High-accuracy AoA-based Localization using Hierarchical ML Classifiers in Outdoor Environments

Bac Trinh-Nguyen*†, Sara Berri*, Sin G. Teo†, Tram Truong-Huu‡†, Asernia Chorti*§
*ETIS UMR 8051 / CY Paris University, ENSEA, CNRS, FR
†Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore
‡Singapore Institute of Technology (SIT), Singapore
§Barkhausen Institut gGmbH, Dresden, DE
Email: {trinh-nguyen.bac, sara.berri, arsenia.chorti}@ensea.fr, teosg@i2r.a-star.edu.sg, truonghuu.tram@singaporetech.edu.sg

***Abstract*—Accurate and reliable localization is a key requirement for 6G network operations, but it can be particularly challenging in outdoor environments. In this paper, we propose a machine learning (ML)–based localization framework that leverages angle of arrival (AoA) as a feature extracted from channel state information (CSI). The proposed approach employs high-resolution AoA estimation algorithms, including multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT), which feed a hierarchical, two-stage classifier to identify specific trajectories (hereafter referred to as tracks) in a given outdoor environment. The first stage of the classifier is a binary line-of-sight (LoS) / non-line-of-sight (NLoS) classifier, followed by region-specific multi-class classifiers for fine-grained identification of the specific LoS or NLoS tracks. We evaluate our approach using a real-world massive multiple-input multiple-output (mMIMO) orthogonal frequency division multiplexing (OFDM) outdoor CSI dataset collected at the Nokia campus in Stuttgart, Germany. Experimental results show that *i)* the LoS / NLoS identification accuracy can reach 100%, and *ii)* the proposed two-stage approach significantly outperforms a single-stage multi-class baseline, achieving accuracy over 98% in LoS regions and 95% in NLoS regions. These findings demonstrate the potential of combining AoA with ML for robust localization in outdoor mMIMO propagation environments.**

***Index Terms*—RF localization, 6G, angle of arrival, machine learning, hierarchical machine learning models, physical layer authentication.**

I. Introduction

As we step into the age of sixth-generation (6G) networks, accuracy, reliability, and robust security have become critical requirements for a wide range of location-based services, spanning both indoor and outdoor scenarios such as smart factories, intelligent autonomous transportation, and augmented / virtual reality. Localization is also becoming a crucial component of physical layer authentication schemes [1].

Unlike traditional positioning methods that rely on GPS, which is often unavailable or unreliable in indoor, dense urban, or underground environments, radio frequency (RF)-based localization systems can operate effectively in both indoor and outdoor settings by leveraging signal propagation laws. Various wireless positioning techniques have been proposed, differing with respect to the signal features used in different algorithms [2], [3]. Received signal strength-based (RSS) methods are widely adopted due to their ease of implementation and low hardware requirements. However, they are sensitive to dynamics in the wireless environment and are prone to small scale fading. Time-based approaches, such as time of flight (ToF), time difference of arrival (TDoA), and return time of flight (RToF), do not require fingerprinting and can provide high localization accuracy. However, they need high accuracy time synchronization between transmitters and receivers, enabling improved robustness against multipath effects, noise and interference.

In addition to methods such as time-based and signal strength-based localization, angle of arrival (AoA)-based methods estimate the direction of incoming signals and can achieve high localization accuracy, particularly under line-of-sight (LoS) conditions. We note that in recent works, AoA in multiple input multiple output (MIMO) digital array systems has been proven to be a robust feature for location-based physical layer authentication (PLA) [4]. In this work, it was also demonstrated that AoA provides resistance against impersonation attacks and can also serve as a valuable feature for training robust ML-based PLA systems. In view of these results, in this paper, we use AoA as a key feature for robust localization against impersonation attacks.

