[1]Abhishek Dixit
[2]Manish Jain

Driving Style Based Trajectory Prediction of the Surrounding Vehicles using LSTM & ARIMA in Autonomous Driving

Journal of Electrical Systems

Abstract: - This research explores the application of advanced time-series forecasting models to predict vehicle trajectories based on driving styles. The study utilizes vehicle trajectory pairs obtained from the I-80 and US-101 freeways, extracted from the NGSIM dataset. Principal Component Analysis (PCA) is employed to simplify characteristic indexes, leading to the identification of three distinct driving styles: aggressive, moderate, and traditional. To facilitate predictive analysis, three datasets are created, each representing a unique driving style cluster. The research employs Long Short-Term Memory (LSTM) and Auto Regressive Integrated Moving Average (ARIMA) models to forecast future trends within each driving style. Evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 value, assess the accuracy and reliability of the forecasting models. The LSTM model, with its capacity to model complex temporal dependencies, delivers impressive results with low error metrics and high R^2 value. This research demonstrates the efficacy of LSTM models in accurately predicting trajectories based on various driving styles. This highlights the effectiveness of employing LSTM algorithms within our study, showcasing their capability to capture complex temporal dependencies inherent in diverse driving behaviors. This underscores not only the strength of the LSTM model itself but also the successful application of our research methodology in leveraging this algorithm to achieve precise trajectory predictions.

Keywords: Trajectory Prediction, Driving Styles, LSTM, ARIMA, NGSIM Dataset, Time-Series Forecasting, ADAS, Traffic Data Analysis, Vehicle Dynamics, Pattern Recognition, Autonomous vehicle.

I. INTRODUCTION

In recent years, the rapid advancements in intelligent transportation systems (ITS) and the increasing demand for safer and more efficient mobility have fueled the development of advanced driver assistance systems (ADAS) and autonomous vehicles (AVs) [1]. One of the critical challenges in realizing these technologies is the accurate prediction of surrounding vehicles' trajectories, which is essential for proactive decision-making and collision avoidance [2]. Trajectory prediction involves estimating the future positions of nearby vehicles based on their historical motion data and contextual information. However, the complex and dynamic nature of traffic environments, coupled with the diverse driving behaviors of individual drivers, makes trajectory prediction a highly challenging task.

Driving style refers to the characteristic ways in which drivers operate their vehicle [3]. Studies have shown that drivers exhibit a wide range of driving styles, from aggressive to conservative, depending on their personality, risk perception, and situational factors. Aggressive drivers tend to exhibit risky behaviors such as speeding, abrupt lane changes, and tailgating, while conservative drivers typically follow traffic rules, maintain safe distances, and have a more relaxed driving manner. These differences in driving styles can significantly impact the trajectory patterns of vehicles and the overall traffic flow dynamics [4].

Integrating intelligent driving capabilities with vehicle connectivity offers a multifaceted approach to enhancing transportation systems. This amalgamation empowers vehicles to seamlessly gather real-time data about their own status as well as that of their surroundings. The advent of intelligent and connected vehicles holds immense promise in bolstering overall traffic safety. A pivotal aspect in ensuring safe driving lies in the accurate prediction of vehicle trajectories, encompassing both the primary vehicle and its neighboring counterparts. Such predictions are indispensable for effective motion planning and preemptive collision avoidance strategies. However, the dynamic nature of driving environments coupled with intricate vehicle dynamics renders precise trajectory prediction a persistent challenge. Addressing this challenge has emerged as a focal point within the realm of research dedicated to intelligent and connected vehicles.

[1]*Corresponding author Email: abhishekdixit2606@gmail.com, JECRC University, Jaipur – 303905, India, ORCID ID : 0000-0003-1294-9654

[2]JECRC University, Jaipur – 303905, India, ORCID ID : 0000-0002-3535-0030

Traditional trajectory prediction approaches often rely on kinematic models, such as the constant velocity (CV) or constant acceleration (CA) models, which assume that vehicles maintain their current speed or acceleration over a short prediction horizon. However, these models fail to capture the complex interactions and dependencies among vehicles, as well as the heterogeneity in driving behaviors. More advanced techniques, such as machine learning and deep learning, have been explored to model the intricate patterns and relationships in vehicle trajectories [5]. Among these techniques, Long Short-Term Memory (LSTM) networks, a specialized type of recurrent neural networks (RNNs), have garnered considerable interest owing to their adeptness in grasping prolonged dependencies and learn complex sequential patterns.

