DOI. No. 10.1109/MAES.2024.3395182

Tutorial:

Machine Learning in GNSS Multipath/NLOS Mitigation: Review and Benchmark

***Penghui Xu*, *Guohao Zhang*, *Bo Yang*, *and Li-Ta Hsu*, The Hong Kong Polytechnic University, Hong Kong**

Global navigation satellite system (GNSS) positioning accuracy is degraded in urban canyons due to signal blockages and reflections, which is still a major challenge. Recently, using machine learning to improve the accuracy of GNSS positioning in urban areas has become a new trend. This article summarizes the works focused on GNSS multipath/non-light-of-sight (NLOS) mitigation using machine learning. The review of the studies is categorized based on the input features, algorithms, and outputs. The categorization shows that the received signal strength, elevation angle, and receiver correlator outputs from a single channel of satellite signal are the most popular input features. For the algorithm selection, the support vector machine and fully connected neural network are the algorithms most widely used. In terms of the outputs, most of the works made improvements in measurement status prediction, namely, LOS, multipath, and NLOS. Besides, this article also provides an open-source dataset with four scenarios for machine learning algorithms for the GNSS multipath/NLOS mitigation. Finally, the benchmarks are established based on the proposed dataset and the FCNN and least-squares estimation to enable performance evaluation in *Kaggle*.

Authors' current address: Penghui Xu, Guohao Zhang, and Li-Ta Hsu are with the Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: peng-hui.xu@connect.polyu.hk, guo-hao.zhang@connect.polyu.hk, lt.hsu@polyu.edu.hk). Bo Yang is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: bo.yang@polyu.edu.hk). *(Corresponding author: Li-Ta Hsu.)*

Manuscript received 19 April 2022, revised 11 March 2024; accepted 24 April 2024, and ready for publication 30 April 2024.

Review handled by William Dale Blair.

INTRODUCTION

Global navigation satellite system (GNSS) signal can be blocked or reflected by buildings in urban areas, leading to degradation of its positioning accuracy. Such interferences can be categorized into three cases [1]. The most straightforward is the complete signal blockage, with no signal acquired and tracked. The second is the non-line-of-sight (NLOS) reception, in which only the reflected signal with a delay can be received. The last is the multipath effect, where both the direct and the reflected signals are received and cause interaction. Many studies have been conducted for different positioning stages to mitigate the multipath or/and NLOS degradation, from signal acquisition to producing ranging-measurement and finally position estimation. A comprehensive introduction to these multipath mitigation methods can be found at [2], such as advanced correlator design and 3D Map Aided (3DMA) GNSS, which could substantially improve the positioning accuracy with the 3D building model. However, the multipath/NLOS effects are still not fully solved by the approaches mentioned in [2]. The correlator design cannot deal with NLOS receptions [3], while the 3DMA GNSS cannot model the interference involving dynamic objects that are not modeled in 3D city models [4]. In recent years, machine learning (ML) is a technology gaining attention in the GNSS realm [5]. ML consists of many computational methods, which use experience data to improve the performance of models [6]. Figure 1 depicts the frequency of the related keywords in Google Scholar over the past few years. The increment in paper related to GNSS ML shows that ML has become a new trend in this field.

ML plays a variety of roles in GNSS positioning [7]. For signal acquisition, the fully connected neural work (FCNN) and convolutional neural network (CNN) are used for signal detection [8]. For GNSS/INS integration, long short-term memory (LSTM), which is a variation of recurrent neural network (RNN), was used to predict the position increment [9], and the random forest (RF) was employed to classify the positioning accuracy level [10]. For ionospheric scintillations detection, the decision tree (DT) [11] or support vector machine (SVM) [12], [13],

Image licensed by Ingram Publishing

[14] is employed. For environmental context awareness, SVM and hidden Markov models are used to classify the scene types [15], [16]. For signal spoofing detection, C-SVM is used [17]. As a result, it is good to have a review on ML in the GNSS discipline. This article aims to review the works related to ML based GNSS multipath and NLOS mitigations. A summary is given in Table 1. This work categorized the reviewed works based on their objectives: improved pseudorange measurement, measurement error prediction, measurement status prediction, and positioning level information.

- **Improved Pseudorange Measurement**: Minimize the multipath error by applying advanced receiver correlator and discriminator techniques to mitigate the effect of multipath.
- **Measurement Error Prediction**: Estimate the NLOS/multipath error.
- **Measurement Status Prediction**: Predict multipath or NLOS signal status, which could help to determine whether to correct, deweight, or exclude the measurement.
- **Positioning Level Information**: Predict positioning error or directing positioning.

The performance evaluation of different ML algorithms helps us to reveal their characteristics, which can help us understand the pros and cons of using a specific ML algorithm. However, this is difficult since the results of different works are hard to be compared. First, the datasets in various works are different, including the measurement quality on different grades of GNSS receivers, the environment complexity in different locations, and the variation of satellite geometry in different time periods. Second, as shown in Table 1, the outputs predicted by ML algorithms are usually in different domains. For example, some works may focus on the measurement error prediction while others may focus on the position level. Therefore, it is necessary to develop a public dataset and benchmarks in ML for the GNSS multipath/NLOS challenges, which facilitates the evaluation and comparison of the increasing studies in this field.

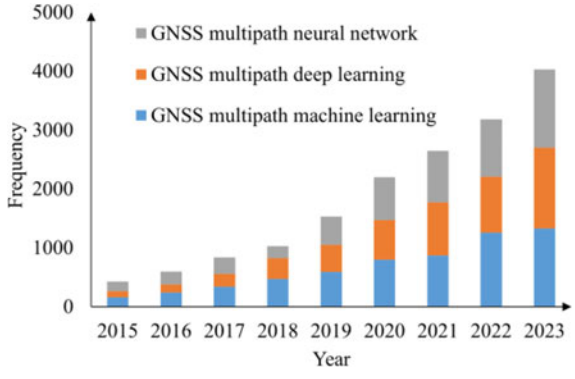


Figure 1.
Frequency of different keywords in *Google Scholar*. Search date: 19 February 2024. Different colors denote different keywords.

In summary, the contributions of this article are threefold:

1. Analyze the works in ML from the perspectives of input features, algorithms, and outputs, including a systematic categorization and detailed comparison.
2. Propose an open-source measurement-level dataset covering most urban environments for the training, validation, and testing of the ML research related to GNSS multipath/NLOS detection and/or mitigation.
3. Provide a benchmark of the three GNSS multipath/NLOS tasks: LOS/NLOS classification, pseudorange error prediction, and positioning error prediction. They are established based on the most popular machine algorithm (FCNN) and the least-squares estimation. Public online competitions are created in *Kaggle* to evaluate and compare the results from different researchers.

The following is the outline for the remainder. The "Features Used in GNSS Multipath Mitigation" section will cover the common GNSS features used in ML for the multipath/NLOS problem. The "Approaches of Machine Learning Based GNSS Multipath and NLOS Mitigation" section reviews and categorizes different ML algorithms for the GNSS multipath/NLOS mitigation and detection. The detailed dataset configuration will be introduced in "Dataset

Table 1.

Output and Method Used for the Reviewed Works			
Output	**Conventional Machine Learning**		**Deep Learning**
	Supervised	**Unsupervised**	
Improved Pseudorange Measurement		[18]	[19], [20], [21], [22]
Measurement Error Prediction	[23], [24], [25] [26], [27]		[28], [29], [30]
Measurement Status Prediction	[31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [30]	[46], [47], [48], [49]	[50], [32], [51], [46], [52], [53], [29], [34], [37], [38], [54], [41], [30]
Positioning Level Information	[55], [56]		[57], [58], [59], [60], [55], [56]

for Machine Learning" section. The establishment of benchmarks will be discussed in "Benchmark" section. Finally, the conclusions are drawn in the "Conclusions" section.

FEATURES USED IN GNSS MULTIPATH MITIGATION

Feature selection is particularly crucial in machine learning. The features will be divided into measurement level and receiver level based on the stage in which the feature is generated. Measurement level denotes the features produced by the receiver, i.e., the RINEX-derived features, while the receiver level denotes the features generated inside the receiver tracking loops. Figure 2 describes the statistics of different features used in the reviewed works. The graph shows the measurement-level features are more diverse than the receiver-level features. For measurement-level features, C/N_0, elevation angle, and pseudorange residual are the most used features, while at the receiver-level ones, the correlator output is the most popular one.

MEASUREMENT-LEVEL FEATURE

Pseudorange (ρ)**:** Pseudorange includes the range between the receiver and satellite, receiver clock bias, satellite clock bias, and error from other sources. It is the most relevant observation for positioning with trilateration. However, using this measurement directly may not be effective in ML because the scale and the numerical variance of the pseudorange are large. Differencing the pseudorange between time epochs or between different receivers are two approaches to utilizing pseudorange. The time differenced pseudorange is the pseudorange contributed by the receiver and satellite movements. In multipath/NLOS, this value can jump because of the sudden appearance/disappearance of the reflection point. The difference of pseudorange in time epochs is used as the feature in the K-means grouping [49]. For the differenced pseudorange between two receivers, it is similar to the differential GNSS. In [44], assuming the locations of the reference station and the rover are known, and the reference station is under open-sky, the multipath error of the rover is estimated through the differencing process and then used to determine whether the multipath exists.

