

Influence of Sequence Length and Geographic Representation on Optimal Prediction Architectures for Stolen Vehicle Geolocation

Gustavo V. I. de Macedo   [Universidade de Brasília | gustavo.isobe@aluno.unb.br]
Geraldo P. R. Filho  [Universidade Estadual do Sudoeste da Bahia | geraldo.rocha@uesb.edu.br]
João K. M. dos Santos  [Universidade de Brasília | joao.kleber@aluno.unb.br]
Arthur R. Neves  [Universidade de Brasília | rodrigues.neves@aluno.unb.br]
Murilo G. Almeida  [Universidade Estadual Paulista | murilo.goes@unesp.br]
Mariana C. Falqueiro  [Universidade de Brasília | mariana.cabral@unb.br]
Rodolfo I. Meneguette  [Universidade de São Paulo | meneguette@icmc.usp.br]
André L. M. Serrano  [Universidade de Brasília | andrelms@unb.br]
Vinícius P. Gonçalves  [Universidade de Brasília | vpgvinicius@unb.br]

 Department of Electrical Engineering, Universidade de Brasília, Campus Universitário Darcy Ribeiro Asa Norte, Brasília, DF, 70910-900, Brazil.

Received: 15 November 2024 • Accepted: 22 March 2025 • Published: 30 June 2025

Abstract When predicting the next geolocation of a stolen vehicle using external sensor data, such as speed radars, the challenge extends beyond the prediction itself to include determining the most suitable prediction architecture. While existing studies provide data that influence prediction performance, there is no consensus on the optimal architecture. Therefore, adopting a broader perspective to identify key criteria influencing the choice of architecture is essential. This study evaluates the shift in the optimal architecture depending on the length of the historical sequence and the format of geographic representation. The results reveal a shift in the optimal architecture, with the shift point being influenced by the type of geographic representation.

Keywords: Vehicle Prediction, External Sensor Trajectory, XGBoost, LSTM, Transformer and LSTM Entangled, TLE, Geographic Representation, Historical Sequence Length, Optimal Architecture Shift.

1 Introduction

Using speed radars with Optical Character Recognition (OCR), as mentioned in [Bernardi *et al.*, 2015], and knowing their geolocation allows for the identification of the momentary geographic location of a stolen vehicle, even if the vehicle does not have any geolocation equipment attached or inside it. This enables the identification of discrete and sparse points along the vehicle's route.

In this context, using historical points to predict the next external sensor (speed radar) where a stolen vehicle is most likely to be detected can assist both in the recovery of the vehicle and in reducing crimes that may follow the theft of the vehicle, as mentioned in [Neto *et al.*, 2021].

However, this scenario presents a trade-off: whether to use the minimum possible number of historical sequence points to predict and act more quickly in the attempt to recover the stolen vehicle, or to wait for additional sequence points to improve the accuracy of the next geolocation prediction, which may never occur, as the stolen vehicles tend to have only a few records.

To illustrate this issue, which is further complicated by the fact that a driver acting outside the law naturally tends to exhibit anomalous driving behavior compared to a typical citizen in their daily life [Haviland and Wiseman, 1974], **Figure 1** shows the number of stolen vehicles by the number of records they have in the non-public anonymized database of the state of São Paulo (SP), Brazil, that we had access from

government to do this study. For example, a vehicle with 3 records also has 2 or 1 record, but not 4 records. Thus, vehicles with at least 9 records represent only 9.2% of the total number of stolen vehicles with at least 1 record. That is, the greater the need for a longer historical input sequence, the fewer vehicles can be located.

Unlike studies such as [Tsiligkaridis *et al.*, 2020] and [Chen *et al.*, 2024], which use the complete trajectory of the vehicle, these records of stolen vehicles, although spread throughout the state of São Paulo, are only made at specific points where speed radars are located. **Figure 2** shows the possible geolocation record points in blue, from the non-public anonymized database of the state of São Paulo, demonstrating the variability and dispersion of these points. It is noteworthy that while some speed radars are kilometers apart, others are positioned side by side.

The Long-Short Term Memory (LSTM) machine learning architecture has been one of the most widely used in studies for prediction trajectory. In some cases, it has proven to be the best, as in [Bae *et al.*, 2022] and [Gaiduchenko *et al.*, 2020], while in other cases, it has not, as in [Xu *et al.*, 2023] and [Chen *et al.*, 2022]. Nevertheless, a specific implementation of Gradient Boosted Regression Trees (GBRT) architecture, also known as Extreme Gradient Boosting (XGBoost), yielded the best results in [Neto *et al.*, 2021]. Moreover, a new architecture called Transformer and LSTM Entangled (TLE) outperformed others when using only two geolocation points to predict the third point, as shown in [Macedo

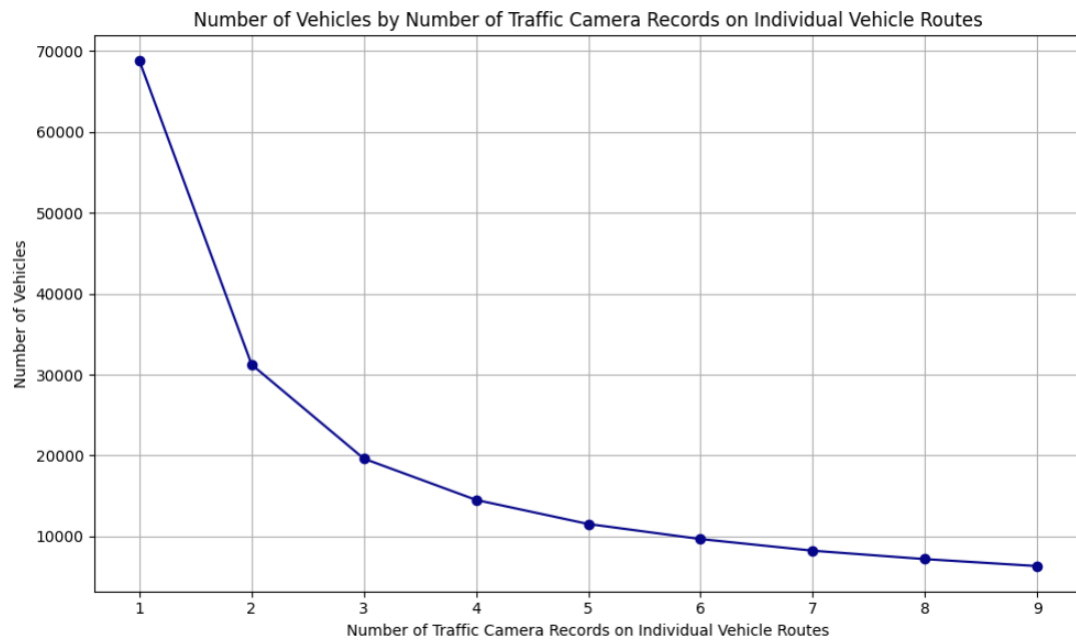


Figure 1. Stolen vehicles from São Paulo, Brazil.

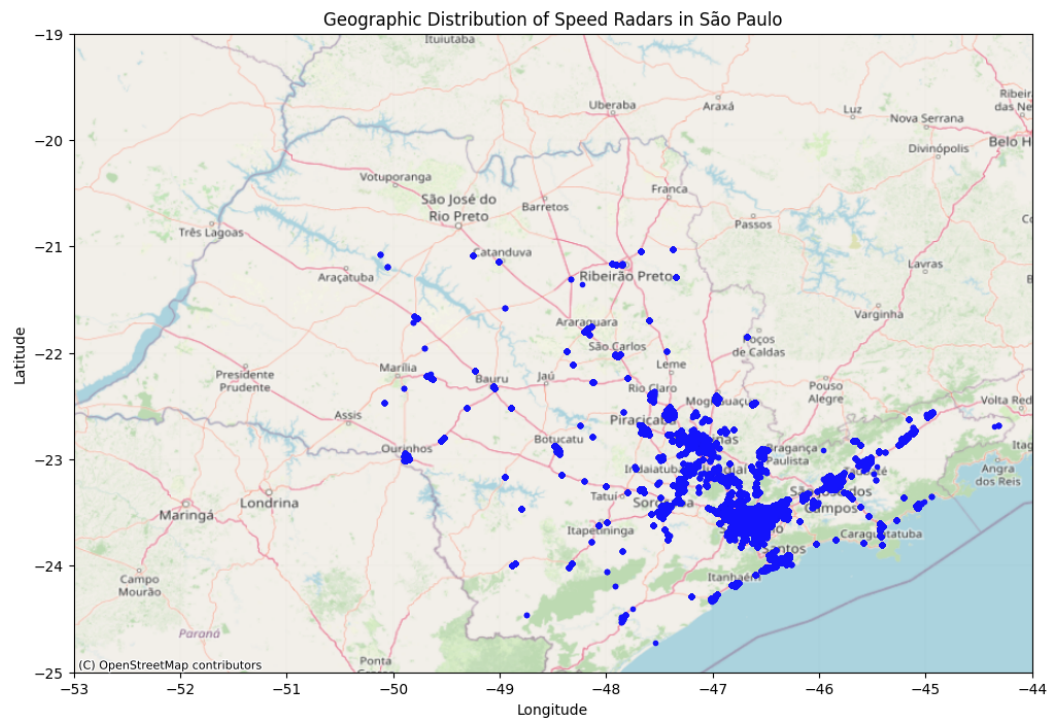


Figure 2. Speed radars mapped in São Paulo, Brazil.

et al., 2024]. Since this study intends to analyze architectures for sequence prediction with few historical points, it will focus on these three main state-of-the-art architectures: LSTM, TLE and XGBoost.