Recent advances in machine learning, particularly in the areas of sequence modeling with recurrent neural networks, have opened up new possibilities for data-driven trajectory prediction. By learning directly from real-world driving data, these methods can implicitly capture the diverse driving styles across a population of drivers. LSTM networks have shown significant results by encoding the sequential patterns in past vehicle trajectories to forecast future positions. However, most existing LSTM-based approaches focus solely on the trajectory history of the target vehicle, ignoring the rich behavioral signals contained in the motion patterns of surrounding traffic participants. LSTM-based models have shown promising results in trajectory prediction tasks, outperforming traditional methods in various scenarios.

Another promising approach for trajectory prediction is the Autoregressive Integrated Moving Average (ARIMA) model, this technique is a widely used for time series forecasting problems. ARIMA models have been successfully applied in various domains, including traffic flow prediction, weather forecasting, and stock market analysis. The ARIMA model captures the temporal dependencies and trends in the data, making it suitable for short-term trajectory prediction. Various researches [6] employed ARIMA models to predict vehicle speeds and demonstrated their effectiveness in capturing the temporal patterns in traffic data. Combining ARIMA with LSTM models can potentially leverage the strengths of both techniques, capturing both the short-term and long-term dependencies in vehicle trajectories [7].

Despite the advancements in trajectory prediction methods, limited research has been conducted on incorporating driving style information into the prediction process. Most existing approaches assume a homogeneous driving behavior among vehicles, neglecting the impact of individual driving styles on trajectory patterns [8].

This research paper aims to bridge the gap between driving style analysis and trajectory prediction. NGSIM dataset is used for the trajectory prediction which is further divided into three distinct datasets based on the driving styles of the vehicles obtained by applying K-means algorithm. Three distinct driving style categories emerge: aggressive, moderate, and traditional, forming the basis of three distinct datasets. Subsequently, advanced time-series forecasting models (LSTM and ARIMA) are applied to each dataset to predict future trends within each driving style cluster. Evaluation metrics such as MAE, MSE, RMSE, and R^2 are utilized to gauge the accuracy and reliability of the forecasting models, providing a comprehensive analysis of their predictive capabilities.

II. LITERATURE SURVEY

In the field of trajectory prediction and positioning, researchers have explored various techniques, including machine learning and deep learning approaches, to accurately model and forecast the motion patterns of different vehicles on the road. This literature review discusses several recent studies that address these challenges using advanced methods.

Li and Fong (2022) proposes a novel approach for trajectory prediction of moving objects by combining data stream learning with a Kalman filter and an evolving correlated horizons feature selection technique (KF-ECH-FS). Traditionally, the Kalman filter has been used as a control-feedback loop mechanism to correct errors from previous predictions and estimate the next position in trajectory prediction. In this fusion model, the Kalman filter and its windowed version are utilized as predictor variables in a multivariate time series forecasting process [9]. The predictor variables act as additional features, aiding in improving trajectory prediction accuracy. However, only relevant features are selected through an incremental learning process using multivariate data stream analytics. It addresses the issue of model overfitting that can occur when evaluating and selecting a large number of features expanded by time-series windowing during the learning process. The authors employ a simple and efficient feature selection heuristic, an auto-encoder, in conjunction with data stream learning using a Gate Recurrent Unit (GRU) model. The proposed KF-ECH-FS approach is evaluated through experiments on a sample case of camera surveillance for accident prevention. The results demonstrate that the proposed KF-ECH-FS method outperforms either the Kalman filter or windowing alone in terms of one-step horizon trajectory prediction accuracy.

Alsanwy et al. (2022) paper focuses on the application of Long Short-Term Memory (LSTM) networks for predicting vehicle motion signals for driving simulators [10]. The primary objective of using driving simulators is to provide a realistic driving experience. However, simulator platforms have physical limitations, and Motion Cueing Algorithms (MCAs) are employed to generate driving sensations for the simulator user while considering these physical and dynamical constraints. Traditional methods for predicting vehicle motion signals struggle with predicting long sequences of time-series data due to the lack of a feedback loop or limited memory size. To address this issue, the authors propose the development of an LSTM model to predict motion signals using Python. The performance of the LSTM model is compared with different traditional methods using several evaluation criteria, including root mean squared error (RMSE), mean absolute error (MAE), and Pearson's correlation coefficient (r). The results indicate that the LSTM model outperforms traditional methods, such as Recurrent Neural Networks (RNNs), by producing more accurate motion predictions. This allows the MCA to deliver realistic motion sensations to the simulator user. Furthermore, the authors suggest that the LSTM model can be employed in a wide range of applications, including autonomous vehicle trajectory prediction and other prediction problems involving time-series data.