Pseudorange residual (ε_{LS})**:** Pseudorange residual is a prevalent feature. It can indicate the consistency between the ranging measurement and the least squares estimation, derived by

$$\boldsymbol{\varepsilon}_{\mathrm{LS}} = \tilde{\boldsymbol{\rho}} - \mathbf{G}\hat{\mathbf{p}} \tag{1}$$

$$\hat{\mathbf{p}} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\tilde{\boldsymbol{\rho}} \tag{2}$$

where $\boldsymbol{\varepsilon}_{\mathrm{LS}}$ is the vector of pseudorange residuals from the satellites, $\tilde{\boldsymbol{\rho}}$ is the pseudorange measurements, $\hat{\mathbf{p}}$ is the estimated position, and receiver clock bias. $\mathbf{G}$ is the geometrical matrix containing the unit LOS vector of each satellite. A larger pseudorange residual indicates the corresponding measurement is inconsistent with the measurements from other satellites, probably under degradation. In [59], the pseudorange residual is concatenated with the LOS vector to construct the 4D vector (one for the pseudorange residual, three for the LOS vector) as the neural network input. The idea behind this is to consider both the satellite geometry and the pseudorange residual, which are the two factors affecting the positioning performance. By integrating the inputs for each satellite measurement through a network structure (known as Transformer [61] in the ML field), the correction on the positioning domain

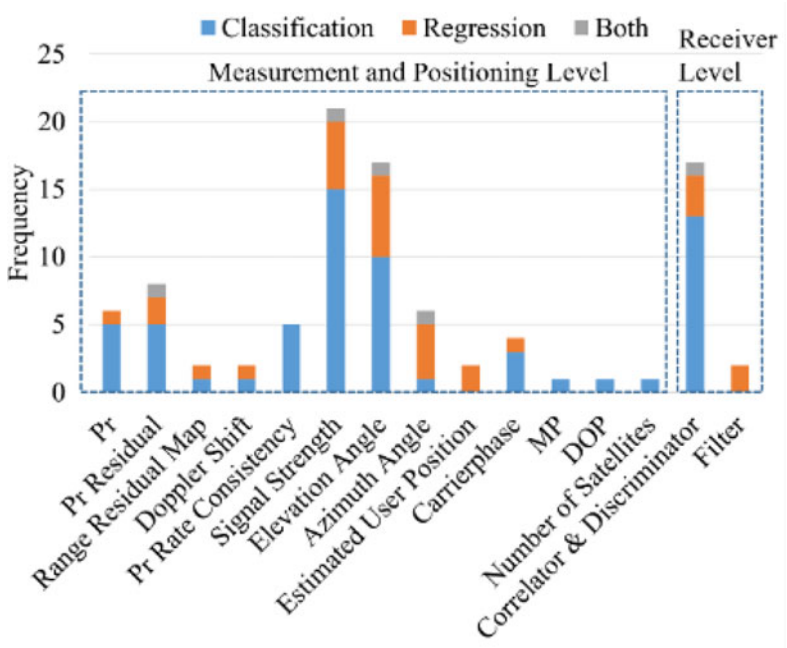


Figure 2.
Frequency of different features used in ML in GNSS multipath and NLOS, where Pr is the abbreviation of pseudorange. The features not listed in the table will be counted as the features used to derive them (e.g., time-differenced pseudorange is regarded as pseudorange). Colors blue, orange, and gray denote the features used for classification, regression, and both tasks, respectively. The classification task outputs the result as a probability (0 to 1), while the regression task can output any arbitrary number.

can be obtained. Besides, because the pseudorange residual is produced by all measurements, it naturally contains the overall information of the positioning solution. The simple combination of the features purely from a single measurement (like C/N_0) and all measurements (like pseudorange residual) can achieve good performance in multipath/NLOS detections [26], [29].

Range Residual Map (ε_{R}): Range residual map describes the residual as a map [56]. The reason for using a map representation is that a map has good compatibility with the neural network. The range residual in a location P can be derived by

$$\varepsilon_{\mathrm{R}}^{\mathrm{P}} = \tilde{\rho} - \hat{\delta}_{clk} - D^{\mathrm{P}} \tag{3}$$

where $\varepsilon_{\mathrm{R}}^{\mathrm{P}}$ is the range residual in a location P, $\tilde{\rho}$ is the pseudorange measurement, whose ionospheric and tropospheric errors have been corrected through convention models, e.g., Klobuchar [62] and Saastamoinen model [63]. $\hat{\delta}_{clk}$ is the receiver clock bias estimated by the least-squares method, D^{P} is the expected range between the location P and a satellite. The range residual map $\boldsymbol{\varepsilon}_{\mathrm{R}}$ can be constructed by sampling different locations P near the least-squares solution. Compared to the mentioned differential method of pseudorange, the residual map contains more information like elevation angle, azimuth angle, and the fitness of a least-squares estimation. The information on the residual map brings the opportunity and the challenge of designing an effective ML algorithm. In [56], the researcher utilized CNN to generate position information with the range residual maps.

Doppler shift ($f_{Doppler}$)**:** Doppler shift is the change in frequency due to relative motion, indicating the pseudorange changing rate. Doppler shift changes continuously since the satellite is orbiting around the Earth. In multipath cases, the Doppler shift is contributed by both the vectors of LOS and the reflecting signals. On the one hand, when the LOS vector is close to the multipath vector, the error of Doppler shift due to multipath will be close to zero. On the other hand, when the LOS vector has an opposite direction with the multipath one, the error can be maximized [64]. Thus, the Doppler shift error can potentially be used to detect the multipath. Currently, researchers tend to use it to construct the pseudorange rate consistency as an input feature instead of using it directly.

Pseudorange rate consistency ($\varepsilon_{\Delta Pr}$)**:** Pseudorange rate consistency $\varepsilon_{\Delta Pr}$ indicates the consistency between time-differenced pseudorange and pseudorange rate derived by

$$\varepsilon_{\Delta Pr} = |\Delta\rho - \dot{\rho}\Delta t| \tag{4}$$

$$\Delta\rho = \tilde{\rho}_t - \tilde{\rho}_{t-1} \tag{5}$$

$$\dot{\rho} = f_{Doppler,t} - \frac{\mathrm{c}}{f_{Carrier}} \tag{6}$$

where $\Delta\rho$ is the time-differenced pseudorange, $\dot{\rho}$ is the pseudorange changing rate calculated by Doppler

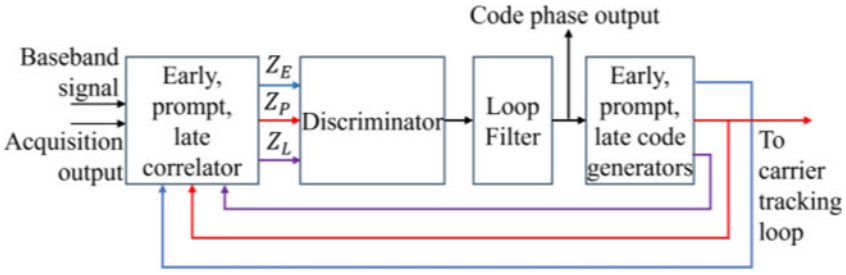


Figure 3.
DLL block diagram. Different colors denote early, prompt, and late correlators, respectively.

frequency $f_{Doppler,\,t}$ at time t, c is the speed of light, $f_{Carrier}$ is the carrier frequency, for example, L1 is 1575.42 MHz. This measurement has been demonstrated to have a high correlation with the multipath and NLOS phenomenon. A small $\varepsilon_{\Delta Pr}$ can represent a healthy signal (i.e., LOS signal), while large $\varepsilon_{\Delta Pr}$ may indicate NLOS or multipath [42].

Carrier-to-noise ratio (C/N_0)**:** C/N_0 highly corresponds to the multipath and NLOS reception since the signal strength can be significantly attenuated by reflections [65]. LOS signal usually has a higher C/N_0 value, while the NLOS or multipath signal usually has a lower C/N_0 value [45], [66], [67], [68]. A high C/N_0 value does not necessarily mean a LOS signal because the contribution of the reflection to the LOS signal can be both constructed or destructive [69].

Elevation angle (θ_{Ele})**:** Although a higher elevation angle means the signal is less likely to be blocked or reflected by the buildings, elevation angle only provides us with one of the criteria to diagnose the visibility of the satellites. Based on the correlation of C/N_0 and elevation angle [70], the signal status (LOS, multipath, or NLOS) can be better diagnosed. The elevation angle can also be represented in different forms depending on the machine algorithm design, such as a LOS vector and the range residual map. It is the same for azimuth angle.

Azimuth angle (θ_{Azi})**:** Signal interference can occur at any azimuth angle, depending on the geolocation of the receiver and the geometry of the satellite and its surrounding buildings. Hence, the azimuth angle is usually used to supplement other features.

Estimated user position ($\mathbf{p}_{user}$)**:** This is the user position estimated by a receiver. In [60], the authors used the temporal series of the estimated position to predict multipath error in a static experiment. The possibility of utilizing the temporal series lies in finding periodic changes in the multipath error, which is caused by the relative phase between the LOS and reflected signals and the unchanged satellite orbits. This is exactly what neural networks are good at. For the applications where the error pattern is changing (e.g., dynamic cases), directly using the temporal series may not help so much.

Carrier phase (ϕ)**:** Carrier-phase measurement describes the distance between a receiver and a satellite by accumulating the cycles of the carrier [71]. Thus, this measurement also includes an unknown cycle, which is well known as ambiguity. Although the carrier phase has a much higher resolution than code measurement, the ambiguity resolution is also more vulnerable to interferences. When multipath or NLOS occurs, the signal strength could drop, and the code tracking will be influenced. Eventually, the chance of losing the lock in the phase-lock-loop will increase, which renders a cycle slip occur more frequently [69]. In [49], the differential carrier phase and pseudorange in the time domain were utilized to detect multipath. When there is a multipath, the differential carrierphase can jump as the phase-lock-loop may lose the lock.

Linear combinations of code and carrierphase measurements (*MP*) [72]: Based on the combination terms, if a signal is suffering from multipath, with the changing satellite position, the multipath is also varying due to the changing geometry and phase difference between the reflection and LOS signal. This will consequently change the *MP* value, which indicates the *MP* is useful for multipath detection. In [41], all the *MPs* in one epoch constitute a *MP* vector to classify the signal status.

Dilution of precision (*DOP*)**:** *DOP* describes the factor of dilution based on the geometry of the satellites corresponding to a receiver [69]. *DOP* can be represented in a local coordinate frame and indicates how the positioning is sensitive to the measurement in different directions. Therefore, it can be a good indicator in ML when we try to have the model at the positioning level.

Number of satellites (N_{SV})**:** The number of satellites observed by the receiver. In [44], the authors used N_{SV} directly as the supplementary feature for the multipath classification. Besides, N_{SV} can help to sense the environmental context [73], and the number of LOS satellites can be used in sky visibility estimation [74].

RECEIVER-LEVEL FEATURE

Features in the receiver will be divided into three categories according to which stage of the delay-lock-loop (DLL) they are generated from, as shown in Figure 3.