Given that [Neto et al., 2021] demonstrated superior performance for XGBoost over LSTM, and that LSTM outperformed TLE with longer sequences yet underperformed compared to TLE with shorter sequences in [Macedo et al., 2024], while also considering the influence of geolocation representation changes highlighted in [Neto et al., 2021], this study aims to explore how these findings apply across diverse conditions. Specifically, we seek to address the following research questions to identify the most suitable architecture under varying sequence lengths and geographic representations:

- **RQ1:** Which architecture performs best with short sequence lengths for prediction?
- **RQ2:** Is there a shift point where the best-performing architecture changes?
- **RQ3:** Does the format of geographic representation impact the prediction results?
- **RQ4:** Can the geographic representation format alter the shift point between the best-performing architectures, if one exists?

Next, Section II presents the related works; Section III provides the problem definition for the prediction of the next external sensor geolocation; Section IV describes the experiments, including data preprocessing, architectures, methodology, and results with analysis; Section V discusses the conclusions and future work, followed by the acknowledgments, funding, author's contributions, competing interests, availability of data and materials, abbreviations, and references.

2 Related Work

Predicting the next external sensor trajectory, due to sparse and discrete data, presents greater challenges than regular trajectory prediction, particularly because the path between sensors is unknown but still important to be represented. **Figure 3** shows a real example of a trajectory mapped using records from external sensors. It can be observed that the actual path between the geolocation points cannot be defined, which is why the vehicle's trajectory is represented as a straight line connecting the speed radar points in sequence, without depicting the actual intermediate route.

Most of the studies reviewed focus on LSTM architecture for trajectory prediction, with less emphasis on other architectures like XGBoost. For instance, [Neto et al., 2021] is notable for being the only work in this set to utilize XGBoost, highlighting its potential in handling sparse and discrete data, particularly in predicting the location of stolen vehicles and achieving the best results. However, this study does not explore the impact of different geographic representations on each architecture, which could potentially cause a shift in the best-performing architecture.

[Macedo et al., 2024] stands out as it introduces the new Transformer and LSTM Entangled (TLE) architecture

alongside LSTM. This study observes a shift in the best-performing architecture, but it lacks a detailed shift point analysis and does not explore the effects of geographic representation variation. This limits a deeper understanding of how changes in input data representation might influence prediction outcomes.

Similarly, although [Xu et al., 2023], [Cruz et al., 2019] and [Tsiligkaridis et al., 2020] observes a shift in the best-performing architecture during the experiments, they do not address the shift with a more detailed analysis. Moreover, those studies do not use XGBoost or examine the impact of different geographic representations. This limits the potential insights into how these variables interact.

All of [Chen et al., 2022], [Bae et al., 2022], [Gaiduchenko et al., 2020], [Tsiligkaridis et al., 2020], [Cruz et al., 2022], [Cruz et al., 2019], [Cruz et al., 2021] and [Chen et al., 2024] use LSTM, but not XGBoost or TLE, and none of them include an analysis of the shift point or geographic representation variation.

In [Capanema et al., 2020], the authors have a significant advantage as the training is based on the driver's own historical data for predicting the next destination, whereas in most other cases, the training data is sourced from other vehicles. In [Hu et al., 2022], the prediction focuses on short time intervals and distances, up to 5 seconds ahead.

Another line of research, including [Brito et al., 2023], [Ladeira et al., 2020], [Ladeira et al., 2019b], and [de Souza et al., 2019], as well as [Almeida et al., 2022] for buses, and [Ladeira et al., 2019a], which focuses on time windows, in addition to [de Souza and Villas, 2020] and [Karimzadeh et al., 2021], which use reinforcement learning architectures, evaluates and predicts safer routes by avoiding areas with a higher probability of crime. Since our aim is to predict where a stolen vehicle will be, these studies take an approach opposite to that of the present work.

An approach worth noting is MEDAVET [Reyna et al., 2024], which focuses on detecting traffic anomalies on highways by utilizing spatial and temporal structures through video monitoring, aiming to ensure road safety. Although MEDAVET's context differs from our work, as it is based on detecting anomalous behavior instead of predicting discrete geolocation points, it presents a valuable perspective on managing complex vehicle movement patterns which could complement prediction tasks in urban traffic security by aiding in the identification of potentially unusual movement patterns.

From a different perspective on trajectory prediction, relevant to Autonomous Vehicles (AVs), studies such as [Geng et al., 2023] and [Liao et al., 2024] predict trajectories over just a few meters, using physics-based and other techniques. However, given the large gaps between geolocation points of speed radars, this approach is not suitable for the context of this study.

In the context of the current study, which considers XGBoost, LSTM, and the TLE architecture, we address these gaps by thoroughly examining the shift in the best-performing architecture, conducting a shift point analysis, and exploring the impact of geographic representation variation. This comprehensive approach aims to improve the prediction of the next external sensor's trajectory, particularly in the challenging context of sparse and discrete data associated

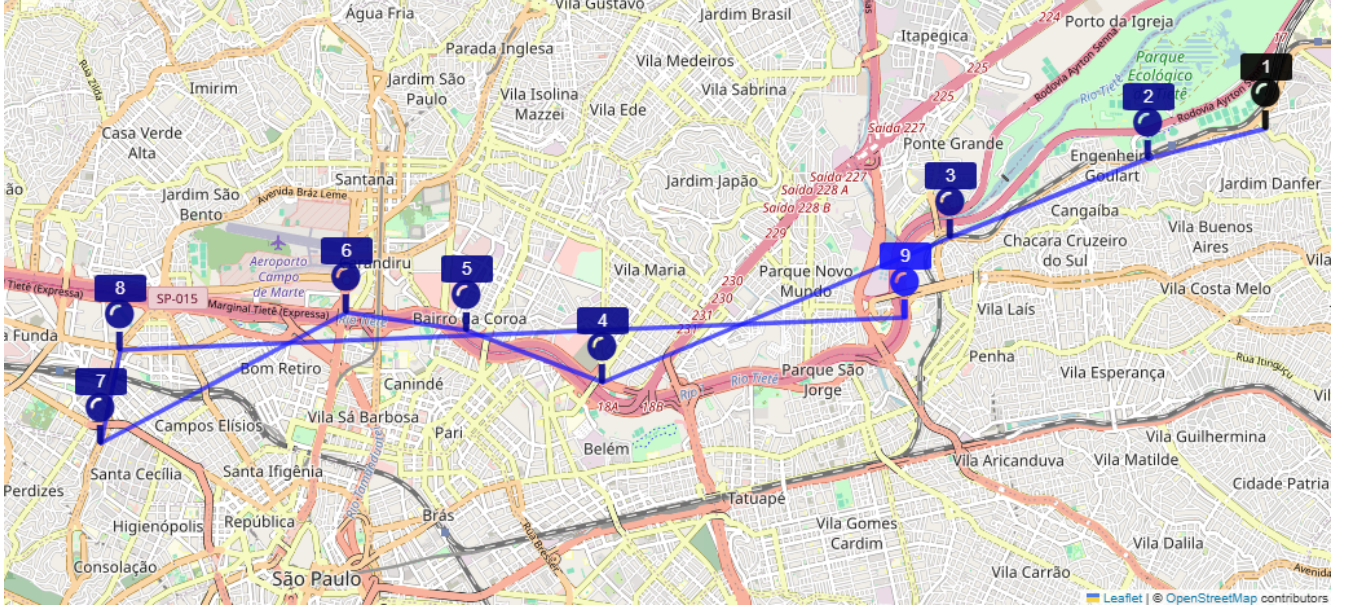


Figure 3. Example of a trajectory with 9 external sensor records in SP.

with stolen vehicles. The comparative summary is presented in Table 1.

3 Problem Definition

For predicting a vehicle’s geolocation in the context of the next external sensor, there is an inherent margin of inaccuracy. On streets with multiple lanes, the same latitude and longitude are recorded regardless of the lane in which the vehicle was identified. For prediction purposes, considering a regression problem for the Artificial Intelligence (AI) model, the predicted location may not correspond to an existing external sensor, requiring adjustment to align with the latitude and longitude of a known external sensor. Additionally, there is the imprecision in the complete actual trajectory, as only discrete and sparse positions are recorded.

Figure 4 presents an example of next external sensor prediction. In this image, the blue vehicle could turn left, where it would be identified again by the external sensor at the top of the image. Alternatively, it could proceed straight and be identified by the sensor to the far right of the image. Lastly, it could turn right and avoid being identified by any other external sensor close to the starting point. In the case illustrated in Figure 4, the prediction concludes that the vehicle is most likely to continue straight. This scenario demonstrates that using street lane information, where the vehicle was identified, may assist in predicting the next external sensor, even without exact knowledge of the intermediate route between points [Macedo et al., 2024].

After visually understanding the problem situation, we can more precisely define the problem. When a vehicle passes by an external sensor capable of identifying it, a record $r = (p, s, t)$ is generated, where p represents the unique identification data of the vehicle, s represents the unique identification data of the sensor and t represents the record’s timestamp. Each sensor s records passages only on a single street lane. However, each geographic location,

$geo = (lat, long)$, can correspond to a single sensor s or a set of sensors $(s_1, s_2, \dots, s_n)$ located on different street lanes in the same area. In other words, each sensor s has a corresponding geographic location, $geo = (lat, long)$, which may or may not be shared with other sensors s . Therefore, s can be replaced with geo , meaning $s = (lat, long)$. The timestamps t are used to ensure the correct chronological order of records’ data. As shown by [Macedo et al., 2024], the best approach is to use both pieces of location information as input: $r = (p, s, lat, long)$. This study will adopt that approach accordingly.

For prediction, a sequence of records $seq = (r_1, r_2, \dots, r_n)$ is used as input, with the goal of predicting the actual location of the next external sensor r_{n+1} as output. Therefore, each architecture model is trained to find a regression function $f(x)$ such that $f(seq) = r_{n+1}$.