Zhang et al. (2022a) proposed CNN model which learn behavior patterns from historical data and is capable of making lane change predictions to assist the decision-making of autopilot systems in real-time [11]. The experiments were conducted on the Next Generation Simulation (NGSIM) dataset, and the model was able to predict lane changes up to 7 seconds before the actual lane change occurred, using observations from only one second of data. The model achieved an impressive accuracy of 99% for a 3-second observation window without the need for extensive feature engineering. Comparisons were made with tree-based and multi-layer perceptron-based methods to evaluate the performance of the proposed CNN model. The ability to accurately predict lane changes is crucial for ensuring the safety and efficient operation of autonomous vehicles. By leveraging the power of CNNs and combining spatial and temporal features, this work aims to improve the decision-making capabilities of autopilot systems, ultimately enhancing public trust and acceptance of autonomous driving technologies.

Li et al. (2022) proposes a hierarchical framework based on Graph Neural Networks (GNNs) to model the interactions between heterogeneous traffic participants (vehicles, pedestrians, and riders) and predict their trajectories [12]. The proposed framework consists of two modules, each employing a GNN: one for Interactive Event Recognition (IER) and another for Trajectory Prediction (TP). The IER module is responsible for recognizing interactive events between traffic participants and the ego vehicle. The recognized results from this module serve as input to the TP module, which is designed for interactive trajectory prediction. To enable multi-step prediction, the TP module combines the GNN with a Long Short-Term Memory (LSTM) network. The LSTM network is integrated with the GNN to capture temporal dependencies and enhance the trajectory prediction capabilities. The proposed hierarchical framework is evaluated using naturalistic driving data collected from urban traffic environments. Comparative results with state-of-the-art methods indicate that the hierarchical GNN framework achieves outstanding performance in both recognizing interactive events and predicting interactive behaviors of traffic participants. By accurately modeling the interactions between heterogeneous traffic participants and predicting their trajectories

Xing et al. (2021) focuses on the correlation between driving behaviors and energy consumption in the context of connected automated vehicles (CAVs). It proposes an energy-aware driving pattern analysis and motion prediction system using deep learning-based time-series modeling approaches [13]. First, the authors statistically analyze energy-aware longitudinal acceleration/deceleration behaviors and lateral lane-change behaviors. They apply a sliding standard deviation (SSD) test to evaluate the smoothness of the trajectory and velocity signals, considering different energy consumption levels. The authors then propose an energy-aware personalized joint time-series modeling (PJTSM) approach based on deep recurrent neural networks (RNNs) and long short-term memory (LSTM) cells. This approach aims to accurately predict the motion (trajectory and velocity) of the leading vehicle. The study compares and discusses the differences in prediction performance across various energy consumption levels. The results show that for heavy energy users, the prediction accuracy is the lowest among the three categories due to the higher randomness of their driving behaviors. This finding suggests that it is more challenging to anticipate the driving behaviors of vehicles exhibiting heavy energy consumption. By accurately recognizing driving behaviors and predicting vehicle motion while considering energy consumption factors, this work aims to contribute to safer automated driving and transportation systems for CAVs. The personalized estimation of driving behaviors can potentially enhance the safety and efficiency of connected automated vehicles on the road.

Zhang et al. (2022b) proposed framework focuses on feature learning to gain a comprehensive understanding of lane change behaviors and achieve high prediction performance based on selected features [14]. A time-step dataset with more than 1,000 features is constructed from vehicle trajectory data. To identify the key features involved in the original feature set, the authors propose an XGBoost-based three-step feature learning algorithm. This algorithm integrates feature importance ranking, metric selection, and recursive feature elimination. After analyzing the accuracy of test data from different time segment positions, the sliding window method is applied to the time-step dataset with filtered features. This approach aims to properly select time segments, which are then flattened into corresponding time-series datasets for model prediction. The authors conducted case studies using a publicly available dataset, the Next Generation Simulation (NGSIM), to perform experiments on feature learning and lane change prediction. They achieved a new state-of-the-art accuracy of 97.6% using a time-series dataset with 75 selected features and a 1-second window size with the XGBoost predictor after adopting the proposed three-step method. This performance is superior to other state-of-the-art feature selection methods. By accurately predicting lane changes and enabling intelligent assistance systems, this work aims to improve road safety and facilitate the integration of autonomous vehicles into the future of transportation.

