Predictive machine learning-based error correction in GPS/IMU localization to improve navigation of autonomous vehicles

Uchenna Charles Onyema[1*], *Mahmoud* Shafik[1]

[1]College of Science and Engineering, University of Derby, DE 22 1GB, England

Abstract. Precise localization is crucial for the safety-critical factor and effective navigation of autonomous vehicles. This applied research examines machine learning models' use to estimate, predict and correct errors in Global Positioning System (GPS)/ Inertial Measurement Unit (IMU) localization for autonomous vehicles indoors and outdoors applications. This ongoing development aims to improve localization accuracy by utilizing exploratory data analysis (EDA) and implementing models such as linear regression, random forest regressor, and decision tree regressor. The assessment is performed with the mean squared error (MSE) metric, yielding values of $1.7069427028104143e^{-05}$ for the decision tree, linear regression, and random forest models. The results showed that the model with the highest performance is determined by evaluating the Mean Squared Error (MSE) values.

1 Introduction

Research in integrating predictive machine learning-based error correction into GPS/IMU localization for autonomous vehicles is focused on improving the accuracy and reliability of autonomous vehicle navigation systems. As autonomous vehicles become more integrated into our transportation networks, it is increasingly important to ensure that these vehicles can accurately perceive and recognize their surroundings to navigate safely and efficiently. Reinforcement learning (RL) is a viable paradigm for creating adaptive control systems for autonomous vehicles [1]. Furthermore, recent advancements in computational methodologies, such as compressive sensing and machine learning have demonstrated promise in improving the spectrum resolution of devices and rectifying mistakes in GPS/IMU localization systems [2]. Moreover, there has been a growing trend in robotics literature to utilize learning-based techniques to tackle the difficulties of perceiving and understanding the environment for autonomous navigation in unorganized surroundings [3].

Deep learning has garnered interest in autonomous driving for its application in radar data analysis. Specifically, it has been recognized for its ability to correct errors and enhance localization accuracy in GPS/IMU systems of autonomous vehicles [4]. Furthermore, incorporating communication-enabling technology and machine learning in autonomous vehicles has been recognized as a crucial factor in delivering connectivity and precise vehicle

*Corresponding author: u.onyema@derby.ac.uk

location [5]. The importance of machine learning in solving the issues related to errors and or mistake correction and localization in autonomous vehicles is highlighted by these achievements.

Moreover, the utilization of machine learning in several domains, including medical informatics and radiation, showed the possibility of employing predictive machine learning-based error correction in GPS/IMU localization for autonomous vehicles (AV). Machine learning techniques have been employed to assist and automate the diagnostic and treatment processes in radiotherapy with great precision. This suggests that similar applications can be explored in autonomous vehicle localization systems [6]. Furthermore, the application of machine learning in combining sensor data for autonomous vehicles is predicted to modify ground transportation significantly, highlighting the significant influence of machine learning in this field of applications [7].

It is evident that incorporating predictive machine learning-based error correction into GPS/IMU localization for autonomous vehicles shows excellent potential in improving the precision and dependability of autonomous vehicle navigation systems. Thus, it is possible to tackle the difficulties of error correction and localization by utilizing the progress made in machine learning, specifically in reinforcement learning, deep learning, and sensor fusion. It will also greatly aid in the connectivity security and effective implementation of autonomous vehicles.

2 Autonomous vehicle localization technology state of the ART

[8] introduced a new method that uses Convolutional Neural Networks (CNN) to solve the error correction problem in inertial sensors, in the field of autonomous vehicle localization. The proposed CNN-based technique accomplishes near real-time error correction by utilizing the responses of an inertial sensor and accounting for intrinsic noise flaws. The authors employ a time-division approach to preprocess IMU output data, facilitating compatibility with the input format of CNN. The CNN technique is specifically designed to achieve superior performance and reduced complexity, making it suitable for use on energy-efficient hardware such as microcontrollers. The experimental results show substantial reductions in errors, with a maximum improvement of 32.5% in straight-path motion and up to 38.69% improvement in oval motion compared to the ground truth data.

[9] introduced an innovative method for dealing with the problem of precise vehicle positioning when faced with unreliable Global Navigation Satellite System (GNSS) information. The proposed method efficiently mitigates the imperfections in the GNSS data by utilizing a multi-step correction filter. Furthermore, incorporating data from many sensors enables compensation for the particular limitations of each sensor, hence improving the overall dependability of the localization system. Moreover, implementing Generalizable Deep Visual Inertial Odometry (GD-VIO) provides a reliable method for accurately determining the vehicle's position, especially when there is a lack of GNSS signal. Real-world experiments confirm the effectiveness of the suggested algorithms and show their capacity to deliver precise and reliable vehicle posture prediction.

