IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 December 2024, accepted 13 January 2025, date of publication 27 January 2025, date of current version 31 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3534659

TOPICAL REVIEW

Smart GNSS Integrity Monitoring for Road Vehicles: An Overview of AI Methods

INÊS VIVEIROS[1], HÉLDER SILVA[2], YURI ANDRADE[1], AND CRISTIANO PENDÃO[3]
[1]Car Multimedia, S.A., Bosch, 4705-820 Braga, Portugal
[2]Algoritmi Center, University of Minho, 4800-058 Guimarães, Portugal
[3]Department of Engineering, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal

Corresponding author: Hélder Silva (hdsilva@dei.uminho.pt)

This work is supported by National funds, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project no 179491; Funding Reference: SIFN-01-9999-FN-179491].

ABSTRACT Integrity monitoring is a key criterion for achieving robust and safe navigation systems. This work explores two integrity frameworks: the classical methods and their respective evolution towards the road vehicle urban scenario, and the artificial intelligence-based methods, where the monitoring process is accomplished by data analysis and learning techniques. In most cases, machine learning outperforms traditional models, which are often observed under controlled, non-real-time conditions, by employing simple algorithms that may have limited success in real-world applications. An overview is provided on how these algorithms have been used, including a comparison of their characteristics and performances, offering insights into how they can evolve and possible future directions to achieve more reliable solutions.

INDEX TERMS Integrity monitoring, GNSS, machine learning, road vehicle.

I. INTRODUCTION

The Global Navigation Satellite System (GNSS) is undergoing an evolutionary phase, owing to the deployment of novel systems for global and regional coverage. These systems have reached a high level of maturity, surpassing originally specified accuracy and precision levels by a significant margin [1]. With the multitude of satellites and multifrequency signals currently available, receivers are required to have higher computational power to take advantage of signal diversity, while maintaining low energy consumption and costs. Correction services such as the Galileo High Accuracy Service broadcast in the E6 signal are currently available, which contributes to lowering the cost of the high accuracy required for road vehicles. Technological advances in antennas capable of mitigating multipath through beamforming will be able to address two major concerns for life safety: multipath mitigation and interference, such as jamming, spoofing, and meaconing [2].

Recently, there has been a growing interest in self-driving capabilities, leading to significant advancements in both advanced driver-assistance systems and fully autonomous vehicles. This trend is supported by empirical evidence that autonomous vehicles have the potential to mitigate the frequency of accidents, diminish fuel/energy consumption and traffic congestion, and enhance the mobility of elderly or disabled individuals. Furthermore, numerous documented occurrences involving automobiles equipped with automated driving capabilities reveal that most of these incidents are attributable to malfunctions in positioning, navigation, and perception functions. To attain these goals, autonomous vehicles must safely execute a series of functions, such as detecting the surrounding environment, determining the most efficient route, regulating velocity, and exhibiting proficiency in navigation and parking [3].

GNSS has the potential to address various automation challenges, such as lane-level maneuvers, vision system oversight, safety through independence, and enabling interoperability through consistent timing and reference frames for vehicle-to-everything cooperation [4]. Although correction services manage to cancel or mitigate most of the factors that make up the GNSS error budget, it is widely known that GNSS performance is greatly affected by errors owing to the presence of multipath and Non-Line-Of-Sight (NLOS) effects in urban and suburban areas, as depicted in Fig. 1. Under these conditions, it is crucial to determine the reliability of the measurements provided by a system rather

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Ni.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/



FIGURE 1. GNSS integrity monitoring application in urban environment.

than solely focusing on achieving high precision [5], that is, it is imperative to devise resilient systems that can withstand flaws and bound errors and continuously monitor their incidence [6].

Receiver-based Integrity Monitoring (IM) is a key criterion for robust navigation systems, as it is designed to detect faults in real-time and provide timely warnings when the system is out of bounds for navigation requirements, allowing users to take necessary actions to mitigate potential harm to any intervening parties. Under certain domains, such as aviation, maritime, and railway, characterized by stringent safety-of-life regulations, integrity is used as a measure of trust in the correctness of the information supplied by the system. Although traditional IM based on statistical methods perform well under good signal conditions, they underperform in challenging multipath and non-line-of-sight environments. Owing to stringent safety-of-life regulations, higher levels of vehicle automation demand that the tight required navigation performance, characterized by low error tolerances and high reliability, be met in complicated urban surroundings.

The IM in GNSS has evolved from single constellation, single frequency, single fault, to multiconstellation, multifrequency, multifault, aided by augmentation systems [7]. Despite this evolution, GNSS alone is unable to achieve the road vehicle requirement target of 10 cm 95% of the time with an integrity risk of 10^{-8} per hour [4], [8]. Furthermore, the automotive field has a higher sensitivity to cost than the aviation field, typically employing low-cost mass market devices, which also exacerbates integrity issues owing to higher noise and lower multipath rejection capabilities [9]. This resulted in IM for road vehicle steadily evolving towards a ''system of systems'', incorporating the use of complementary sensors such as inertial, wheel speed sensors, cameras, and Light Detection and Ranging (LiDAR), as well as mapping assistance and wireless signal augmentation [6], [10].

Data-driven techniques, particularly Artificial Intelligence (AI), are being explored for IM tasks, demonstrating the potential to complement or even replace existing integrity estimation techniques in challenging environments. In contrast to the traditional statistical methods used in Receiver Autonomous Integrity Monitoring (RAIM), AI algorithms uncover intricate data dependencies by learning the relationship between input features and the corresponding output. In addition, AI serves as a valuable tool for investigating the interplay among various dataset features.

The goal of this work is to introduce the basis performance concepts, along with a brief overview of classic GNSS IM methods used in aviation and automotive. Further information regarding classic approaches can be found in [6], [7], [8], [9], and [10] and the works cited in Sections II and III. This provides a basis that allows the reader to better understand recent developments in GNSS IM using AI-based methods, which are the focus of this work. To the best of our knowledge, the use of AI to estimate protection levels has not been reported, and an overview focusing on GNSS IM using AI is lacking in the literature. Some notable works performed position error estimation using AI, which can be used as a proxy for protection level. Many attempts have been made using AI-based systems in Fault Detection and Exclusion (FDE), measurement error modeling, including multipath and NLOS detection, adjacent areas such as atmospheric error prediction, and detection of interference such as jamming, spoofing, and meaconing. The use of real signals in dynamic scenarios under challenging GNSS environments, and the need to comply with highly demanding requirements due to the safety-of-life, is a highly complex task to be accomplished.

The remainder of this paper is organized as follows. The navigation performance concepts that include integrity are presented in Section II. The traditional statistical methods for IM in automotive applications are discussed in Section III. A review of promising state-of-the-art AI algorithms that either are or can be used for IM, categorized by application area, is presented in Section IV. A table summarizing the selected studies for each application area was provided, highlighting the most important features, advantages, and limitations. The results of the review process are discussed in Section V, which provide insights into potential future AI proposals. Finally, conclusions and future work are presented in Section VI.

II. NAVIGATION PERFORMANCE CONCEPTS

The performance of a positioning system is typically evaluated through four interrelated criteria: accuracy, integrity, continuity, and availability [11], as depicted in Fig. 2.

The foundation is sustained by accuracy which is given by the difference between the estimated and true position and velocity and is commonly provided as a statistical value obtained from the error distribution (e.g., 95th percentile). Integrity is a measure of the trust that can be placed in the

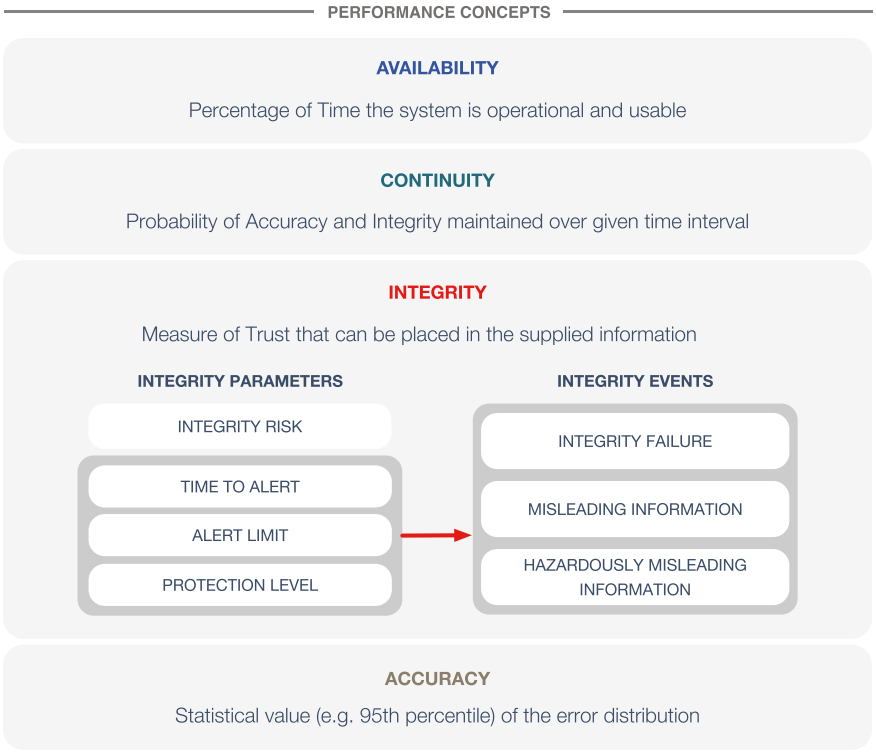


FIGURE 2. Overview of navigation performance concepts.

correctness of information provided by the positioning system; it also includes the ability to provide timely warnings to the user when the system deviates from safe navigation requirements. The third level is continuity, which is a measure of reliable operation that expresses the probability that accuracy and integrity will be maintained over a given time interval. The top level is availability, which is defined as the percentage of time a system is operational and usable.

The concept of integrity is further defined using a set of IM parameters [12]: Alert Limit (AL), Time To Alert (TTA), Integrity Risk (IR), and Protection Level (PL). IR represents the maximum probability of providing out of tolerance signals without a warning. AL represents the largest Positioning Error (PE) which can be tolerated for a safe operation. Because the PE is not known, it must be estimated through the PL, which is a statistical upper bound on PE computed as a guarantee that the probability of the PE being greater than the PL is below the IR. PE, PL, and AL can be further separated into horizontal and vertical components, as well as along-track and cross-track for automotive applications. TTA is the maximum elapsed time since a failure occurs until the system announces the alert. A set of IM events is defined based on the relationship between AL, PL, TTA, and PE as follows [13]:

- Integrity failure when the TTA is exceeded;
- System unavailable when $PL \geq AL$
- Nominal operations when $PE \leq PL \leq AL$;
- Misleading operations when $PL \leq PE \leq AL$;
- Hazardously misleading operations when $PL \leq AL \leq PE$;

The Stanford plot is a commonly used representation for accuracy and integrity analysis [14].

In the aviation domain, the IR for operational safety is typically defined with a probability of occurrence of 10^{-7} per 150 s. In the case of autonomous driving, the target requirement for integrity risk is much stricter, with a probability of occurrence of 10^{-8} per hour [8], considering automation level 4 or higher or Automotive Safety Integrity Level (ASIL) D. Additionally, the challenging environments of suburban and urban roads make this target requirement even harder to achieve and prove, making this an open research topic.

III. CLASSIC INTEGRITY AND AUTOMOTIVE APPLICATIONS

The RAIM technique is a well-established approach that leverages redundant measurements to identify and mitigate outliers. A minimum of four measurements are required to estimate the position (three coordinates and one clock bias). The availability of additional measurements enables consistency verification of the solution. The primary components of RAIM include FDE, position estimation, and protection level estimation.