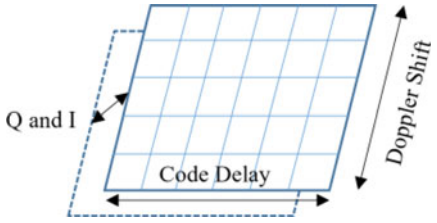


Figure 4.
Code-frequency searching map representation of an ACF output.

Correlator: In the correlator, the local replicas will be used to correlate with the incoming signal. A conventional DLL has three correlators: early, prompt, and late. Each correlator corresponds to a specific correlation value. Therefore, more correlators can retrieve a more complete autocorrelation function (ACF), which means a higher resolution. Taro, etc., defined features like the number of local maxima and the distribution of the delay of the maximum correlation [34]. This is because the characteristics (e.g., mean value, peak value, and standard deviation) of the ACF perform differently on LOS and multipath/NLOS cases [75]. Besides, an ACF can also be represented as the code-frequency searching map in Figure 4. The signal acquisition search a proper Doppler shift and the code delay in a 2D space, which could be regarded as an image. As shown in Figure 4, the two dimensions are the searched code delay and Doppler shift. In addition, in-phase (I) and quadrature (Q) components of the signal can be regarded as different channels in image representation, like the RGB color channels. The correlation peak can be constructed or destructed by the reflection signal, which makes it possible for multipath detection [37].

Discriminator: A discriminator calculates the code delay based on the correlation values. Discriminator output is also possible for NLOS detection. In [75], the authors found that the NLOS signal showed different properties in discriminator output under the vector tracking framework.

Filter: In the filter, the noise in the discriminator output will be reduced, resulting in a more accurate and robust result [76]. Assuming that the estimated code delay is stabilized when the loop is steady, the author in [48] developed the new features using the mean and the standard deviation of the code delay. Features such as the difference between the means, the maximum change, and the difference between the standard deviation are used in the self-organizing map (SOM) for further classification of the signal status transitions.

FEATURE SELECTION

In most cases, a single feature does not provide enough information for the model. Therefore, multiple features should be selected to achieve better performance based on their supplementary. The author investigates how the mixture of features at the measurement level affects the classification accuracy [42]. When the features are properly incorporated, the model performance can be improved. The incorporation can also happen at the receiver level and measurement level. For example, the correlator output can combine with the signal strength and elevation angle to achieve a better performance [28], [34], [43], [48]. Usually, the measurement-level data can be used as features directly. However, the receiver-level data are a bit more complicated in feature extraction. With traditional ML algorithms, the features in the correlator always need to be extracted deliberately with considerable domain expertise. For example, the number of local maxima and distribution of the delay of the maximum correlation are extracted from the raw correlator output [43]. A study has shown that receiver-level features outperform those using measurement-level features with SVM in the signal classification [39]. Finding new available features can be another approach to improve the ML model for the multipath and NLOS problem. The range residual map proposed in [56] has good compatibility with the CNN. Pseudorange rate consistency was proposed by Hsu, which can serve as the NLOS/multipath indicator by differencing pseudorange rate and time-differenced pseudorange [42]. Ziedan combined the mean and standard deviation of code delay errors from DLL to form a new feature set [48].

In summary, for the feature selection, C/N_0, elevation angle, pseudorange residual, and correlator output are the most popular features. Feature integration benefits machine learning. Therefore, exploring the reasonable feature combination based on the geometrical and physical relation of features is a direction to improve the ML performance. Table 2 provides the paper taxonomy according to the features categorization. The features from the correlator and single satellite measurement are most widely used because of their effectiveness in signal status prediction, which is given the most attention in multipath/NLOS mitigation (see the "Improvement Using Machine Learning" section).

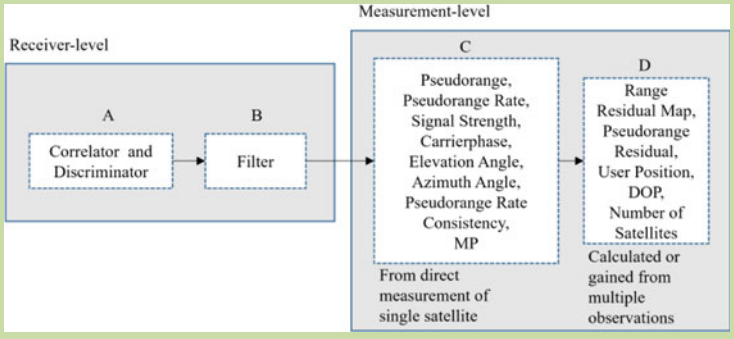
APPROACHES OF ML BASED GNSS MULTIPATH AND NLOS MITIGATION

ALGORITHM SELECTION

This section discusses how researchers utilize deep learning algorithms to better fit the data and obtain better performance. As the popularity of different algorithms shown in Figure 6, SVM and FCNN are currently the most

Table 2.

(Left) Level of the Features Used in ML Based GNSS Multipath Mitigation and Their relationships. A–D are the Indexes of Different boxes. (Right) Summary of the Features Used in the Related Works



Block	Reference
A	[28], [34], [37], [38], [21], [43], [22], [18], [50], [19], [53], [32], [51], [52], [20], [30],
B	[48], [33]
C	[28], [48], [34], [24], [29], [56], [36], [54], [39], [49], [40], [41], [42], [43], [44], [45], [26], [27], [58], [31], [23], [57], [46], [25], [33], [47], [35]
D	[59], [60], [55], [24], [29], [56], [36], [42], [44], [45], [31], [57], [35]

popular methods for conventional ML and deep learning, respectively. In the early stage of dealing with multipath or NLOS problems with machine learning, people are in favor of conventional algorithms like SVM, DT, linear regression, etc. These few years, with the successful deployment of deep learning in other topics such as image processing and neural language processing, deep learning has become popular in GNSS. Deep learning involves techniques like network structures (CNN, RNN, Transformer, etc.), optimization methods, loss functions, and regularization [77]. Figure 5 describes the accumulative frequency of the conventional ML methods and the deep learning (supervised ML only) over the years. The trend is that researchers are leveraging deep learning more frequently.

Conventional ML algorithms have their advantages, including being simple to build, requiring less computing resources, being interpretable, working well on small datasets, and being robust. All these benefits make the conventional method still have a place in machine learning. The increase in the usage of deep learning does not necessarily mean poor performance of the conventional method. Taro et al. have investigated the performance of the classifiers using FCNN and SVM with the correlator output. With the manually selected features, the accuracy of SVM is similar to that of FCNN, or even more, in some of the test cases, the performance of SVM outperforms the performance of FCNN [34]. However, deep learning has a better capacity to extract features than conventional methods. Instead of elaborately selecting the features, deep learning digs and reconstructs the inner feature by layers of the network, for example, dealing with the data from correlator output. The ACF data that feed to SVM should be elaborately engineered to achieve better performance [34], [37], while the data used to train the deep learning model are manipulated more directly—serialized correlation output for FCNN in [34], and even more, 3D correlator output (IQ channels, Doppler, code delay) for CNN in [37]. Besides, a suitable deep learning network structure can accommodate the diversity of data structures. A typical example is the satellite visibility of each epoch of data. The varying visibility will produce a changing number of observations and consequently produce difficulties in applying the traditional ML technique. So, it is normal to resort to other methods (RNN, Transformer, etc.), which will not be constrained; or even more, they could take advantage of this kind of data structure [29], [59]. And for the challenge, besides the well-known one, which is dataset labeling in deep learning, there is another problem in GNSS: how to generate and integrate the features from different sources. The cause of NLOS and multipath lies in the geometry of the receiver, buildings, and satellites. To model the interaction better, extra

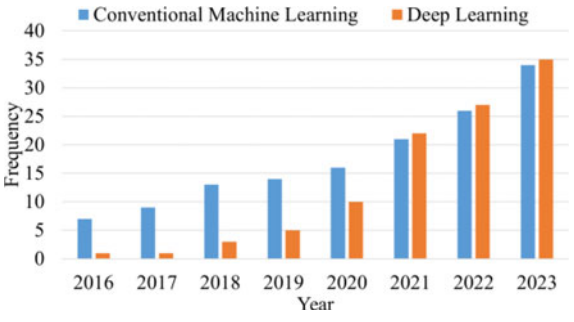


Figure 5.
Accumulative frequency of using conventional ML and deep learning in GNSS multipath/NLOS mitigation across the years. Scope: reviewed paper in Table 1.

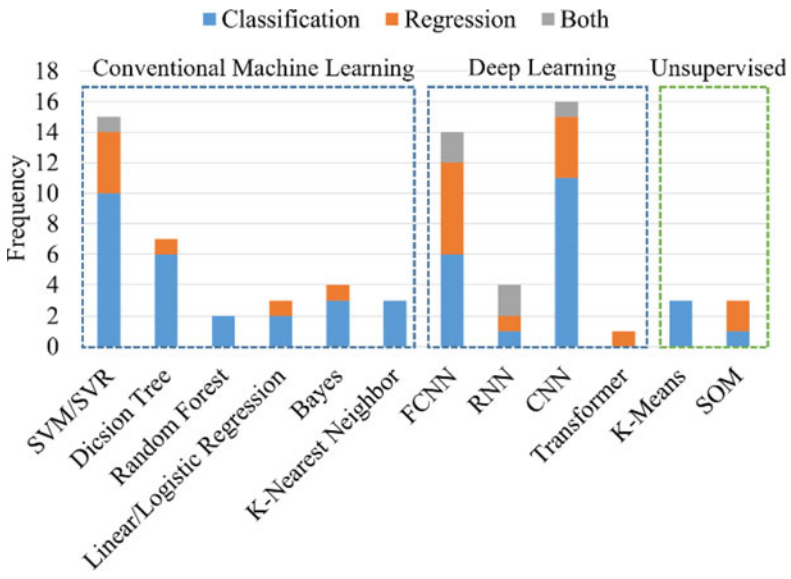


Figure 6.
Frequency of different ML algorithms used in GNSS multipath/NLOS mitigation. The blue box denotes the supervised ML approach, while the green one stands for the unsupervised ML method. Colors blue, orange, and gray represent the algorithms are used for classification, regression, and both tasks, respectively. The classification task outputs the result as a probability (0 to 1), while the regression task can output any arbitrary number.

information like the 3D building model can be essential, which brings challenges to the model design. Therefore, to achieve better performance, the network structures always need to be tailored according to the insight of the multimodal data from the network designer.