[10] presented a comprehensive analysis of the latest map-based localization approaches. The authors categorize, describe, and evaluate these techniques in order to assess their strengths and weaknesses. The research introduced methods and strategies that align pre-existing maps with observations collected from different sensor modalities on-board. It examines approaches that consider localization as a probabilistic issue, providing understanding into the management of uncertainty and the utilization of Bayesian inference methods. In addition, it examined the developing field of deep-learning localization algorithms, specifically in relation to autonomous vehicles' potential and consequences.

[11] presented a new method to improve the accuracy of GPS localization for autonomous vehicles using a reinforcement learning framework. This methodology differs from traditional methods by not making inflexible assumptions about the hardware specifications of GPS

devices or motion models, and it does not depend on reference locations provided by the infrastructure. Instead, it utilized a reinforcement learning model to learn the best method for improving raw GPS observations. It does this by using a reward mechanism based on confidence, which is not dependent on geolocation. This improved the ability to apply the model in different locations. Moreover, incorporating a map matching-based regularisation term assisted in reducing reward variance. The implementation of the reinforcement learning model utilized the asynchronous advantage actor-critic (A3C) method, which allowed for simultaneous training and supported shorter training sessions to enhance robustness. The suggested approach has been evaluated against an extended Kalman filter benchmark, showing its effectiveness in improving the accuracy of GPS localization for autonomous vehicles.

[12] provided a comprehensive analysis of modern methods designed to improve the effectiveness of autonomous vehicle (AV) systems in close-range or localized contexts. This text explored new research efforts that utilize advanced deep-learning sensor fusion algorithms. The focus was on how these algorithms are applied in important tasks such as perception, localization, and mapping, which are critical for the operation of autonomous vehicles. The research highlights the need to use advanced approaches to improve the performance of autonomous vehicles by combining information from several investigations. Moreover, it outlines developing patterns in the domain and proposes new paths for future investigation, thereby offering significant perspectives into the continuous development of AV technology. [13] The initial work examined the application of computational methods, such as machine learning, to improve the spectrum resolution of devices and rectify mistakes and faults. It demonstrated the promise of these techniques in enhancing localization accuracy. Furthermore, [14] emphasized the capacity of machine learning in rectifying techniques for underwater hyperspectral image processing, demonstrating its suitability in various fields. These preliminary investigations showcase the extensive applicability of machine learning in mistake correction tasks and its capacity to improve localization accuracy in diverse situations. In addition, the third publication [15] highlighted the significance of employing reinforcement learning, a framework within machine learning, for developing adaptive control systems for autonomous vehicles, emphasizing its relevance in this specific situation. This approach signified a divergence from conventional procedures and proposed a transition towards more flexible and adjustable control mechanisms. The fourth research [16] comprehensively examined machine and deep learning techniques for recognizing sport-specific movements utilizing IMU data inputs. This study showcased the practicality of machine learning in analyzing movements, which is relevant to the localization of autonomous vehicles. These works demonstrate the adaptability of machine learning techniques in addressing various elements of autonomous vehicle operation, including control systems and movement identification. Although machine learning showed potential in enhancing localization accuracy, [17] warned about its limitations in comprehending fundamental ideas, which must be considered when addressing error correction in autonomous vehicle localization. This observation highlights the significance of combining the advantages of machine learning with a more profound comprehension of the fundamental principles that govern localization processes. [18] examined the application of machine learning in lidars for object detection, a crucial component for precise localization in autonomous vehicles. This work emphasizes the practical uses of machine learning in solving important problems related to autonomous vehicle localization by specifically focusing on object recognition.

3 Introduction CAD model of predictive machine learning-based error correction for autonomous vehicles

This architectural framework outlines the essential elements involved in using predictive machine learning for error and mistake correction in GPS/IMU localization in autonomous vehicles. The framework (see figure 1 below) consists of interconnected components that are essential for achieving precise and dependable localization in dynamic situations. The methodology starts with data acquisition and preprocessing, which involves collecting and initially processing sensor data from GPS and IMU units. Afterwards, the data goes through feature extraction and selection, where important information is condensed to aid in efficient model training. The learned predictive models are the foundation of the error correction system. It uses machine learning algorithms to predict and correct IMU drift as time passes. Ultimately, rigorous evaluation and validation methods ensure the error correction system's robustness and effectiveness in enhancing the localization performance of autonomous vehicles.