Furthermore, in order to address issues related to localization accuracy in NLoS conditions, we investigate the use of state-of-the-art ML-based classifiers. In recent years, a wide range of ML techniques have been applied to RF-based localization tasks due to their ability to address the limitations of traditional methods, such as sensitivity to environmental changes, hardware variability, and synchronization issues [5]. k-nearest neighbors (KNN) model presented in [6] achieved high accuracy for CSI-based indoor localization. Yang *et al.* [7] provided a comprehensive survey on ML-based indoor localization, where supervised methods such as support vector machine (SVM), KNN, and neural networks are used for offline fingerprinting. MobIntel [8] demonstrates the effectiveness of ML methods for passive outdoor localization using RSSI data. Multipath-based CSI fingerprinting localization has been proposed in [9], achieving meter-level accuracy for outdoor localization. These studies demonstrate the effectiveness of ML-based approaches in various localization scenarios.

In this paper, we develop a hierarchical ML-based framework for outdoor localization in mMIMO OFDM systems, validated on a real outdoor dataset provided by Nokia [10].

More specifically, our main contributions include:

- Extensive experiments to compare MUSIC- and ESPRIT-based AoA estimation in a real-world outdoor mMIMO OFDM dataset in terms of accuracy and computational efficiency, targeting low-latency and real-time application requirements.
- A robust AoA-based binary classification approach to distinguish between LoS and NLoS regions, achieving 100% accuracy.
- A robust region-specific localization approach based on AoA, leveraging hierarchical ML classifiers to classify specific trajectories within LoS and NLoS regions separately, enhancing localization accuracy.
- A comprehensive evaluation of six base ML models, including logistic regression (LR), KNN, random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), as well as stacking ensemble model [11], providing insights into accuracy and computational efficiency.

The rest of the paper is organized as follows: Section II presents our proposed approach for a two-stage hierarchical classifier, along with a background on MUSIC and ESPRIT algorithms for feature extraction. Section III presents the outdoor mMIMO OFDM dataset and the analysis of the results of our models against the baseline single-stage classifier. Finally, we conclude the paper in Section IV and discuss possible perspectives for future work.

II. Proposed Hierarchical Two-Stage Classification for Outdoor Localization

In this paper, we focus on outdoor AoA-based localization in an mMIMO OFDM setting, with potential for application to both FR1 and FR2 frequency bands. In this work, we are interested in identifying different users over distinctive trajectories (tracks). The proposed approach is presented in Section II-A, while the feature distillation using the MUSIC and ESPRIT algorithms is presented in Section II-B.

A. Proposed Approach

To address the problem of CSI-based localization in complex outdoor environments while ensuring robustness against impersonation attacks [4], we propose an ML-based framework that leverages AoA features as robust inputs for track identification under both LoS and NLoS conditions. Our approach is designed as a two-stage classifier with the training and inference processes illustrated in Fig. 1:

1) Training phase - Model development and evaluation: The CSI data are first preprocessed and segmented using a sliding window technique. Subsequently, the AoA features are extracted using the MUSIC and the ESPRIT algorithms. Then, a hierarchical classifier is trained, consisting of a binary LoS / NLoS classifier, followed by two specific multi-class classifiers: one for LoS tracks and one for NLoS tracks. Finally, the ML-based models are trained and evaluated to select the best-performing and most robust among them.

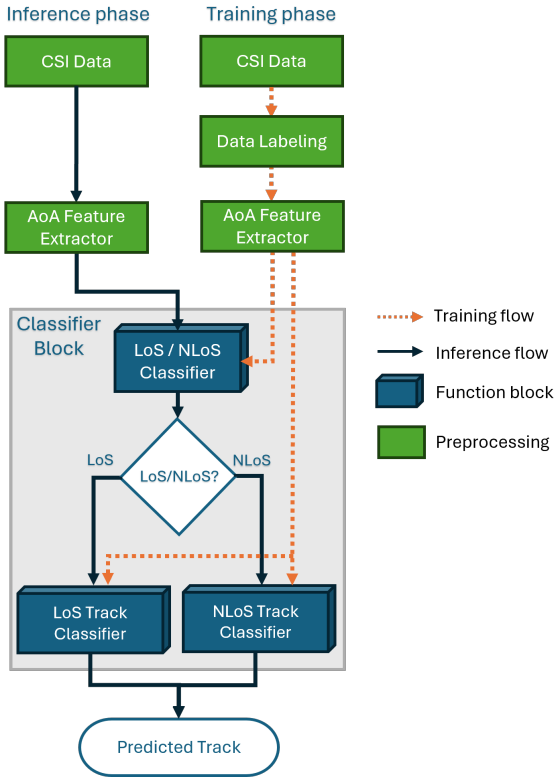

Fig. 1: The proposed hierarchical two-stage classifiers.