For unsupervised algorithms, the work is considerably less than that of the supervised technique in the multipath/NLOS problem. The challenge for traditional unsupervised ML is the model capacity since most of the unsupervised approaches in conventional ML are the clustering method. For deep learning unsupervised methods, the problem lies in how to design the tasks without manual labeling (pretext task). The aim of the pretext task is to learn the features that can be transferred and used in other tasks. For example, the task of regenerating the input in image processing enables the autoencoder to learn to extract features. In natural language processing, the network is able to be aware of the context by learning from the cloze, which does not need manual labeling either [78]. Due to the unlimited unlabeled GNSS data in the real world, the future of unsupervised learning for GNSS is promising yet full of challenges.

The performance of ML is always the aim. Research in [34] examined the performance of FCNN and SVM using correlator output and concluded that FCNN performed marginally better. Authors in [37] used SVM as the benchmark algorithm for their CNN multipath classifier, the result of which outperforms that in SVM, particularly for short coherent integration time (1 ms). Many research studies have highlighted the superior performance of deep learning compared to conventional machine learning. The most significant improvement occurs in regression tasks like error prediction or position correction, whose input can be high dimensional. The potential of deep learning in GNSS still needs lots of effort to be explored, especially for an effective feature representation. This is because studies in other areas [79] indicate that deep learning is more powerful when expanding the dimension and the volume of the data.

IMPROVEMENT USING MACHINE LEARNING

The pseudorange measurement can be described as :

$$\rho_k^i = R_k^i + \delta_k + ATM_k^i + \tau^i + \varepsilon_{k,Mp/NLOS}^i \quad (7)$$

where R_k^i is the geodetic range between the k receiver and the i satellite, δ_k is the receiver clock bias, ATM_k^i is the atmospheric error, τ^i is the satellite clock bias, and $\varepsilon_{k,Mp/NLOS}^i$ is the error due to multipath or NLOS.

In multipath and NLOS cases, researchers focus on $\varepsilon_{k,multipath/NLOS}^i$ most. So, we categorize the output domains of ML in GNSS multipath/NLOS topic as Table 1. Improved pseudorange measurement aims to minimize $\varepsilon_{k,Mp/NLOS}^i$ by applying neural networks inside the receiver correlators to mitigate the effect of multipath. Measurement error prediction tries to estimate $\varepsilon_{k,Mp/NLOS}^i$ in the pseudorange domain. Measurement status prediction predicts the existence of multipath or NLOS phenomenon.

Table 3.

Summarization of Supervised/Unsupervised Conventional Machine and Deep Learning With Respect to GNSS				
Type	**Well-known Algorithms**	**Labeled Data Requirement**	**Benefits**	**Challenges**
Supervised (Conventional)	SVM, DT, RF	Yes	Easy to implement; Require less computation power; Interpretation friendly	Performance is limited by the model capacity; For complicated data, a manual feature extraction is needed
Unsupervised (Conventional)	K-Means, SOM	No	Work well on the small dataset; Robust	Validation of the model; Performance is limited by the model capacity
Supervised (DL)	FCNN, CNN, RNN, transformer	Yes	Good fitting to various data; Does not need manual feature extraction	Multimodal data integration; Well-labeled dataset
Unsupervised (DL)	Auto-encoder, self-supervised learning	No	Good fitting to various data; Features extracted from the unsupervised mission can transfer to other missions	Determine the pretext task (which does not need manual labels) in the GNSS field so that the network can learn to extract the feature for downstream missions

Positioning level information could correct the positioning error or conduct direct positioning.

Both the improved pseudorange measurement and the measurement error prediction can generate a better pseudorange measurement, which can help to improve the positioning accuracy. For the measurement status prediction, the most common status in urban canyons is LOS/NLOS/Multipath. The status has a high correlation with the error model, and it can serve as a good criterion for the strategy in positioning (like weighted least squares, removing NLOS) and 3DMA GNSS shadow matching [36], [80]. In 3DMA GNSS shadow matching, predicting the signal status is important. Because the scoring scheme of shadow matching is based on the prediction of LOS/NLOS probabilities using 1) the measurement and 2) the 3D building model. If the prediction based on the measurement fits the prediction using a 3D model, then this location will obtain a high score; otherwise, it will obtain a low score. Table 1 is the categorization according to the output of the ML of different studies, indicating most of the current studies focus on signal status classification.

SUMMARY

For algorithm selection, the most popular ML techniques are the SVM and FCNN. Traditional ML has a role in GNSS because of its benefits, including easiness of implementation, low computational power, interpretability, ability to work with small datasets, and robustness. At the same time, deep learning has higher model complexity, is good at extracting features, and is more adaptive to different data structures. Based on the discussion in the "Algorithm Selection" section, Table 3 summarizes the benefits and challenges of both supervised and unsupervised conventional/deep learning algorithms with respect to GNSS. For the GNSS multipath/NLOS problem, most studies focus on the measurement status prediction.

DATASET FOR MACHINE LEARNING

This section will first introduce the data collection scenarios by comparison of the ratio of sky visibility. Then, the properties of the ML training dataset will be shown, including the class imbalance status, root-mean-square error of the pseudorange error, and the urbanization level.

DATA COLLECTION SETUP

GNSS measurement is used to feed the ML for multipath/NLOS problems. *UrbanNav* [81], an urban localization kinematic dataset, provides a good starting point for machine learning. Currently, we provide four scenarios data for machine learning, which are Tsim Sha Tsui (TST), Whampoa (WP), Mong Kok (MK), and Kowloon

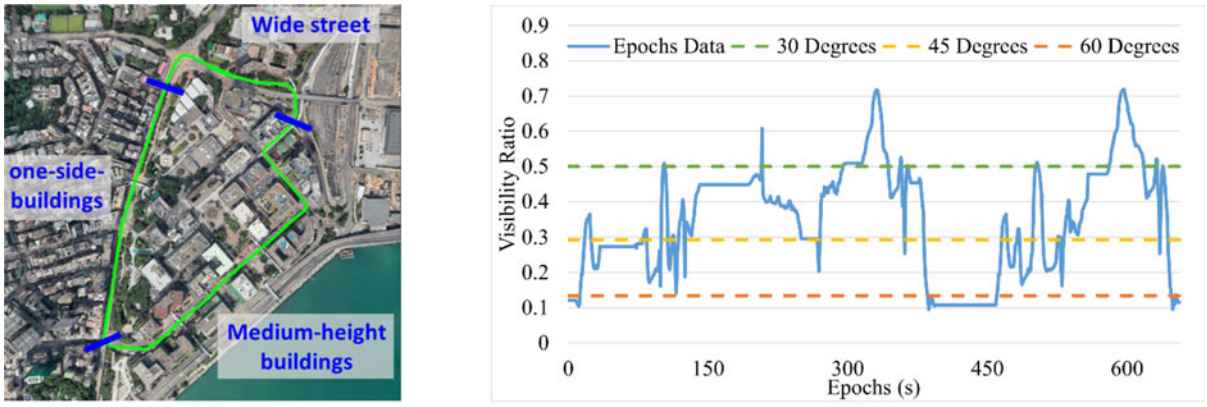


Figure 7.
(a) Ground truth trajectory of TST dataset [81]. (b) Visibility ratio of the ML dataset in TST. 30°, 45°, and 60° equivalent to the situation in which the receiver is surrounded by the buildings with 30°, 45°, and 60° elevation angle, respectively. Same for Figures 8–10.

Bay (KLB), respectively. Details of the data collection settings can be found in [81]. Each scenario is the composition of various types of urban canyons. The urbanization level can be represented by the Skymask [82] to some extent, which indicates the visible area at a certain location. The ratio of sky visibility plotting will be shown to indicate the urbanization level for different areas. Assuming the skyplot is half of the unit spherical surface with radius 1, the ratio of the visibility η is calculated as

$$\eta = 1 - \frac{\sum_{Az=1}^{Az=360} \sin\theta_{bb,Az}}{360} \tag{8}$$

where $\theta_{bb,Az}$ is the elevation angle of the building boundary at each azimuth angle. The ratio is estimated along the epoch data to indicate the urbanization level.

Scenario 1 – TST (training and validation): This urban canyon includes wide streets, one-side buildings, and medium-height buildings [see Figure 7(a)]. The urbanization level of the collected data in TST is the lightest among the three training scenarios, with an average visibility ratio of 0.351. The visibility ratio changes with epochs are shown in Figure 7(b).

Scenario 2 – WP (training and validation): This urban canyon includes wide streets, open-sky, tall buildings, medium-height buildings, and narrow streets [see Figure 8(a)]. WP dataset has better visibility in the starting and ending epochs, and in the middle of the epochs, most of the collection has a visibility ratio lower than 0.5. The average visibility ratio of the WP dataset is 0.322. The visibility ratio changes with epochs are shown in Figure 8(b).

Scenario 3 – MK (training and validation): This urban canyon includes narrow streets, tall buildings, narrow streets, and bridges. MK has the longest and densest data epochs among these three training datasets, with a mean visible area ratio of 0.222. The trajectory and the visibility ratio are shown in Figure 9.

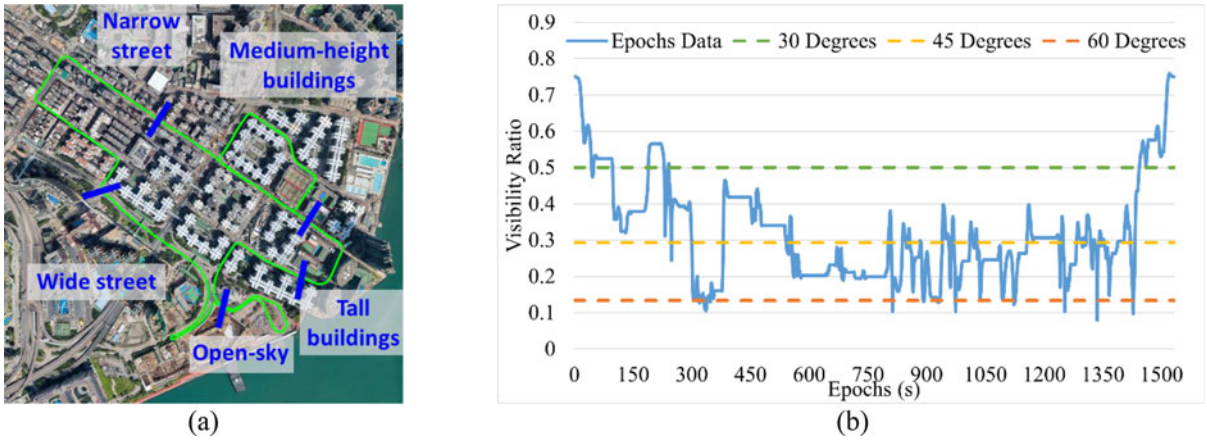


Figure 8.
(a) Ground truth trajectory of WP dataset [81]. (b) Visibility ratio of the ML dataset in WP.