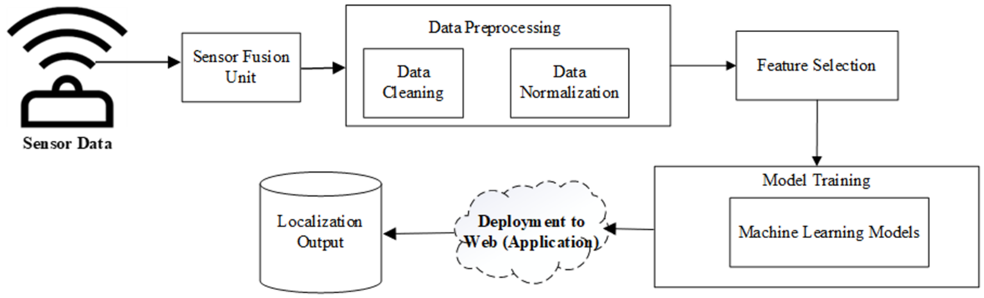


Fig. 1. CAD Model of ML-Based Error Correction System Architecture

A. Sensor Data: Diverse sensors are employed to gather data on the vehicle's environment and internal condition. The following items are included:

GPS Receiver: Offers worldwide positioning data.

An IMU, or Inertial Measurement Unit, is a device that measures a vehicle's acceleration, angular rate, and, occasionally, its magnetic field. It provides valuable information on the vehicle's movement and orientation.

B. Sensor Fusion Unit: The Kalman Filter is a computational algorithm that combines information from several sensors, such as GPS and IMU, to enhance the precision of determining the vehicle's position and orientation. It aids in minimizing mistakes and enhancing dependability.

C. Data Preprocessing: This task entails identifying and rectifying faults or discrepancies in the sensor data. For instance, outliers or inaccurate measurements can be eliminated or rectified in GPS data. IMU data can be effectively denoised by employing median or Kalman filtering techniques.

Data normalization ensures that each characteristic in the sensor data contributes equally to the learning process and prevents particular features from dominating the process due to their scale. For example, GPS coordinates and IMU measurements may vary in terms of their scales. Normalization approaches, such as min-max scaling or z-score normalization, can standardize all features to a uniform scale.

D. Feature Selection: Feature selection entails determining the most pertinent features from the sensor data that provide a major contribution to the localization task, while simultaneously minimizing the number of dimensions and computing complexity.

To choose pertinent features, methods such as correlation analysis, feature importance ratings derived from machine learning models, or domain expertise can be employed.

In the context of autonomous vehicle localization, potential features for selection include GPS coordinates, IMU acceleration, angular rate, and data from supplementary sensors such as LiDAR or cameras.

After identifying the pertinent characteristics, they are then transferred to the machine/deep learning module for additional analysis.

E. Model Training: Various strategies are utilized to record and rectify IMU's drift over time while training machine learning models for Predictive Machine Learning-Based Error Correction in GPS/IMU Localization for Autonomous Vehicles. By employing libraries such as scikit-learn, it is possible to train models like Linear Regression, Random Forest Regressor, and Decision Tree Regressor on past sensor data to identify drift patterns and predict future errors. These models are essential to error correction methods, allowing autonomous vehicles to accurately determine their location even when trustworthy GPS signals are unavailable.

F. Localization Output: The final output from the system, providing the vehicle's accurate position and orientation (pose)

4 Testing and validation: results and discussions

This section focuses on the utilization of machine Learning algorithms for predictive Machine Learning-based Error Correction in GPS/IMU Localization for Autonomous Vehicles. The process comprises two distinct phases: exploratory data analysis and implementing three key machine learning algorithms. These algorithmes including Linear Regression, Random Forest Regressor, and Decision Tree Regressor from scikit-learn, are integral in capturing IMU drift patterns and facilitating error correction, thereby ensuring precise vehicle localization in autonomous navigation scenarios.