2) Inference phase - Realtime track classification: The incoming CSI samples are processed in real-time to extract the AoA features. To this end, first, the LoS / NLoS classifier determines whether the AoA (and consequently the corresponding CSI sample) originates from a LoS or NLoS region. Based on the output of the LoS / NLoS classifier in the first stage, the following actions are taken:

- If classified as LoS, the sample is forwarded to a LoS track classifier, which performs multi-class classification among tracks located in LoS regions.
- If classified as NLoS, the sample is forwarded to an NLoS track classifier, trained for NLoS conditions.

To develop a robust classifier, the training phase involves training and evaluating multiple ML algorithms, including: LR, KNN, RF, GBM, XGBoost, LightGBM, and a stacking ensemble model (combining the top n best performing ML models, $n = 2, \ldots, 6$). The final classification decision of the stacking ensemble model is determined by a meta-learner that combines the outputs of the individual base models. Based on the model performance metrics (accuracy, precision, recall, F1-score), the most accurate and robust model is selected for deployment in the inference phase, demonstrated in Fig. 2.

B. Feature Extraction

In our work, we adopted two well-known AoA estimation algorithms to estimate the azimuth angle of the CSI signal. MUSIC [12] and ESPRIT [13], [14] are eigenvalue decomposition-based, high-resolution subspace algorithms commonly used to estimate AoA of emitted signals in antenna array systems.

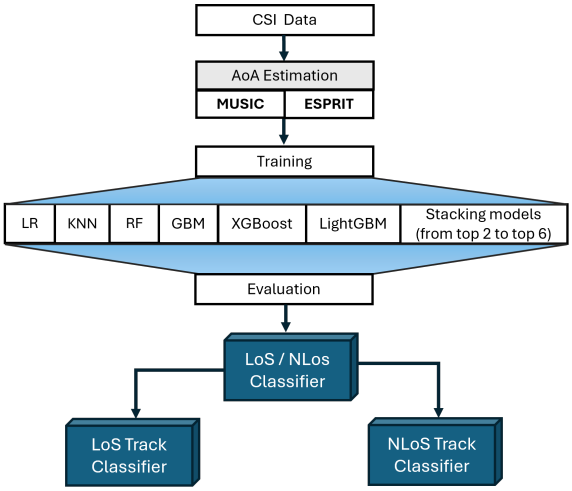

Fig. 2: ML models training and selection workflow.

As the primary feature used in our classification framework, AoA provides spatial information about the signal's propagation path. AoA features were selected because they are robust against both environmental variations and impersonation attacks [4]. We briefly describe MUSIC and ESPRIT below.

1) MUSIC: MUSIC first computes the covariance matrix of the received signal vectors and performs eigenvalue decomposition to separate it into signal and noise subspaces. Exploiting the orthogonality between these two subspaces, the MUSIC spectrum is constructed using the array steering vector and the noise subspace and is formulated as:

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)}, \tag{1}$$

where $\mathbf{a}(\theta)$ is the steering vector corresponding to the angle θ, $\mathbf{E}_n$ is the matrix of eigenvectors spanning the noise subspace, and $(\cdot)^H$ denotes the conjugate transpose operator. The estimated AoAs correspond to the peaks in the calculated MUSIC spectrum $P_{\text{MUSIC}}(\theta)$.