The goal of FDE is to find outliers in the measurements according to a predefined error model by comparison to a given threshold computed based on the application required probabilities of false alarm (i.e., probability of being in a

fault-free case and declaring a fault) and missed detection (i.e., probability of being in a faulty case and declaring fault-free). The measurement residuals follow a chi-square distribution with $n-4$ degrees of freedom. The distribution is central in the fault-free case and non-central when a fault is present, with the non-centrality parameter representing the bias introduced by the faulty measurement. This relationship is illustrated in Fig. 3.

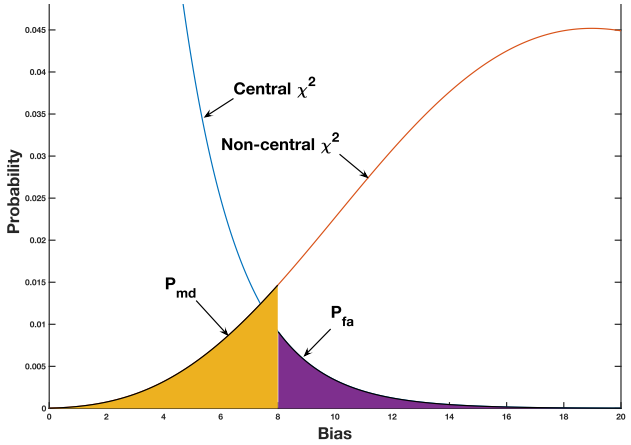


FIGURE 3. Example of central and non-central χ^2 probability distribution for two degrees of freedom, non-centrality λ of 20 meters and threshold of 8 meters.

The threshold defines a critical value between the probabilities of false alarm and missed detection, which are given as follows:

$$P_{fa} = 1 - \int_0^T f_{\chi^2}(x)\mathrm{dx} \tag{1}$$

$$P_{md} = \int_0^T f_{\chi^2_\lambda}(x)\mathrm{dx} \tag{2}$$

There are multiple methods for computing the test statistic. In the least-squares range residuals method [15], the test statistic r is computed using:

$$r = \frac{s \cdot \sigma}{\sqrt{n-4}} \tag{3}$$

where s is the root squared sum of the normalized range residuals of the n satellites, and σ is the standard deviation of the pseudo-range errors. In the solution separation algorithm [16], the solutions for n sets of $n-1$ satellites are computed, and the test statistic is given by the maximum of all possible distances between the solutions. In the parity space method [17], the test statistic D is computed as:

$$D = (S \cdot z)^T \cdot (S \cdot z) \tag{4}$$

where z is the difference between the observed and predicted measurements and S is the idempotent matrix obtained using:

$$S = I_M - H \cdot H^* \tag{5}$$

where H is the design matrix relating the observations to the state estimates, I_M is an identity matrix of size M, and H^* is the generalized inverse of H given by:

$$H^* = (H^T \cdot H)^{-1} \cdot H^T \tag{6}$$

By applying the above procedures to solutions obtained with different subsets of satellites, faulty measurements can be identified and either excluded or down-weighted to remove their influence from the positioning solution.

The goal of PL is to provide an estimate that over-bounds the position error, which cannot be observed directly. A well-known method for computing PL is the slope method [18], which is a function of the design matrix H, and matrix S from Eq. (5), and for the horizontal component is given as:

$$\mathrm{Slope}(i) = \frac{\sqrt{H^{*\,2}_{1,i} + H^{*\,2}_{1,i}}}{\sqrt{S_{i,i}}} \tag{7}$$

The slope provides insight into the effect that a bias in satellite measurement can produce in position estimation. The satellite with the largest slope will have the smallest contribution to the test statistic; therefore, it is difficult to detect measurement biases or faults. The horizontal PL is computed by determining the maximum slope from the satellite set and applying:

$$\mathrm{HPL} = \mathrm{Slope}_{max} \cdot \sigma \cdot \sqrt{\lambda} \tag{8}$$

A similar procedure is applied for the weighted least squares version.

Several improvements to RAIM methods have been proposed. Relative RAIM introduced the time differencing technique used to mitigate constant (i.e., phase ambiguity) or slowly changing parameters in time (i.e., atmospheric errors and satellite hardware biases) and to avoid the need for ambiguity resolution in the carrier phase [19]. Further developments occurred naturally with the evolution of GNSS by making use of multiconstellation, multifrequency signals, and considering multiple simultaneous faults. Advanced RAIM [20], [21] explored these features while also taking advantage of an integrity message broadcast by the ground segment monitoring GNSS satellites. The Advanced RAIM method was further developed from snapshot to a Kalman Filter (KF) solution in [22], providing tighter bounds in the protection levels estimated with increased computational load, owing to the bank of KFs used for each satellite subset. Comprehensive information on aviation domain methods can be found in [19], [23], [24], and [25].

For the road vehicle use case, the highway can be seen as the best-case scenario; even so, large-dimension vehicles in the proximity of the receiver can disturb the propagation channels. Urban canyons are one of the worst possible scenarios for GNSS signal reception, second only to tunnels where signals are completely blocked and navigation is not possible using only GNSS signals. With a propagation channel highly degraded due to signal blockages, multipath and interferences, and with much tighter protection bounds needed, integrity estimation in urban scenarios is a highly challenging and open research topic.

For static, Line-Of-Sight (LOS) open-sky scenarios, measurement rejection is the typical procedure used to deal with faults and the respective PL estimation, because measurements are abundant and highly stable. For urban applications, error characterization is used, where the goal is to compute an effective PL without the need to identify and remove the affected measurements [13]. Some examples of the RAIM methods used in the aviation and automotive are provided in Fig. 4.

In urban and dynamic scenarios where multipath and NLOS are common, the GNSS receiver should extract as much information possible from all measurements, while minimizing the dilution of precision. Multipath reception is sometimes the only form of signal reception in urban canyons; therefore such affected measurements should be identified, and ideally, the respective bias estimated/removed, to avoid contaminating the positioning solution.

Much research has been applied in techniques to mitigate the effect of multipath interference and NLOS signals at different levels, including antenna design, baseband algorithms, positioning engines, and post-processing techniques [26], [27], helping to improve the accuracy and reliability of GNSS positioning in urban environments. In [28], a signal quality monitoring technique and crowdsourced map containing Carrier-to-Noise Ratio (C/N0) measurements were employed for multipath prediction and detection. The map information is used to control the stochastic model in a loosely coupled and tightly coupled KF architectures using GNSS and inertial measurement unit sensors. Improvements of 17% and 13% were achieved in the horizontal accuracy when applying loosely and tightly coupled architectures, respectively. Mapping information is also used in [26] and [29] to identify which satellites are in the NLOS condition.

The work in [30] proposed the use of a vehicular area network to collect the ensemble average of effects that can occur over a specific location and time to help compute the effective error for each satellite. This error is then used to compute the protection levels, which are usually conservative values when compared to theoretical values obtained from the geometry matrices involved in the process. A higher bound on the PL leads to a lower level of continuity and availability. It is a key factor that a PL algorithm provides correct bounds, not overly confident, because these can result in misleading or hazardous information, but also not overly protective, since the system may then be often declared unusable, especially for liability-critical and safety-of-life applications with strict requirements.

The isotropy-based PL is a method patented by GMV [31], and its main idea is that the measurement error vector can point in any direction with the same probability. The PL is as follows:

$$\mathrm{PL} = k \cdot ||r|| \cdot \sqrt{\lambda^{h}_{max}(H^{T} \cdot H)^{-1}} \tag{9}$$

Parameter k is precomputed for a given target integrity risk using the number of measurements and the number of states to be estimated (a table is created to be used in real-time). This value is then multiplied by the residual r obtained from a least-squares solution and the maximum eigenvalue obtained from the dilution of precision matrix. The results obtained in [12] show that its application under driving conditions requires a least squares approach, restricting its achievable performance.

Techniques based on the KF appear to have a higher potential to reach the protection levels needed for driving conditions, as can also be seen in the results from [12]. The work in [22] applied solution separation to an Extended Kalman Filter (EKF) in kinematic precise point positioning, achieving promising error bounds for stable signal conditions. The results obtained in the suburban scenario highlight the issues of using only GNSS data in environments that are prone to interference. With a solution that requires convergence, the problem is exacerbated owing to partial or full measurement outages, as the filter constantly resets the ambiguity estimates.

Time-correlated errors have also been studied in [32], which uses a frequency domain approach to the overbounding problem, assuming the stationarity of the auto correlation sequence of the noise over a predefined duration. The KF covariance matrix overbounds the error distribution when the noise models overbound the Fourier transform of a windowed version of the auto correlation sequence. The results presented used only 10 min of simulation data in an apparently benign baseline estimation environment; therefore, further testing is necessary.

Recently, u-Blox patented an IM method [33], [34] that applies multipath mitigation techniques, measurement uncertainty modeling and tight position bounding schemes, known as single epoch position bounding. The distribution probability relating the error states, GNSS measurements, and quality parameters was obtained by integrating all possible ambiguity solutions from the carrier phase. The method aims to identify and remove individual signals subject to NLOS or severe multipath conditions and entire epochs of data in degraded environments affecting a large number of satellites (e.g., no solution is reported when data are unsuitable, just as would occur inside a tunnel). The method is applied to a moving receiver/vehicle, and preliminary results indicate performance with error bounds mostly below 1 m, satisfying a target integrity risk of 10^{-5} per hour.

IV. INTEGRITY IN AUTOMOTIVE APPLICATIONS USING ARTIFICIAL INTELLIGENCE

AI corresponds to computer software that mimics human cognition in performing complex tasks such as analysis, reasoning, learning, and planning. AI can be divided into several fields, the most common of which are Machine Learning (ML), Robotics, and Computer Vision. ML is a class of methods through which AI can acquire cognition by enabling computers to learn from historical data, resorting to statistical tools, without being programmed explicitly. This results in algorithms that can be used to make accurate predictions of past, current, and future events.

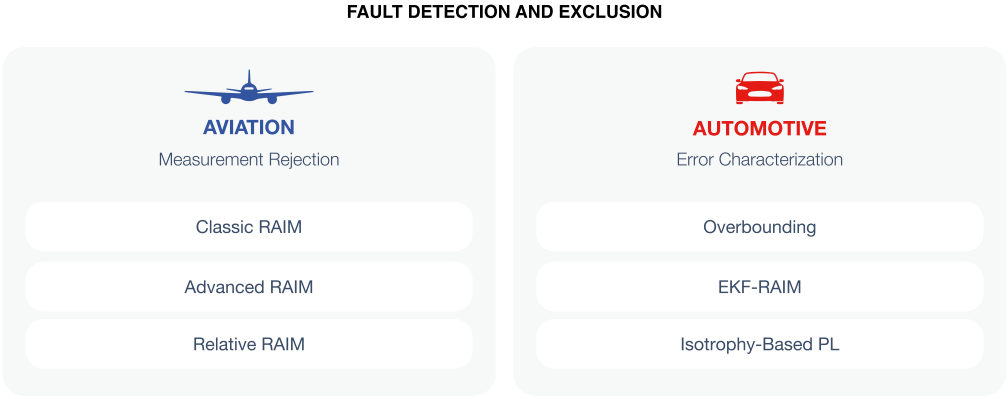


FIGURE 4. Examples of RAIM methods in aviation and automotive.