Moreover, in addition to the information about the street lane, another factor that can impact prediction results is how the geolocation of the external sensor is represented. For instance, for the same geographic point, instead of using latitude and longitude, it is possible to employ Northing and Easting, potentially utilizing different reference systems. This is because AI models, especially in regression problems such as this one, process numerical data, and depending on the relationships inherent to each geographic representation, as well as the distribution and distances between geolocation records, the model may capture patterns in the input data sequences better, equally well, or worse.

In the next section, the data preprocessing, AI architectures (XGBoost, LSTM, and TLE), metrics, and experimental methodology will be defined, followed by the presentation of the results and analysis.

4 Experiments

To address the four primary research questions of this study, the experiments were conducted as outlined below.

Table 1. Comparative summary between Studies.

#	Paper	XGBoost	LSTM	Transformer and LSTM Entangled	Sparse/ Discrete Data	Stolen Vehicle Data	Street Lane Information	Shift in best-performing Architecture	Shift Point Analysis	Geographic Representation Analysis
1	[Gaiduchenko et al., 2020]	×	✓	×	✓	×	×	×	×	×
2	[Xu et al., 2023]	×	✓	×	×	×	×	✓	×	×
3	[Bae et al., 2022]	×	✓	×	×	×	×	×	×	×
4	[Cruz et al., 2019]	×	✓	×	✓	×	×	✓	×	×
5	[Cruz et al., 2021]	×	✓	×	✓	×	×	×	×	×
6	[Neto et al., 2021]	✓	✓	×	✓	✓	×	×	×	✓
7	[Cruz et al., 2022]	×	✓	×	✓	✓	×	×	×	×
8	[Tsiligkaridis et al., 2020]	×	✓	×	×	×	×	✓	×	×
9	[Chen et al., 2022]	×	✓	×	×	×	×	×	×	×
10	[Chen et al., 2024]	×	✓	×	×	×	×	×	×	×
11	[Macedo et al., 2024]	×	✓	✓	✓	✓	✓	✓	×	×
12	This Study	✓	✓	✓	✓	✓	✓	✓	✓	✓

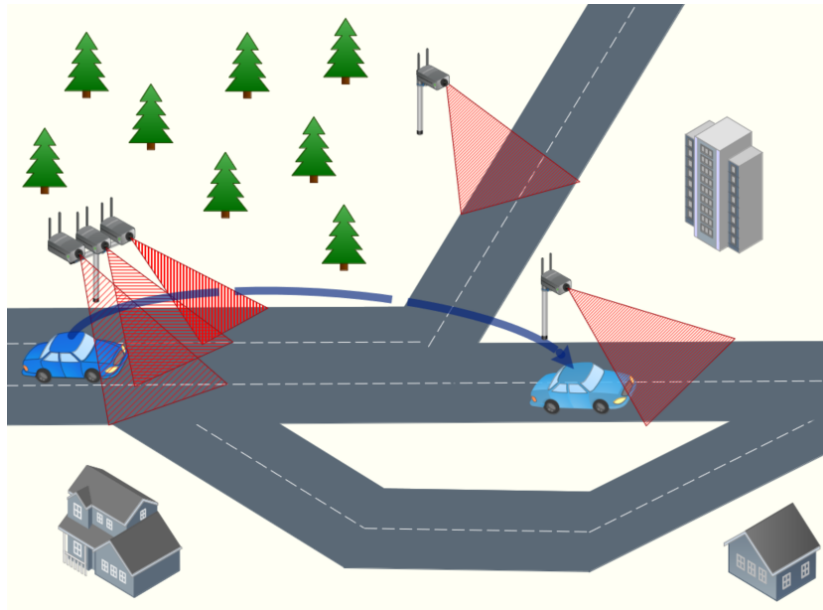


Figure 4. Illustration of the prediction of the next external sensor.

4.1 Data Preprocessing

A database of speed radar records, from a non-public anonymized database that we had access from government for this study, containing data of external sensor records that generated stolen or robbed vehicle alerts (e.g., cars, motorcycles, trucks) in the state of São Paulo, Brazil, covering the period from April 30, 2021, to May 23, 2024 was utilized. This dataset includes anonymized license plate numbers, timestamps, and unique speed radar identification codes. By merging this data with another database containing the latitude and longitude coordinates corresponding to each unique speed radar identification code, we were able to determine the geolocation for each record. The resulted dataset by merging those two databases was called “df_all”.

One of the objectives of this study is to investigate the impact of changes in geographic representation. In addition to latitude and longitude, we transformed the data and created columns with two additional types of geographic representations for the same latitude and longitude points. The three geographic representations used are as follows:

- **EPSG:4326**, also known as Latitude and Longitude in the latest version World Geodetic System 1984 (WGS84), which uses geographic coordinates latitude

and longitude in degrees.

- **EPSG:3857**, also known as Web Mercator, which uses Easting and Northing measured in meters, rather than degrees as in latitude and longitude.
- **EPSG:32723**, also known as Universal Transverse Mercator (UTM), which also uses Easting and Northing measured in meters, but with different reference systems.

The European Petroleum Survey Group (EPSG) was the original name of the organization responsible for establishing a standardized set of coordinate reference systems. Each EPSG code serves as a unique identifier used to define a specific coordinate reference system. These codes specify how geographic data is projected onto a flat surface, such as a map, ensuring that different datasets align correctly when combined [Wikipedia, 2024a].

The geographic representation systems EPSG:4326 (WGS84), EPSG:3857 (Web Mercator), and EPSG:32723 (UTM) differ in how they project Earth’s surface, which can impact vehicle geolocation predictions. EPSG:4326 (WGS84) represents locations using latitude and longitude in degrees, making it ideal for global coverage but less precise for local predictions due to Earth’s curvature. EPSG:3857 (Web Mercator) projects geographic coordinates onto a flat plane using meters, which distorts distances, especially near

the poles, and is commonly used for web mapping, displaying geographic data on interactive digital two-dimensional maps to visualize geographic information. EPSG:32723 (UTM) also uses meters but divides the Earth into zones, providing more accurate local measurements, making it better suited for regional-scale predictions. These differences in precision and projection can influence the accuracy of vehicle geolocation predictions [IBGE, 2016][Wikipedia, 2024b][GISGeography, 2016].

Although this study uses three different geographic representations, they share a common structure: two components of information are used, one for horizontal distance, either Longitude or Easting, and one for vertical distance, either Latitude or Northing. Notably, Easting is typically mentioned first because, in projected coordinate systems, it follows the convention of treating Easting as the x-coordinate and Northing as the y-coordinate, following the (x, y) format. In contrast, for Latitude and Longitude, geographic coordinates are generally listed as (Latitude, Longitude), with latitude appearing first, in accordance with the geographic convention of (Latitude, Longitude).

4.2 Architectures

Three architectures were employed for training and predictions: XGBoost; LSTM; and Transformer and LSTM Entangled (TLE). These architectures are presented and detailed in subsections 4.2.1, 4.2.2, and 4.2.3, respectively.

Since the machine learning process is highly experimental [Lima, 2022], the “Root Mean Squared Error” (RMSE) loss function was adopted for the LSTM and TLE architectures, as it yielded better results in several exploratory analysis tests compared to “Mean Squared Error” (MSE), “Mean Absolute Error” (MAE), and other custom metrics tested. For XGBoost, however, “Mean Squared Error” (MSE) was used as the default and optimal loss function.

4.2.1 XGBoost Architecture

The XGBoost (Extreme Gradient Boosting) architecture is a highly efficient and scalable implementation of Gradient Boosted Regression Trees (GBRT), designed for structured data, including sequential data. It constructs an ensemble of decision trees in a sequential manner, where each tree aims to correct the errors made by the previous ones [Kundu, 2023]. XGBoost is effective for tasks such as classification and regression, where it can handle large datasets and complex feature interactions with remarkable performance.

The XGBoost architecture used was based on the GBRT architecture described in [Neto et al., 2021] configured with the following parameters: $max_depth = 3$; $learning_rate = 1$; $n_estimators = 100$; $min_child_weight = 1$; and $scale_pos_weight = 1$. One instance of this model is used to predict part 1, for example latitude or northing, and another is used for part 2, for example longitude or easting. Two instances of the XGBoost architecture were employed because this architecture is designed to handle two-dimensional input data, while LSTM and TLE can handle three-dimensional input data ($batch_size$, $time_steps$, $n_features$). This XGBoost model, when created, consists

of approximately 3 thousand splits (the equivalent a about 3 thousand trainable parameters) and is illustrated in **Figure 5**.

4.2.2 LSTM Architecture

The Long Short-Term Memory (LSTM) architecture is specifically designed to tackle the challenges of sequential learning. The key feature of LSTM is its ability to “remember” information over long sequences, enabling it to capture temporal dependencies and patterns over time. By processing data step by step, it ensures that the order of events is considered during learning. This makes it highly effective for tasks involving temporal data, such as time series and trajectory predictions.

The Long Short-Term Memory (LSTM) architecture used in this study follows the state-of-the-art model described in [Gaiduchenko et al., 2020] and [Macedo et al., 2024], configured with the following parameters: AdamW optimizer with the $learning_rate = 0.001$, $beta1 = 0.9$, $beta2 = 0.98$, and $epsilon = 10^{-8}$. This LSTM model, when created, comprises approximately 3.4 million trainable parameters and is illustrated in **Figure 6**.

4.2.3 TLE Architecture

Self-Attention Learning, as utilized in the Transformer architecture [Vaswani et al., 2017] [Macedo et al., 2024], allows the model to capture global dependencies across the entire sequence without requiring to process the data in a sequential order. This mechanism enables the model to focus on different parts of the sequence simultaneously, capturing long-range relationships and complex patterns that may occur at any point in the data series. On the other hand, Sequence Learning, represented by LSTMs, specializes in modeling temporal dependencies by processing data step by step. LSTMs retain information over time through internal memory, enabling the model to learn patterns dependent on event order and maintain a history within its memory cells.