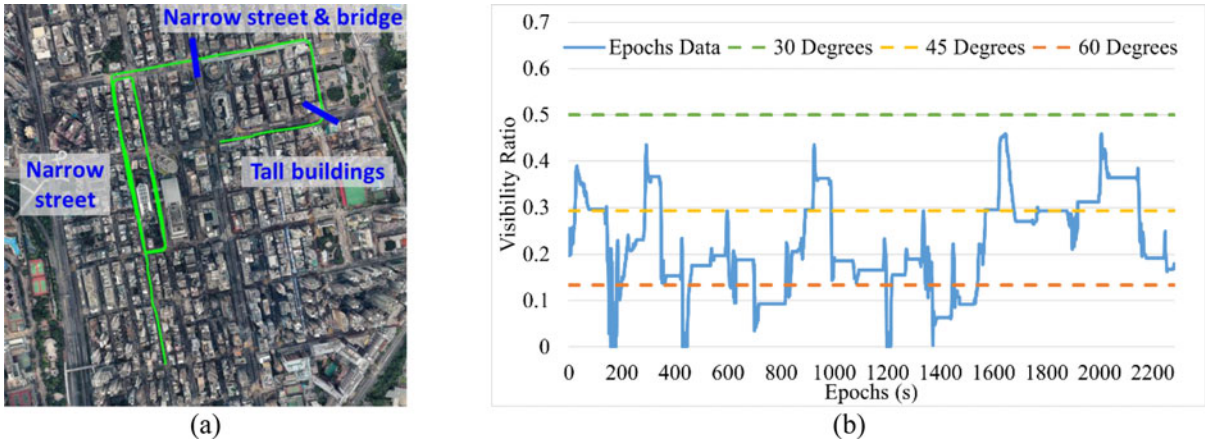


Figure 9.
(a) Ground truth trajectory of MK dataset [81]. (b) Visibility ratio of the ML dataset in MK.

Scenario 4 – KLB (testing): This urban canyon includes low buildings, medium-height buildings, one-side tall buildings, and tall buildings [see Figure 10(a)]. The average visibility ratio of the KLB dataset is 0.338. The visibility ratio changes with epochs are shown in Figure 10(b). The data from this scenario are aimed at testing the model's generalization ability. The route is selected to cover most of the visibility cases and keep the average visibility ratio at a medium level.

INTRODUCTION TO ML DATASET

UrbanNav is an open-source dataset containing multiple localization sensors, including LiDAR, stereo camera, inertial navigation system, and GNSS. The position ground truth is provided by the NovAtel SPANCPT. However, the position label solely is not sufficient to train a model that caters to the multipath/NLOS problem. Training a model related to multipath or NLOS problems requires more specific labels. The dataset for ML includes two parts: 1) training and validation set, and 2) testing set. The first three scenes are for training and validation datasets, and the KLB scenario is for testing purposes. For supervised machine learning, datasets serve three purposes, as in Table 4. Both training, validation, and testing datasets are available in *Kaggle*. For the training and validation set, the measurement and label are provided in a CSV format. Currently, three types of labels are prepared, as shown in Table 5. Only raw measurements will be provided for the testing set.

Both four scenarios use the u-blox F9P receiver data. The training dataset is a CSV file containing the information from RINEX raw measurement and the labeling, in which users can read the L1 and L2 code measurement, carrier-phase measurement, Doppler, C/N_0, and three types of labels. The zero pseudorange error indicates it is used as a master satellite in the pseudorange differencing process. A total of 4,471 epochs of data will be provided based on the ground truth location for training and validation, while 2,281 epochs of data will be used as the testing set (labels will not be provided).

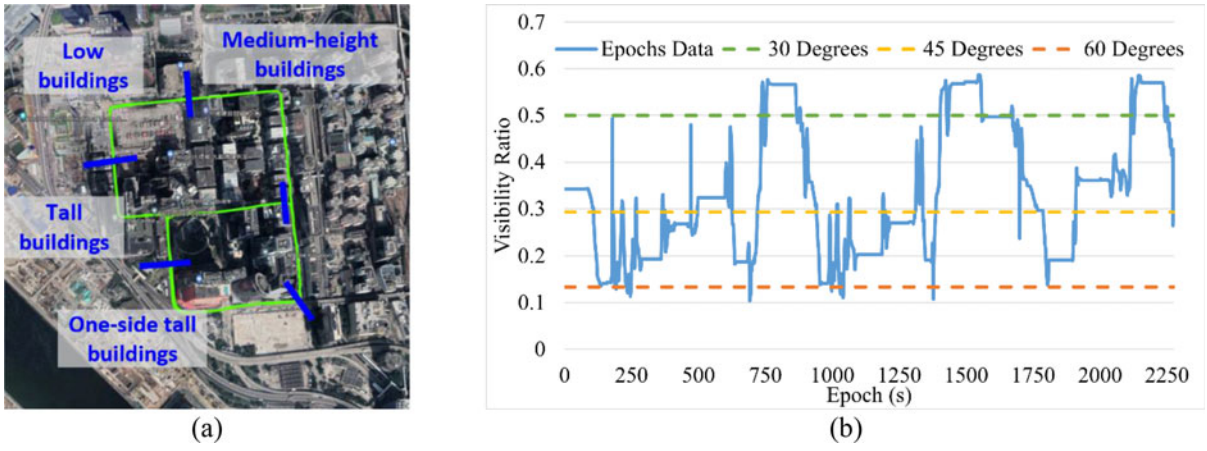


Figure 10.
(a) Ground truth trajectory of KLB dataset. (b) Visibility ratio of the ML dataset in KLB.

Table 4.

Difference of the Training, Validation, and Testing Dataset		
Dataset	**Purpose**	**Used to Update the Model Parameter**
Training set	Train the model	Yes
Validation set	Monitor the model performance to avoid overfitting or underfitting, and help to decide the better model	No
Testing set	Test the model generality performance	No

Table 5.

Ground Truth Labeling Rules	
Labels	**Generation Method**
True position	Real-time kinematic solution from NovAtel SPANCPT optimized by NovAtel Inertial Explore
Pseudorange error*	Differencing receiver pseudorange measurement with the closest known ground station and different satellites.*
LOS/NLOS labeling	If a signal is blocked from the LOS vector according to the 3D city model,** but it can still be tracked, it is regarded as NLOS. Otherwise, it will be labeled as LOS.

**The calculation is based on [39, eq. (17)]. The pseudorange error is assumed mostly contributed by multipath/NLOS.*
***The 3D city model is obtained from the open-source data of the Hong Kong Land Department (*https://data.gov.hk/tc-data/dataset/hk-landsd-openmap-development-hkms-digital-3d-bit00*).*

One of the approaches to evaluate the coverage of the dataset is by looking deeply into the data distribution. Therefore, two views of the data will be provided. The first one is the class imbalance status, which, in our case, is the NLOS/LOS imbalance. Table 6 shows for TST and WP scenarios, the ratio of LOS to NLOS is about 2:1, while for the dense urban area MK, the ratio is close to 1:1. Besides, for LOS and NLOS satellites, the mean absolute error (MAE) of pseudorange is calculated for different scenarios. The second one is the visibility ratio. Figure 11 depicts the distribution of the visibility ratio of the whole training and validation dataset, revealing that most of the visibility ratio of the collected dataset lies between 0.1 to 0.5 (equivalent to being surrounded by buildings with elevation angles between 64° to 30°), which cover lots of urban canyons. A training dataset covering most of the urban canyons can better ensure the model's generalization ability.

BENCHMARK

TASK DESIGN AND EVALUATION METRIC

Performance evaluation of ML consists of two elements: one is the common dataset, and another is the

Table 6.

Details of the Training and Validation Dataset in the Proposed Kaggle Challenge								
Scenario	**Mean Visibility Ratio**	**Data Epoch**	**Sample of NLOS**	**Sample of LOS**	**MAE of Pseudorange in NLOS Satellites (m)**	**MAE of Pseudorange in LOS Satellites (m)**	**MAE of Pseudorange (m)**	**Ordinary Least Square RMSE (m)**
TST	0.351	657	3,467	6,924	16.45	2.92	7.43	14.1
WP	0.322	1,533	9,274	18,478	13.20	1.95	5.71	10.35
MK	0.222	2,281	17,208	18,734	22.76	3.50	12.72	19.88
TOTAL	0.275	4,471	29,949	44,136	19.07	2.76	9.35	16.34

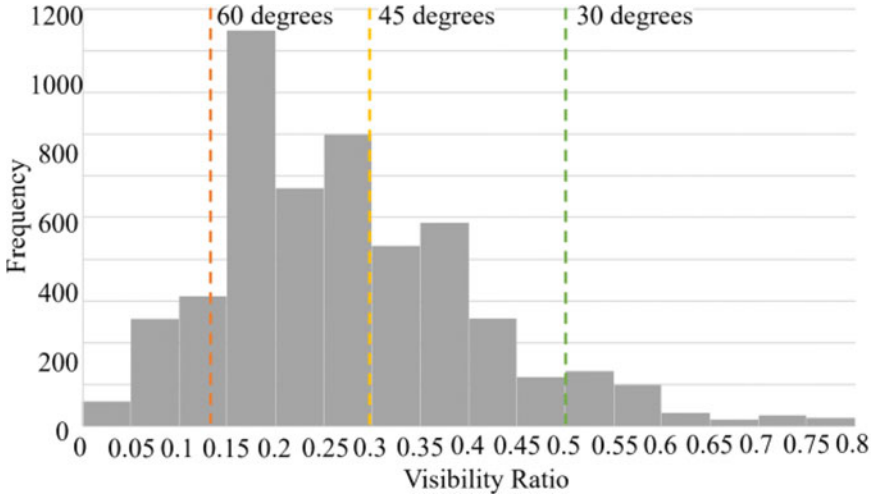


Figure 11.
Visibility ratio distribution of the training and validation dataset. The dashed lines of 60, 45, 30 degrees represent the visibility ratio surrounded by the building with an elevation angle of 60° (0.134), 45° (0.293), and 30° (0.5), respectively.

benchmark. Benchmarks will be established for three tasks, including:

1) LOS/NLOS classification:
   https://www.kaggle.com/c/gnss-classification
2) Pseudorange error prediction:
   https://www.kaggle.com/c/PrError-Prediction
3) Positioning error prediction:
   https://www.kaggle.com/c/gnss-urban-positioning

For the LOS/NLOS classification, the accuracy is used as the metric derived by

$$\text{Accuracy} = \frac{TP + TN}{P + N} \tag{9}$$

where TP is true positive samples, TN is true negative samples, P and N represent positive samples and negative samples, respectively.

