Figure 6.10: Comparative PM emissions (Metric tons)

Table 6.1: Summary of total emissions under scenario 01

Year-Span	Pollutant-Specific Emissions (Metric Tons)					
	CO	NMVOC	NO_x	N_2O	NH_3	PM
2023-11 – 2024-10	28,271.43	8,047.87	3,911.48	143.38	444.40	251.83
2023-12 – 2024-11	30,812.35	8,352.51	4,242.89	139.05	446.74	254.96
2024-01 – 2024-12	30,526.64	7,873.42	4,238.22	156.42	471.77	288.22
2024-02 – 2025-01	29,660.85	7,715.09	4,006.93	175.10	494.84	292.90
2024-03 – 2025-02	29,467.22	7,834.80	4,007.74	188.43	505.61	284.34
2024-04 – 2025-03	30,363.62	8,226.81	4,197.61	226.28	543.40	309.29
2024-05 – 2025-04	26,651.28	7,288.67	3,754.02	178.31	478.82	256.01
2024-06 – 2025-05	26,177.38	7,252.18	3,634.23	168.49	471.79	243.47
2024-07 – 2025-06	26,139.74	7,412.55	3,611.47	159.10	462.88	237.19
2024-08 – 2025-07	27,857.84	7,811.43	3,875.00	163.00	474.30	244.19
2024-09 – 2025-08	24,875.47	7,599.98	3,419.75	207.02	505.25	274.22
2024-10 – 2025-09	25,768.46	7,598.92	3,590.86	192.41	491.26	284.65

Table 6.2: Summary of total emissions under scenario 02

Year-Span	Pollutant-Specific Emissions (Metric Tons)					
	CO	NMVOC	NO_x	N_2O	NH_3	PM
2023-11 – 2024-10	28,325.64	8,078.59	3,919.83	145.21	447.41	253.71
2023-12 – 2024-11	30,961.95	8,424.63	4,265.26	142.87	453.33	259.04
2024-01 – 2024-12	30,774.66	7,986.14	4,275.21	162.37	482.16	294.73
2024-02 – 2025-01	29,963.11	7,858.26	4,051.70	183.01	508.44	301.40
2024-03 – 2025-02	29,792.50	8,000.34	4,056.73	198.02	521.73	294.35
2024-04 – 2025-03	30,686.88	8,407.66	4,247.47	237.43	561.59	320.65
2024-05 – 2025-04	26,976.57	7,483.85	3,805.12	190.67	498.86	268.45
2024-06 – 2025-05	26,507.26	7,464.06	3,686.64	182.32	493.94	257.30
2024-07 – 2025-06	26,467.61	7,640.34	3,664.37	174.46	487.16	252.39
2024-08 – 2025-07	28,171.98	8,051.59	3,926.94	179.90	500.57	260.76
2024-09 – 2025-08	25,196.23	7,857.30	3,473.48	225.47	533.74	292.21
2024-10 – 2025-09	26,114.87	7,879.18	3,649.12	212.57	522.35	304.35

The results showcase how the obtained annual average mileage values produce significant shifts in total emissions under the two scenarios. At the final yearly span of the forecast horizon, an increase of 1.34% (CO), 3.68% (NMVOC), 1.62% (NO_x), 10.5% (N_2O), 6.33% (NH_3) and 6.92% (PM) can be observed, which has a significant impact on emissions on a national scale, as observed by the two tables. These graphs can directly be incorporated into policy decisions depending on the emissions target for each pollutant. These results also indicate the achievement of the fourth objectives of the study.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 Conclusions

In this study, a combination of predictive time-series modelling and scenario analysis is used to gain insight into how transport emissions in Sri Lanka are likely to evolve under different policy conditions. By combining LSTM-based mileage forecasting with Tier 02 emissions estimation, the study presents a scalable framework that connects data-driven forecasting and national-scale policy strategies, implemented for ICE passenger cars as a case study. The developed methodology is economical, as it sources raw data from an existing national database, which is otherwise underutilised. The developed model has the capability of predicting annual spanning emissions on a monthly temporal depth, up to a forecast horizon of 12 months from the available data.

The LSTM model developed showcases high predictive capability through evaluation and validation measures, with low validation error and stable learning curves. The forecasted mileage trends capture realistic divergences under the considered policy scenarios. The model has been able to produce reference month-based estimates for six key pollutants, CO, NMVOC, NO_x, N_2O, NH_3, and PM, identified under the EMEP/EEA guidelines. Lead (Pb) has been disregarded as Sri Lanka uses 100% lead-free fuel. In terms of mass, CO constitutes the largest share of total emissions, followed by NMVOC and NO_x as products of partial and high-temperature combustion processes, respectively. Other pollutants, such as PM, although comparatively lower in terms of contribution towards emissions, are linked to respiratory and cardiovascular diseases [77]. The noticeable increase in these pollutants under Scenario 02 raises concerns about fleet expansion without substantial emission reduction strategies. Observing the emission predictions obtained under the two scenarios, lifting of the import ban, which introduces newer Euro 6 vehicles, consistently shows higher emissions across all pollutants considered. Although these vehicles are cleaner, increased fleet size combined with higher vehicle activity effects indicate that fleet modernisation doesn't necessarily lead to reduced emissions. This also highlights the necessity of policy-making decisions to incorporate total vehicle activity and travel behaviour of vehicles for better effective results.