Li et al. (2022) addresses the challenge of accurate trajectory prediction, which is essential for driving decision-making and local motion planning of smart vehicles [15]. To address these issues, the authors propose a Two-stream LSTM Network with a hybrid attention mechanism (TH-Net). Specifically, they construct a Two-stream LSTM structure (TS-LSTM) to build independent information transmission links for inter-vehicle interactions and vehicle motion states while maintaining their coupling relationship. Additionally, the authors introduce a Hybrid Attention Mechanism (H-AM) to explore the importance of hidden states from the dimensions of time and feature. This mechanism guides the TH-Net to selectively reuse the hidden states, enhancing the model's ability to capture relevant information. The results demonstrate that TH-Net remarkably outperforms state-of-the-art methods in terms of long-term trajectory prediction performance. By effectively modeling inter-vehicle interactions, vehicle motion states, and selectively attending to important historical information, the TH-Net model aims to improve the accuracy of trajectory prediction for smart vehicles. Accurate trajectory prediction is crucial for enabling intelligent driving decision-making and local motion planning, ultimately enhancing the safety and efficiency of smart vehicle operations.

These studies demonstrate the application of various machine learning and deep learning techniques, including LSTM, CNN, XGBoost, and graph neural networks, to address the challenges of trajectory prediction and positioning. The researchers explored innovative approaches to handle complex temporal patterns, incorporate relevant features, and model interactions between multiple entities, aiming to improve prediction accuracy and support safer and more efficient transportation systems.

Despite recent advancements, integrating driving style information into trajectory prediction models remains an ongoing challenge in the literature. While existing methods have shown promise, they often suffer from limitations in terms of generalizability and real-time performance. This research endeavors to confront these challenges head-on by introducing a novel approach that compares LSTM and ARIMA models, augmented with driving style classification. By doing so, our study aims to achieve both enhanced accuracy and efficiency in trajectory prediction, offering a significant step forward in addressing the shortcomings of existing methodologies.

III. METHODOLOGY

As shown in figure 1, in this study, vehicle trajectory pairs were obtained from the data collected on the I-80 and US-101 freeways. The NGSIM dataset provides valuable information about vehicle trajectories under both congested and moderate traffic conditions, as the data were gathered from specific locations at different times. To simplify the characteristic indexes, Principal Component Analysis (PCA) was employed, resulting in the extraction of two components that captured all the relevant features. The optimal number of clusters was determined using the "Elbow rule" and Silhouette methods, and subsequently, the K-means algorithm was applied to group the vehicle driving styles. As a result, three distinct driving style categories were identified: aggressive, moderate, and traditional which is termed as driving style 1, driving style 2 and driving style 3 respectively in this research. The subsequent step involves the creation of three distinct datasets based on these clusters, each representing a unique driving style.

To forecast future trends within each driving style cluster, the paper employs two advanced time-series forecasting models: ARIMA & LSTM. These models are applied to each of the three datasets, enabling a comprehensive analysis of their predictive capabilities. The assessment of accuracy and reliability in forecasting models is conducted through the utilization of evaluation metrics such as R square, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