2) ESPRIT: Unlike MUSIC, ESPRIT does not require spectral search to estimate the AoA. Instead, the algorithm determines the rotation operator Φ by exploiting the rotational invariance property between two overlapping subarrays. After computing the covariance matrix and performing eigenvalue decomposition, the signal subspace is extracted and partitioned into two subarrays E_1 and E_2. By forming the matrix $\mathbf{C} = \begin{bmatrix} E_1^H \\ E_2^H \end{bmatrix} [E_1 \quad E_2]$ and performing its eigen-decomposition, the rotation matrix can be estimated as $\Phi = -E_{12}E_{22}^{-1}$. The angles of arrival θ_i are estimated from the eigenvalues λ_i of Φ as:

$$\theta_i = \sin^{-1}\left(\frac{\arg(\lambda_i)}{k\Delta}\right), \tag{2}$$

where $k = \frac{2\pi}{\lambda}$ is the wave number and Δ is the spacing between subarrays in wavelengths.

In the following section, we showcase the performance enhancement of the proposed localization approach on a real mMIMO OFDM outdoor dataset.

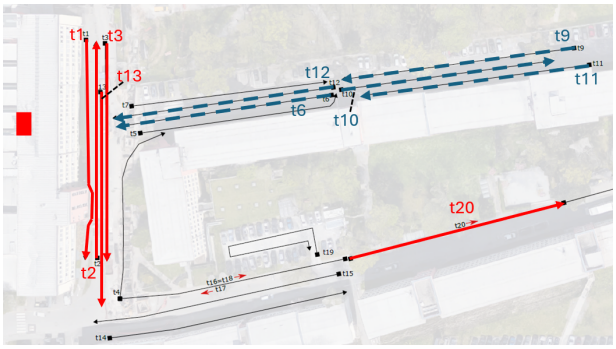

Fig. 3: Nokia campus in Stuttgart, Germany. The red rectangle denotes the mMIMO antenna array mounted on top of a building, while the lines with arrows represent the trajectories (tracks) and their respective directions. Red solid lines indicate NLoS tracks, while blue dashed lines represent LoS tracks.

TABLE I: Data shapes of the original CSI and estimated AoAs (MUSIC) for LoS tracks (6, 9, 10, 11, 12) and NLoS tracks (1, 2, 3, 13, 20). The dataset generated using ESPRIT follows the same format.

Tracks	CSI data shape	AoA data shape by window size		
		500	1000	2000
1	(60, 50, 120000)	(479, 200)	(239, 200)	(119, 200)
2	(60, 50, 120000)	(479, 200)	(239, 200)	(119, 200)
3	(60, 50, 122000)	(487, 200)	(243, 200)	(121, 200)
13	(60, 50, 122000)	(487, 200)	(243, 200)	(121, 200)
20	(60, 50, 118000)	(471, 200)	(235, 200)	(117, 200)
6	(60, 50, 96000)	(383, 200)	(191, 200)	(95, 200)
9	(60, 50, 120000)	(479, 200)	(239, 200)	(119, 200)
10	(60, 50, 120000)	(479, 200)	(239, 200)	(119, 200)
11	(60, 50, 120000)	(479, 200)	(239, 200)	(119, 200)
12	(60, 50, 92000)	(367, 200)	(183, 200)	(91, 200)

III. Experimental Results

In this section, we present experimental results to evaluate the proposed method using a real-world 64-antenna mMIMO 50-subcarrier OFDM outdoor dataset at FR1 (2.18 GHz) collected at Nokia campus, in Stuttgart, Germany [10]. All experiments were carried out on a server equipped with an Intel Xeon X7550 CPU (4 sockets, 64 threads, 2.00 GHz), 188 GiB of DDR3 RAM and running Ubuntu 22.04 LTS. Additionally, we used the Python multiprocessing module to distribute tasks in parallel across all 64 logical cores.