These results provide the first Tier 02 level emissions outlook under Sri Lanka's post-2019 import ban conditions, with no prior studies to compare vehicle-class specific results. The framework implemented for ICE passenger cars can be replicated and scaled up for other vehicle classes, and for hybrid and electric vehicles if a comprehensive database is available. Further, the study delivers a methodology that produces more accurate emission estimations with an ever-evolving and readily available dataset, which has previously been underutilised. Additionally, the methodology allows the mapping of EFs related to any vehicle within the active vehicle fleet in Sri Lanka, based on the fuel-technology-category characteristics. The web-scraping algorithm that has been developed can be used to import any additional vehicle characteristics as per the requirement (power output in this case). The study has managed to complete the six objectives that have been set initially, with necessary steps validated through standard metrics. This places the approach used as an accurate and validated methodology for tier 02 emissions estimation on a national scale.

7.3 Limitations and Future Work

One of the main limitations of the study is the lack of data for EVs and hybrids, as they are omitted from the VET test procedure. This is why the scope has been defined to only include ICE passenger cars. However, the EMEP/EEA guideline has specific EFs for hybrid vehicles if a potential database on a national scale may be created. The same methodology used in the study can be implemented to accommodate those vehicle types as well. The lack of detailed active vehicle fleet data for Sri Lanka can also be considered as a limitation, and would be a better alternative input variable in place of the national fleet. This also negates the possibility of removing the EV and hybrid vehicles from the national fleet count. Even the VET database itself is still in its relative infancy, with properly maintained records spanning less than a decade. In future iterations of this study, the dataset will get stronger from a time-series standpoint, with more data points for the prediction model, resulting in higher sequence lengths, leading to longer forecast horizons.

The primary improvement that can be made to the study is the localisation of the EFs. This involves the development of driving cycles and for required vehicle classes, and chassis dynamometer testing with pollutant measurement for fuel-technology-category-specific vehicles in order to obtain a comprehensive set of EFs for each combination. With the comparison of EFs of similar vehicle segments under subsection 2.1.3, where EMEP/EEA and ICAP EFs revolve in a similar range for CO and NO_x, it is safe to assume that the Sri Lankan EFs to be developed will also be in a similar range, thus giving similar results for emission estimation by the multiplication with the annual mileage values. Another economical option is to convert and match global driving cycles such as the World Harmonised Light Vehicle Test Cycle (WLTC) [78], and interpolate the EFs developed based on them. Tier 03 level estimations can also be explored through the integration of mobile phone/GPS data to model trip-level data. Although unrealistic to be implemented on a national scale, sampling strategies can be used to identify patterns of various vehicle types and travel activity. Improvement of the model itself is another avenue to be explored, as newer evolutions of the LSTM concept, such as Gated Recurrent Units (GRUs), have now been utilised. GRUs, which have a two-gate structure (reset and update), are computationally efficient and have been found to give better results with limited datasets [79]. Direct emission estimation, in place of the two-step estimation implemented in the study, can also be explored. This requires training of pollutant-specific LSTM models, which may showcase better predictive accuracy. Other options include transformer-based models for long horizon predictions, and bayesian LSTMs for uncertainty quantification. Finally, the inclusion of the binary scenario flag can be extended into a multi-factor policy simulation layer, which allows combinations of incentives, restrictions, or investments to be tested.

References

[1] N. Singh, T. Mishra, and R. Banerjee, 'Projection of Private Vehicle Stock in India up to 2050', *Transp. Res. Procedia*, vol. 48, pp. 3380–3389, Jan. 2020, doi: 10.1016/j.trpro.2020.08.116.