For the pseudorange error prediction, the root-mean-square error (RMSE) is used as the metric calculated by

$$\text{Score} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Err_{gt}^{i} - Err_{pre}^{i}\right)^2} \tag{10}$$

where Err_{gt}^{i} is the pseudorange error of the ground truth in the i epoch, Err_{pre}^{i} is the predicted pseudorange error, and n is the number of samples.

For the positioning error prediction, the RMSEs of East and North directions are used as the metric derived by

$$\text{Score} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left(E_{gt}^{i} - E_{pre}^{i}\right)^2 + \left(N_{gt}^{i} - N_{pre}^{i}\right)^2}{2}} \tag{11}$$

where E_{pre}^{i} and N_{pre}^{i} are the prediction location in East and North directions based on a known location P, respectively, and E_{gt}^{i} and N_{gt}^{i} are the ground truth location corresponding to a known position P, in East and North directions, respectively.

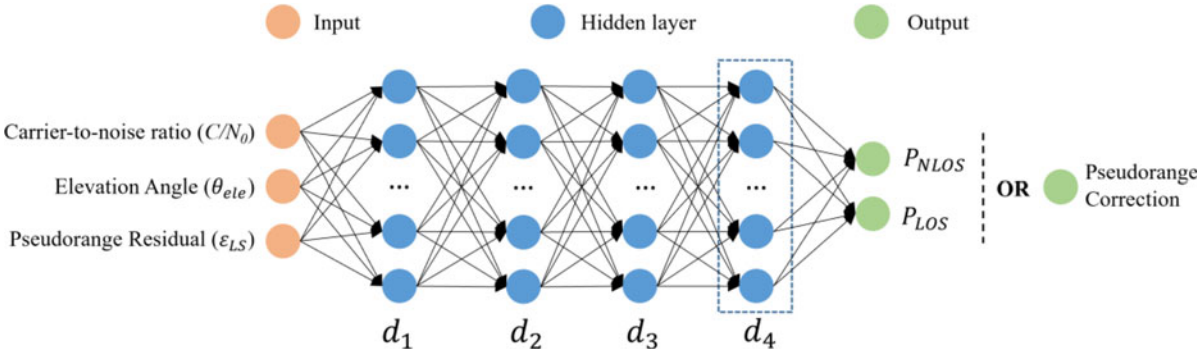


Figure 12.
FCNN structure of the benchmark method, where d_1, d_2, d_3, and d_4 represent the dimensions. P_{NLOS} and P_{LOS} represent the probability of NLOS or LOS. In this work, the network has three or four hidden layers.

MODEL DESIGN AND VALIDATION RESULT

Benchmarks of LOS/NLOS classification and pseudorange error prediction will be provided using the most prevalent network—FCNN, which is expected to represent the performance of a basic model. Five variations of FCNN will be created for each task, and the model with the best validation performance will be used to infer the testing set and serve as the benchmark. For the feature selection, the most popular features, C/N_0, elevation angle, and pseudorange residual, will be used as input features for training. Table 7 describes the training details, including the network structure and hyperparameter. For the positioning error prediction, the conventional least-squares method is used to estimate the RMSE ranging error.

BENCHMARKS OF TASKS

Table 9 shows the benchmark results (testing performance). For classification, model #2, with the highest validation performance, is used to predict the LOS and NLOS. The testing classification result output performs the validation because the KLB scenario is easier to classify. Figure 13 shows the distribution of LOS and NLOS in the training set and testing set, from which the scattering of LOS and NLOS can be observed. For the pseudorange error prediction, model #4 performed best, and it is used to predict the pseudorange error of the testing set. The RMSE of the testing set is 13.01 m. This small value does not mean the model performs better in the testing. This is because the RMSE of the ground truth pseudorange error in the testing dataset is lower, at 18.1 m. For positioning error prediction, the least-squares estimation is general

Table 7.

Training Setting Using FCNN		
	LOS/NLOS Classification	**Pseudorange Error Prediction**
Training Data	Scenarios 1, 2 & 3. Training 80%, Validation 20%	
Input Feature	C/N_0, elevation angle, and pseudorange residual	
Optimizer	Adam, learning rate = 0.01, learning rate decay = 0.99/epoch	
Network Structure	Figure 12	
Network Parameter	Table 8.	
Normalization of Input Feature	Linearly scales features between 0 to 1	Linearly scales features between–1 to 1
Loss Function	Negative log-likelihood loss	MAE
Activation Function	Softmax for the last layer, ReLU for the rest	None for the last layer, ReLU for the rest
Output	NLOS/LOS classification	Pseudorange correction

Table 8.

Variations of FCNN Architecture and Performance. Accuracy Means the Accuracy for LOS/NLOS Classification. RMSE-Ground Truth is the RMSE of the Pseudorange Error of the Ground Truth in the Validation set. RMSE-Validation is the RMSE of the Corrected Pseudorange Error in the Validation set. d is the Layer Dimension in Figure 12

Parameter Set	d_1	d_2	d_3	d_4	**Accuracy Training**	**Accuracy Validation**	**RMSE - Ground Truth (m)**	**RMSE - Validation (m)**
#1	64	128	64	–	0.815	0.758	23.87	15.16
#2	64	256	64	–	0.820	0.771		15.55
#3	64	128	32	–	0.816	0.766		15.35
#4	128	128	64	–	0.814	0.768		15.14
#5	64	128	64	64	0.812	0.761		15.51

Table 9.

Benchmark Results of Three Tasks in the Proposed Kaggle Challenge			
Tasks	**LOS/NLOS Classification**	**Pseudorange Error Prediction**	**Positioning Error Prediction**
Model	#2 parameter set	#4 parameter set	Least-squares estimation
Metric	Accuracy	RMSE (m)	RMSE of east and north (m)
Validation Result	0.771	15.14	17.76
Benchmark	0.853	13.01	21.49

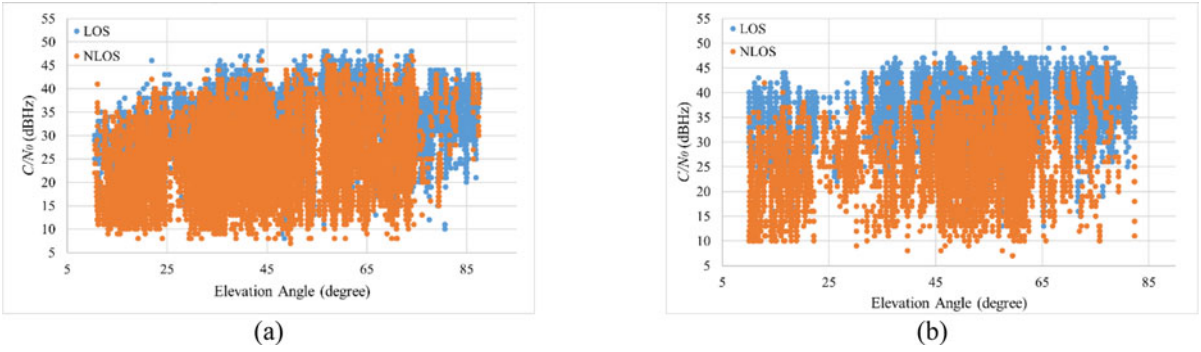


Figure 13.
(a) LOS and NLOS distribution in C/N_0 and elevation angle of the training set. (b) LOS and NLOS distribution in C/N_0 and elevation angle of the testing set.

enough to serve as the benchmark. The higher RMSE of the testing dataset is mainly due to the extreme cases (where the number of satellites is just enough to estimate the positions) and high DOP at some epochs.

CONCLUSION

This article reviewed the works in GNSS multipath/NLOS mitigation using machine learning. The ML algorithms are discussed and combined with their application in this problem. Through the related works, it is clear that ML can help mitigate the GNSS multipath and NLOS effect. Especially with the increasing data volume and diversity, deep learning has a superior ability to deal with diverse data structures and fit the input data compared to conventional machine learning. The reviewed papers are categorized based on the input features and output of machine learning. The most prevalent input features are the direct measurement of a single satellite, such as C/N_0 and elevation angle. For the output of machine learning, most of the work lies in the measurement status prediction. The most used algorithms are SVM for conventional ML and FCNN for deep learning.

Besides, an open-source ML dataset containing GNSS raw measurements and three types of labels are provided. The dataset covers four scenarios, and three will be used as training and validation sets, while the remaining one will serve as the testing set. Based on the training dataset, benchmarks of LOS/NLOS classification, pseudorange error estimation, and positioning error prediction are established using FCNN and the least-squares method. The benchmarks and testing dataset in Kaggle will be a great boost in helping researchers develop the ML algorithm related to multipath/NLOS problems.

ACKNOWLEDGEMENT

This work was supported in part by the University Grants Committee of Hong Kong under the Scheme Research Impact Fund on the project R5009-21 “Reliable Multi-agent Collaborative Global Navigation Satellite System Positioning for Intelligent Transportation Systems,” and in part by the National Natural Science Foundation of China under Grant 62303391.

REFERENCES

[1] P. Groves, “Multipath vs. NLOS signals,” *Inside GNSS*, vol. 8, no. 6, pp. 40–42, 2013.