A. Exploratory data analysis

The exploratory data analysis (EDA) provided valuable insights into the properties and linkages present in the dataset. Summary statistics offer a thorough summary of each feature's primary tendencies and variability, which helps in comprehending the distribution of the dataset. The utilization of visualization tools, such as time series plots (see figure 3 below), aided probable trends within the dataset. Furthermore, the correlation matrix (See figure 2) revealed noteworthy connections among many variables, guiding subsequent research and modelling choices. It can be claimed that the EDA approach revealed important dataset characteristics, providing a solid basis for future tasks such as predictive modelling and error correction in GPS/IMU localization for autonomous vehicles.

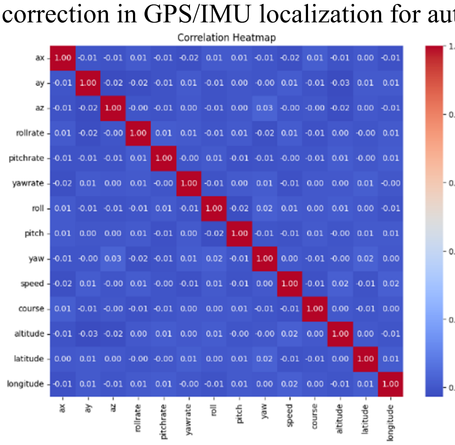


Fig. 2. Correlation Matrix.

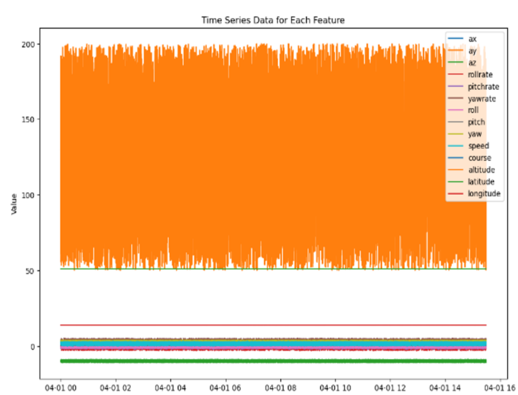


Fig. 3. Time Series data of each feature

B. Machine learning algorithms

The study employed three machine learning algorithms for predicting localization for autonomous vehicles. These are:

a. Linear Regression

The linear regression model offers a direct method for predicting longitude and latitude using the information provided. Nevertheless, it assumes a linear correlation between the input characteristics and the goal variables. In this scenario, while it may successfully identify certain straightforward patterns in the data, it may encounter difficulties in detecting more intricate connections, particularly when there are non-linear interdependencies. Consequently, the model's performance, as measured by the Mean Squared Error (MSE), could be greater compared to more adaptable models such as the Random Forest Regressor and Decision Tree Regressor.

Table 1 below shows a sample (first 5 rows of the dataset) comparison between the Actual and predicted vehicle positions when a simple Linear regression algorithm is employed.

Table 1. Actual values vs predicted values for linear regression for the first five rows

	latitude	longitude	predicted_latitude	predicted_longitude
0	51.027833	13.730935	51.024790	13.734908
1	51.026803	13.737911	51.025017	13.735114
2	51.021957	13.734974	51.025143	13.735137
3	51.023659	13.732241	51.024941	13.735181
4	51.023638	13.734061	51.024963	13.735029

b. Random forest regressor

The Random Forest Regressor model offers a more flexible approach by using a collective of decision trees to make predictions. It can capture nonlinear relationships and interactions between features, leading to better performance, especially when dealing with complex datasets like those encountered in GPS/IMU localization for autonomous vehicles. As a result, the Random Forest Regressor may achieve a lower MSE compared to the Linear Regression model, indicating improved prediction accuracy. Table 2 shows Random Forest regressors tracking smaller changes in predicted positions as they move from one point to another.

Table 2. Actual values vs predicted values for random forest regressor for the first five rows

	latitude	longitude	predicted_latitude	predicted_longitude
0	51.027833	13.730935	51.024715	13.734640
1	51.026803	13.737911	51.025520	13.735094
2	51.021957	13.734974	51.025547	13.734580
3	51.023659	13.732241	51.024899	13.734481
4	51.023638	13.734061	51.025229	13.735313

c. Decision tree regressor

The Decision Tree Regressor model is an algorithm that effectively divides the feature space into distinct areas according to the input feature values. Although decision trees can identify intricate relationships within the data, they are susceptible to overfitting, particularly when dealing with noisy or high-dimensional data. Consequently, the Decision Tree Regressor may demonstrate increased variability and may not effectively apply to new data compared to the Random Forest Regressor. Table 3 below shows actual and predicted position co-ordinates when a stand-alone decision tree algorithm is employed on the same dataset.