[2] 'Statistics Report 2023', National Transport Commission NTC. Accessed: May 16, 2024. [Online]. Available: https://www.ntc.gov.lk/corporate/pdf/2023/statistics_Report/Stat_2023_EN.pdf

[3] '1.A.3.b.i-IV road transport', European Environment Agency, 2024. Accessed: Nov. 16, 2024. [Online]. Available: https://www.eea.europa.eu/publications/emep-eea-guidebook-2023/part-b-sectoral-guidance-chapters/1-energy/1-a-combustion/1-a-3-b-i/view

[4] *Carbon Net Zero 2050 Roadmap and Strategic Plan*. Ministry of Environment, 2023. Accessed: Jan. 15, 2025. [Online]. Available: https://env.gov.lk/web/images/pdf/divisions/climate_change_division/publications/2023/Synthesis_Report_for_Carbon_Net_Zero_2050__Roadmap_and_Strategic_Plan.pdf

[5] E. Urda, B. Daniel, and N. Victoria, 'COMPENDIUM ON GREENHOUSE GAS BASELINES AND MONITORING: PASSENGER AND FREIGHT TRANSPORT', United Nations Framework Convention on Climate Change, 2. Accessed: Nov. 16, 2024. [Online]. Available: https://unfccc.int/sites/default/files/resource/Transport_0.pdf

[6] V. Marion, 'Bottom-Up GHG Inventory and MRV of Measures', Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, May 2017. Accessed: June 11, 2024. [Online]. Available: https://changing-transport.org/wp-content/uploads/2017_ViewegMersman_Bottom-Up_GHG_Inventory.pdf

[7] Department of Motor Traffic, Ministry of Transport and Highways, Ministry of Environment, and Ministry of Health, 'VET – Vehicular Emission Test Trust Fund'. Accessed: Sept. 17, 2025. [Online]. Available: https://vet.lk/

[8] K. R. Gurney, Y. Song, J. Liang, and G. Roest, 'Toward Accurate, Policy-Relevant Fossil Fuel CO_2 Emission Landscapes', *Environ. Sci. Technol.*, vol. 54, no. 16, pp. 9896–9907, Aug. 2020, doi: 10.1021/acs.est.0c01175.

[9] A. R. M. Arshard and M. S. M. Imthiyas, 'Assessing the Consequences of Vehicle Import Bans on Sri Lanka's', *Int. J. Bus. Law Polit. Sci.*, vol. 2, no. 2, pp. 17–23, Feb. 2025, doi: 10.61796/ijblps.v2i2.271.

[10] Ministry of Finance Sri Lanka, 'Imports and Exports (Control) Regulations No. 02 of 2025', Mar. 11, 2025. Accessed: Apr. 23, 2025. [Online]. Available: https://www.gazette.lk/dl/Gazette/01/822436012-Extraordinary-Gazette-Imports-and-Exports-Control-Regulations-No-02-of-2025.pdf

[11] S. S. Ravi, S. Osipov, and J. W. G. Turner, 'Impact of Modern Vehicular Technologies and Emission Regulations on Improving Global Air Quality', *Atmosphere*, vol. 14, no. 7, p. 1164, July 2023, doi: 10.3390/atmos14071164.

[12] H. Zhong, K. Chen, C. Liu, M. Zhu, and R. Ke, 'Models for predicting vehicle emissions: A comprehensive review', *Sci. Total Environ.*, vol. 923, p. 171324, May 2024, doi: 10.1016/j.scitotenv.2024.171324.

[13] B. Ghobadian, H. Rahimi, A. M. Nikbakht, G. Najafi, and T. F. Yusaf, 'Diesel engine performance and exhaust emission analysis using waste cooking biodiesel fuel with an artificial neural network', *Renew. Energy*, vol. 34, no. 4, pp. 976–982, Apr. 2009, doi: 10.1016/j.renene.2008.08.008.

[14] J. Mohammadhassani, Sh. Khalilarya, M. Solimanpur, and A. Dadvand, 'Prediction of NO x Emissions from a Direct Injection Diesel Engine Using Artificial Neural Network', *Model. Simul. Eng.*, vol. 2012, pp. 1–8, 2012, doi: 10.1155/2012/830365.

[15] I. Arsie, D. Marra, C. Pianese, and M. Sorrentino, 'Real-Time Estimation of Engine NOx Emissions via Recurrent Neural Networks', *IFAC Proc. Vol.*, vol. 43, no. 7, pp. 228–233, July 2010, doi: 10.3182/20100712-3-DE-2013.00117.

[16] L. Alfaseeh, R. Tu, B. Farooq, and M. Hatzopoulou, 'Greenhouse gas emission prediction on road network using deep sequence learning', *Transp. Res. Part Transp. Environ.*, vol. 88, p. 102593, Nov. 2020, doi: 10.1016/j.trd.2020.102593.

[17] Y. Wang, P. Liu, C. Xu, C. Peng, and J. Wu, 'A deep learning approach to real-time CO concentration prediction at signalized intersection', *Atmospheric Pollut. Res.*, vol. 11, no. 8, pp. 1370–1378, Aug. 2020, doi: 10.1016/j.apr.2020.05.007.

[18] S. Khurana, S. Saxena, S. Jain, and A. Dixit, 'Predictive modeling of engine emissions using machine learning: A review', *Mater. Today Proc.*, vol. 38, pp. 280–284, 2021, doi: 10.1016/j.matpr.2020.07.204.