The common parts of an ML model are the decision process (used to compute the prediction), error function (used to evaluate the predictions made), and model optimization process (adjustment of the parameters of the model to approximate the predictions to the real values). ML models can be used to perform regression (continuous output) or classification (discrete and categorized output), and depending on the characteristics of the training data, the three learning types are:

- Supervised Learning: uses labeled datasets to train the algorithms. In this approach, the training data feed the model, which updates its weights until it is properly fitted to the expected output. The accuracy of the model is measured using task-specific metrics. The majority of the ML models analyzed in this study fall into this category.
- Unsupervised Learning: the datasets are unlabeled; therefore, the models are trained to find implicit patterns or structures without any guidance, as in clustering. These algorithms can also be employed for association and dimensionality reduction/feature selection, such as in Principal Component Analysis. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), K-Nearest Neighbors (KNN), and Auto-Regressive (AR) fall in this category.
- Reinforcement Learning: the algorithm learns to make decisions by interacting with an environment and receiving feedback in terms of rewards or penalties without labeled data. The algorithm learns from trial and error, with the goal of maximizing the cumulative reward over time. Currently, this type of learning is most commonly employed in the robotics field, but there are some studies employing this type of learning to augment GNSS accuracy, such as [35] and [36].

Currently, GNSS IM strategies are rough adaptations of the classic RAIM technique. It is also possible to find a few curious approaches such as vision-aided RAIM, which employs computer vision models to match landmarks with photographs to obtain additional measurements [7], [37]. However, none of these methods can fully solve the existing urban environment challenges. Therefore, it is necessary to develop new methods to increase the reliability of integrity estimation for each type of scenario, focusing on autonomous land vehicles.

The use of ML in GNSS is promising, and several methods have been developed to mitigate issues originating from the use of GNSS measurements. ML models have been employed in FDE. Integrity-adjacent areas have also been pursued, including multipath and NLOS detection, atmospheric error prediction, and interference detection, such as jamming, spoofing and meaconing. These solutions could potentially improve the integrity of GNSS-based positioning systems for automotive applications in challenging environments. To the best of our knowledge, PL estimation using ML/AI has not been reported in the literature. A few recent works target position error estimation, which can theoretically be used as a basis to infer a protection level.

This section provides an overview of AI-based approaches applied to GNSS use cases. To analyze the IM in GNSS, AI applications were divided into four main categories: (1) position error estimation (as a proxy for PL), (2) fault detection and exclusion, dealing directly with IM, and integrity-adjacent areas, namely (3) measurement error modeling encompassing the classification of signal reception, estimation of atmospheric errors, and detection of interference, and (4) environment detection and classification, as depicted in Fig. 5.

A. POSITION ERROR ESTIMATION

The position error estimation category, which belongs to the accuracy of GNSS performance, is explored in this section. Accuracy is the basis of integrity, and position error estimation can also be used as an input for PL estimation.

The thesis in [38], studies the usage of two ML algorithms by exploring several different features to predict positioning errors using a Decision Tree (DT) and a Support Vector Machine (SVM). DT did not present satisfactory results but

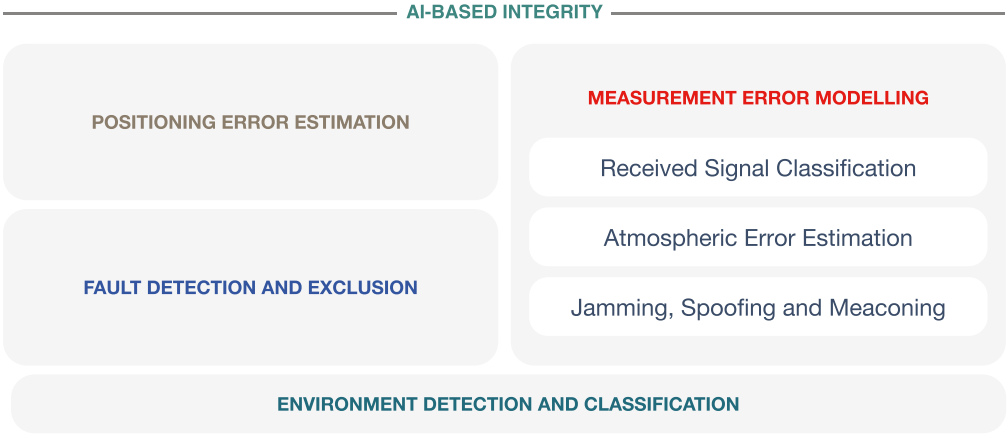


FIGURE 5. Main application areas of AI-based approaches for integrity monitoring and adjacent areas.

enabled feature ranking, which showed that C/N0, number of used satellites and elevation explain most of the location error variance. The ranking obtained was used for feature selection prior to SVM development. On the other hand, the SVM reached a Root Mean Square Error (RMSE) of 3.54 cm, using the top eight features (C/N0, number of satellites, elevation, difference east-north-up, least-square residuals, azimuth, innovation east-north-up and position Dilution of Precision (DOP)), with a good match between predicted and true errors. From an IM perspective, it provided approximately 50% correct bounding PL (the other half was misleading or hazardous). To augment the performance of the model, a camera was used to estimate the environmental obstruction. However, the results deteriorated when obstruction information was used in the models. This could be explained by errors in the sky clearance input and multipath presence for LOS satellites (e.g., via ground reflection). The lack of statistical significance when using this new input with SVM and the high variance obtained in the ranking of the sky clearance feature using DT also supports this possibility.

According to [39], awareness of the environment surrounding the receiver is important for the generation of better error models. The authors suggest that the environment be classified into suburban, urban, and urban canyons based on satellite visibility, minimum elevation angle, and DOP; and multipath reflections estimated from C/N0 variation. Two approaches can be employed for this classification task: using the observables directly or by employing the difference between the observables and the values under optimal conditions. Although the literature states that the second approach leads to better environment recognition, the first approach is followed in [39] because it can be performed in real-time, without the need for initial calibration. In addition, three models were introduced to predict the positioning error: a KNN, a SVM and a Random Forest (RF), which were optimized using a grid search. The feature vector includes the Horizontal Dilution of Precision (HDOP)/Vertical Dilution of Precision (VDOP), the number of satellites, the minimum elevation angle and the mean signal strength. The best-performing model is RF, and it is shown that the model errors increase with environmental complexity, mostly due to multipath. Overall, the authors maintain that an adaptive selection of training parameters according to environment recognition allows for better error prediction and thus safer conditions. The tests were conducted using artificial GNSS data under stationary conditions in well-characterized environments, which is hardly the case in real applications.

Karlson and Mohammadiha [40] presented an approach to model the GNSS error based on the AR and Gaussian Mixture Model (GMM). The pipeline described is dedicated to generating artificial GNSS positioning errors and considers different environment types surrounding the receiver, such as urban and highway, with two different levels of ionospheric activity. The error is approximated as several stochastic processes because it is known that the error behavior changes over time; therefore, it cannot be properly modeled by a single stationary process. Unsupervised clustering was performed to group points with similar properties. The residual error of each cluster was modeled as a GMM. Each group is used for supervised model fitting of an AR process to obtain a set of parameters that describe each stochastic behavior. These two steps are iterated until convergence, using 40 hours of Volvo driving data. The inference framework also has two steps: first, a sub-model is selected based on the optional input of surrounding conditions for the desired artificial data; and second, an AR model with the GMM residual error from the set of models within the sub-model is employed to make the predictions. Non-stationary error modeling is enforced by an auto-regressive model of order p.

Recently, the approach used in [41] applied a Deep Neural Network (DNN) based on a set transformer to estimate a correction to an initial position estimate obtained using a traditional least-squares model. The set transformer enables the model to adapt to the variability of satellite measurements

in both order and number. Two issues affect solutions using pseudoranges and satellite positions: (1) numerically ill-conditioned inputs, owing to very small variations within very large values for pseudorange (receiver-satellite distance) and satellite position in the Earth-Centered-Earth-Fixed frame, and (2) geographic sparsity due to the dataset being obtained typically from a specific location on Earth. These issues are mitigated by formulating the problem in terms of pseudorange residuals and LOS vectors and by outputting a correction to the initial position in the local-tangent-plane North East Down (NED) frame. Compared to a weighted least-squares solution, this method was able to improve the mean absolute positioning error by more than 50% under simulated data and more than 40% in real data from the Android raw GNSS measurement dataset. The results indicated a high sensitivity to errors present in the initial position estimate, and an improvement of up to 10% was obtained from the data augmentation strategy.

The authors in [42] attempted to reduce the sensitivity to the error in the initial position estimate by using a hybrid architecture, employing a KF that estimates the initial position and a Graph Convolutional Neural Network (GCNN) to estimate the position correction. Pseudorange, carrier phase, and Doppler measurements are used in the KF, where pseudorange smoothing is applied using the latter two measurement types. The inputs to the GCNN are the pseudorange residuals and LOS vectors. The GCNN can handle the variability of the sets of measurements, achieving a mean absolute error improvement of 26% over weighted least-squares and KF-based methods, and also outperforming [41]. All pseudorange errors that are typically modeled in GNSS positioning are corrected in both [41] and [42], the latter including inter-system and inter-frequency biases due to the use of multifrequency and multiconstellation (Global Positioning System (GPS) L1, L2, L5; GLONASS L1, L2; Galileo E1, E5) measurements. This might be a key contributing factor to achieve better performance since the former used only single-frequency GPS L1 measurements.

A summary of the studies analyzed is presented in Table 1.

B. FAULT DETECTION AND EXCLUSION

Fault detection refers to the system's capability to acknowledge the occurrence of abnormal behavior, and fault exclusion refers to the ability of the system to remove the cause of said behavior. Particularly, in the case of IM, these terms are usually related to the pseudorange measurement quality, which is classically determined by the dispersion of residuals.

AI-based methods have demonstrated good performance in FDE. In [43] a Neural Network (NN) was used to improve the weight definition of a particle filter used in the classic RAIM approach, reducing fault detection time, and improving the mitigation of the faulted measurement, even when the data are affected by non-Gaussian noise. Similarly, in [44], a Chaos Particle Swarm Optimization Particle Filter (CPSOF) was employed to reduce the detection time of faulty measurements in a non-Gaussian error scenario. This algorithm mitigates particle degeneracy and sample impoverishment, typical of a particle filter, by adding a CPSOF that moves the samples towards regions in the state space where the likelihood is more significant, in addition to reducing the number of iterations and run time. Moreover, the FDE is accomplished by a log-likelihood ratio statistical test between the outputs of several CPSOF (if m measurements are performed, then $m + 1$ CPSOF are used), one employed using all m measurements, and one employed for each set of $m - 1$ measurements.

RAIM-Net, introduced in [45], utilizes a multilayer Long-Short Term Memory (LSTM) DNN to perform FDE. The network is fed with precomputed features at different time steps to output a classification of each signal in terms of fault or fault-free. Position solutions are obtained with all measurements in view, as well as with each subset of observations using the iterated least-squares method. Input features are obtained by taking the difference between the all-in-view and the subset solutions. After the number of layers was iterated, the results of this algorithm were compared with those of classic RAIM, and it was concluded that RAIM-Net has a higher fault detection rate and a lower minimal detection bias. Performance was assessed using real data with simulated errors under static conditions.

Supervised learning-based methodologies require a labeled dataset in which the model can be trained; however, obtaining the said GNSS dataset can be difficult, particularly in urban scenarios. Xia et al. [46] introduced HDBSCAN, a clustering methodology capable of labeling anomalous satellite observations to construct a dataset for offline training of a binary classifier to detect measurement faults in real-time. The system can achieve positioning accuracy improvements of the order of 39% to 63% by removing faulty measurements. A comparison was performed with traditional RAIM FDE, which is a single fault test known to be inadequate for urban scenarios. These clustering algorithms are promising approaches for fault isolation, although additional tests are necessary to compare ML-based methods with traditional techniques that are capable of handling multiple faults.