Table 3. Actual values vs predicted values for decision tree

	latitude	longitude	predicted_latitude	predicted_longitude
0	51.027833	13.730935	51.020921	13.733100
1	51.026803	13.737911	51.029283	13.730275
2	51.021957	13.734974	51.026344	13.737986
3	51.023659	13.732241	51.029798	13.738259
4	51.023638	13.734061	51.025614	13.735306

d. Actual and predicted position co-ordinates

Figures 4 and 5 below shows a scatter plot comparing predicted longitude and latitudes of all three algorithms explored in this experiment. Notice the Random forest algorithm and Linear regressor producing similar clusters, this is owing to their very similar MSE (see figure 6 in next section)

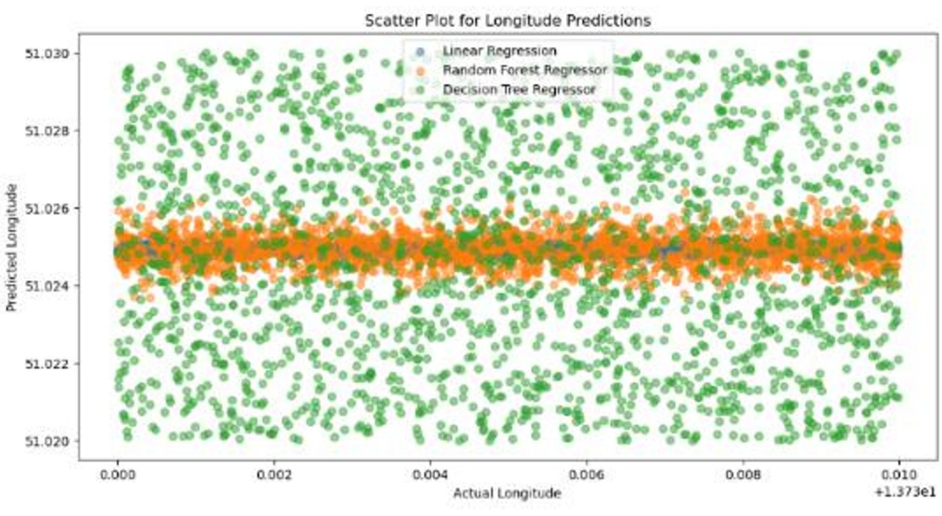


Fig. 4. Actual and predicted longitude

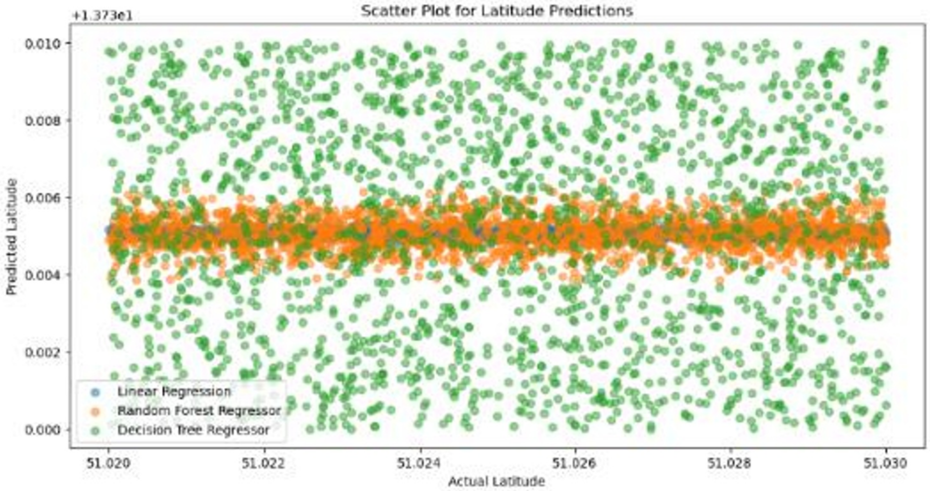


Fig. 5. Actual and predicted latitude

e. Evaluation of model performance using mean square error

By comparing the MSE values and examining the scatter plots between actual and predicted values for longitude and latitude, enable to assess the performance and interpretability of each model. A lower MSE and tighter clustering of points around the diagonal line in the scatter plot indicate better model performance and accuracy in predicting the target variables. [19][20]