[19] G. Geilenkirchen, M. Bolech, J. Hulskotte, S. Dellaert, N. Ligterink, and E. van Eijk, 'Methods for calculating the emissions of transport in the Netherlands', Rijksinstituut voor Volksgezondheid en Milieu RIVM, 2024. doi: 10.21945/RIVM-2024-0023.

[20] V. Franco, M. Kousoulidou, M. Muntean, L. Ntziachristos, S. Hausberger, and P. Dilara, 'Road vehicle emission factors development: A review', *Atmos. Environ.*, vol. 70, pp. 84–97, May 2013, doi: 10.1016/j.atmosenv.2013.01.006.

[21] G. Fontaras, V. Franco, P. Dilara, G. Martini, and U. Manfredi, 'Development and review of Euro 5 passenger car emission factors based on experimental results over various driving cycles', *Sci. Total Environ.*, vol. 468–469, pp. 1034–1042, Jan. 2014, doi: 10.1016/j.scitotenv.2013.09.043.

[22] P. Adak, R. Sahu, and S. P. Elumalai, 'Development of emission factors for motorcycles and shared auto-rickshaws using real-world driving cycle for a typical Indian city', *Sci. Total Environ.*, vol. 544, pp. 299–308, Feb. 2016, doi: 10.1016/j.scitotenv.2015.11.099.

[23] J. Prakash and G. Habib, 'A technology-based mass emission factors of gases and aerosol precursor and spatial distribution of emissions from on-road transport sector in India', *Atmos. Environ.*, vol. 180, pp. 192–205, May 2018, doi: 10.1016/j.atmosenv.2018.02.053.

[24] J. Davison *et al.*, 'Distance-based emission factors from vehicle emission remote sensing measurements', *Sci. Total Environ.*, vol. 739, p. 139688, Oct. 2020, doi: 10.1016/j.scitotenv.2020.139688.

[25] P. J. Pérez-Martínez *et al.*, 'Emission factors of air pollutants from vehicles measured inside road tunnels in São Paulo: case study comparison', *Int. J. Environ. Sci. Technol.*, vol. 11, no. 8, pp. 2155–2168, Nov. 2014, doi: 10.1007/s13762-014-0562-7.

[26] X. Li, T. R. Dallmann, A. A. May, and A. A. Presto, 'Seasonal and Long-Term Trend of on-Road Gasoline and Diesel Vehicle Emission Factors Measured in Traffic Tunnels', *Appl. Sci.*, vol. 10, no. 7, p. 2458, Apr. 2020, doi: 10.3390/app10072458.

[27] B. Giechaskiel *et al.*, 'Vehicle Emission Factors of Solid Nanoparticles in the Laboratory and on the Road Using Portable Emission Measurement Systems

(PEMS)', *Front. Environ. Sci.*, vol. 3, Dec. 2015, doi: 10.3389/fenvs.2015.00082.

[28] 'Emission Factor development for Indian Vehicles', The Automotive Research Association of India, AFL/2006-07/IOCL, Aug. 2007. Accessed: Mar. 15, 2024. [Online]. Available: https://learning-cleanairasia.org/lms/library/ga2/07-Air-Quality-Monitoring-Project-Indian-Clean-Air-Programme-(ICAP).pdf

[29] T. V. Ramachandra and Shwetmala, 'Emissions from India's transport sector: Statewise synthesis', *Atmos. Environ.*, vol. 43, no. 34, pp. 5510–5517, Nov. 2009, doi: 10.1016/j.atmosenv.2009.07.015.

[30] N. Raparthi and H. C. Phuleria, 'Real-world vehicular emissions in the Indian megacity: Carbonaceous, metal and morphological characterization, and the emission factors', *Urban Clim.*, vol. 39, p. 100955, Sept. 2021, doi: 10.1016/j.uclim.2021.100955.

[31] C. A. Alves, D. J. Lopes, A. I. Calvo, M. Evtyugina, S. Rocha, and T. Nunes, 'Emissions from Light-Duty Diesel and Gasoline in-use Vehicles Measured on Chassis Dynamometer Test Cycles', *Aerosol Air Qual. Res.*, vol. 15, no. 1, pp. 99–116, 2015, doi: 10.4209/aaqr.2014.01.0006.

[32] X. Shen *et al.*, 'Development of database of real-world diesel vehicle emission factors for China', *J. Environ. Sci.*, vol. 31, pp. 209–220, May 2015, doi: 10.1016/j.jes.2014.10.021.

[33] Q. Zhang *et al.*, 'The effects of deterioration and technological levels on pollutant emission factors for gasoline light-duty trucks', *J. Air Waste Manag. Assoc.*, vol. 67, no. 7, pp. 814–823, July 2017, doi: 10.1080/10962247.2017.1301275.