By entangling these two mechanisms in the Transformer and LSTM Entangled (TLE) architecture, the model has the potential to leverages the strengths of both approaches. The Transformer captures global dependencies within the sequence, while the LSTM adds a layer of temporal learning that preserves the order of events. This combination enables TLE to learn both global patterns and structural relationships as well as detailed temporal dependencies, resulting in a model capable of capturing more complex nuances in sequential data.

The Transformer and LSTM Entangled (TLE) architecture, which the results have been compared against the other two architectures, is the innovative model described in [Macedo et al., 2024]. It uses the following configuration: $num_layers = 4$, $embed_dim = 64$, $num_heads = 4$, and $ff_dim = 64$. In that study, LSTM and TLE exhibited a shift point where the best-performing architecture changed, which is the focus of this study. The TLE model, when created, has approximately 1 million trainable parameters and is illustrated in **Figure 7**.

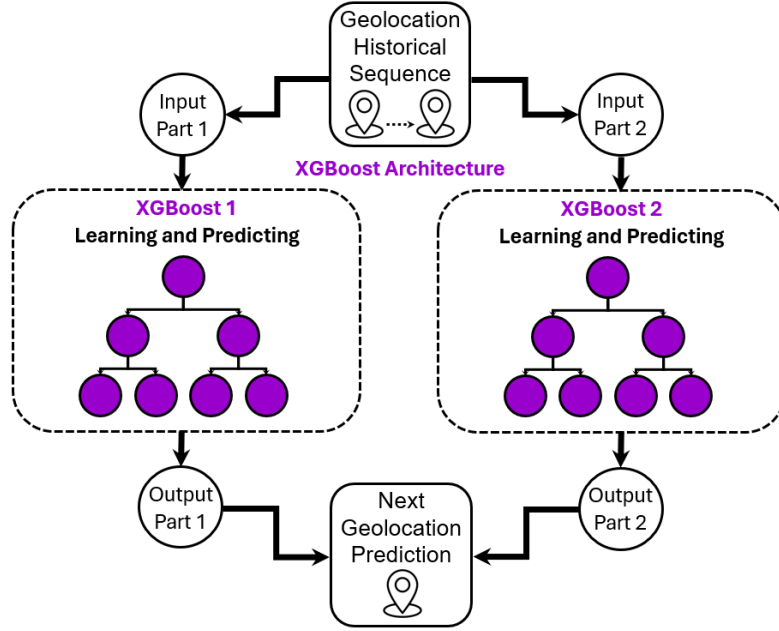


Figure 5. XGBoost Architecture.

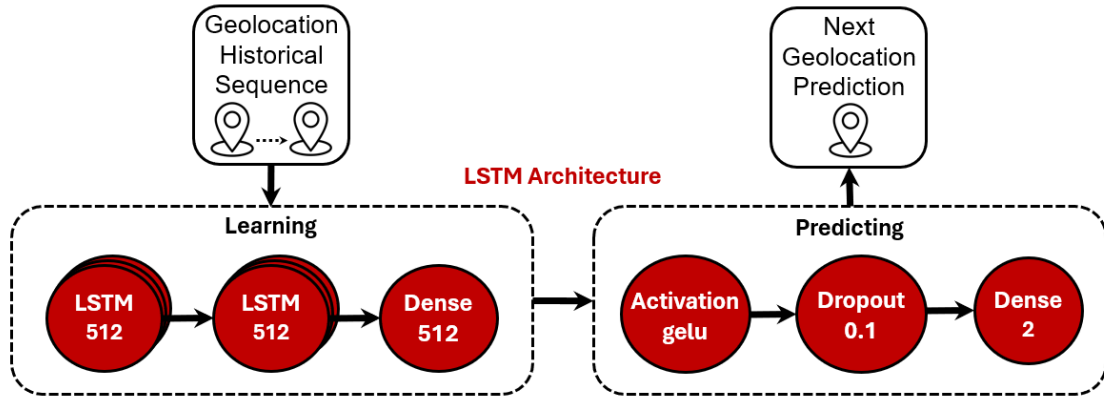


Figure 6. LSTM Architecture.

4.3 Methodology

In this study, three different architectures are compared: the state-of-the-art LSTM (Figure 6), an XGBoost architecture (Figure 5), and a TLE architecture (Figure 7). In addition to these architectures, three different geographic representations are also compared: WGS84, Web Mercator, and UTM. The data and architectures are applied within this context as a regression problem.

Each of these architectures was trained using the resulted dataset “df_all” mentioned in 4.1. Considering this information, the strategy developed to address the four Research Questions is outlined as follows.

To address **RQ1** (Which architecture performs best with short sequence lengths for prediction?) and **RQ2** (Is there a shift point where the best-performing architecture changes?), and based on findings from [Macedo et al., 2024], where TLE and LSTM exhibited a performance shift between sequence lengths of 2 and 10, this study, using WGS84 geographic representation, began with a sequence length of 1 and continued incrementally to find a possible shift point and to ensure it did not reverse. This process covered sequence

lengths from 1 to 8 historical points, training, predicting, and adding one point at a time.

Next, to address **RQ3** (Does the format of geographic representation impact the prediction results?) and **RQ4** (Can the geographic representation format alter the shift point between the best-performing architectures, if one exists?), the study trained and predicted using sequence lengths at the identified shift point boundaries (observed in **RQ1** and **RQ2**), but this time using Web Mercator and UTM representations. The primary goal was to determine if the geographic representation consistently influenced the results, while the secondary goal was to assess whether the shift point remained between the same sequence lengths, which were identified as 5 and 6.

To assess the results, the metrics used were Mean Squared Error (MSE) and R-Squared (R^2). MSE is a standard metric for regression problems [Brownlee, 2021] [Botchkarev, 2018], while R-Squared, or the coefficient of determination, is commonly used in machine learning as an evaluation metric for regression models. It helps determine how well a model fits the data [Onose, 2023a]. The value of R^2 ranges from 0 to 1, where 0 indicates that the model explains none

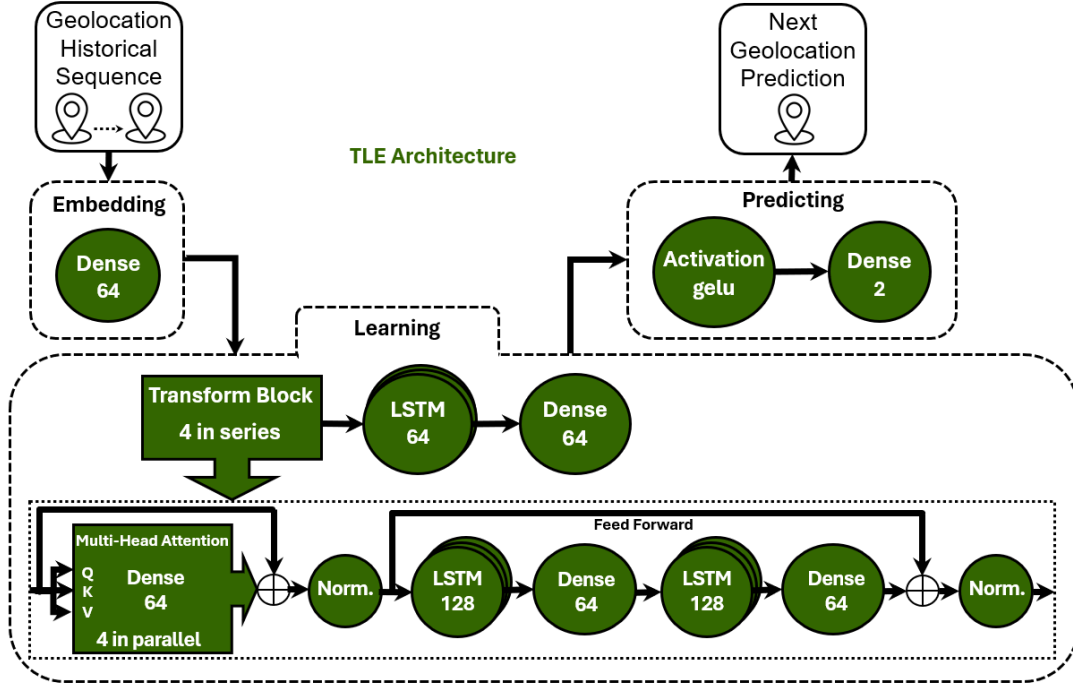


Figure 7. TLE Architecture.

of the variation, and 1 indicates that the model explains all observed variation [Onose, 2023b] [Oliveira, 2021]. Therefore, for MSE, lower values indicate better performance, and for R^2 , higher values indicate better performance.

The dataset “df_all” were split into 80% for training and 20% for testing, which is a common division for training machine learning models [Raschka, 2018] [Brownlee, 2020]. Unlike studies such as [Cruz et al., 2022] and [Macedo et al., 2024], which used sliding windows to increase the number of sequences, this study, with a greater focus on **RQ1** (Which architecture performs best with short sequence lengths for prediction?), used only the first n_{seq} records of each vehicle for each day as the input sequence for the models, where n_{seq} represents the sequence length. Additionally, to ensure robust result analysis, the dataset were first ordered by the unique vehicle identifier and then by timestamp to reduce potential temporal biases.

Evaluations were conducted twice for each set of (architecture, n_{seq} , geographic representation): first using the last 20% of the dataset as the test set, and then using the first 20%, with the final metric value taken as the average of results obtained in each test. This approach is like a partial k-fold cross-validation, ensuring a robust yet efficient experiment analysis. The full k-fold cross-validation was not performed due to its high computational cost. However, we believe our approach is sufficient for the study’s objectives. Additionally, we have made the full databases and source code available in the “Availability of Data and Material” section, allowing researchers who wish to conduct further validations to do so. Considering all aspects, a total of 168 executions were performed in this study, comprising: 42 scenarios, 84 training executions, and an additional 84 prediction executions.