To perform fault detection, the authors of [47] delineated two models: an SVM and an NN trained to classify GNSS measurements as safe or unsafe for navigation. The input feature vector is a concatenation of the range residuals and received Signal-to-Noise Ratio (SNR), and the performance is higher than standard RAIM, both in terms of recall and precision, with the added advantage of handling multiple fault scenarios. The results were obtained in a simulated environment with synthetic signals. Furthermore, [48] employed a Time Delayed Neural Network (TDNN), known for its ability to detect faulty characteristics expressed in irregular sequences of data, to perform the same classification task. In this methodology, the feature vector is the time sequence of the test statistics computed by RAIM from the pseudoranges

TABLE 1. Specifications, advantages and limitations of the selected studies for Position Error Estimation.

Model	Conditions	Input Features	Advantages / Limitations
DT, SVM [38]	Real data, dynamic conditions, post-processing, dataset size of 150K+19K samples	C/N0, constellation, least square residuals, sat. number, elevation and azimuth, Position Dilution of Precision (PDOP), Ambiguity Dilution of Precision (NDOP), tracking type, difference ENU, innovation ENU, camera sky view	DT: - Prone to over-fitting. - Lower predictive power than classical statistical methods. - Useful for feature ranking/selection. SVM: - Only tested for a single SVM kernel. - Does not scale with larger datasets. - Provided roughly 50% correct bounding PL.
SVM, KNN, RF [39]	Synthetic data, static conditions, post-processing, dataset size of 3x12 hour simulations	HDOP/HDOP, sat. number, minimum elevation angle, mean signal strength, environment recognition	- Tests conducted under stationary conditions with well-characterized environments using synthetic data. - Performance degradation in urban-canyon environments with LOS blockage. - Applies error-free environment recognition. - RF achieved the best error prediction compared to KNN and SVM.
AR GMM [40]	Real data, dynamic conditions, post-processing, dataset size of 40 hours	Pre-set selection of environmental conditions, position estimates and respective ground truth	- Accounts for non-stationary conditions. - Used to generate realistic GNSS error distributions for simulation and not for real-time error prediction.
DNN [41]	Real and synthetic data, dynamic conditions, post-processing, dataset size of 120K samples	Pseudorange measurements, satellite positions	- The correction in NED frame avoids ill-conditions and makes output independent of location. - Sensitive to the initial position error. - Data augmentation strategy improved performance marginally. - Atmospheric, clock and orbit error sources are a-priori corrected.
GCNN [42]	Real data, dynamic conditions, post-processing, dataset size of 120K samples	Pseudorange, carrier phase, Doppler measurements, satellite positions	- The correction in NED frame avoids ill-conditions and makes output independent of location. - Hybrid architecture is able to attenuate sensitivity to initial position error. - Mean performance improvement of 26% over KF. - Atmospheric, clock and orbit error sources are a-priori corrected.

and the TDNN computes the dissimilarity between the current trend (resulting from online measurements) and past trend (computed offline with GNSS satellites in the normal state). Based on the simulated tests, the AI-based model can also detect faults more reliably and rapidly than standard RAIM.

The studies analyzed in this category are summarized in Table 2.

C. MEASUREMENT ERROR MODELING

GNSS error sources can be categorized into clock, signal propagation, system, and other sources, as shown in Fig. 6.

Most satellite-based errors can be removed by applying correction services, mitigated by estimating the respective bias in the positioning solution, or removed by means of calibration. The remaining GNSS error sources add up in creating a measurement error that is a function of many variables which can be complex to model and predict. The combination of several factors, such as multipath, ionospheric delay, and jamming, to name a few, add up to a positioning error that is intricate and non-deterministic, as well as non-Gaussian. This behavior of the error suggests that employing an ML model to learn it will likely result in a simpler and more accurate pipeline than trying to define static models for each error source, since this kind of models can handle complex behaviors very well and with a high degree of disregard for the causes.

An example of an ML application is found in [49], which details a Vehicle-to-Infrastructure (V2I) approach for estimating pseudorange residuals under real-time dynamic conditions. Assuming that neighboring receivers are affected by similar errors, this method is based on the collection of several measurements obtained by different receivers within a given area through collaboration between vehicles. A central unit gathers measurements and uses them to train a backpropagation NN capable of predicting pseudorange errors, which are then broadcast to all receivers within the area. The NN weights are updated periodically by the central unit, depending on the dataset size or update interval. The pseudorange residuals were used to compute a weighting factor for each satellite, which, when used with the slope method, was able to improve the envelope ratio of the horizontal protection level, particularly in areas with poor satellite visibility. This approach has intrinsic environment detection owing to the

TABLE 2. Specifications, advantages and limitations of the selected studies for Fault Detection and Exclusion.

Model	Conditions	Input Features	Advantages / Limitations
NN [43]	Real (clean, fault-free) data with manual fault injection, other conditions not specified, dataset size unavailable	Receiver position estimate and pseudo-ranges	- Simulation of just one satellite failure, which can already be successfully detected by classic RAIM. - Increases particle filter sensitivity and reduces response time to faults. - Effective using non-Gaussian measurements.
NN [45]	Real data, static conditions with manual fault injection, post-processing, dataset size of 1M samples (56 hours @ 5 Hz)	RAIM time-series features, difference between the positioning solutions of the full observations and the subset observations	- Model is hard to tune, takes longer to train and requires more memory. - Achieves higher fault detection rate and a lower minimal detection bias than classic RAIM.
HDB SCAN [46]	Real data, dynamic conditions, real-time, dataset size of 38K samples	C/N0, sat. number and elevation angle, PDOP, HDOP, HDOP, pseudorange rate consistency and residual	- Training set with few scenarios. - Uses only single frequency data. - Improved positioning accuracy by up to 87%, highly dependent on dataset. - Useful for data labeling.
SVM, NN [47]	Synthetic data, dynamic conditions, post-processing, dataset size of 120K samples	Range residuals and SNR	- Better performance than standard RAIM. - Linear kernel used in SVM struggles to develop a linearly separating decision boundary. - RBF kernel achieved similar performances to those of a small NN. - Fast training and inference of NN due to using small model with few neurons. - Similar results between NN and SVM with RBF kernel. - Test in multiple fault scenarios.
NN [48]	Real and synthetic data with manual fault injection, static conditions, post-processing, dataset size of 24 hours (simulated) and 3x24 hours (real)	Time-series of RAIM statistics obtained from pseudoranges	- Solution with noise suppression capabilities. - No comparison made with the best classical methods or with the best time-series ML models.

division of the map into areas and training a model for each area. Performance improvements may also be partly attributed to the cooperative aspects of the implementation. The size of the dataset used to train the NN is not specified. The input features were satellite elevation and azimuth, SNR, and pseudorange residuals. Some baseline assumptions are made to simplify the implementation, namely speed, direction, and other parameters of the vehicles are considered not to affect the quality of corrections. It is also highly dependent on V2I wireless communication.

Most of the studies using ML in the area of GNSS found in the literature belong to the category of GNSS received signal classification, which is described next.

1) RECEIVED SIGNAL CLASSIFICATION: LOS, NLOS AND MULTIPATH

Multipath reception occurs when multiple replicas of the signal are received through different paths, including direct LOS. In the case of NLOS reception, a direct path replica is absent. In urban areas, GNSS signals often encounter obstacles such as trees and buildings, which have a detrimental impact on GNSS positioning and integrity performance. Although receiver design improvements can partially mitigate multipath effects, they are ineffective in addressing the NLOS case. Several studies have investigated the application of ML techniques to classify the received GNSS signals into three categories: LOS, NLOS, and multipath. This classification is used to enhance positioning accuracy and integrity.

One such study [45] focused on achieving accurate classification with high computational efficiency by employing the Gradient Boosting Decision Tree (GBDT) algorithm using three key features: C/N0, pseudorange residuals, and satellite elevation angle. Importantly, the study collected raw measurements in various urban environments, achieving 89% classification accuracy for static data. External testing in nearby locations verified the robustness of the algorithm, with GBDT outperforming the other classification techniques. Among GBDT variants, XGBoost and LightGBM could be useful for further improving performance of the referred approach.

In [50], a collaborative approach addressed multipath challenges by leveraging in-car navigation systems. Nearby vehicles shared information such as pseudorange corrections and map matching. ML algorithms, including Logistic Regression, SVM, Naïve Bayes, and DT, were used in this context. The authors selected six predictor variables and conducted binary classification experiments under various conditions. The results demonstrated prediction correctness

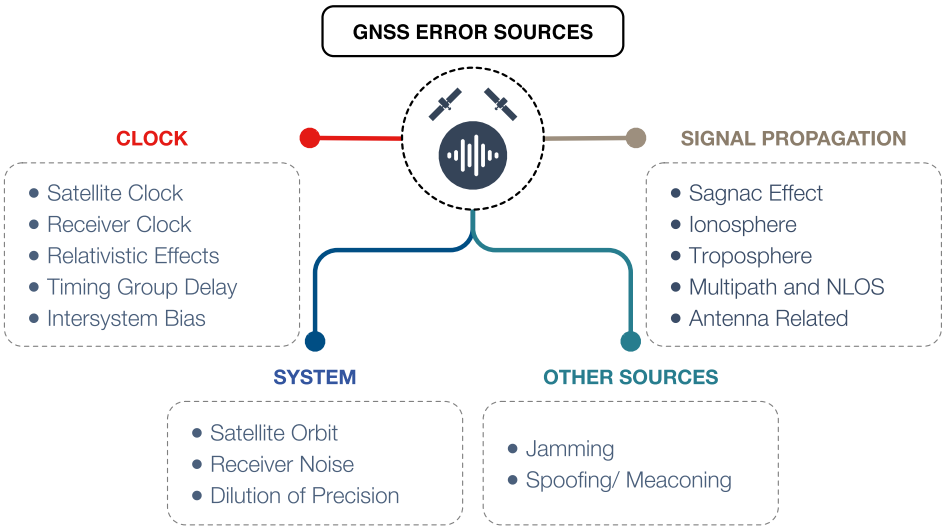


FIGURE 6. Error sources affecting GNSS measurements.

rates close to 90%, suggesting the transferability of the models across different locations and times.

Another study [51] introduced a GNSS signal classifier based on a DT algorithm that can predict satellite visibility in dense urban regions with over 85% certainty. Their success was attributed to the use of a substantial number of raw measurements and sophisticated feature-selection techniques. A similar approach was found in [52], where GNSS signals were classified into NLOS, LOS, and multipath signals using two antennas with different polarizations (left- and right-hand circular polarization). This unique classifier achieved a remarkable accuracy of 99%, surpassing that of the KNN and SVM techniques. In a previous study [53], a fuzzy logic-based algorithm was explored using satellite elevation angles and C/N0-R-L, effectively distinguishing between NLOS and multipath signal-reception states.

A Convolutional Neural Network (CNN) was demonstrated to play a vital role in signal classification [54]. The authors proposed a CNN-based multipath detection method for kinematic positioning, achieving significant accuracy improvements using a sparse auto-encoder. Softmax, and random forest classifiers were compared, with both excelling in static and kinematic data scenarios. The results indicated an 18–30% improvement in accuracy after applying this algorithm. In addition, [55] introduced a CNN-based model for identifying multipath signals. This model treats raw data from the correlator as images, using a CNN to extract features. A binary multipath classifier achieved superior results compared to an SVM, taking advantage of the CNNs' capabilities to process video and image data.