[34] J. Park, J. Seo, and S. Park, 'Development of vehicle emission rates based on vehicle-specific power and velocity', *Sci. Total Environ.*, vol. 857, p. 159622, Jan. 2023, doi: 10.1016/j.scitotenv.2022.159622.

[35] D. Wu, B. Deng, M. Li, J. Fu, and K. Hou, 'Improvements on performance and emissions of a heavy duty diesel engine by throttling degree optimization: A steady-state and transient experimental study', *Chem. Eng. Process. - Process Intensif.*, vol. 157, p. 108132, Nov. 2020, doi: 10.1016/j.cep.2020.108132.

[36] M. Cai, Y. Yin, and M. Xie, 'Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach', *Transp. Res. Part Transp. Environ.*, vol. 14, no. 1, pp. 32–41, Jan. 2009, doi: 10.1016/j.trd.2008.10.004.

[37] L. Salazar, O. Nicolis, F. Ruggeri, J. Kisel'ák, and M. Stehlík, 'Predicting hourly ozone concentrations using wavelets and ARIMA models', *Neural Comput. Appl.*, vol. 31, no. 8, pp. 4331–4340, Aug. 2019, doi: 10.1007/s00521-018-3345-0.

[38] Q. Zhang, F. Li, F. Long, and Q. Ling, 'Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network', *IEEE Access*, vol. 6, pp. 56984–56994, 2018, doi: 10.1109/ACCESS.2018.2874068.

[39] D. Brezak, T. Bacek, D. Majetic, J. Kasac, and B. Novakovic, 'A comparison of feed-forward and recurrent neural networks in time series forecasting', in *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, New York City, NY, USA: IEEE, Mar. 2012, pp. 1–6. doi: 10.1109/CIFEr.2012.6327793.

[40] J. Connor and L. Atlas, 'Recurrent neural networks and time series prediction', in *IJCNN-91-Seattle International Joint Conference on Neural*

Networks, Seattle, WA, USA: IEEE, 1991, pp. 301–306. doi: 10.1109/IJCNN.1991.155194.

[41] A. Hamdan, A. Al-Salaymeh, I. M. AlHamad, S. Ikemba, and D. R. E. Ewim, 'Predicting future global temperature and greenhouse gas emissions via LSTM model', *Sustain. Energy Res.*, vol. 10, no. 1, p. 21, Dec. 2023, doi: 10.1186/s40807-023-00092-x.

[42] V. Shepelev, A. Glushkov, I. Slobodin, and Y. Cherkassov, 'Measuring and Modelling the Concentration of Vehicle-Related PM2.5 and PM10 Emissions Based on Neural Networks', *Mathematics*, vol. 11, no. 5, p. 1144, Feb. 2023, doi: 10.3390/math11051144.

[43] R. C. Staudemeyer and E. R. Morris, 'Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks', 2019, *arXiv*. doi: 10.48550/ARXIV.1909.09586.

[44] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[45] Z. Sun, C. Wang, Z. Ye, and H. Bi, 'Long short-term memory network-based emission models for conventional and new energy buses', *Int. J. Sustain. Transp.*, vol. 15, no. 3, pp. 229–238, Jan. 2021, doi: 10.1080/15568318.2020.1734887.

[46] H. Xie *et al.*, 'Parallel attention-based LSTM for building a prediction model of vehicle emissions using PEMS and OBD', *Measurement*, vol. 185, p. 110074, Nov. 2021, doi: 10.1016/j.measurement.2021.110074.

[47] D. C. Montgomery, L. A. Johnson, and J. S. Gardiner, *Forecasting and time series analysis*, 2nd ed. New York: McGraw-Hill, 1990.

[48] Z. Xu, Y. Kang, Y. Cao, and L. Yue, 'Residual Autoencoder-LSTM for City Region Vehicle Emission Pollution Prediction', in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, Anchorage, AK: IEEE, June 2018, pp. 811–816. doi: 10.1109/ICCA.2018.8444183.

[49] A. Graves, 'Long Short-Term Memory', in *Supervised Sequence Labelling with Recurrent Neural Networks*, vol. 385, in Studies in Computational Intelligence, vol. 385. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 37–45. doi: 10.1007/978-3-642-24797-2_4.

[50] J. L. Elman, 'Finding Structure in Time', *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990, doi: 10.1207/s15516709cog1402_1.

[51] P. J. Werbos, 'Backpropagation through time: what it does and how to do it', *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990, doi: 10.1109/5.58337.

[52] J. F. Kolen and S. C. Kremer, Eds, *A field guide to dynamical recurrent networks*. New York: IEEE Press, 2001.

[53] F. A. Gers, J. Schmidhuber, and F. Cummins, 'Learning to Forget: Continual Prediction with LSTM', *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.