We did not perform a runtime and computational cost comparison because the significant differences occur primarily during the training phase, which is typically conducted in a

non-critical period. For prediction, the time difference is negligible, as all architectures complete the next external sensor prediction in less than a second.

Data processing, point plotting, and trajectory visualization on maps were carried out using the Python programming language and OpenStreetMap within a JupyterLab platform, running on a notebook configured with a 12th-generation i7 processor, 16GB RAM, and a 1TB solid-state drive (SSD).

4.4 Results and Analysis

For the analysis, this study investigates four main research questions based on the results from three architectures (LSTM, XGBoost, and TLE) combined with three geographic representations (WGS84, Web Mercator, and UTM). The aim was to evaluate each architecture’s performance with various short sequence lengths and geographic representations, as well as to identify potential shift points between the best-performing architectures.

For the first research question **RQ1**: *Which architecture performs best with short sequence lengths for prediction?*, the models were tested with sequence lengths ranging from 1 to 8 using the WGS84 geographic representation. The results indicated that the TLE architecture performed best with the shortest sequence lengths (1 to 5), while LSTM began to outperform TLE as the sequence length increased beyond 5 (6 to 8). This suggests that TLE is more effective at capturing information from shorter sequences, whereas LSTM benefits from longer historical data. The XGBoost architecture consistently underperformed compared to both LSTM and TLE in all these scenarios. **Table 2** demonstrates these results using MSE and R^2 metrics. These values represent the output results of the AI models without denormalization, ensuring better comparability, considering that variations in geographic representation formats will be applied.

Table 2. Comparison of MSE and R^2 values across different sequence lengths for XGBoost, LSTM, and TLE architectures.

Sequence Length	XGBOOST		LSTM		TLE	
	MSE	R^2	MSE	R^2	MSE	R^2
1	0,0004848	0,9034	0,0004809	0,9047	0,0004741	0,9059
2	0,0003932	0,9104	0,0003841	0,9141	0,0003473	0,9217
3	0,0003845	0,9061	0,0003468	0,9172	0,0003353	0,9197
4	0,0003725	0,9083	0,0003490	0,9150	0,0003237	0,9208
5	0,0004016	0,8995	0,0003278	0,9153	0,0003121	0,9200
6	0,0004310	0,8918	0,0002852	0,9223	0,0002948	0,9193
7	0,0004227	0,8797	0,0002884	0,9186	0,0003099	0,9127
8	0,0004395	0,8788	0,0003001	0,9114	0,0003286	0,9030

Additionally, for the second research question, **RQ2**: *Is there a shift point where the best-performing architecture changes?*, in line with the findings from [Macedo et al., 2024], TLE performed better for sequence lengths 1 to 5, but a shift point was identified between sequence lengths 5 and 6, where LSTM began to outperform TLE. XGBoost showed comparatively lower performance than the other two architectures across both sequence length ranges, 1 to 5 and 6 to 8. **Figure 8** also demonstrates the shift point in the results, using the WGS84 geographic representation along with MSE and R^2 metrics, showing where the LSTM line crosses the TLE line between sequence lengths 5 and 6.

To answer **RQ3**: *Does the format of geographic representation impact the prediction results?*, in addition to WGS84 geographic representation, the performance of each architecture was evaluated converting geolocations from WGS84 to Web Mercator and to UTM representations. Using the same metrics, MSE and R^2 , this study compares Web Mercator over WGS84, UTM over WGS84, and UTM over Web Mercator. For interpreting the graphics, it is important to note that lower MSE indicates better performance, as does a higher R^2 . Considering this, and analyzing the comparisons for each architecture, we identify three possible scenarios for each metric when comparing the results of two distinct geographic representations:

1. MSE representation A over MSE representation B < 1.0 : This indicates that representation A has an advantage over representation B in terms of MSE.
2. MSE representation A over MSE representation B > 1.0 : This indicates that representation A has a disadvantage compared to representation B in terms of MSE.
3. MSE representation A over MSE representation B $= 1.0$: This indicates that changing the geographic representation had no impact on MSE.
4. R^2 representation A over R^2 representation B > 1.0 : This indicates that representation A has an advantage over representation B in terms of R^2 .
5. R^2 representation A over R^2 representation B < 1.0 : This indicates that representation A has a disadvantage compared to representation B in terms of R^2 .
6. R^2 representation A over R^2 representation B $= 1.0$: This indicates that changing the geographic representation had no impact on R^2 .

Table 3 shows that across all comparisons, none of the geographic representation conversions maintained the same MSE or R^2 values. This finding indicates that the geolocation representation format does indeed impact the results. Although this was not a primary research question, we further analyzed whether the impact was positive or negative, as it remains an intriguing and relevant aspect of the study. One of the reasons that may explain these impacts is that, unlike converting a measurement from inches to meters, for example, where only the unit of measurement changes while maintaining the same way of measuring, changing the geographic representation format alters not only the unit (e.g., from degrees to meters) but also the way measurements are taken and their reference systems. This fundamental difference potentially leads to the observed impacts.

For this in-depth evaluation, it is essential to combine MSE and R^2 to ensure consistent conclusions. If a change in geographic representation led to improvements in both MSE and R^2 , it was classified as a positive change. Conversely, if the change resulted in declines in both metrics, it was deemed a negative change. Finally, if one metric improved while the other declined, regardless of which one, the impact was considered inconclusive. Using this approach and observing **Figure 9**, **Figure 10**, and **Figure 11**, which illustrate the comparison impact of each geographic representation change, we can construct the summary **Table 4** that presents the results of the comparisons.

In conclusion, changing to the Web Mercator format did not exhibit a consistently positive or negative impact in this study, whereas adopting UTM over WGS84 or Web Mercator had a positive effect on XGBoost and LSTM across the three sequence lengths analyzed (4, 5, and 6). For TLE, UTM showed a positive impact at $n_{seq} = 4$, but a negative impact at sequence lengths 5 and 6. These results demonstrate that geographic representation does impact the prediction outcomes and that the effect may vary with sequence length, resulting in either a positive or negative influence.

For the final research question, **RQ4**: *Can the geographic representation format alter the shift point between the best-performing architectures, if one exists?*, this study analyzed the shift point identified in RQ1 and RQ2. Using the same range of sequence lengths (4, 5, and 6) was sufficient to observe that, with the WGS84 representation format, the shift point between LSTM and TLE, initially occurring from sequence lengths 5 to 6, was displaced to between sequence lengths 4 and 5 with the adoption of UTM representation for-

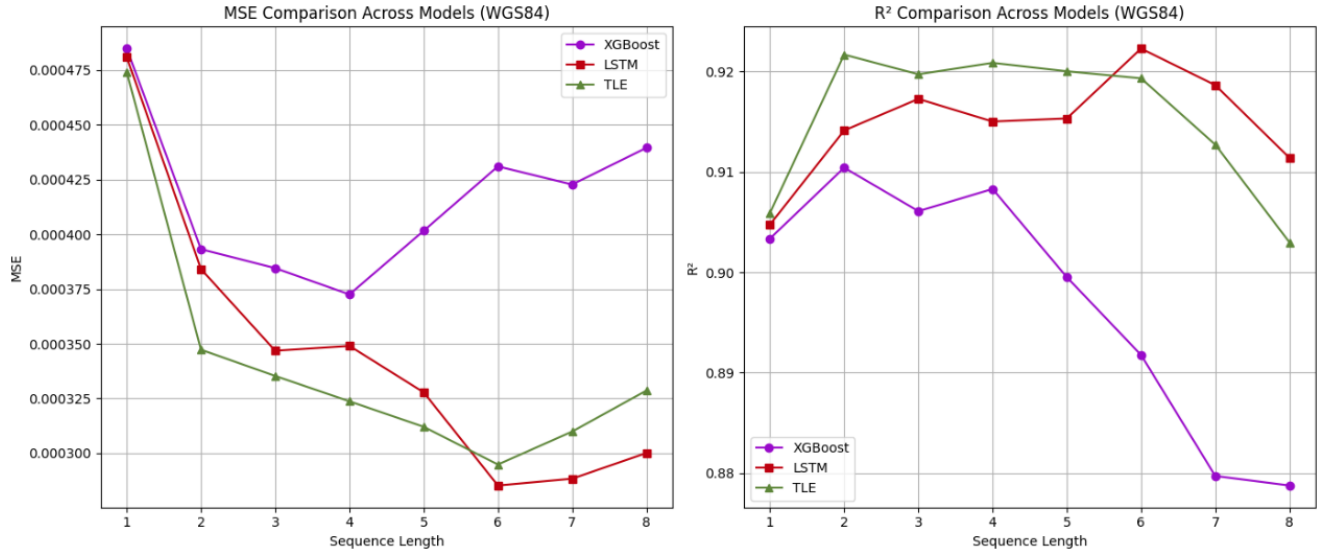


Figure 8. Comparisons of XGBoost, LSTM, and TLE Architectures using MSE and R² Metrics.

Table 3. Comparison of changes in geolocation representation formats.