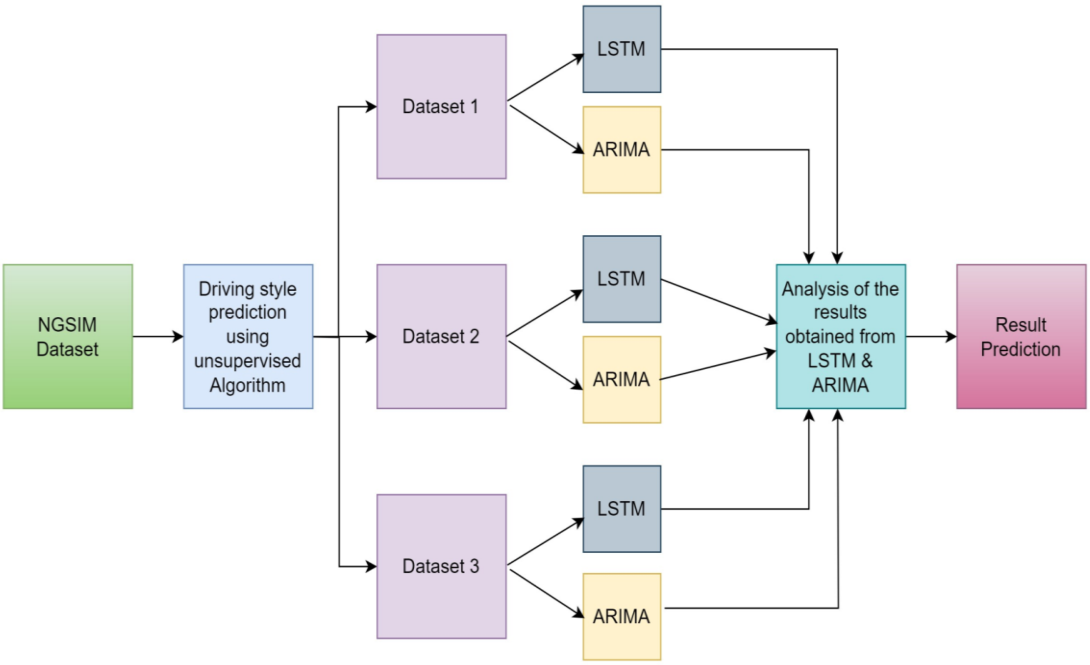

Fig. 1: Methodology adopted for the trajectory prediction

The results obtained from the LSTM and ARIMA models are meticulously analyzed to provide insights into the effectiveness of each model in capturing the intricate temporal dependencies within the driving style clusters. The comparison involves assessing the models' performance across all three datasets, considering both X and Y coordinates of the vehicle. This thorough analysis serves to highlight the strengths and weaknesses of each model in predicting driving behaviors under different circumstances. Finally, the research paper concludes with result predictions, extrapolating the findings from the LSTM and ARIMA models to offer insights into future driving style trends. This approach not only contributes to a deeper understanding of the dynamics of driving behaviors but also provides a valuable foundation for the development of intelligent transportation systems and predictive models for safer and more efficient traffic management.

3.1 ARIMA:

ARIMA (Autoregressive Integrated Moving Average) stands out as a widely-used approach in time series forecasting. Integrating auto regression, differencing, and moving averages, ARIMA effectively models and predicts time series data. Characterized by three essential parameters p, d, and q. The ARIMA model delineates the autoregressive order, differencing order, and moving average order, respectively [16].

The autoregressive component of the ARIMA model of order p is given by:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

Where Y_t is the time series at time t, $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients, and ϵ_t is the white noise error term.

Integrated (I) Component: The integrated component of the ARIMA model of order d involves differencing the time series d times to achieve stationarity.

$$Y_{t'} = Y_t - Y_{t-1}$$

Repeat this differencing process d times until stationarity is achieved.

Moving Average (MA) Component: The moving average component of the ARIMA model of order q is given by:

$$Y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients and ϵ_t is the white noise error term.

ARIMA Equation: Combining the autoregressive, integrated, and moving average components, the ARIMA (p, d, q) model is represented as:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1 - B)^d Y_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\epsilon_t$$

Where, B is the backshift operator.

In this paper, we are predicting the trajectory of a surrounding vehicle by forecasting its x & y coordinate using ARIMA model based on various features provided in the dataset. Initially, the necessary libraries are imported, including pandas for data manipulation, numpy for numerical operations, SARIMAX from statsmodels for time series modeling, and scikit-learn for evaluating model performance. The dataset containing the vehicle's data is loaded, ensuring that the 'Frame_ID' column is sorted in ascending order and set as the index. The data is then split into training and testing sets with an 80-20 split. Next, a multivariate ARIMA model is defined. The order of the ARIMA model is set as (1, 1, 1), which may require adjustment based on the data's characteristics. Exogenous variables (features) are included in the model to enhance prediction accuracy. The model is fitted using the training data. Then, it is used to forecast the x & y coordinate of the vehicle for the test set. After forecasting, error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared Error (R^2) are calculated to evaluate the model's performance in predicting the trajectory accurately.

Finally, the actual and predicted values of the x-coordinate are plotted against the 'Frame_ID' to visualize the model's performance. As shown in figure 2, the green line represents the actual values, while the red line represents the predicted values. This plot helps in assessing how well the model captures the trajectory of the surrounding vehicle.

3.2 LSTM

Long Short-Term Memory (LSTM) networks, belonging to the category of recurrent neural networks (RNNs), are engineered to capture extended dependencies within sequential data. Renowned for their prowess in tasks spanning sequence prediction to natural language processing, LSTM networks excel owing to their adeptness in retaining and updating memory across prolonged sequences. At the core of LSTM architecture lie memory cells and three essential gates: the input gate, forget gate, and output gate. [17].

The memory cell, denoted as c_t is responsible for storing information over time. It can be updated through a combination of the input gate, forget gate, and output gate.

Input Gate (i_t): Controls the information to be stored in the memory cell.

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi})$$

Forget Gate (f_t): Determines the information to be discarded from the memory cell.

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$$

Output Gate (o_t): Controls the information to be output from the memory cell.