[54] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever, 'An Empirical Exploration of Recurrent Network Architectures', in *Proceedings of Machine Learning Research*, PMLR, 2015, pp. 2342–2350. [Online]. Available: https://proceedings.mlr.press/v37/jozefowicz15.html

[55] A. Graves and J. Schmidhuber, 'Framewise phoneme classification with bidirectional LSTM and other neural network architectures', *Neural Netw.*, vol. 18, no. 5–6, pp. 602–610, July 2005, doi: 10.1016/j.neunet.2005.06.042.

[56] A. Graves, A. Mohamed, and G. Hinton, ‘Speech Recognition with Deep Recurrent Neural Networks’, 2013, *arXiv*. doi: 10.48550/ARXIV.1303.5778.
[57] N. Kalchbrenner, I. Danihelka, and A. Graves, ‘Grid Long Short-Term Memory’, 2015, *arXiv*. doi: 10.48550/ARXIV.1507.01526.
[58] A. Graves, S. Fernández, and J. Schmidhuber, ‘Multi-dimensional Recurrent Neural Networks’, in *Artificial Neural Networks – ICANN 2007*, vol. 4668, J. M. De Sá, L. A. Alexandre, W. Duch, and D. Mandic, Eds, in Lecture Notes in Computer Science, vol. 4668. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 549–558. doi: 10.1007/978-3-540-74690-4_56.
[59] K. Cho *et al.*, ‘Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation’, 2014, *arXiv*. doi: 10.48550/ARXIV.1406.1078.
[60] I. Sutskever, O. Vinyals, and Q. V. Le, ‘Sequence to Sequence Learning with Neural Networks’, in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds, Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/5a18e133cbf9f257297f410bb7eca942-Paper.pdf
[61] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, ‘Attention-Based Models for Speech Recognition’, in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds, Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/1068c6e4c8051cfd4e9ea8072e3189e2-Paper.pdf
[62] V. R. Khare, P. Bandyopadhyay, and M. B. Waldo, ‘Automotive field failure analysis based on mileage — Feasibility & benefits’, in *2012 IEEE Conference on Prognostics and Health Management*, Denver, CO, USA: IEEE, June 2012, pp. 1–8. doi: 10.1109/ICPHM.2012.6299513.
[63] Q. Zhang *et al.*, ‘Influences of accumulated mileage and technological changes on emissions of regulated pollutants from gasoline passenger vehicles’, *J. Environ. Sci.*, vol. 71, pp. 197–206, Sept. 2018, doi: 10.1016/j.jes.2018.03.021.
[64] A. Alberini, L. Teja Burra, C. Cirillo, and C. Shen, ‘Counting vehicle miles traveled: What can we learn from the NHTS?’, *Transp. Res. Part Transp. Environ.*, vol. 98, p. 102984, Sept. 2021, doi: 10.1016/j.trd.2021.102984.
[65] ‘Car engine sizes explained | Carwow’. Accessed: Dec. 15, 2024. [Online]. Available: https://www.carwow.co.uk/guides/choosing/what-engine-size-do-i-need-0521
[66] ‘Technical specs, data and fuel consumption of all makes and models automobiles’. Accessed: Dec. 15, 2024. [Online]. Available: https://www.auto-data.net/en/allbrands
[67] M. Opetnik *et al.*, ‘The Impact of Vehicle Technology, Size Class, and Driving Style on the GHG and Pollutant Emissions of Passenger Cars’, *Energies*, vol. 17, no. 9, p. 2052, Apr. 2024, doi: 10.3390/en17092052.
[68] B. S. McMullen and N. Eckstein, ‘Relationship between Vehicle Miles Traveled and Economic Activity’, *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2297, no. 1, pp. 21–28, Jan. 2012, doi: 10.3141/2297-03.
[69] ‘Monthly Economic Indicators | Central Bank of Sri Lanka’. Accessed: Dec. 10, 2024. [Online]. Available: https://www.cbsl.gov.lk/en/statistics/economic-indicators/monthly-indicators