Representation Comparison	Sequence Length	XGBOOST		LSTM		TLE	
		MSE	R ²	MSE	R ²	MSE	R ²
Web Mercator / WGS84	4	0,9940054	1,0007	1,0036207	1,0001	1,0160543	0,9989
Web Mercator / WGS84	5	1,0040891	1,0002	1,0116758	0,9993	1,0040741	1,0009
Web Mercator / WGS84	6	1,0950706	0,993	1,0023664	1,0003	0,9985159	1,0008
UTM / WGS84	4	0,9630129	1,0015	0,9769478	1,0005	0,9723658	1,0009
UTM / WGS84	5	0,9333241	1,0037	0,9772796	1,0009	1,0482563	0,9932
UTM / WGS84	6	0,9882064	1,0001	0,9724874	1,0011	1,0024658	0,9995
UTM / Web Mercator	4	0,9688205	1,0008	0,9734234	1,0005	0,9570018	1,0021
UTM / Web Mercator	5	0,9295231	1,0035	0,9660008	1,0015	1,044003	0,9924
UTM / Web Mercator	6	0,9024135	1,0072	0,9701915	1,0007	1,0039557	0,9987

mat. Based on these results, it is concluded that, depending on the geographic representation used, the shift point between the best-performing architectures can indeed vary, as illustrated in Figure 12.

Comparing this study's results with prior research, it is observed that the Transformer and LSTM Entangled (TLE) architecture performs best with short sequences, while LSTM becomes more effective as the sequence length increases. This finding aligns with [Macedo *et al.*, 2024], which identified a shift point between these architectures for certain sequence lengths.

In terms of geographic representation, this study provides a more detailed and comprehensive analysis of how different coordinate systems impact prediction accuracy, a topic less extensively discussed in prior studies, such as [Neto *et al.*, 2021]. These comparisons are presented in Table 2 and Table 3 and illustrated in Figure 9, Figure 10, and Figure 11, which synthesize the performance of the XGBoost, LSTM, and TLE architectures across various geographic representations and sequence lengths.

Unlike previous works that focus solely on using different architectures or additional information, this study contributes significantly by demonstrating that both sequence length and representation format play essential roles in defining the ideal architecture. This broader analysis can translate into cost savings for law enforcement agencies and increased

effectiveness in police interventions.

5 Conclusion

In this study, two criteria were evaluated for their influence on selecting the best-performing architecture for predicting the next external sensor in the context of stolen vehicles: the length of historical sequence data and the format of geolocation representation.

The results highlight the importance of considering these two criteria, as there is a shift point where the optimal architecture changes depending on the length of the historical sequence data. Additionally, geographic representation format changes were observed to have a positive, negative or inconclusive impact on predictions, depending on the architecture and the historical sequence length. Furthermore, a change in geolocation representation format displaced the position of this shift point.

In the context of this study, with geolocation represented in latitude and longitude values (WGS84), the Transformer and LSTM Entangled (TLE) architecture demonstrated superior performance over LSTM and XGBoost for historical sequence lengths from 1 to 5. The LSTM architecture began to outperform for historical sequence lengths from 6 to 8. This indicates a shift in the optimal architecture as the his-

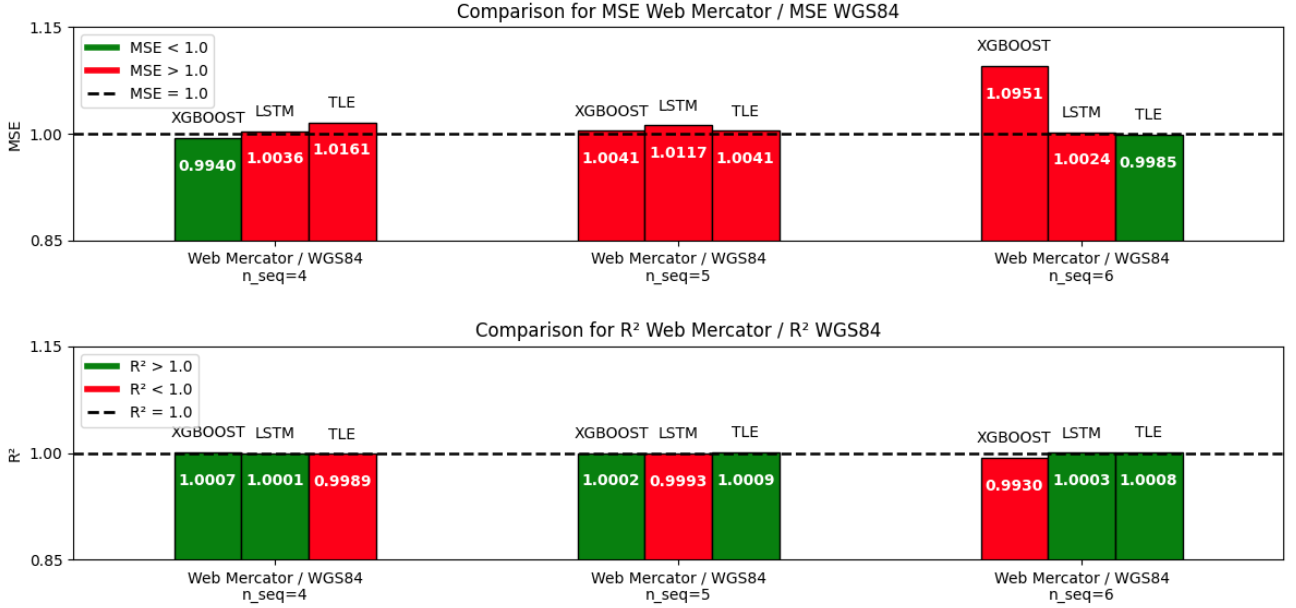


Figure 9. Comparisons of Web Mercator over WGS84 in geographic representation formats using MSE and R² Metrics.

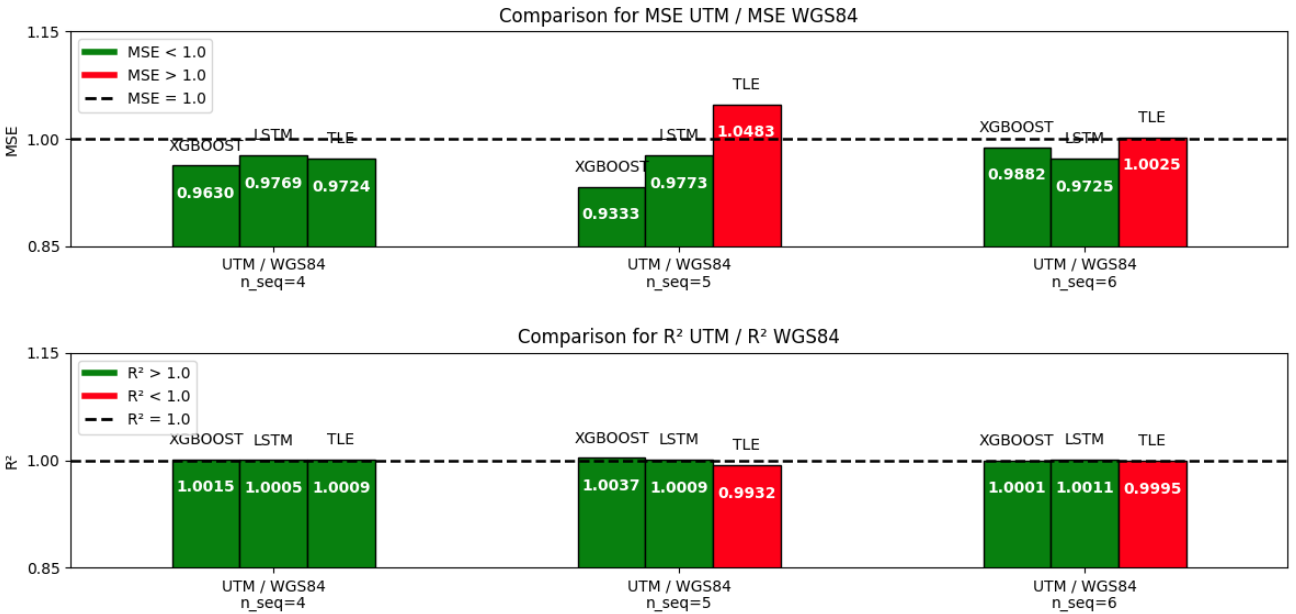


Figure 10. Comparisons of UTM over WGS84 in geographic representation formats using MSE and R² Metrics.

torical sequence length increases from 5 to 6 for prediction purposes. When using the Universal Transverse Mercator (UTM) geographic representation, this shift point between the best-performing architectures occurs at sequence lengths 4 and 5.

Therefore, it is concluded that when selecting the most suitable architecture for predicting the next external sensor in stolen vehicles, the length of the historical sequence and the format of geographic representation format should be considered to achieve optimal results.

From a broader perspective, the results achieved in the context of stolen vehicles can provide insights for optimizing the allocation of public resources and refining highway monitoring strategies, as well as enhancing the deployment of law enforcement forces in public security operations. In this regard, this study presents a potential efficiency gain for public se-

curity institutions and contributes to the recovery of citizens' vehicles, as well as mitigating various subsequent crimes. Additionally, insurance companies may benefit from this predictive capability to assess risk more accurately, refine policy pricing, and develop proactive fraud detection mechanisms. By integrating these advancements into decision-making processes, both sectors can improve their strategic planning and contribute to broader public safety and financial sustainability.