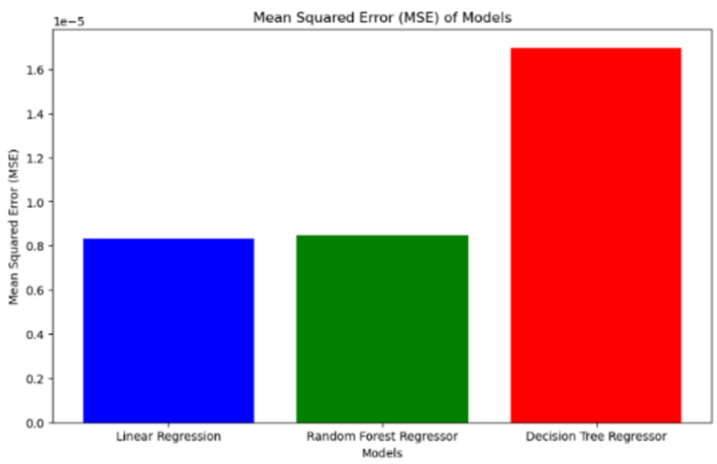
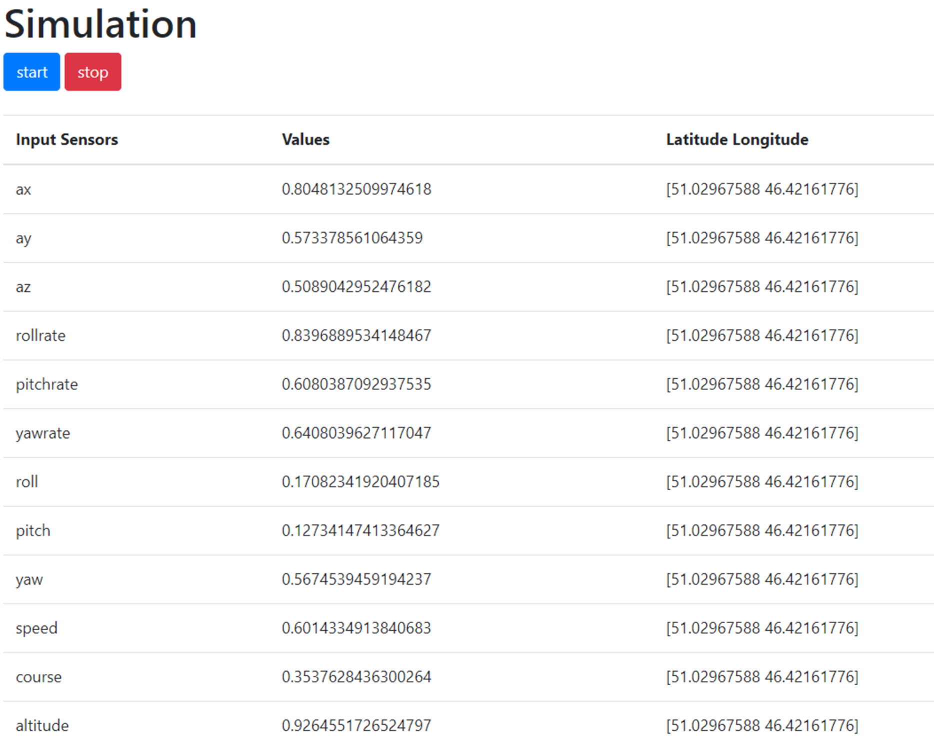


Fig. 6. Mean Square Error

f. Deployment to web

The best model was deployed to an online web application for continuous, close-to-real-time prediction of localization for autonomous vehicles. A CAD simulation was carried out to generate sensor values, which were passed to the model to predict the longitude and latitude values. Figure 7 shows an example of the data sensor, values, and latitude and longitude values.



Simulation

start stop

Input Sensors	Values	Latitude Longitude
ax	0.8048132509974618	[51.02967588 46.42161776]
ay	0.573378561064359	[51.02967588 46.42161776]
az	0.5089042952476182	[51.02967588 46.42161776]
rollrate	0.8396889534148467	[51.02967588 46.42161776]
pitchrate	0.6080387092937535	[51.02967588 46.42161776]
yawrate	0.6408039627117047	[51.02967588 46.42161776]
roll	0.17082341920407185	[51.02967588 46.42161776]
pitch	0.12734147413364627	[51.02967588 46.42161776]
yaw	0.5674539459194237	[51.02967588 46.42161776]
speed	0.6014334913840683	[51.02967588 46.42161776]
course	0.3537628436300264	[51.02967588 46.42161776]
altitude	0.9264551726524797	[51.02967588 46.42161776]