[70] K. D. Makatjane and N. D. Moroke, ‘Comparative study of holt-winters triples exponential smoothing and seasonal Arima: Forecasting short term seasonal car sales in South Africa’, *Risk Gov. Control Financ. Mark. Inst.*, vol. 6, no. 1, pp. 71–82, 2016, doi: 10.22495/rgcv6i1art8.
[71] G. Cheng, V. Peddinti, D. Povey, V. Manohar, S. Khudanpur, and Y. Yan, ‘An Exploration of Dropout with LSTMs’, in *Interspeech 2017*, ISCA, Aug. 2017, pp. 1586–1590. doi: 10.21437/Interspeech.2017-129.
[72] Safwan Mahmood Al-Selwi, Mohd Fadzil Hassan, Said Jadid Abdulkadir, and Amgad Muneer, ‘LSTM Inefficiency in Long-Term Dependencies Regression Problems’, *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 30, no. 3, pp. 16–31, May 2023, doi: 10.37934/araset.30.3.1631.
[73] D. P. Kingma and J. Ba, ‘Adam: A Method for Stochastic Optimization’, 2014, *arXiv*. doi: 10.48550/ARXIV.1412.6980.
[74] F. Bin and J. He, ‘A short-term photovoltaic power prediction model using cyclical encoding and STL decomposition based on LSTM’, in *2023 3rd International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, Wuhan, China: IEEE, Dec. 2023, pp. 260–266. doi: 10.1109/EIECC60864.2023.10456727.
[75] Y.-S. Kim, M. K. Kim, N. Fu, J. Liu, J. Wang, and J. Srebric, ‘Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models’, *Sustain. Cities Soc.*, vol. 118, p. 105570, Jan. 2025, doi: 10.1016/j.scs.2024.105570.
[76] P. Charilaou and R. Battat, ‘Machine learning models and over-fitting considerations’, *World J. Gastroenterol.*, vol. 28, no. 5, pp. 605–607, Feb. 2022, doi: 10.3748/wjg.v28.i5.605.
[77] S. Y. Kyung and S. H. Jeong, ‘Particulate-Matter Related Respiratory Diseases’, *Tuberc. Respir. Dis.*, vol. 83, no. 2, p. 116, 2020, doi: 10.4046/trd.2019.0025.
[78] J. Lasocki, ‘The WLTC vs NEDC: A Case Study on the Impacts of Driving Cycle on Engine Performance and Fuel Consumption’, *Int. J. Automot. Mech. Eng.*, vol. 18, no. 3, pp. 9071–9081, Sept. 2021, doi: 10.15282/ijame.18.3.2021.19.0696.
[79] I. D. Jordan, P. A. Sokół, and I. M. Park, ‘Gated Recurrent Units Viewed Through the Lens of Continuous Time Dynamical Systems’, *Front. Comput. Neurosci.*, vol. 15, p. 678158, July 2021, doi: 10.3389/fncom.2021.678158.

APPENDIX – TIER 02 EXHAUST EMISSIONS FOR PASSENGER CARS (EMEP/EEA)

Type	Technology	CO	NMVOC	NO_x	N_2O	NH3	PM
Units		g/km	g/km	g/km	g/km	g/km	g/km
Notes			Given as THC-CH_4	Given as NO2 equivalent			PM2.5 =PM10 =TSP
Petrol Mini	Euro 4	1.050	0.633	0.065	0.002	0.031	3.20E-07
	Euro 5	1.042	0.625	0.047	0.001	0.013	3.20E-07
	Euro 6 a/b/c	0.397	0.549	0.032	0.001	0.013	3.20E-07
	Euro 6 d-temp	0.397	0.549	0.032	0.001	0.013	3.20E-07
	Euro 6 d	0.397	0.549	0.032	0.001	0.013	3.20E-07
Petrol Small	PRE ECE	40.792	5.386	1.795	0.008	0.002	4.80E-07
	ECE 15/00-01	32.242	4.756	1.795	0.008	0.002	4.80E-07
	ECE 15/02	32.242	4.756	1.795	0.008	0.002	4.80E-07
	ECE 15/03	24.779	4.274	2.213	0.008	0.002	4.80E-07
	ECE 15/04	14.286	3.313	1.971	0.008	0.002	4.80E-07
	Improved Conventional	12.130	2.903	1.853	0.008	0.002	4.80E-07
	Open Loop	12.597	2.143	1.467	0.008	0.002	4.80E-07
	Euro 1	5.463	1.473	0.456	0.012	0.087	3.20E-07
	Euro 2	3.359	0.817	0.248	0.007	0.105	3.20E-07
	Euro 3	3.148	0.436	0.106	0.003	0.031	3.20E-07
	Euro 4	1.050	0.277	0.065	0.002	0.031	3.20E-07
	Euro 5	1.042	0.210	0.047	0.001	0.014	3.20E-07
	Euro 6 a/b/c	0.397	0.091	0.032	0.001	0.014	3.20E-07
	Euro 6 d-temp	0.397	0.078	0.032	0.001	0.013	3.20E-07
	Euro 6 d	0.397	0.044	0.032	0.001	0.013	3.20E-07
Petrol Medium	PRE ECE	40.792	6.055	2.402	0.008	0.002	4.80E-07
	ECE 15/00-01	32.242	4.574	2.402	0.008	0.002	4.80E-07
	ECE 15/02	24.210	4.339	2.296	0.008	0.002	4.80E-07
	ECE 15/03	24.779	4.393	2.406	0.008	0.002	4.80E-07
	ECE 15/04	14.286	3.449	2.539	0.008	0.002	4.80E-07
	Improved Conventional	5.712	2.859	2.406	0.008	0.002	4.80E-07
	Open Loop	7.178	1.936	1.235	0.008	0.002	4.80E-07
	Euro 1	4.962	0.992	0.454	0.012	0.088	3.20E-07
	Euro 2	2.998	0.528	0.247	0.007	0.107	3.20E-07
	Euro 3	2.827	0.300	0.105	0.003	0.032	3.20E-07
	Euro 4	0.954	0.187	0.064	0.002	0.031	3.20E-07