In [56], SVMs were employed to distinguish between three categories of received GNSS signals: LOS, multipath, and NLOS. They extracted four features from raw measurements, achieving classification accuracies of approximately 67% for individual features and about 75% when using them together in training. The study provided validation results but lacked information about the feature distribution in each fold.

Lastly, [57] introduced a classification technique using variables derived from raw GPS measurements and an adaptive neuro fuzzy inference system. This method achieved high accuracy rates, outperforming SVM, DT, and C/N0-based classifiers. The work in [58] explored a Gated Recurrent Unit (GRU) model for detecting multipath in GNSS signals, surpassing DT, KNN classifiers, and SVM. GRUs are particularly effective in capturing non-linear relationships and using previous data to enhance accuracy.

It is important to note that most of these studies relied on simulated data because of the complexity in determining real-life signal states. Dynamic factors such as passing vehicles or pedestrians can create NLOS conditions. Notably, strategies including DTs, GBDTs, CNN, and GRU, consistently outperformed other strategies in these signal-reception classification tasks.

The selected studies for received signal classification are summarized in Table 3.

2) IONOSPHERE AND TROPOSPHERE ERROR ESTIMATION

The ionosphere and troposphere are two well-known sources of error in GNSS positioning that introduce delays of different magnitudes. The ionospheric delay is given by the integral of the electron density along the ray path, known as Total Electron Content (TEC), which causes an advance in the phase measurements and a delay in the pseudorange measurements. Atmospheric delay is typically divided into two components: hydrostatic and water vapor. The former is estimated using a model that depends on temperature, pressure, humidity, and transmitting and receiving antenna position, while the latter is typically estimated in the positioning solution, known as the Zenith Wet Delay (ZWD), by applying an elevation-dependent mapping function to convert between zenith and slant delay.

TABLE 3. Specifications, advantages and limitations of the selected studies for Received Signal Classification.

Model	Conditions	Input Features	Advantages / Limitations
DT [51]	Real data, static and dynamic conditions, post-processing, dataset size of 11K samples	C/N0, sat. elevation and clock bias, delta carrier phase and pseudorange residuals	- Tested in unseen data from another nearby location. - Limited dynamic conditions condensed in a small square region.
DT [52]	Real data, static conditions, post-processing, dataset size of 100 samples	Satellite elevation, C/N0-R-L ratio	- Performed better than SVM, KNN, with 99% accuracy. - Needs right-hand and left-hand circular polarized antennas.
Fuzzy Logic [53]	Real data, static conditions, real-time, dataset size unavailable	Satellite elevation, C/N0-R-L ratio	- Does not test for NLOS conditions. - The reflection scenario is not representative of an urban scenario with many different reflection properties. - Needs right-hand and left-hand circular polarized antennas.
NN [54]	Real and synthetic data, dynamic conditions, post-processing, dataset size of 51 hours	Time-series of C/N0 and multipath estimate, computed from pseudorange and carrier-phase for each frequency	- Requires a large amount of data. - Achieved accuracy of around 60% in kinematic data.
NN [55]	Real and synthetic data, static conditions, post-processing, dataset size unavailable	GNSS correlator in time-frequency domain as tensor images from Doppler and code delay offsets	- Performed better than SVM. - Can capture geometrical dependencies in the data.
SVM [56]	Real data, static conditions, post-processing, dataset size of 48 hours	C/N0, ΔC/N0, pseudorange residual, and the difference between Δpseudorange and pseudorange rate	- Does not account for non-stationarities in data. - Avoids data imbalance issues by using identical number of LOS, multipath and NLOS measurements.
Fuzzy Logic [57]	Real data, static conditions, post-processing, dataset size of 134K samples	C/N0, temporal difference in C/N0, HDOP, VDOP, elevation and azimuth angle, pseudorange residual	- Used static conditions from the same locations of the training data. - Higher accuracies than DT, SVM, and C/N0-based classifiers.
NN [58]	Synthetic data, static conditions, post-processing, dataset size of 66K samples	Multiconstellation time-series pseudorange, ephemerides, Doppler shift, C/N0 and elevation angle from each satellite	- Uses past error dependencies to influence current state outputs. - Larger training time compared to DT, KNN and SVM. - Performed better than DT, KNN, and SVM, especially for multipath affected signals for which classification accuracy was improved.
DT, KNN, GBDT [59]	Real data, static conditions, post-processing, dataset size of 134K samples	C/N0, pseudorange residual, and sat. elevation angle, 3D city model	- GBDT required less computational power and performed better than KNN and DT. KNN performed better than DT, and DT had lowest performance. - Difficulties in mitigating NLOS and multipath, despite use of 3D city model and ray-tracing in labeling process. - The model's sensitivity to the environment/location of the training data prevented signal classification in a different location.

Given their higher flexibility, ML-based approaches have shown great performance in estimating the complex behavior of these delays. The works in [60] and [61] tackled the task of predicting TEC using Gaussian Process Regression (GPR). In [60] a GPR model was trained to predict the vertical TEC at each location to generate a TEC map. This algorithm is designed to be updated in real-time in an EGNOS station network based on the dual-frequency measurements conducted by this correction system. In contrast, the work in [61] focused on the TEC content over an Indian GNSS station, and the results showed that the model had 100% accuracy. IONONet, a fully connected NN used to predict the vertical TEC, was described in [62]. An intensive study of the network architecture was conducted, using the solar flux 10.7 cm, Kp-index, day of the year, time of day, longitude, and latitude as the input features. Although the model showed promise, its results were worse than those derived from the NeQuick model. This could be explained by the 9-year gap between the training and testing datasets. The model could not predict anomalies because they were not present in the training dataset. Additionally, the network architecture was selected based on a sensitivity analysis using approximately 4% of the available dataset, which may not be representative of the complexity of the ionospheric TEC being modeled.

The authors in [63] implemented an NN to forecast the ZWD. The study focused on determining the precipitable water vapor, which is highly important in meteorological applications. Since the precipitable water vapor is obtained from the ZWD the conclusions of the article can be applied towards augmenting positioning accuracy, by reducing the tropospheric effect on GNSS measurements. In fact, for a Precise Point Positioning - Real-Time Kinematic (PPP-RTK)

TABLE 4. Specifications, advantages and limitations of the selected studies for Ionospheric and Tropospheric delay.

Model	Conditions	Input Features	Advantages / Limitations
GPR [60]	Real data, static conditions, real-time, dataset size of 26K samples	Vertical TEC at piercing point and respective latitude and longitude coordinates	- Convenient for incorporating prior knowledge. - Less training data than NN. - Half the accuracy of current state-of-art.
GPR [61]	Real data, static conditions, post-processing, dataset size of 11M samples (11 years)	Vertical TEC measurements, solar flux index and geo-magnetic index	- Better performance than ANN and auto regressive moving average. - Requires less training data compared to ANN.
NN [62]	Real data, static conditions, post-processing, dataset size of 123M samples	Solar flux, Kp-index, day of year, time of day, longitude and latitude	- Discontinuity of longitude input (180^{o} and -180^{o}) creates difficulties. - Could be used for real time estimation, but highly computation intensive.
NN [63]	Real data, static conditions, post-processing, dataset size of 105K samples	Day of year, time of day, temperature, pressure and relative humidity	- The non-linear relation between temperature and wet delay was not correctly predicted by the ANN.

solution, the wet delay can cause difficulties in solving integer ambiguity, leading to increased convergence time. After the feed-forward NN was trained on GNSS data obtained from two ground stations in Austria, the algorithm was tested, and it was found that it is capable of predicting ZWD with an RMSE of 1.50 cm up to six hours in advance.

A summary of the selected studies for atmospheric error estimation is presented in Table 4.

3) JAMMING, SPOOFING AND MEACONING

Radio frequency interference in GNSS signals is growing in amount and complexity, given current technological advancements. These have exposed GNSS vulnerabilities that need to be addressed, particularly for safety-critical applications, such as autonomous driving [64].

Jamming can be defined as an interference signal over the frequency band of interest that originates from intentional or unintentional actions. Although jamming is generally accepted to lead to signal loss, its effects can vary from having no effect (e.g., if the jammer is out of range) to severely deteriorating GNSS transmissions. Satellites with the lowest C/N0, which typically also have lower elevation angles, are impacted first, affecting the dilution of precision and causing a loss of lock and front-end saturation.

The study in [65] classified the type of jamming affecting the receiver by generating a black-and-white image of the spectrogram obtained from the incoming radio frequency in-phase/quadrature samples. These images were then processed with two different classifiers, SVM and CNN, achieving 94% and 91% average classification accuracy, respectively. This type of solution is not possible when using measurement-level GNSS receiver data, but could be implemented, for example, in a software-defined radio solution, where there is access to raw samples.

Technological growth has enabled other positioning methods based on the Signals of Opportunity (SOP). Cellular Long Term Evolution (LTE) towers exist in abundance in urban areas and transmit over a wide range of frequencies and directions. The C/N0 at the receiver is typically tens of dBs higher than typical GNSS signals, and they can be exploited with minimal effort. A jamming attack requires the entire spectrum of emissions to be targeted using high-power signals. Cellular SOPs span 700 MHz to 6 GHz; therefore, these signals present good robustness to interference and can be used to augment positioning reliability. The work in [64] exploits these characteristics and uses pseudorange measurements obtained from cellular SOP towers along with GNSS pseudoranges in a Simultaneous Localization and Mapping (SLAM) approach to improve GNSS robustness in challenging or denied environments. This approach successfully reduces the jamming impact.

Spoofing can be defined as a radio frequency signal interference attack with the intention of deceiving a GNSS receiver to track misleading signals, leading to invalid position, velocity, and timing solutions. Meaconing is a data-repetition attack that can lead to the same practical result as spoofing. GNSS vulnerability to spoofing and meaconing is of high importance because it can lead the receiver to compute a completely different solution without the user system detecting it or raising an alarm. This imposes an increase in hazardous misleading information and deterioration of the system's IM capabilities, deeming it unsuitable for safety-of-life applications. The work in [66] studied the use of an SVM with a radial basis function kernel to separate measurements between authentic and manipulated. The resulting high number of support vectors of the SVM reflects the high complexity of this task. Despite the short interference period used in the dataset, the system achieved a 98% detection accuracy under static conditions.

Table 5 summarizes the selected works on jamming, spoofing, and meaconing.

D. ENVIRONMENT DETECTION/CLASSIFICATION

Multipath and NLOS signals are among the most severe phenomena affecting GNSS signals. They are location- and time-dependent and most frequent in urban environments owing to the aforementioned issues. Considering these effects is crucial for improving GNSS-based positioning

TABLE 5. Specifications, advantages and limitations of the selected studies for Jamming, Spoofing and Meaconing.

Model	Conditions	Input Features	Advantages / Limitations
SVM, NN [65]	Synthetic data for GNSS and each jammer type, dataset size of 61800 black-and-white images	Black-and-white spectogram images with 512x512 pixels and 600 DPI.	- Interference detection accuracy nearly 99% for both SVM and CNN. - Interference classification accuracy of 94% for SVM and 91% for CNN.
SVM [66]	Real and synthetic data, static conditions, post-processing, dataset size of 8K epochs	Lock time, C/N0, pseudorange, Doppler, carrier-phase, multipath correction, code variance	- Dataset used had very short interference period. - High complexity due to high number of support vectors.