While this study contributes significantly to understanding how historical sequence length and geolocation representation affect the selection of the optimal prediction architecture, certain limitations must be noted. Firstly, the dataset is limited to records of stolen vehicles, potentially restricting the generalizability of results to other trajectory prediction contexts. Additionally, the analysis is based on discrete

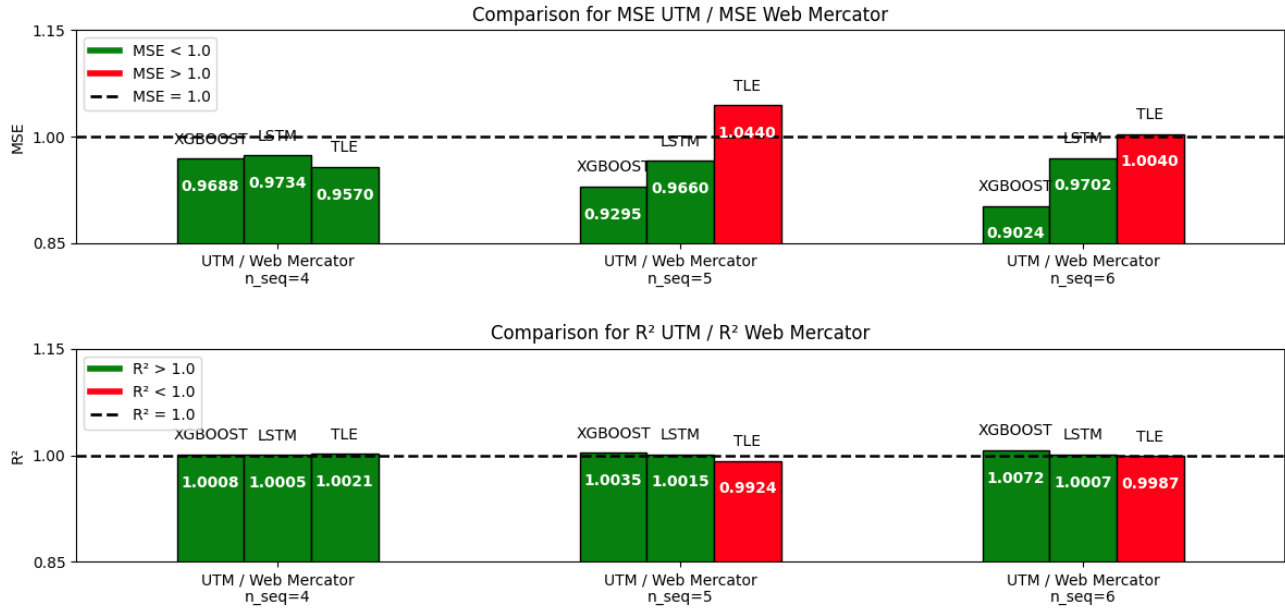


Figure 11. Comparisons of UTM over Web Mercator in geographic representation formats using MSE and R² Metrics.

Table 4. Impact of relative changes in geographic representation formats across sequence lengths.

Geographic Representation Comparison	XGBoost Impact			LSTM Impact			TLE Impact		
	n_seq=4	n_seq=5	n_seq=6	n_seq=4	n_seq=5	n_seq=6	n_seq=4	n_seq=5	n_seq=6
Web Mercator / WGS84	Positive	Inconclusive	Negative	Inconclusive	Negative	Inconclusive	Negative	Inconclusive	Positive
UTM / WGS84	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Negative	Negative
UTM / Web Mercator	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Negative	Negative

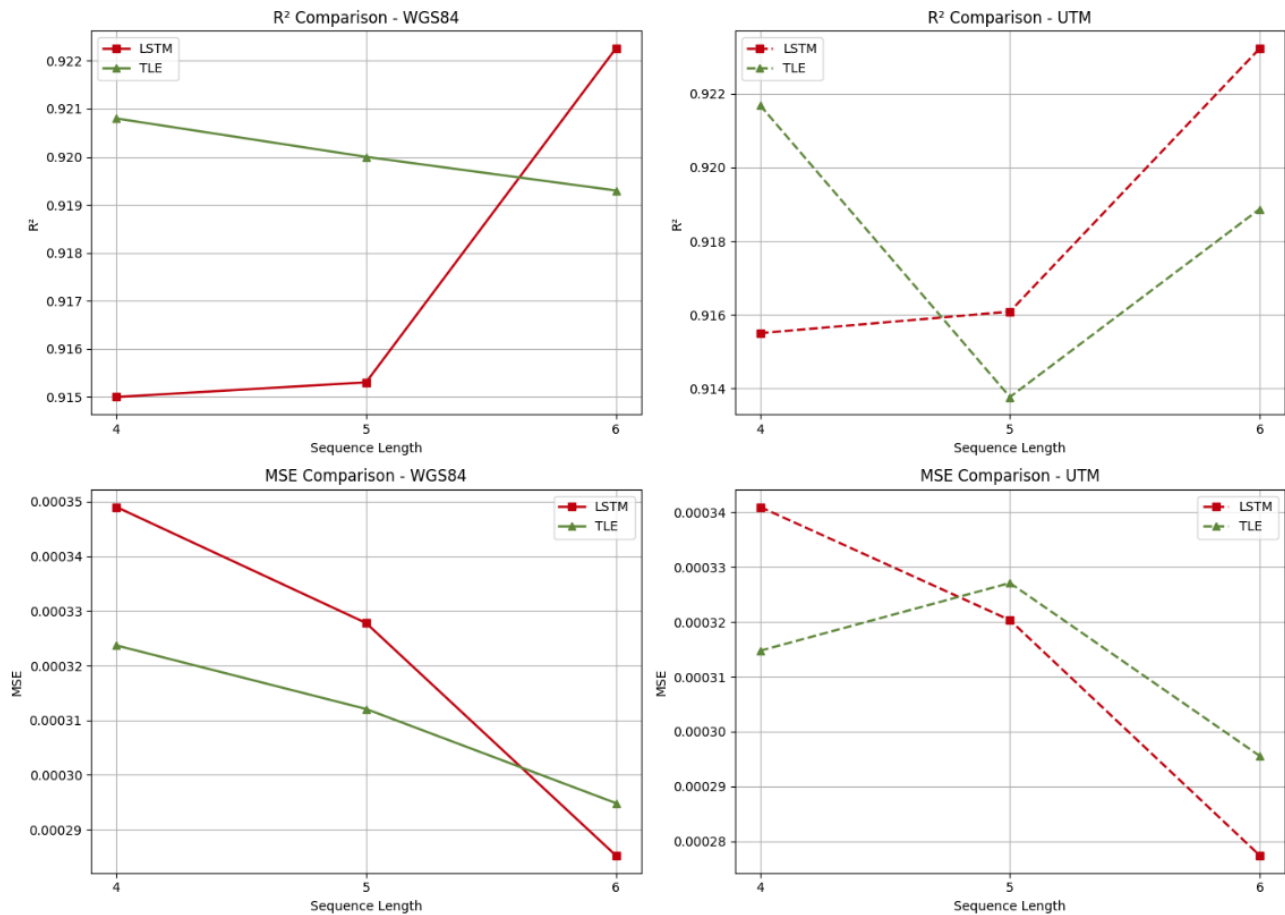


Figure 12. Shift point of best-performing architectures displaced by geographic representation format change.

and sparse sequences from external sensors, a context that may not capture the complex movement patterns present in continuous trajectories. Another limitation is related to the impact of changes in geographic representation and the architecture transition point. Depending on the dataset and other geographic representation formats, the transition may occur at different historical sequence lengths than those identified in this study, considering a possible margin of variability.

In future research, restricting the time frames of records for training and prediction could enhance the accuracy of predicting the next external sensor by capturing more significant patterns in the flow of stolen vehicles. Furthermore, we recommend expanding the dataset to include other vehicle tracking situations and exploring the feasibility of adapting the model to continuous trajectories. These adjustments could enhance the model's applicability and prediction accuracy, supporting its implementation across a broader array of vehicle tracking contexts. In addition, a more comprehensive analysis of the data in the databases would be valuable. Also, it would be worthwhile to conduct further validation experiments to verify whether the shift point observed between LSTM and TLE remains consistent with broader sequence lengths.

Acknowledgements

The authors thank the São Paulo Research Foundation (FAPESP), the National Council for Scientific and Technological Development (CNPq), the Decision Technologies Laboratory (LATITUDE/UnB), the National Confederation of Industry (CNI), and the University of Brasília (UnB). This paper is an extension of our previous work [Macedo et al., 2024].

Funding

This research was partially funded by the Scientific Initiation Program ProIC/DGP through the PIBIC 2023/2024 call for proposals from the University of Brasília (UnB).

Authors' Contributions

Conception, G.V.I.M., G.P.R.F., J.K.M.S., A.R.N. and V.P.G.; Data curation, G.V.I.M. and M.G.A.; Methodology, G.V.I.M., G.P.R.F. and V.P.G.; Coding, G.V.I.M.; Experimental execution, G.V.I.M.; Literature review, G.V.I.M., G.P.R.F., J.K.M.S., A.R.N. and V.P.G.; Supervision, G.P.R.F. and V.P.G.; Writing, G.V.I.M.; Reviewing, G.P.R.F., M.C.F., A.L.M.S. and V.P.G.. G.V.I.M. is the main contributor and writer of this manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing and reviewing of the manuscript, or in the decision to publish the results.

Availability of data and materials

The datasets generated and analyzed during the current study are

available on Github at <https://github.com/gustavomacedo20/jisa-2024-vehicle-prediction>.