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$$

The memory cell is updated using the input, forget, and output gates:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc})$$

Where, $\odot$ represents element-wise multiplication.

The hidden state, h_t , is the output of the LSTM cell and is computed as:

$$h_t = o_t \odot \tanh(c_t)$$

LSTM networks provide an effective solution for learning and retaining information over long sequences. The architecture's ability to manage memory through gates makes it well-suited for various applications in sequential data analysis.

In this research, LSTM neural network model is used to predict the trajectory of a surrounding vehicle by forecasting its x & y coordinate of the surrounding vehicle. Initially, the necessary libraries are installed and imported, including TensorFlow for deep learning, pandas for data manipulation, numpy for numerical operations, and matplotlib for visualization. We also import tools from scikit-learn for preprocessing and evaluating model performance, such as MinMaxScaler for data normalization and functions to calculate mean absolute error, mean squared error, and R-squared error. The dataset containing the vehicle's data is loaded, and features are selected along with the target variable x & y.

Next, the data is normalized using MinMaxScaler to ensure that all features are on a similar scale. The dataset is then split into training and testing sets with an 80-20 ratio using train_test_split. To prepare the data for LSTM, it is reshaped into a 3D array with dimensions (samples, time steps, features). The LSTM model is constructed with one LSTM layer containing 100 units followed by a dense layer with one output unit. The model is compiled using the Adam optimizer and mean squared error loss function. The model is trained on the training data with 100 epochs and a batch size of 32. Validation data is provided to monitor the model's performance during training. After training, predictions are made on the test set. The predicted values are then inverse transformed using the MinMaxScaler to obtain the actual values. Evaluation metrics including mean absolute error, mean squared error, root mean squared error, and R-squared error are calculated using the inverse transformed predictions and actual values to assess the model's performance.

Finally, the actual and predicted values of the x & y coordinate are plotted against the sample index to visualize the model's performance. As shown in figure 3, the blue line represents the actual values, while the red line represents the predicted values. This plot allows for a visual comparison of the actual and predicted trajectories of the surrounding vehicle.

IV. Results

Figure 2 & 3 depict the relationship between actual and predicted value of x & y coordinates of the vehicles' using ARIMA & LSTM. The presented table-1 & table-2 showcases the various errors comparing the performance of two forecasting models, ARIMA and LSTM, across three distinct datasets (Dataset-1, Dataset-2, and Dataset-3) of driving style 1, driving style 2, driving style 3 & overall NGSIM for X and Y coordinates of the vehicles. The evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R squared value.