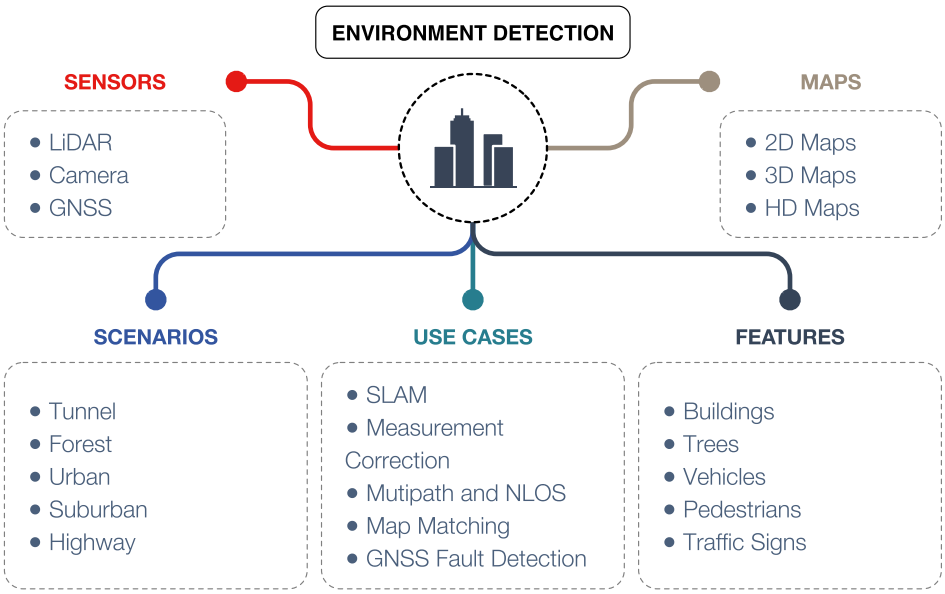


FIGURE 7. Environment detection and classification: sensors, features, scenarios and use cases.

performance. However, estimating the respective contributions is a challenging task because it depends on the complex interactions between the GNSS signal and elements in the surrounding environment. The type of sensors, features, maps, typical scenarios, and use cases for environment detection and classification are depicted in Fig. 7.

Different approaches can be used to obtain information about the surrounding environment. Using maps is a straightforward approach in which, in addition to the regular information found in 2D maps, 3D city models are becoming widely available. These maps are being explored to support 3D mapping-aided positioning methods, helping to address multipath and NLOS issues [67]. There are also proposals to utilize NLOS signals in a constructive manner [68].

While the information in regular maps can be used to determine the type of scenario in which a vehicle is located, more advanced approaches can use 3D information to model the expected effects at the location of the vehicle, ranging from models based on ray-tracing techniques to more sophisticated models based on wave propagation theory. The estimated path is employed to correct for the bias present in the pseudorange measurement. However, this approach has limitations as the level of detail and accuracy of the 3D surrounding environment is typically limited, and the models require significant processing power, which is usually not available in low-cost receivers.

More advanced approaches are required to address multipath and NLOS effects, such as the use of high-resolution 3D models or ML algorithms to predict their impact on signals. In [56], the authors trained a classifier using supervised learning to classify GNSS pseudorange measurements into clean, multipath, and NLOS categories. The authors created basic 3D building models of the Hung Hom area for data labeling by manually referencing a 3D model available on Google Earth. Various features from the GNSS raw data were evaluated, and a new feature was proposed to indicate the consistency between pseudorange and Doppler shift measurements. The SVM classifier achieved approximately 75% accuracy using the proposed feature and received signal strength.

Cameras and LiDAR sensors are frequently explored to implement perception systems to support automated driving. In a perception solution, computer vision and ML techniques are applied to video streams, usually from multiple cameras

or to a point-cloud from LiDAR, in order to detect and classify relevant elements surrounding the vehicle. The elements detected included different types of obstacles, other vehicles, pedestrians or traffic signs.

This type of system can also be explored to assist positioning in different ways (e.g., relative positioning). Scene detection/classification applied to the GNSS positioning domain can be used to detect and classify elements that can affect GNSS signals, such as buildings, trees, or tunnels. This approach has the advantage of using a system that may already be available in the vehicle and that can provide real-time information about the surrounding environment. The level of detail that can be obtained is limited by the sensor's field of view at a specific moment, providing only information about nearby elements.

In [69], a fish-eye camera was used to detect NLOS scenarios caused by urban canyons that affect the reception of the GNSS signal in the vehicle. The authors used this information to correct some of the GNSS pseudorange measurements, thereby improving the overall GNSS positioning accuracy.

Perception systems based on LiDAR and cameras have various advantages and drawbacks. The high cost of LiDAR sensors limits their large-scale deployment. In contrast, camera technology is cheaper and readily available on the mass market, making it a viable option. Cameras can detect both moving and static objects in the surrounding environment, and have an advantage over other perception sensors because of their ability to capture colors and textures, which is not possible with LiDAR sensors. Both solutions require high computing power for image-processing algorithms.

LiDAR sensors can also be used to create High Definition (HD) maps. Point cloud data can be used for SLAM [70]. When combined with GNSS, LiDAR can be used for 3D mapping of outdoor environments. HD maps are 3D representations of the environment that can be captured by survey vehicles. These maps can then be used in the online phase to obtain absolute positioning as an alternative or to complement the absolute positioning provided by GNSS (especially useful in urban environments).

HD maps can also be used in vehicles without LiDAR sensors as a source of information regarding the surrounding environment. In [71], map-matching was used to estimate the horizontal uncertainty level of the GNSS position fixes for autonomous vehicle navigation. This method uses an HD map to match the position based on proprioceptive sensors and GNSS fixes, and a particle filter is used to manage multiple hypotheses. The map-aided horizontal uncertainty level was computed by determining the maximum distance between a map-matched hypothesis and a given GNSS position, thereby providing a confidence indicator for integrity monitoring. Using HD maps has the disadvantage of not providing real-time information; large surveys are required to create these maps, and as the maps become outdated, the surveys must be repeated.

Selecting the best approach from those described above depends on striking a balance between complexity, cost, and the level of detail. In a vehicle that is already equipped with perception systems, information regarding nearby surrounding elements can be obtained at a relatively low cost. However, combining multiple approaches is the best way to achieve optimal results, with perception systems providing real-time detailed information about the nearby environment and maps providing a "big-picture" overview. HD maps may become an interesting source of information in the future, particularly if they are widely available and frequently updated.

V. DISCUSSION

From the studies related to position error estimation using AI, it is clear that this is a novel application, with very few studies currently available. The results obtained across this category were heterogeneous given the differences in the datasets and particular conditions. These ranged from different ground vehicles in real environments with nearby building, simulation of urban air mobility using synthetic data, and real data from urban scenarios, both from automotive-grade and smartphone-grade GNSS receivers. RF was shown to be better than KNN and SVM in directly predicting the positioning error using simulation data. Using real data, while DT was mainly useful for feature ranking and a-priori feature selection, SVM was able to provide 50% correct bounding PL, despite additional information from an external camera to estimate environmental obstruction not improving overall results. DNN and GCNN were able to handle GNSS measurement variability due to satellite visibility and improved both accuracy and precision over typical least-squares and KF implementations.

In the fault detection and exclusion category, most ML models were mainly used in supporting roles, providing improved parameters for RAIM implementations. Most studies used manual fault injection, with step and ramp errors used to prove the effectiveness of each implementation against traditional RAIM. In addition, these are also single faults; therefore, it is important for new works targeting RAIM implementation comparison to move in the direction of real and multiple faults in urban environments. The models applied in this category included NN, LSTM, HDBSCAN, and TDNN.

ML models can assist in the IM process by classifying the received signals in LOS, NLOS, or multipath conditions. These methods use various features such as C/N0, pseudorange residuals, satellite elevation angles, Doppler shift, and different variables obtained from in-car navigation systems. The models employed for this task were DT, GBDT, fuzzy logic, CNN, and GRU. The last model outperformed some standard models, which can be sustained by its ability to use information from the previous states.

Different ML models have been tested and compared for atmospheric error estimation applications, such as TEC and ZWD prediction. For example, the GPR model has been used to predict the TEC with real-time updates, whereas a fully connected NN has been used to predict the vertical TEC. Another NN-based approach was used to forecast ZWD,

which showed promising results in reducing the tropospheric effect on GNSS positioning. These studies proved that ML models can be used to predict this type of error. However, in a dynamic setting within urban areas, this type of application is unlikely to have a significant impact, because the quantities being estimated can be orders of magnitude below other prominent error sources.

Jamming, spoofing, and meaconing are becoming increasingly prevalent and complex, which is a significant concern for safety-critical GNSS applications. The solutions analyzed for this category operated at different levels: intermediate frequency samples for jamming detection, and GNSS observables for spoofing/meaconing detection. The SVM and CNN models were used for these tasks. The detection accuracy spanned between 90% and 99% among the multiple jamming types, with SVM performing slightly better than the CNN. For the spoofing/meaconing case, the detection accuracy is 98%. A spoofed satellite typically provides inconsistent positioning data; therefore, traditional RAIM techniques can detect this as an outlier and remove the affected satellite from the positioning solution.

The detection and classification of the surrounding environment is another component of the equation that is essential for improving the performance of GNSS-based positioning systems. Maps, including 3D city models, can provide useful information about the environment to support 3D mapping-aided positioning methods, but their accuracy is limited when modeling the impact of multipath and NLOS effects. Cameras and LiDAR sensors are being increasingly explored as perception systems to support automated driving and provide real-time information about the surrounding environment. The level of detail about the surrounding environment that can be obtained is limited by the capabilities of the sensors and environmental changes; however, once this information is provided to the GNSS system, it will enable performance improvements. Using the surrounding environment information may allow the use of a multi-model system, depending on the location of the receiver. This would then pose a different set of questions, such as when the transition from one model to the other should be performed, and what level of detail would be necessary to distinguish between, for example, an urban setting and a suburban one.

Some of the works analyzed in the position error category either included environmental detection or tried to infer it. It is clear from the literature that accounting for the environment is important, yet care must be taken since this is not a trivial task, and attempts to do so may become counterproductive, given that the estimation itself is typically time- and resource-consuming and not error-free.

In general, ML can consider more complex correlations between error sources that may not be captured by traditional statistical models. Owing to the ability to detect patterns in complex datasets, it is able to estimate errors without considering sources individually, which comprises in a simpler approach compared to classic models. ML models can help detect and correct errors, which is important for applications that require high levels of accuracy and reliability. However, despite multiple attempts to use different learning algorithms, the problem of integrity in the automotive context remains unsolved. This could be explained by the scarcity of large, balanced, and diverse datasets in dynamic environments. Consequently, simpler ML models are often implemented instead of advanced and data-hungry applications. Consequently, the limited success of DT, RF, SVM, and clustering algorithms such as HDBSCAN or KNN may be constrained by the lack of generalization and the presence of noisy data environments. Some state-of-the-art algorithms can be used to improve integrity if a proper dataset is available. For instance, the GBDT family can be effective in noisy data environments, is less prone to overfitting than RF and DT, and can handle different types of GNSS features without extensive pre-processing. NN can handle complex patterns in high-dimensional feature spaces, whereas CNNs can exploit and automatically learn the spatial correlation of data. LSTM and transformers are particularly effective in handling time-series sequential data by capturing global dependencies in the input sequence and are therefore appropriate for non-stationary conditions.

VI. CONCLUSION AND FUTURE WORK

Traditional techniques such as RAIM have been used for years to mitigate errors affecting GNSS measurements. With the shift from airplanes to road vehicles, urban scenarios, and safety-of-life requirements as the main drivers, the increasing complexity and diversity of error sources in GNSS systems require more advanced approaches. ML models can potentially provide solutions for improving integrity of GNSS-based positioning systems. However, further testing is required to compare these methods with the traditional techniques under realistic and fair conditions. The performance of ML models depends on various factors, including the size, shape, quality, and complexity of the dataset, the number of features, and the target problem. Datasets should be large and represent all intended scenarios and error classes to prevent overfitting. Additionally, the models should use inputs that do not depend on the global position of the receiver. This allows the model to generalize from the data and effectively provide better performance.