Fig. 7. Web interface with predicted localization

Conclusion

This ongoing applied research programme has developed a predictive machine learning-based error correction model for GPS/IMU localization for autonomous vehicles. The exploratory data analysis and evaluation of various models have shown that the decision tree, linear regression, and random forest models are the most effective in forecasting GPS/IMU localization errors. These models have achieved a mean squared error of 1.7069427028104143e-05. The testing and validation results of various cases have showed the effectiveness of machine learning models in enhancing the precision of determining the location for autonomous vehicles (i.e., self-driving vehicles). The results also emphasize the capability of error correction mechanisms based on machine learning to improve the dependability and accuracy of autonomous vehicle navigation systems, thus facilitating a safer and more efficient mobility in smart cities applications and beyond. [21][22][23]

References

1. P. Gelder, P. Klaassen, B. Taebi, B. Walhout, J. Ommen, I. Poelet al., Int. J. Environ. Res. Public Health, **18** (12), 6329 (2021).
2. Z. Yang, T. Albrow-Owen, W. Cai, & T. Hasan, Science, **371** (6528), (2021).
3. D. Guastella and G. Muscato, Sensors, **21** (1), 73, (2020).
4. T. Zhou, M. Yang, K. Jiang, H. Wong, & D. Yang, Sensors, **20** (24), 7283 (2020).
5. M. Ahangar, Q. Ahmed, F. Khan, & M. Hafeez, Sensors, **21** (3), 706 (2021).
6. M. Köcher, M. Ruge, N. Galldiks, & P. Lohmann, Strahlentherapie Und Onkologie, **196** (10), 856-867 (2020).
7. J. Fayyad, M. Jaradat, D. Gruyer, & H. Najjaran, Sensors, **20** (15), 4220 (2020).

8. H. Chen, T. M. Taha, and V. P. Chodavarapu, Appl. Sci. **12** (7), 3645 (2022).
9. A. Yusefi, A. Durdu, F. Bozkaya, Ş. Tığlıoğlu, A. Yılmaz, and C. Sungur, IEEE Trans. Intell. Veh. **8** (2), 1574-1596 (2023).
10. Chalvatzaras, I. Pratikakis, and A. A. Amanatiadis, IEEE Trans. Intell. Veh. **8** (2), 1574-1596 (2022).
11. E. Zhang and N. Masoud, IEEE Trans. Intell. Transp. Syst. **22** (5), 2615-2626 (2020).
12. J. Fayyad, M. A. Jaradat, D. Gruyer, and H. Najjaran, Sensors, **20** (15), 4220, (2020).
13. Z. Yang, T. Albrow-Owen, W. Cai, & T. Hasan, Science, **371**, 6528 (2021).
14. M. Shiga, S. Muto, High spatial resolution hyperspectral imaging with machine-learning techniques, In *Nanoinformatics*, 179–203 (2018).
15. Liu, Z. Liu, S. Men, Y. Li, Z. Ding, J. Heet al., Sensors, **20** (17), 4962 (2020).
16. P. Gelder, P. Klaassen, B. Taebi, B. Walhout, J. Ommen, I. Poel et al, Int. J. Environ. Res. Public Health **18** (12), 6329 (2021).
17. E. Cust, A. Sweeting, K. Ball, & S. Robertson, Journal of Sports Sciences, **37** (5), 568-600 (2018).
18. F. Eitel, M. Schulz, M. Seiler, H. Walter, & K. Ritter, Experimental Neurology, **339**, 113608 (2021).
19. J. Brownlee, *Regression Metrics for Machine Learning*, Available at: https://machinelearningmastery.com/regression-metrics-for-machine-learning/ (2021)
20. Plevris, V. *et al. 8th European Congress on Computational Methods in Applied Sciences and Engineering* [Preprint]. (2022)
21. U. Charles, M. Shafik, T. Dobrev, and J. Hardy, , DOI: 10.3233/ATDE210022, 18th International Conference on Manufacturing Research ICMR2021, (2021).
22. Y. Jadeja, M. Shafik, and P. Wood, 20th International Conference on Manufacturing Research ICMR2023 [Accepted for Publications] (2023).
23. S. Alshebli, M. Shafik, and F. Kurugollu, 19th International Conference on Manufacturing Research ICMR2022, (2022).