A. Dataset Description

The dataset consists of uplink CSI measurements from a mMIMO digital antenna array consisting of 64 elements, which was mounted on top of a building at a height of 20 meters with a mechanical downtilt of 10 degrees. The array was arranged in 4 rows $\times$ 16 columns of single-polarized patch antennas. The horizontal spacing between antennas was $\lambda/2$, and the vertical spacing was λ, at a central frequency of 2.18 GHz. 50 OFDM subcarriers over a 10 MHz bandwidth for each antenna and pilot bursts were transmitted every 0.5 ms. The receiver used for experiments was a monopole antenna placed on a cart at an approximate height of 2 meters. This cart moved

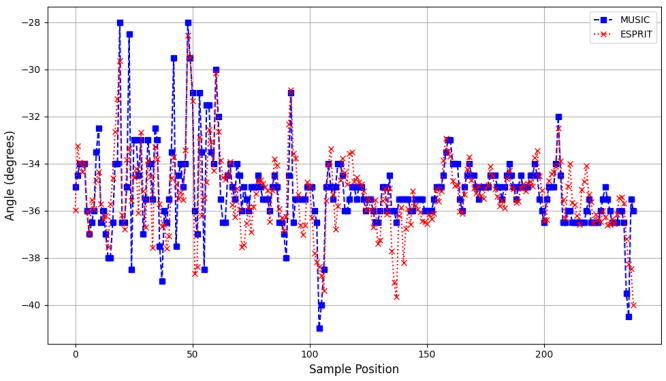

Fig. 4: Example of AoA estimation results for track 11, using antenna row 0 at subcarrier 10, with a window size of 1000.

TABLE II: Performance of MUSIC and ESPRIT in AoA estimation across track 11.

Algorithm	AoA estimation (°)		Processing time (s)	
	Mean	Std Dev	Total	Mean
MUSIC	-35.17	1.71	7.1672	0.0300
ESPRIT	-35.30	1.63	0.2777	0.0012

along predefined trajectories (tracks) with an average speed of 3.5 km/h, and the spatial sampling interval was 0.5 mm. The dataset size for each individual track is listed in Table I.

Due to the presence of obstacles such as buildings and trees, this area included both LoS and NLoS regions. In this work, we focused on five tracks located in the LoS region (track 6, track 9, track 10, track 11, track 12) and five tracks located in the NLoS region (track 1, track 2, track 3, track 13, track 20), depicted in the campus map in Fig. 3.

B. AoA Feature Extraction

We used the CSI dataset described above to distill vectors of AoAs over different rows of the antenna array and different subcarriers. Let $\mathbf{H} \in \mathbb{C}^{N \times K \times T}$ denote the CSI matrix, where $N = 64$ is the total number of antennas, arranged as 4 rows and 16 columns, $K = 50$ is the number of subcarriers, and T is the total number of consecutive measurement locations on each track, spaced at intervals of 0.5 mm.

As presented in Section II-B, MUSIC and ESPRIT algorithms are used for each individual azimuth AoA estimation. Given that the precision of both algorithms depends on the size of the input vectors, we have investigated the impact of this on the classification performance by using a vector $W \in \{500, 1000, 2000\}$ of consecutive measurements for AoA estimation. As a result, CSI measured over ranges of $\{0.25, 0.5, 1\}$ meters is used for each AoA estimate. Furthermore, to augment the AoA dataset, we slided the windows W over each of the tracks. Each window overlaps by 50% ($shift\_ratio$), ensuring a smooth transition between consecutive data frames on tracks and increasing the number of data samples for training ML models. Consequently, three different configurations of input datasets were produced as described in Table I.

Furthermore, each of the four rows of the uniform linear array (ULA) was used independently to estimate four AoAs. Additionally, we utilized the 50 OFDM subcarriers to obtain 50 AoA estimates, resulting in a vector of $4 \times 50 = 200$ AoA values for each individual window of W values on each track. For each row $r \in \{0, 1, 2, 3\}$, $k \in \{0, \ldots, K-1\}$, and measurement window index $l \in \{0, \ldots, L-1\}$, the input matrix to AoA estimation algorithm was defined as:

$$\mathbf{x}^{(l)} = [\mathbf{H}_{n,\,k,\,t}]\,, \quad \begin{matrix} n \in \{rM, \ldots, (r+1)M - 1\}, \\ t \in \{l \cdot S, \ldots, l \cdot S + W - 1\}, \end{matrix} \tag{3}$$

where W is the window size, S is the step size between consecutive windows, and L is the total number of overlapping windows. L is defined by Eq. (4) given below

$$L = \left\lfloor \frac{N - W}{S} \right\rfloor + 1, \tag{4}$$

where N is the number of measurements in the whole track, W is the window length, $S = W \cdot shift\_ratio$ (with $shift\_ratio = 0.5$) is the step size between windows.