A few challenges still need to be resolved before IM systems based on machine learning can be implemented in road vehicles, particularly with automation levels 4 and above. First, it is necessary to construct a model that can estimate integrity in complicated urban contexts while accounting for many sources of errors, including jamming, spoofing, and meaconing. To accomplish this, the model must be able to estimate the integrity state, protection level (or the position error as a proxy of). Based on this research, it is anticipated that a learner that can deduce the fundamental behavior of GNSS error in urban settings will require a higher level of complexity than that of classic approaches (SVM, RF, and DT). Indeed, most current GCNN implementations for positioning error prediction have shown promising results.

It is crucial to provide an open-source baseline dataset to promote research in this field and achieve an already specified model that meets the required navigation performance. The aim of this dataset is to enable unbiased comparisons of techniques, even those built with expensive or confidential information. In addition, it enables broader validation of the features of interest when performing the IM task. The Android raw GNSS measurement dataset is a good first step towards this goal, although it only has 39 features, reducing the research scope of the features that may impact GNSS performance the most. In addition, this dataset should ideally contain samples from other complementary sensors, following the trend observed in the classic approaches towards the urban scenario.

The goal of the positioning system is to provide accurate position estimations that fall within $n \cdot \sigma$ range. An integrity monitoring system is interested in the tail of the error distribution (e.g., $1 - 1e^{-7}$) to estimate a protection level that correctly bounds the positioning error without hindering continuity and availability. This poses a serious issue regarding data imbalance because the number of epochs with faults is likely to be a very small percentage of the dataset. If the system sees a few points in fault conditions, it will not be able to learn to identify them. Data augmentation is an option that can be used to mitigate data imbalance. However, this must be approached with caution because the errors follow the laws of physics and creating artificial data points could lead to unmeaningful samples.

Even after a model has been successfully trained and validated, demonstrating that it is capable of complying with the requirements established for autonomous vehicles of level 4 or higher is far from trivial. The topics addressed in ISO/IEC TR 24029-1 (if NN are employed), ISO 23894, and ISO/IEC TR 24027 govern the usage of AI systems. The first sets the baseline for the robustness of NN systems, which are typically difficult to prove and test, necessary for their usage in safety-critical applications. The second addresses the risks that must be considered during model development and validation, namely, the complexity of the environment, lack of transparency, data quality, and hardware issues. The last focuses on one of the main ML bottlenecks, namely data bias. This bias can be caused by many factors, such as deficiencies in the system architecture, cognitive bias of the developer, and bias inherent to the data generation process, among others. It is easily learned by the model and can be propagated or even augmented while performing inferences.

The hardware in which the model will be deployed also poses a challenge to this research topic. This limits the size of the model and can cause latency problems or even loss of accuracy. These constraints typically translate to the use of small low-complexity models, which are not appropriate for modeling GNSS integrity in challenging environments. The computational power of small-sized hardware is increasing, and in future years, the computational power of road vehicles will increase significantly, thereby extending the implementation options.

In the future, our goal is to develop ML-based PL estimation for GNSS IM system, targeting road vehicles in urban and suburban scenarios.

ACRONYMS

AI	Artificial Intelligence.
AL	Alert Limit.
AR	Auto-Regressive.
ASIL	Automotive Safety Integrity Level.
C/N0	Carrier-to-Noise Ratio.
CNN	Convolutional Neural Network.
CPSOF	Chaos Particle Swarm Optimization Particle Filter.
DNN	Deep Neural Network.
DOP	Dilution of Precision.
DT	Decision Tree.
EKF	Extended Kalman Filter.
FDE	Fault Detection and Exclusion.
GBDT	Gradient Boosting Decision Tree.
GCNN	Graph Convolutional Neural Network.
GMM	Gaussian Mixture Model.
GNSS	Global Navigation Satellite System.
GPR	Gaussian Process Regression.
GPS	Global Positioning System.
GRU	Gated Recurrent Unit.
HD	High Definition.
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise.
HDOP	Horizontal Dilution of Precision.
IM	Integrity Monitoring.
IR	Integrity Risk.
KF	Kalman Filter.
KNN	K-Nearest Neighbors.
LiDAR	Light Detection and Ranging.
LOS	Line-Of-Sight.
LSTM	Long-Short Term Memory.
LTE	Long Term Evolution.
ML	Machine Learning.
NDOP	Ambiguity Dilution of Precision.
NED	North East Down.
NLOS	Non-Line-Of-Sight.
NN	Neural Network.
PDOP	Position Dilution of Precision.
PE	Positioning Error.
PL	Protection Level.
PPP-RTK	Precise Point Positioning - Real-Time Kinematic.
RAIM	Receiver Autonomous Integrity Monitoring.
RF	Random Forest.
RMSE	Root Mean Square Error.
SLAM	Simultaneous Localization and Mapping.
SNR	Signal-to-Noise Ratio.
SOP	Signals of Opportunity.
SVM	Support Vector Machine.

TDNN	Time Delayed Neural Network.
TEC	Total Electron Content.
TTA	Time To Alert.
V2I	Vehicle-to-Infrastructure.
VDOP	Vertical Dilution of Precision.
ZWD	Zenith Wet Delay.

REFERENCES

[1] *EUSPA EO and GNSS Market Report. 2022/Issue 1*, Publications Office of the European Union, European Union Agency for the Space Programme, Prague, Czech Republic, 2022.

[2] D. S. Maier and T. Pany, "Multipath and attitude estimation phase lock loop for antenna array signal processing," in *Proc. I_{ON} GNSS+, Int. Tech. Meeting Satell. Division Inst. Navigat.*, Oct. 2021, pp. 3402–3421.

[3] D. Parekh, N. Poddar, A. Rajpurkar, M. Chahal, N. Kumar, G. P. Joshi, and W. Cho, "A review on autonomous vehicles: Progress, methods and challenges," *Electronics*, vol. 11, no. 14, p. 2162, Jul. 2022.

[4] N. Joubert, T. G. R. Reid, and F. Noble, "Developments in modern GNSS and its impact on autonomous vehicle architectures," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 2029–2036.

[5] E. D. Tijero, L. M. Fernandez, J. I. H. Zarzosa, J. Garcia, J. Ibanez-Guzman, E. Stawiarski, P. Xu, G. Avellone, F. Pisoni, E. Falletti, and M. Ortiz, "High accuracy positioning engine with an integrity layer for safety autonomous vehicles," in *Proc. I_{ON} GNSS+, Int. Tech. Meeting Satell. Division Inst. Navigat.*, Oct. 2018, pp. 1566–1572.

[6] H. Jing, Y. Gao, S. Shahbeigi, and M. Dianati, "Integrity monitoring of GNSS/INS based positioning systems for autonomous vehicles: State-of-the-art and open challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14166–14187, Sep. 2022.

[7] P. Zabalegui, G. De Miguel, A. Pérez, J. Mendizabal, J. Goya, and I. Adin, "A review of the evolution of the integrity methods applied in GNSS," *IEEE Access*, vol. 8, pp. 45813–45824, 2020.

[8] T. G. R. Reid, S. E. Houts, R. Cammarata, G. Mills, S. Agarwal, A. Vora, and G. Pandey, "Localization requirements for autonomous vehicles," *SAE Int. J. Connected Automated Vehicles*, vol. 2, no. 3, pp. 173–190, Sep. 2019.

[9] L. Norman, E. Infante, and L. de Groot, "Integrity performance for precise positioning in automotive," in *Proc. I_{ON} GNSS+, Int. Tech. Meeting Satell. Division Inst. Navigat.*, Oct. 2019, pp. 1653–1663.

[10] T. Hassan, A. El-Mowafy, and K. Wang, "A review of system integration and current integrity monitoring methods for positioning in intelligent transport systems," *IET Intell. Transp. Syst.*, vol. 15, no. 1, pp. 43–60, Jan. 2021.

[11] *2021 Federal Radionavigation Plan*, Dept. Transp., Southeast, WA, USA, 2021.

[12] A. Welte, "Protection levels for high integrity localization for autonomous driving," M.S. thesis, Université Angers, ENSTA Bretagne, Angers, France, 2017.

[13] N. Zhu, "GNSS propagation channel modeling in constrained environments: Contribution to the improvement of the geolocation service quality," Ph.D. thesis, Université de Lille, IFSTTAR, Lille, France, 2018.

[14] M. Tossaint, J. Samson, F. Toran, J. Ventura-Traveset, M. Hernandez-Pajares, J. M. Juan, J. Sanz, and P. Ramos-Bosch, "The stanford–ESA integrity diagram: A new tool for the user domain SBAS integrity assessment," *Navigation*, vol. 54, no. 2, pp. 153–162, Jun. 2007.

[15] B. W. Parkinson and P. Axelrad, "Autonomous GPS integrity monitoring using the pseudorange residual," *Navigation*, vol. 35, no. 2, pp. 255–274, Jun. 1988.

[16] R. G. Brown and P. W. McBurney, "Self-contained GPS integrity check using maximum solution separation," *Navigation*, vol. 35, no. 1, pp. 41–53, Mar. 1988.

[17] M. A. Sturza, "Navigation system integrity monitoring using redundant measurements," *Navigation*, vol. 35, no. 4, pp. 483–501, Dec. 1988.

[18] E. Kaplan and C. Hegarty, *Understanding GPS: Principles and Applications*, 3rd ed., Bedford, MA, USA: Artech House, May 2017.

[19] L. Gratton, M. Joerger, and B. Pervan, "Carrier phase relative RAIM algorithms and protection level derivation," in *Proc. Int. Tech. Meeting Inst. Navigat.*, vol. 63, Feb. 2010, pp. 215–231.

[20] I. Nikiforov, "Advanced RAIM algorithms for safe navigation based on the constrained GLR test," in *Proc. Amer. Control Conf.*, New York, NY, USA, Jul. 2007, pp. 1840–1845.

[21] J. Blanch, T. Walter, and P. Enge, "Gaussian bounds of sample distributions for integrity analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 4, pp. 1806–1815, Aug. 2019.

[22] K. Gunning, J. Blanch, T. Walter, L. de Groot, and L. Norman, "Design and evaluation of integrity algorithms for PPP in kinematic applications," in *Proc. I_{ON} GNSS+, Int. Tech. Meeting Satell. Division Inst. Navigat.*, Oct. 2018, pp. 1910–1939.

[23] J. Blanch, T. Walter, C. Milner, M. Joerger, B. Pervan, and D. Bouvet, "Baseline advanced RAIM user algorithm: Proposed updates," in *Proc. Int. Tech. Meeting Inst. Navigat.*, Feb. 2022, pp. 229–251.

[24] Y. Jiang and J. Wang, "A-RAIM and R-RAIM performance using the classic and MHSS methods," *J. Navigat.*, vol. 67, no. 1, pp. 49–61, Jan. 2014.

[25] S. Pullen and M. Joerger, *GNSS Integrity and Receiver Autonomous Integrity Monitoring (RAIM)*. Hoboken, NJ, USA: Wiley, Dec. 2020, pp. 591–617.

[26] P. D. Groves and Z. Jiang, "Height aiding, C/N_0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas," *J. Navigat.*, vol. 66, no. 5, pp. 653–669, Sep. 2013.

[27] Z. Xue, Z. Lu, Z. Xiao, J. Song, and S. Ni, "Overview of multipath mitigation technology in global navigation satellite system," *Frontiers Phys.*, vol. 10, pp. 1–18, Dec. 2022.