Abbreviations

AI	Artificial Intelligence
EPSG	European Petroleum Survey Group
GBRT	Gradient Boosted Regression Trees
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
RQ	Research Question
SP	São Paulo
TLE	Transformer and LSTM Entangled
UTM	Universal Transverse Mercator
WGS84	World Geodetic System 1984
XGBoost	Extreme Gradient Boosting

References

- Almeida, V. G., Silva, T. R., and Silva, F. A. (2022). Se for, vá na paz: Construindo rotas seguras para veículos coletivos urbanos. In *Anais do XL Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 140–153. SBC. DOI: 10.5753/sbrc.2022.221978.
- Bae, K., Lee, S., and Lee, W. (2022). Transformer networks for trajectory classification. In *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 331–333. IEEE. DOI: 10.1109/big-comp54360.2022.00074.
- Bernardi, E., Marte, C. L., Yoshioka, L. R., Ribeiro, P. C. M., and Fontana, C. F. (2015). Modelo sistêmico e classificação de falhas associadas ao sistema de reconhecimento de placas para fiscalização automática de veículos. In *Anais do XXIX Congresso de Pesquisa e Ensino em Transportes*, pages 1510–1521. Available at: http://146.164.5.73:20080/ssat/interface/content/anais_2015/TrabalhosFormatados/AC834.pdf.
- Botchkarev, A. (2018). Evaluating performance of regression machine learning models using multiple error metrics in azure machine learning studio. Available at SSRN 3177507. DOI: 10.2139/ssrn.3177507.
- Brito, M., Martins, B., Santos, C., Medeiros, I., Araújo, F., Seruffo, M., Oliveira, H., Cerqueira, E., and Rosário, D. (2023). Personalized experience-aware multi-criteria route selection for smart mobility. In *Anais do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 504–517. SBC. DOI: 10.5753/sbrc.2023.513.
- Brownlee, J. (2020). Train-test split for evaluating machine learning algorithms. Available at: <https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/> Online; Accessed on: 09-04-2024.
- Brownlee, J. (2021). Regression metrics for machine learning. Available at: <https://machinelearningmastery.com/regression-metrics-for-machine-learning/> Online; Accessed on: 09-04-2024.

- Capanema, C. G. S., Silva, F. A., and Silva, T. R. d. M. B. (2020). Mfa-rnn: Uma rede neural recorrente para previsão de próximo local de visita com base em dados esparsos. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 127–140. SBC. DOI: 10.5753/sbrc.2020.12278.
- Chen, J., Feng, Q., and Fan, D. (2024). Vehicle trajectory prediction based on local dynamic graph spatiotemporal-long short-term memory model. *World Electric Vehicle Journal*, 15(1):28. DOI: 10.3390/wevj15010028.
- Chen, X., Zhang, H., Zhao, F., Cai, Y., Wang, H., and Ye, Q. (2022). Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12. DOI: 10.1109/tim.2022.3192056.
- Cruz, L. A., Coelho da Silva, T. L., Magalhães, R. P., Melo, W. C. D., Cordeiro, M., de Macedo, J. A. F., and Zeitouni, K. (2022). Modeling trajectories obtained from external sensors for location prediction via nlp approaches. *Sensors*, 22(19):7475. DOI: 10.3390/s22197475.
- Cruz, L. A., Zeitouni, K., da Silva, T. L. C., de Macedo, J. A. F., and Silva, J. S. d. (2021). Location prediction: a deep spatiotemporal learning from external sensors data. *Distributed and Parallel Databases*, 39:259–280. DOI: 10.1007/s10619-020-07303-0.
- Cruz, L. A., Zeitouni, K., and de Macedo, J. A. F. (2019). Trajectory prediction from a mass of sparse and missing external sensor data. In *2019 20th IEEE International conference on mobile data management (MDM)*, pages 310–319. IEEE. DOI: 10.1109/mdm.2019.00-43.
- de Souza, A. M., Braun, T., Botega, L. C., Villas, L. A., and Loureiro, A. A. (2019). Safe and sound: Driver safety-aware vehicle re-routing based on spatiotemporal information. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3973–3989. DOI: 10.1109/tits.2019.2958624.
- de Souza, A. M. and Villas, L. A. (2020). Vem tranquilo: Rotas eficientes baseado na dinâmica urbana futura com deep learning e computação de borda. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 351–364. SBC. DOI: 10.5753/sbrc.2020.12294.
- Gaiduchenko, N. E., Gritsyk, P. A., and Malashko, Y. I. (2020). Multi-step ballistic vehicle trajectory forecasting using deep learning models. In *2020 International Conference Engineering and Telecommunication (En&T)*, pages 1–6. IEEE. DOI: 10.1109/ent50437.2020.9431287.
- Geng, M., Li, J., Xia, Y., and Chen, X. M. (2023). A physics-informed transformer model for vehicle trajectory prediction on highways. *Transportation research part C: emerging technologies*, 154:104272. DOI: 10.1016/j.trc.2023.104272.
- GISGeography (2016). How universal transverse mercator (utm) works. Available at: <https://gisgeography.com/utm-universal-transverse-mercator-projection/> Online; Accessed on: 09-30-2024.
- Haviland, C. V. and Wiseman, H. (1974). Criminals who drive. In *Proceedings: American Association for Automotive Medicine Annual Conference*, volume 18, pages 432–439. Association for the Advancement of Automotive Medicine. Available at: [https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds\[\]=citjournalarticle_97342_1](https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_97342_1).
- Hu, H., Wang, Q., Du, L., Lu, Z., and Gao, Z. (2022). Vehicle trajectory prediction considering aleatoric uncertainty. *Knowledge-Based Systems*, 255:109617. DOI: 10.1016/j.knosys.2022.109617.
- IBGE (2016). Sistemas de referência. Available at: https://geofp.ibge.gov.br/informacoes_sobre_posicionamento_geodesico/sirgas/sisref_2.pdf Online; Accessed on: 09-30-2024.
- Karimzadeh, M., Aebi, R., de Souza, A. M., Zhao, Z., Braun, T., Sargento, S., and Villas, L. (2021). Reinforcement learning-designed lstm for trajectory and traffic flow prediction. In *2021 IEEE wireless communications and networking conference (WCNC)*, pages 1–6. IEEE. DOI: 10.1109/wcnc49053.2021.9417511.
- Kundu, N. (2023). Random forests vs gradient boosting: An overview of key differences and when to use each method. Available at: <https://medium.com/@nitishkundu1993/random-forests-vs-gradient-boosting-an-overview-of-key-differences-and-when-to-use-each-method-1dab19fcc283> Online; Accessed on: 03-05-2025.
- Ladeira, L. Z., de Souza, A. M., Rocha Filho, G. P., Silva, T. H., Sanches, M. F., and Villas, L. A. (2019a). Martini: Towards a mobile and variable time window identification for spatio-temporal data. In *2019 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6. IEEE. DOI: 10.1109/latincom48065.2019.8937868.
- Ladeira, L. Z., de Souza, A. M., Rocha Filho, G. P., Silva, T. H., and Villas, L. A. (2019b). Serviço de sugestão de rotas seguras para veículos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 608–621. SBC. DOI: 10.5753/sbrc.2019.7390.
- Ladeira, L. Z., de Souza, A. M., Silva, T. H., Rocha Filho, G. P., Peixoto, M. L. M., and Villas, L. A. (2020). Cerva: Roteamento contextual para veículos com risco espaço-temporal. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 379–392. SBC. DOI: 10.5753/sbrc.2020.12296.
- Liao, H., Wang, C., Li, Z., Li, Y., Wang, B., Li, G., and Xu, C. (2024). Physics-informed trajectory prediction for autonomous driving under missing observation. Available at *SSRN 4809575*. DOI: 10.2139/ssrn.4809575.
- Lima, M. (2022). 6 passos para criar seu primeiro projeto de machine learning. Available at: <https://www.insightlab.ufc.br/6-passos-para-criar-seu-primeiro-projeto-de-machine-learning/> Online; Accessed on: 09-04-2024.
- Macedo, G., Filho, G. R., Santos, J., Neves, A., Almeida, M., Falqueiro, M., Meneguette, R., Serrano, A., Mendonça, F., and Gonçalves, V. (2024). Predição de geolocalização de veículo com alerta de roubo usando lstm, trans-

- former e tle. In *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 61–70, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbcup.2024.2568.
- Neto, J. S. D. S., Da Silva, T. L. C., Cruz, L. A., de Lira, V. M., de Macêdo, J. A. F., Magalhaes, R. P., and Peres, L. G. (2021). Predicting the next location for trajectories from stolen vehicles. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 452–456. IEEE. DOI: 10.1109/ictai52525.2021.00073.
- Oliveira, C. (2021). Métricas para regressão: Entendendo as métricas r^2 , mae, mape, mse e rmse. Available at: <https://medium.com/data-hackers/prevendo-n%C3%BAmoros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70> Online; Accessed on: 09-04-2024.
- Onose, E. (2023a). R squared: Understanding the coefficient of determination. Available at: <https://arize.com/blog-course/r-squared-understanding-the-coefficient-of-determination/#:~:text=R%2Dsquared%2C%20also%20known%20as,explained%20by%20the%20independent%20variables.> Online; Accessed on: 03-05-2025.
- Onose, E. (2023b). R squared: Understanding the coefficient of determination. Available at: <https://arize.com/blog-course/r-squared-understanding-the-coefficient-of-determination/> Online; Accessed on: 09-04-2024.
- Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*. DOI: 10.48550/arXiv.1811.12808.
- Reyna, A. R. H., Farfán, A. J. F., Filho, G. P. R., Sampaio, S., de Grande, R., Nakamura, L. H. V., and Meneguette, R. I. (2024). Medavet: Traffic vehicle anomaly detection mechanism based on spatial and temporal structures in vehicle traffic. *Journal of Internet Services and Applications*, 15(1):25–38. DOI: 10.5753/jisa.2024.3809.
- Tsiligkaridis, A., Zhang, J., Taguchi, H., and Nikovski, D. (2020). Personalized destination prediction using transformers in a contextless data setting. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE. DOI: 10.1109/ijcnn48605.2020.9207514.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. DOI: <https://doi.org/10.48550/arxiv.1706.03762>.
- Wikipedia (2024a). Epsg geodetic parameter dataset. Available at: https://en.wikipedia.org/wiki/EPSS_Geodetic_Parameter_Dataset Online; Accessed on: 09-30-2024.
- Wikipedia (2024b). Web mercator projection. Available at: https://en.wikipedia.org/wiki/Web_Mercator_projection Online; Accessed on: 09-30-2024.
- Xu, Y., Wang, Y., and Peeta, S. (2023). Leveraging transformer model to predict vehicle trajectories in congested urban traffic. *Transportation research record*, 2677(2):898–909. DOI: 10.1177/03611981221109594.