	Euro 5	0.946	0.163	0.047	0.001	0.014	3.20E-07
	Euro 6 a/b/c	0.397	0.055	0.032	0.001	0.014	3.20E-07
	Euro 6 d-temp	0.397	0.052	0.032	0.001	0.013	3.20E-07
	Euro 6 d	0.397	0.036	0.032	0.001	0.013	3.20E-07
Petrol Large-SUV-Executive	PRE ECE	40.792	5.574	3.708	0.008	0.002	4.80E-07
	ECE 15/00-01	32.242	4.565	3.708	0.008	0.002	4.80E-07
	ECE 15/02	22.210	4.417	2.581	0.008	0.002	4.80E-07
	ECE 15/03	24.779	4.116	3.355	0.008	0.002	4.80E-07
	ECE 15/04	14.286	3.793	2.715	0.008	0.002	4.80E-07
	Euro 1	3.974	0.815	0.423	0.013	0.089	3.20E-07
	Euro 2	2.287	0.403	0.225	0.007	0.108	3.20E-07
	Euro 3	2.195	0.227	0.095	0.003	0.032	3.20E-07
	Euro 4	0.765	0.146	0.058	0.002	0.032	3.20E-07
	Euro 5	0.757	0.122	0.041	0.001	0.015	3.20E-07
	Euro 6 a/b/c	0.397	0.045	0.032	0.001	0.014	3.20E-07
	Euro 6 d-temp	0.397	0.041	0.032	0.001	0.013	3.20E-07
	Euro 6 d	0.397	0.289	0.032	0.001	0.013	3.20E-07
Diesel Mini	Euro 4	0.092	0.017	0.599	0.008	0.001	1.74E-06
	Euro 5	0.035	0.001	0.562	0.008	0.002	1.74E-06
	Euro 6 a/b/c	0.052	0.001	0.507	0.006	0.007	1.74E-06
	Euro 6 d-temp	0.041	0.001	0.086	0.006	0.007	1.74E-06
	Euro 6 d	0.040	0.001	0.074	0.006	0.007	1.74E-06
Diesel Small	Conventional	0.667	0.180	0.567	0.000	0.001	1.74E-06
	Euro 1	0.415	0.057	0.697	0.003	0.001	1.74E-06
	Euro 2	0.295	0.040	0.729	0.005	0.001	1.74E-06
	Euro 3	0.093	0.024	0.786	0.008	0.001	1.74E-06
	Euro 4	0.092	0.017	0.599	0.008	0.001	1.74E-06
	Euro 5	0.035	0.001	0.562	0.008	0.002	1.74E-06
	Euro 6 a/b/c	0.052	0.001	0.507	0.006	0.007	1.74E-06
	Euro 6 d-temp	0.041	0.001	0.086	0.006	0.007	4.80E-07
	Euro 6 d	0.040	0.001	0.074	0.006	0.007	4.80E-07
Diesel Medium	Conventional	0.667	0.180	0.567	0.000	0.001	1.74E-06
	Euro 1	0.415	0.057	0.697	0.003	0.001	1.74E-06
	Euro 2	0.295	0.040	0.729	0.005	0.001	1.74E-06
	Euro 3	0.093	0.024	0.786	0.008	0.001	1.74E-06
	Euro 4	0.092	0.017	0.599	0.008	0.001	1.74E-06
	Euro 5	0.035	0.001	0.562	0.008	0.002	1.74E-06
	Euro 6 a/b/c	0.052	0.001	0.507	0.006	0.007	1.74E-06
	Euro 6 d-temp	0.041	0.001	0.086	0.006	0.007	4.80E-07

	Euro 6 d	0.040	0.001	0.074	0.006	0.007	4.80E-07
Diesel Large-SUV-Executive	Conventional	0.667	0.180	0.905	0.000	0.001	1.74E-06
	Euro 1	0.415	0.085	0.697	0.003	0.001	1.74E-06
	Euro 2	0.295	0.119	0.729	0.005	0.001	1.74E-06
	Euro 3	0.093	0.047	0.786	0.008	0.001	1.74E-06
	Euro 4	0.092	0.017	0.599	0.008	0.001	1.74E-06
	Euro 5	0.035	0.001	0.562	0.008	0.002	1.74E-06
	Euro 6 a/b/c	0.052	0.001	0.507	0.006	0.007	1.74E-06
	Euro 6 d-temp	0.041	0.001	0.086	0.006	0.007	1.74E-06
	Euro 6 d	0.040	0.001	0.074	0.006	0.007	1.74E-06