[28] I. Smolyakov, M. Rezaee, and R. B. Langley, "Resilient multipath prediction and detection architecture for low-cost navigation in challenging urban areas," *Navigation*, vol. 67, no. 2, pp. 397–409, Jun. 2020.

[29] L. Wang, P. D. Groves, and M. K. Ziebart, "Smartphone shadow matching for better cross-street GNSS positioning in urban environments," *J. Navigat.*, vol. 68, no. 3, pp. 411–433, May 2015.

[30] D. Margaria and E. Falletti, "A novel local integrity concept for GNSS receivers in urban vehicular contexts," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, May 2014, pp. 413–425.

[31] A. Saenz, "Method for autonomous determination of protection levels for GNSS positioning based on navigation residuals and an isotropic confidence ratio," U.S. Patent 8 203 482, Jun. 19, 2012.

[32] S. Langel, O. G. Crespillo, and M. Joerger, "A new approach for modeling correlated Gaussian errors using frequency domain overbounding," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 868–876.

[33] I. Sheret, C. D. Hide, O. Julien, and R. Bryant, "Method and apparatus of single epoch position bound," U.S. Patent 11 885 890, Jan. 30, 2024.

[34] O. Julien, R. Bryant, and C. Hide, *Tight Position Bounding for Automotive Integrity*. Red Bank, NJ, USA: Inside GNSS, Jun. 2021.

[35] H. Zhao, J. Tang, Z. Li, Z. Wu, S. Xie, Z. Wu, M. Liu, and B. T. G. S. Kumara, "Efficient graph neural network driven recurrent reinforcement learning for GNSS position correction," in *Proc. I_{ON} GNSS+, Int. Tech. Meeting Satell. Division Inst. Navigat.*, Oct. 2023, pp. 216–230.

[36] E. Zhang and N. Masoud, "Increasing GPS localization accuracy with reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2615–2626, May 2021.

[37] L. Fu, J. Zhang, R. Li, X. Cao, and J. Wang, "Vision-aided RAIM: A new method for GPS integrity monitoring in approach and landing phase," *Sensors*, vol. 15, no. 9, pp. 22854–22873, Sep. 2015.

[38] A. Kuratomi, "GNSS position error estimated by machine learning techniques with environmental information input," M.S. thesis, School Ind. Eng. Manage., KTH Roy. Inst. Technol., Stockholm, Sweden, May 2019.

[39] O. K. Isik, I. Petrunin, G. Inalhan, A. Tsourdos, R. V. Moreno, and R. Grech, "A machine learning based GNSS performance prediction for urban air mobility using environment recognition," in *Proc. IEEE/AIAA 40th Digit. Avionics Syst. Conf. (DASC)*, Oct. 2021, pp. 1–5.

[40] E. Karlsson and N. Mohammadiha, "A statistical GPS error model for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Changshu, China, Jun. 2018, pp. 754–759.

[41] A. V. Kanhere, S. Gupta, A. Shetty, and G. Gao, "Improving GNSS positioning using neural-network-based corrections," *Navigation, J. Inst. Navigat.*, vol. 69, no. 4, pp. 1–20, 2022, Art. no. navi.548.

[42] A. Mohanty and G. Gao, "Learning GNSS positioning corrections for smartphones using graph convolution neural networks," *Navigation, J. Inst. Navigat.*, vol. 70, no. 4, pp. 1–15, 2023, Art. no. navi.622.

[43] E. Wang, T. Pang, M. Cai, and Z. Zhang, "Application of neural network aided particle filter in GPS receiver autonomous integrity monitoring," in *Proc. China Satell. Navigat. Conf. (CSNC)*, J. Sun, W. Jiao, H. Wu, and M. Lu, Eds., Berlin, Germany: Springer, Jan. 2014, pp. 147–157.

[44] E. Wang, C. Jia, G. Tong, P. Qu, X. Lan, and T. Pang, "Fault detection and isolation in GPS receiver autonomous integrity monitoring based on chaos particle swarm optimization-particle filter algorithm," *Adv. Space Res.*, vol. 61, no. 5, pp. 1260–1272, Mar. 2018.
[45] Y. Sun, "RAIM-NET: A deep neural network for receiver autonomous integrity monitoring," *Remote Sens.*, vol. 12, no. 9, p. 1503, May 2020.
[46] Y. Xia, S. Pan, X. Meng, W. Gao, F. Ye, Q. Zhao, and X. Zhao, "Anomaly detection for urban vehicle GNSS observation with a hybrid machine learning system," *Remote Sens.*, vol. 12, no. 6, p. 971, Mar. 2020.
[47] S. Tanwar and T. Y. Mina, "Integrity-based fault detection for GNSS navigation safety," Stanford Univ., Stanford, CA, USA, Tech. Rep. 26645897, 2019.
[48] D. Kim and J. Cho, "Improvement of anomalous behavior detection of GNSS signal based on TDNN for augmentation systems," *Sensors*, vol. 18, no. 11, p. 3800, Nov. 2018.
[49] S.-l. Tan, J. Liu, B.-g. Cai, and J. Wang, "Integrity monitoring of GNSS-based vehicle positioning based on cooperative weighting strategy," in *Proc. 6th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Oct. 2021, pp. 674–680.
[50] M. Socharoentum, H. A. Karimi, and Y. Deng, "A machine learning approach to detect Non-Line-Of-Sight GNSS signals in Nav2Nav," in *Proc. 23rd ITS World Congr.*, Oct. 2016, pp. 10–14.
[51] R. Yozevitch, B. B. Moshe, and A. Weissman, "A RobustGNSS-LOS/NLOS signal classifier," *Navigation*, vol. 63, no. 4, pp. 429–442, Dec. 2016.
[52] B. Guermah, H. E. Ghazi, T. Sadiki, and H. Guermah, "A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems," in *Proc. IEEE Int. Conf. Technol. Manage., Oper. Decisions (ICTMOD)*, Nov. 2018, pp. 94–100.
[53] B. Guermah, T. Sadiki, and H. El Ghazi, "Fuzzy logic approach for GNSS signal classification using RHCP and LHCP antennas," in *Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2017, pp. 203–208.
[54] Y. Quan, L. Lau, G. W. Roberts, X. Meng, and C. Zhang, "Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning," *Remote Sens.*, vol. 10, no. 12, p. 2052, Dec. 2018.
[55] E. Munin, A. Blais, and N. Couellan, "Convolutional neural network for multipath detection in GNSS receivers," in *Proc. Int. Conf. Artif. Intell. Data Anal. Air Transp. (AIDA-AT)*, Feb. 2020, pp. 1–10.
[56] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
[57] R. Sun, L.-T. Hsu, D. Xue, G. Zhang, and W. Y. Ochieng, "GPS signal reception classification using adaptive neuro-fuzzy inference system," *J. Navigat.*, vol. 72, no. 3, pp. 685–701, May 2019.
[58] P. Geragersian, I. Petrunin, W. Guo, and R. Grech, "Multipath detection from GNSS observables using gated recurrent unit," in *Proc. IEEE/AIAA 41st Digit. Avionics Syst. Conf. (DASC)*, Sep. 2022, pp. 1–7.
[59] R. Sun, G. Wang, W. Zhang, L.-T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft Comput.*, vol. 86, pp. 1–12, Jan. 2020, Art. no. 105942.
[60] B. Lupsic and B. Takacs, "Gauss process regression for real-time ionospheric delay estimation from GNSS observations," *Acta Geodaetica et Geophys.*, vol. 57, no. 1, pp. 107–127, Mar. 2022.
[61] L. Mallika I, D. V. Ratnam, S. Raman, and G. Sivavaraprasad, "Machine learning algorithm to forecast ionospheric time delays using global navigation satellite system observations," *Acta Astronautica*, vol. 173, pp. 221–231, Aug. 2020.
[62] R. O. Perez, "Using TensorFlow-based neural network to estimate GNSS single frequency ionospheric delay (IONONet)," *Adv. Space Res.*, vol. 63, no. 5, pp. 1607–1618, Mar. 2019.
[63] M. O. Selbesoglu, "Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 5, pp. 967–972, Oct. 2020.
[64] Z. M. Kassas, J. Khalife, A. A. Abdallah, and C. Lee, "I am not afraid of the GPS jammer: Resilient navigation via signals of opportunity in GPS-denied environments," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 7, pp. 4–19, Jul. 2022.
[65] R. M. Ferre, A. de la Fuente, and E. S. Lohan, "Jammer classification in GNSS bands via machine learning algorithms," *Sensors*, vol. 19, no. 22, p. 4841, Nov. 2019.
[66] S. Semanjski, I. Semanjski, W. De Wilde, and A. Muls, "Use of supervised machine learning for GNSS signal spoofing detection with validation on real-world meaconing and spoofing data—Part I," *Sensors*, vol. 20, no. 4, p. 1171, Feb. 2020.
[67] P. D. Groves, "Shadow matching: A new GNSS positioning technique for urban canyons," *J. Navigat.*, vol. 64, no. 3, pp. 417–430, Jul. 2011.
[68] S. Miura, L.-T. Hsu, F. Chen, and S. Kamijo, "GPS error correction with pseudorange evaluation using three-dimensional maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3104–3115, Dec. 2015.
[69] X. Bai, W. Wen, and L. Hsu, "Using sky-pointing fish-eye camera and LiDAR to aid GNSS single-point positioning in urban canyons," *IET Intell. Transp. Syst.*, vol. 14, no. 8, pp. 908–914, Aug. 2020.
[70] G. He, X. Yuan, Y. Zhuang, and H. Hu, "An integrated GNSS/LiDAR-SLAM pose estimation framework for large-scale map building in partially GNSS-denied environments," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[71] F. Li, P. Bonnifait, and J. Ibañez-Guzmán, "Using high definition maps to estimate GNSS positioning uncertainty," in *Proc. Eur. Navigat. Conf. (ENC)*, Lausanne, Switzerland, May 2017, pp. 1–6.



INÊS VIVEIROS received the M.Sc. degree in aerospace engineering with a minor in artificial intelligence and control systems, from the Instituto Superior Técnico, University of Lisbon. She is currently an AI Software Developer with Car Multimedia, S.A., Bosch, Braga. She specializes in the development of data driven AI algorithms for the automotive sector, particularly for safety critical applications.



HÉLDER SILVA received the M.Sc. degree in communications engineering and the Ph.D. degree in electronics and computer engineering from the University of Minho. He is currently an Assistant Researcher with the Algoritmi Center, University of Minho. His research interests include GNSS positioning, indoor positioning, wireless sensor networks, and communication protocols.



YURI ANDRADE received the M.Sc. degree in remote sensing from the University of Porto, Portugal, in 2022. He is currently a Systems Engineer with Bosch, Portugal, with a focus on GNSS and navigation systems. His work emphasizes developing and testing innovative GNSS integrity solutions to support autonomous land vehicles, with a strong focus on applying machine learning models to improve system reliability in challenging environments, such as urban canyons. His expertise spans GNSS integrity, UAV operations, geoprocessing, and remote sensing technologies.



CRISTIANO PENDÃO received the integrated Ph.D. degree in telecommunications/computer science engineering from the University of Minho, University of Aveiro, and University of Porto, in 2019. He is currently a Professor with the Department of Engineering, University of Trás-os-Montes and Alto Douro, Portugal. He is also a Full Research Member with the ALGORITMI Research Centre, University of Minho. He has co-authored several scientific publications and has actively participated in innovation and commercial projects within the industry. His research interests include mobile computing, positioning and navigation, computer vision/ perception, and machine learning.