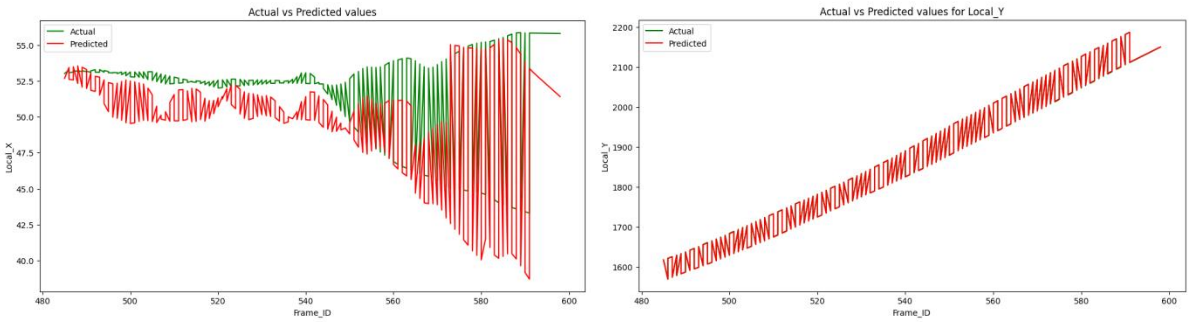

Fig 2: Actual vs Predicted value of x & y coordinate using ARIMA

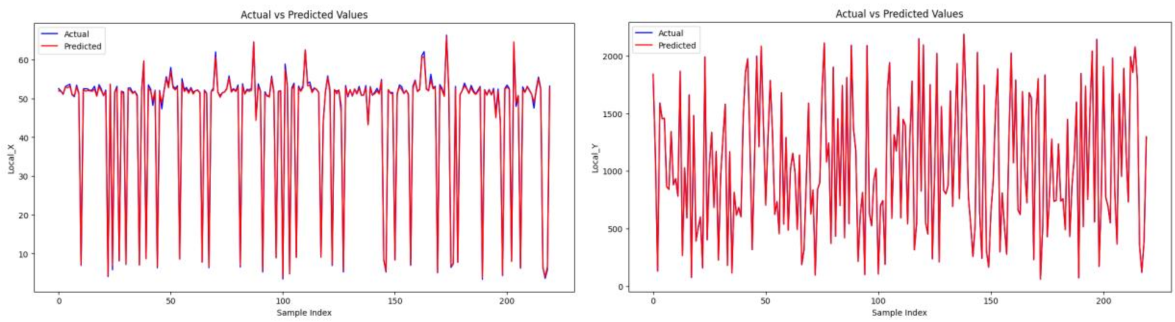

Fig 3: Actual vs Predicted value of x & y coordinate using LSTM

While comparing the results obtained through three distinct driving styles with the overall NGSIM dataset results, for the ARIMA model, NGSIM dataset have lower MAE, MSE, and RMSE values and high R^2 value for the X coordinate compared to all individual datasets, indicating better performance. For the Y coordinate, the NGSIM dataset has a higher MAE, MSE and RMSE and equal R^2 values compared to individual dataset of various driving styles indicating poor performance of NGSIM on predicting the Y coordinate of the vehicle.

Table: 1 Model trajectory prediction errors comparison using dataset of driving style 1, 2 & 3

		Dataset-1/Driving style-1				Dataset-2/ Driving style-2				Dataset-3/ Driving style-3			
	Coordinate	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
ARIMA	X	3.09	21.53	4.64	0.92	2.22	10.1	3.17	0.94	2.71	14.97	3.87	0.94
	Y	0.66	0.81	0.9	0.99	1.85	5.47	2.33	0.99	1.38	3.2	1.78	0.99
LSTM	X	0.02	0.001	0.03	0.99	0.02	0.002	0.04	0.99	0.07	0.007	0.08	0.99
	Y	1.32	2.18	1.47	0.99	0.81	1.03	1.01	0.99	0.44	0.32	0.57	0.99

Table: 2 Model trajectory prediction errors comparison using overall NGSIM dataset

	Coordinate	MAE	MSE	RMSE	R^2
ARIMA	X	1.69	5.46	2.33	0.98
	Y	1.57	4.06	2.01	0.99
LSTM	X	0.08	0.009	0.09	0.99
	Y	2.46	7.29	2.7	0.99

For the LSTM model, for both X and Y coordinate NGSIM dataset results have higher MAE, MSE, and RMSE values compared to all individual datasets, indicating poorer performance of NGSIM dataset in predicting of X and Y coordinates of the vehicle as compared to the various driving styles obtained in our research. The R^2 values for both X and Y coordinates are 0.99 for all datasets, suggesting good fit to the data across all cases. The detailed comparison graph obtained from table 1 & 2 showcasing the comparison among various driving style and NGSIM data through various error metrics using ARIMA and LSTM model is shown in figure 4.

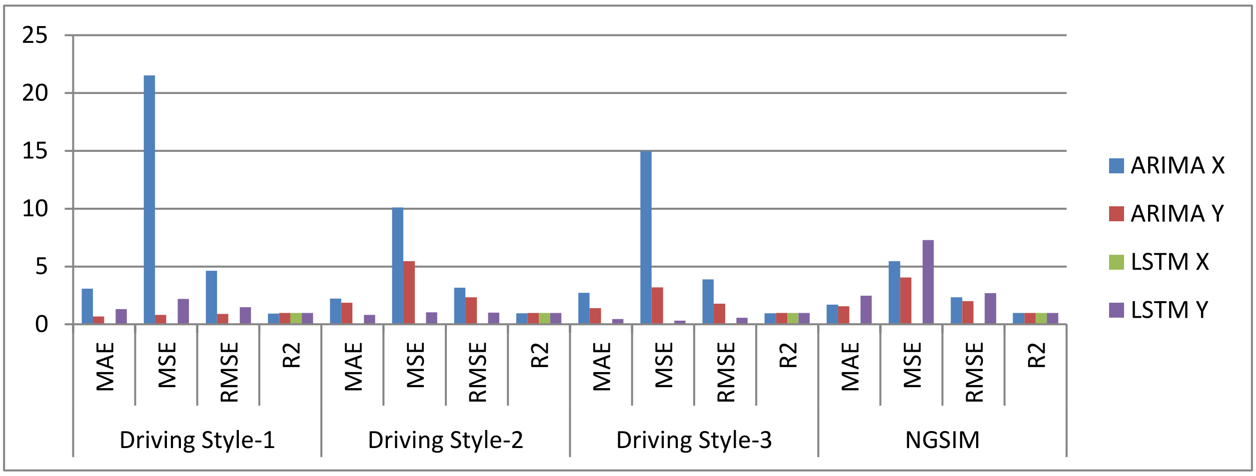

Fig 4: Model trajectory prediction errors comparison using dataset of driving style 1, 2 & 3