[2] G. A. McGraw, P. D. Groves, and B. W. Ashman, "Robust positioning in the presence of multipath and NLOS GNSS signals," in *Position, Navigation, and Timing Technologies in the 21st Century: Integrated. Satellite Navigation, Sensor Systems, and Civil Applications*. Hoboken, NJ, USA: Wiley-IEEE Press, 2020, pp. 551–589.

[3] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, vol. 22, no. 1, 2017, Art. no. 7, doi: 10.1007/s10291-017-0667-9.

[4] W. W. Wen, G. Zhang, and L. T. Hsu, "GNSS NLOS exclusion based on dynamic object detection using LiDAR point cloud," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 853–862, Feb. 2021, doi: 10.1109/TITS.2019.2961128.

[5] HSU L.-T., "What are the roles of artificial intelligence and machine learning in GNSS positioning?," *Inside GNSS*, no. 6, pp. 20–22, Dec. 8, 2020. [Online]. Available: https://insidegnss.com/what-are-the-roles-of-artificial-intelligence-and-machine-learning-in-gnss-positioning/

[6] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2018.

[7] A. Siemuri, H. Kuusniemi, M. S. Elmusrati, P. Välisuo, and A. Shamsuzzoha, "Machine learning utilization in GNSS—Use cases, challenges and future applications," in *Proc. Int. Conf. Localization*, Jun. 1–3, 2021, pp. 1–6, doi: 10.1109/ICL-GNSS51451.2021.9452295.

[8] P. Borhani-Darian and Pau Closas, "Deep neural network approach to GNSS signal acquisition," in *Proc. IEEE/ION Position, Location Navig. Symp.*, Portland, OR, USA, 2020, pp. 1214–1223.

[9] C. Guo and Weijuan Tu, "A novel self-learning GNSS/INS integrated navigation method," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021, pp. 168–179, doi: 10.33012/2021.17873.

[10] G. Zhang and L.-T. Hsu, "Intelligent GNSS/INS integrated navigation system for a commercial UAV flight control system," *Aerosp. Sci. Technol.*, vol. 80, pp. 368–380, Sep. 2018, doi: 10.1016/j.ast.2018.07.026.

[11] N. Linty, A. Farasin, A. Favenza, and F. Dovis, "Detection of GNSS ionospheric scintillations based on machine learning decision tree," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 1, pp. 303–317, 2019, doi: 10.1109/TAES.2018.2850385.

[12] C. Savas and F. Dovis, "The impact of different kernel functions on the performance of scintillation detection based on support vector machines," *Sensors*, vol. 19, no. 23, Feb. 2019, Art. no. 5219. [Online]. Available: https://www.mdpi.com/1424-8220/19/23/5219

[13] Y. Jiao, J. J. Hall, and Y. T. Morton, "Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 405–418, Feb. 2017, doi: 10.1109/TAES.2017.2650758.

[14] Y. Jiao, J. J. Hall, and Y. T. Morton, "Performance evaluation of an automatic GPS ionospheric phase scintillation detector using a machine-learning algorithm," *Navigation*, vol. 64, no. 3, pp. 391–402, 2017, doi: 10.1002/navi.188.

[15] Y. Wang, P. Liu, Q. Liu, J. Qian, X. Jin, and R. Ying, "Urban environment recognition based on the GNSS signal characteristics," *Navigation*, vol. 66, pp. 211–225, 2019, doi: 10.1002/navi.280.

[16] H. Gao and P. D. Groves, "Improving environment detection by behavior association for context-adaptive navigation," *Navigation*, vol. 67, no. 1, pp. 43–60, 2020.

[17] 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, 2020, Art. no. 1171, [Online]. Available: https://www.mdpi.com/1424-8220/20/4/1171

[18] M. Musso, G. Gera, A. Cattoni, and C. S. Regazzoni, "Neural networks based approach for fine tracking in satellite navigation systems," in *Proc. 2nd Int. Conf. Recent Adv. Space Technol.*, Jun. 9–11, 2005, pp. 369–373, doi: 10.1109/RAST.2005.1512594.

[19] P. Borhani-Darian, H. Li, P. Wu, and P. Closas, "Deep learning of GNSS acquisition," *Sensors*, vol. 23, no. 3, 2023, Art. no. 1566.

[20] H. Li, P. Borhani-Darian, P. Wu, and P. Closas, "Deep neural network correlators for GNSS multipath mitigation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 1249–1259, Apr. 2023.

[21] M. Orabi, J. Khalife, A. A. Abdallah, Z. M. Kassas, and S. S. Saab, "A machine learning approach for GPS code phase estimation in multipath environments," in *Proc. IEEE/ION Position, Location Navig. Symp.*, Apr. 20–23, 2020, pp. 1224–1229, doi: 10.1109/PLANS46316.2020.9110155.

[22] M. Musso, A.F. Cattoni, and C.S. Regazzoni, "Neural networks based approach for data fusion in multi-frequency navigation receivers," in *Proc. Nat. Tech. Meeting Inst. Navig.*, Monterey, CA, USA, 2006, pp. 319–326.

[23] Y. Lee, P. Wang, and B. Park, "Nonlinear regression-based GNSS multipath dynamic map construction and its application in deep urban areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 5082–5093, May 2023.

[24] R. Sun et al., "Improving GPS code phase positioning accuracy in urban environments using machine learning," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 7065–7078, Apr. 2021, doi: 10.1109/JIOT.2020.3037074.

[25] Y. Lee and B. Park, "Nonlinear regression-based GNSS multipath modelling in deep urban area," *Mathematics*, vol. 10, no. 3, 2022, Art. no. 412.

[26] Q.-H. Phan, S.-L. Tan, I. McLoughlin, and D.-L. Vu, "A unified framework for GPS code and carrier-phase multipath mitigation using support vector regression," *Adv. Artif. Neural Syst.*, vol. 2013, Mar. 2013, Art. no. 240564, doi: 10.1155/2013/240564.

[27] Q.-H. Phan, S.-L. Tan, and I. McLoughlin, "GPS multipath mitigation: A nonlinear regression approach," *GPS Solutions*, vol. 17, no. 3, pp. 371–380, Jul. 2013, doi: 10.1007/s10291-012-0285-5.

[28] M. Maaref, Lionel Garin, and Paul McBurney, "Leveraging machine learning to mitigate multipath in a GNSS pure L5 receiver," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021, pp. 3740–3748, doi: 10.33012/2021.18014.

[29] G. Zhang, P. Xu, H. Xu, and L. T. Hsu, "Prediction on the urban GNSS measurement uncertainty based on deep learning networks with long short-term memory," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20563–20577, Sep. 2021, doi: 10.1109/JSEN.2021.3098006.

[30] S. Cho, H.-W. Seok, and S.-H. Kong, "MPCNet: GNSS multipath error compensation network via multi-task learning," in *Proc. IEEE Intell. Veh. Symp.*, 2023, pp. 1–6.

[31] S. J. Koiloth, D. S. Achanta, and P. R. Koppireddi, "ML-based LOS/NLOS/multipath signal classifiers for GNSS in simulated multipath environment," *Aerosp. Syst.*, pp. 1–14, Nov. 2023.

[32] R. R. Yakkati, B. Pardhasaradhi, J. Zhou, and L. R. Cenkeramaddi, "A machine learning based GNSS signal classification," in *Proc. IEEE Int. Symp. Smart Electron. Syst.*, 2022, pp. 532–535.

[33] N. I. Ziedan, "Optimized position estimation in mobile multipath environments using machine learning," *J. Inst. Navig.*, vol. 70, no. 2, 2023, Art. no. 569.

[34] T. Suzuki and Y. Amano, "NLOS multipath classification of GNSS signal correlation output using machine learning," *Sensors*, vol. 21, no. 7, 2021, Art. no. 2503, [Online]. Available: https://www.mdpi.com/1424-8220/21/7/2503

[35] Y. Wang, J. Xu, R. Yang, and X. Zhan, "GNSS multipath detection based on decision tree algorithm in urban canyons," in *Proc. China Satell. Navig. Conf., Volume II*, 2021, pp. 375–383.

[36] H. Xu, A. Angrisano, S. Gaglione, and L.-T. Hsu, "Machine learning based LOS/NLOS classifier and robust estimator for GNSS shadow matching," *Satell. Navig.*, vol. 1, pp. 1–12, 2020.

[37] E. Munin, A. Blais, and N. Couellan, "Convolutional neural network for multipath detection in GNSS receivers," in *Proc. Int. Conf. Artif. Intell. Data Analytics Air Transp.*, 2020, pp. 1–10, doi: 10.1109/AIDA-AT48540.2020.9049188.

[38] T. Suzuki, K. Kusama, and Y. Amano, "NLOS multipath detection using convolutional neural network," in *Proc. 33rd Int. Tech. Meeting Satell. Division Inst. Navig.*, 2020, pp. 2989–3000, doi: 10.33012/2020.17663.

[39] B. Xu, Q. Jia, Y. Luo, and L.-T. Hsu, "Intelligent GPS L1 LOS/multipath/NLOS classifiers based on correlator-, RINEX- and NMEA-level measurements," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1851. [Online]. Available: https://www.mdpi.com/2072-4292/11/16/1851

[40] 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*, Nov. 21–23, 2018, pp. 94–100, doi: 10.1109/ITMC.2018.8691272.

[41] 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, 2018, Art. no. 2052. [Online]. Available: https://www.mdpi.com/2072-4292/10/12/2052

[42] L. Hsu, "GNSS multipath detection using a machine learning approach," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, Oct. 16–19, 2017, pp. 1–6, doi: 10.1109/ITSC.2017.8317700.

[43] T. Suzuki, Y. Nakano, and Y. Amano, "NLOS multipath detection by using machine learning in urban environments," in *Proc. 30th Int. Tech. Meeting Satell. Division Inst. Navig.*, Portland, OR, USA, 2017, pp. 3958–3967, doi: 10.33012/2017.15291.

[44] 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.*, 2016, pp. 10–14.

[45] R. Yozevitch, B. B. Moshe, and A. Weissman, "A robust GNSS LOS/NLOS signal classifier," *Navigation*, vol. 63, no. 4, pp. 429–442, 2016, doi: 10.1002/navi.166.

[46] R. Zawislak et al., "GNSS multipath detection aided by unsupervised domain adaptation," in *Proc. 35th Int. Tech. Meeting Satell. Division Inst. Navig.*, 2022, pp. 2127–2137.