V. CONCLUSION & FUTURE SCOPE

In conclusion, both the ARIMA and LSTM models exhibit strong forecasting capabilities across all datasets. The ARIMA model consistently performs well, particularly in capturing the patterns in the Y coordinates which is continuously increasing in nature. The LSTM model, with its ability to model complex temporal dependencies, also delivers impressive results with notably low error metrics and high R^2 value. The results of ARIMA model for the NGSIM Dataset are better for the X coordinate compared to all individual datasets, but the performance for the Y coordinate is mixed, with some individual datasets performing better or worse depending on the evaluation metric. For the LSTM model, the results obtained through the dataset of various driving styles of the vehicle showcase better performance as compared with the NGSIM Dataset which showcase the effectiveness of our research to predict trajectory of the Surrounding Vehicles' based on driving style using LSTM & ARIMA. The integration of hybrid models presents an exciting prospect for future research. Combining the strengths of ARIMA and LSTM or incorporating other sophisticated forecasting approaches could potentially yield more accurate and robust trajectory predictions.

REFERENCES

[1] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3303-3315, 2021.

[2] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION Dataset: An international, Adversarial and Cooperative motion Dataset in Interactive Driving Scenarios with Semantic Maps," 2020.

[3] J. Wang, K. Li, S. Zhu, and T. Xu, "Driving Style Recognition Based on Driver Behavior Questionnaire and Driving Features," IEEE Access, vol. 9, pp. 40467-40479, 2021.

[4] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 2, pp. 315-329, 2020.

[5] Y. Xiao and C. Yin, "Hybrid LSTM Neural Network for Short-Term Traffic Flow Prediction," Information, vol. 10, no. 3, p. 105, 2019.

[6] R. Mohan and G. Ramadurai, "Real-Time Vehicle Speed Prediction Using ARIMA and GARCH Models," in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-6.

[7] Y. Yuan, X. Lv, and S. Hu, "Short-term traffic flow prediction based on LSTM-ARIMA hybrid model," in 2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE), 2020, pp. 194-198.

[8] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, "Generalizable intention prediction of human drivers at intersections," in 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 1665-1670.

[9] T. Li and S. Fong, "Trajectory Prediction Using Multivariate Time-series Data Stream Learning with Fused Kalman-filter and Evolving Correlated Horizons Feature Selection," 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Prague, Czech Republic, 2022, pp. 663-670.

[10] S. Alsanwy et al., "Prediction of Vehicle Motion Signals for Motion Simulators Using Long Short-Term Memory Networks," 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Prague, Czech Republic, 2022, pp. 34-39.

[11] Zhang, S. Zhang and R. Luo, "Lane Change Intent Prediction Based on Multi-Channel CNN Considering Vehicle Time-Series Trajectory," 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 287-292.

[12] Z. Li, C. Lu, Y. Yi and J. Gong, "A Hierarchical Framework for Interactive Behaviour Prediction of Heterogeneous Traffic Participants Based on Graph Neural Network," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 9102-9114, July 2022.

[13] Y. Xing, C. Lv, X. Mo, Z. Hu, C. Huang and P. Hang, "Toward Safe and Smart Mobility: Energy-Aware Deep Learning for Driving Behavior Analysis and Prediction of Connected Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 4267-4280, July 2021.

[14] Y. Zhang, X. Shi, S. Zhang and A. Abraham, "A XGBoost-Based Lane Change Prediction on Time Series Data Using Feature Engineering for Autopilot Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 19187-19200, Oct. 2022.

[15] C. Li, Z. Liu, J. Zhang, Y. Wang, F. Ding and X. Zhao, "Two-stream LSTM Network with Hybrid Attention for Vehicle Trajectory Prediction," 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 1927-1934.

[16] L. Chen, H. Xiao and Z. Dai, "Computerized Intelligent Pricing and Replenishment Model for Vegetable Products Based on ARIMA Prediction," 2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE), Changchun, China, 2023, pp. 1238-1243.

[17] K. S. Gill, V. Anand, R. Chauhan, A. Choudhary and R. Gupta, "CNN, LSTM, and Bi-LSTM based Self-Attention Model Classification for User Review Sentiment Analysis," 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 2023, pp. 1-6.