[47] A. K. Shukla and S. A. Sinha, "Unsupervised machine learning approach for multipath classification of NavIC signals," in *Proc. 35th Int. Tech. Meeting Satell. Division Inst. Navig.*, 2022, pp. 2618–2624.

[48] N. I. Ziedan, "Optimized position estimation in multipath environments using machine learning," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021 pp. 3437–3451, doi: 10.33012/2021.17880.

[49] S. Caner and D. Fabio, "Multipath detection based on K-means clustering," in *Proc. 32nd Int. Tech. Meeting Satell. Division Inst. Navig.*, Miami, FL, USA, 2019, pp. 3801–3811.

[50] A. Guillard, P. Thevenon, C. Milner, and C. Macabiau, "Benefits of CNN-based multipath detection for robust GNSS positioning," in *Proc. 36th Int. Tech. Meeting Satell. Division Inst. Navig.*, 2023, pp. 283–297.

[51] A. Blais, N. Couellan, and E. Munin, "A novel image representation of GNSS correlation for deep learning multipath detection," *Array*, vol. 14, 2022, Art. no. 100167.

[52] C. Jiang et al., "Convolutional neural networks based GNSS signal classification using correlator-level measurements," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 46, pp. 61–66, 2022.

[53] A. Guillard, P. Thevenon, and C. Milner, "Using convolutional neural networks to detect GNSS multipath," *Front. Robot. AI*, vol. 10, 2023, Art. no. 1106439.

[54] S. J. Cho, B. S. Kim, T. S. Kim, and S. Kong, "Enhancing GNSS performance and detection of road crossing in urban area using deep learning," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Oct. 27–30, 2019, pp. 2115–2120, doi: 10.1109/ITSC.2019.8917224.

[55] A. Siemuri, Kannan Selvan, Heidi Kuusniemi, Petri Välisuo, and Mohammed S. Elmusrati, "Improving precision GNSS positioning and navigation accuracy on smartphones using machine learning," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021, pp. 3081–3093, doi: 10.33012/2021.18004.

[56] V. D. Frank, "End game for urban GNSS: Google's use of 3D building models," *Inside GNSS*, no. 2, pp. 42–49, 2021. [Online]. Available: https://insidegnss.com/end-game-for-urban-gnss-googles-use-of-3d-building-models/

[57] P. Xu, G. Zhang, B. Yang, and L.-T. Hsu, "PositionNet: CNN-based GNSS positioning in urban areas with residual maps," *Appl. Soft Comput.*, vol. 148, Sep. 2023, Art. no. 110882, doi: 10.1016/j.asoc.2023.110882.

[58] A. Mohanty and G. Gao, "Learning GNSS positioning corrections for smartphones using graph convolution neural networks," *J. Inst. Navig.*, vol. 70, no. 4, 2023, Art. no. 622.

[59] A. V. Kanhere, S. Gupta, A. Shetty, and G. Gao, "Improving GNSS positioning using neural network-based corrections," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021, pp. 3068–3080, doi: 10.33012/2021.17999.

[60] Y. Tao et al., "Real-time multipath mitigation in multi-GNSS short baseline positioning via CNN-LSTM method," *Math. Problems Eng.*, vol. 2021, Jan. 2021, Art. no. 6573230, doi: 10.1155/2021/6573230.

[61] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[62] J. A. Klobuchar, "Ionospheric time-delay algorithm for single-frequency GPS users," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 3, pp. 325–331, May 1987.

[63] J. Saastamoinen, "Atmospheric correction for the troposphere and stratosphere in radio ranging satellites," in *The Use of Artificial Satellites for Geodesy*. Hoboken, NJ, USA: Wiley, 1972, vol. 15, pp. 247–251.

[64] P. Xie and M. G. Petovello, "Measuring GNSS multipath distributions in urban canyon environments," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 366–377, Feb. 2015, doi: 10.1109/TIM.2014.2342452.

[65] P. D. Groves, "Shadow matching: A new GNSS positioning technique for urban canyons," *J. Navig.*, vol. 64, no. 3, pp. 417–430, 2011, doi: 10.1017/S0373463311000087.

[66] R. Yozevitch, B. B. Moshe, and H. Levy, "Breaking the 1 meter accuracy bound in commercial GNSS devices," in *Proc. IEEE 27th Conv. Elect. Electron. Eng, Isr.*, Nov. 14–17, 2012, pp. 1–5, doi: 10.1109/EEEI.2012.6376966.

[67] A. T. Irish, J. T. Isaacs, F. Quitin, J. P. Hespanha, and U. Madhow, "Belief propagation based localization and mapping using sparsely sampled GNSS SNR measurements," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 1977–1982, doi: 10.1109/ICRA.2014.6907121.

[68] L. Wang, P. D. Groves, and M. K. Ziebart, "GNSS shadow matching: Improving urban positioning accuracy using a 3D City model with optimized visibility scoring scheme," *Navigation*, vol. 60, no. 3, pp. 195–207, 2013, doi: 10.1002/navi.38.

[69] K. Elliott and H. Christopher, *Understanding GPS/GNSS: Principles and Applications*, 3rd ed. Norwood, MA, USA: Artech, 2017.

[70] G. Zhang, B. Xu, H.-F. Ng, and L.-T. Hsu, "GNSS RUMS: GNSS realistic urban multiagent simulator for collaborative positioning research," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 544. [Online]. Available: https://www.mdpi.com/2072-4292/13/4/544

[71] C. O'Driscoll, "Carrier phase and its measurement for GNSS," *InsideGNSS*, no. 4, pp. 18–22, 2010. [Online]. Available: https://www.insidegnss.com/auto/julaug10-solutions.pdf

[72] L. H. Estey and C. M. Meertens, "TEQC: The Multi-purpose toolkit for GPS/GLONASS data," *GPS Solutions*, vol. 3, no. 1, pp. 42–49, Jul. 1999, doi: 10.1007/PL00012778.

[73] H. Gao and P. D. Groves, "Environmental context detection for adaptive navigation using GNSS measurements from a smartphone," *Navigation*, vol. 65, no. 1, pp. 99–116, 2018, doi: 10.1002/navi.221.

[74] H. Xu, L.-T. Hsu, D. Lu, and B. Cai, "Sky visibility estimation based on GNSS satellite visibility: An approach of GNSS-based context awareness," *GPS Solutions*, vol. 24, no. 2, Mar. 2020, Art. no. 59, doi: 10.1007/s10291-020-0973-5.

[75] B. Xu, Q. Jia, and L.-T. Hsu, "Vector tracking loop-based GNSS NLOS detection and correction: Algorithm design and performance analysis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4604–4619, Jul. 2020, doi: 10.1109/TIM.2019.2950578.

[76] Y. J. Morton, F. van Diggelen, J. J. Spilker Jr, B. W. Parkinson, S. Lo, and G. Gao, *Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications*. New York, NY, USA: Wiley, 2021.

[77] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[78] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

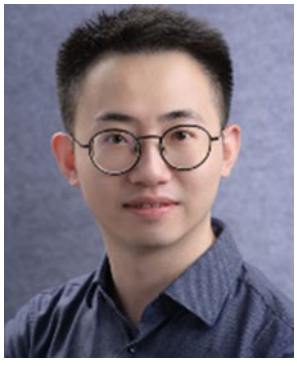
[79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[80] L. Wang, P. D. Groves, and M. K. Ziebart, "Smartphone shadow matching for better cross-street GNSS positioning in urban environments," *J. Navig.*, vol. 68, no. 3, pp. 411–433, 2015, doi: 10.1017/S0373463314000836.

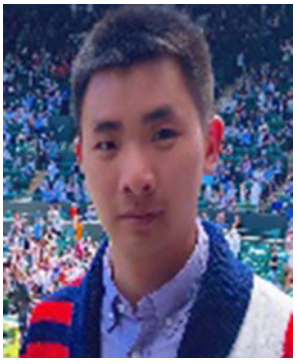
[81] L.-T. Hsu et al., "UrbanNav:An open-sourced multisensory dataset for benchmarking positioning algorithms designed for urban areas," in *Proc. 34th Int. Tech. Meeting Satell. Division Inst. Navig.*, St. Louis, MO, USA, 2021, pp. 226–256, doi: 10.33012/2021.17895.

[82] H.-F. Ng, G. Zhang, and L.-T. Hsu, "A computation effective range-based 3D mapping aided GNSS with NLOS correction method," *J. Navig.*, vol. 73, no. 6, pp. 1202–1222, 2020, doi: 10.1017/S037346332000003X.

Penghui Xu received the B.S. degree from South China Agricultural University, Guangzhou, China, in 2015 and the M.Sc. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2017. He is currently working toward the Ph.D. degree with The Hong Kong Polytechnic University. After that, he mainly works in machine learning algorithm development. His research interests include machine learning, GNSS urban localization, and multisensor integration for positioning.



Guohao Zhang received the bachelor's degree in mechanical engineering and automation from University of Science and Technology Beijing, Beijing, China, the master's degree in mechanical engineering, and the Ph.D. degree in aeronautical and aviation engineering from the Hong Kong Polytechnic University, Hong Kong. He is currently a Research Assistant Professor with the Department of Aeronautical and Aviation Engineering of the same university. His research interests include GNSS urban positioning, collaborative positioning, machine learning aided GNSS, and signal propagation modeling.



Bo Yang received the D.Phil. degree from the University of Oxford, Oxford, U.K., in 2020. He is an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. His research interests lie in machine learning, computer vision and robotics.

Li-Ta Hsu (Member, IEEE) received the B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Taiwan, in 2007 and 2013, respectively. He is an Associate Professor with the Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong. He is Limin Endowed Young Scholar in Aerospace Navigation. He was a Visiting Researcher with the Faculty of Engineering, University College London, London, U.K., and Tokyo University of Marine Science and Technology, in 2012 and 2013, respectively. He won a Student Paper Award and two Best Presentation Awards from the Institute of Navigation in 2013. He was selected as a Japan Society for the Promotion of Sciences Postdoctoral Fellow with the Institute of Industrial Science, The University of Tokyo and worked from 2014 to 2016. He is an Associate Fellow in the Royal Institute of Navigation. He is currently member of ION and a member of the editorial board and reviewer in professional journals related to GNSS.