Fig. 4 illustrates an example of AoA estimation results for track 11, using antenna row 0 at subcarrier 10, with a window size of 1000. It can be observed that the mean estimated AoA values of both algorithms are nearly identical, $-35.17°$ and $-35.30°$ for MUSIC and ESPRIT, respectively, with standard deviations of approximately $1.71°$ and $1.63°$. Besides, a comparison of processing times between the MUSIC and ESPRIT algorithms shows a significant difference, as shown in Table II. In this result, we can see that the total processing time across the entire track using MUSIC is 7.1672 seconds, approximately 26 times longer than that of ESPRIT (0.2777 seconds). The average processing time per estimate is 30 milliseconds for MUSIC and 1.2 milliseconds for ESPRIT.

C. Classification Performance

We carried out experiments that included the training and evaluation of three classifiers: LoS / NLoS classifier, LoS track classifier, and NLoS track classifier. In addition, we performed experiments using a single multi-class classifier to distinguish all tracks to compare their performance with our proposed hierarchical architecture, as detailed in Section II-A. During the experimental process, we collected evaluation metrics such as accuracy, precision, recall, and F1-score to assess and compare the performance of these models. Additionally, we employed 5-fold cross-validation, where the dataset is split into five subsets: one fold is used for validation, while the remaining folds are used for training. The results were averaged and used in the evaluation to ensure statistical reliability.

1) Performance of LoS / NLoS Classifier: The LoS / NLoS classifier is a binary classifier that distinguishes between LoS and NLoS tracks. Before training, we divided the tracks into two groups, LoS and NLoS, and assigned labels. The classifier was then trained using these labeled samples and achieved *perfect classification with accuracy equal to* 1. This result is extremely important because, as previously discussed, the LoS / NLoS classifier plays a critical role in our approach as it is positioned at the first stage of the hierarchical architecture. Any misclassification at this stage can misroute samples to

Approach	W	LR		KNN		RF		GBM		XGBoost		LightGBM		stacking top 2		stacking top 3		stacking top 4		stacking top 5		stacking top 6	
		MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT	MUSIC	ESPRIT
NLOS	2000	50.1	60.4	72.4	90.3	87.1	87.4	71.2	68.3	75	73.7	74.2	75.4	83.7	93.5	85.6	93.5	86.8	93.5	89.4	93.5	89.4	95
	1000	33.9	40.9	48.4	72.9	78.6	78.6	60.7	63.7	71.4	74.3	74.9	77.1	80.3	81.4	80.9	84.7	80	89.2	80.4	89.8	81.5	89.7
	500	34.6	36.3	30.4	45	65.8	64.3	57.1	51.1	67	64.2	70.5	66.5	71.8	69.5	73.5	70.5	74.3	70.5	75	73.6	75.4	75.7
LOS	2000	71.1	86.7	76.6	85.1	94.3	92.6	89.5	83.4	89.1	84.7	86	82.3	94.3	93.9	93.5	94.1	93.5	95	97.6	95.4	98	95.4
	1000	48.3	59	65.8	75.7	87.4	76.8	77.3	69.6	81.2	72.5	83.2	74.6	88.1	88	87.8	88.5	88.1	89.9	89.7	89.6	90.1	89.9
	500	34.9	42.6	44.4	56.2	71.1	63.5	70.6	61.7	72.8	64.6	77.4	66.8	78.1	67.5	78.2	68	78.3	68	79.1	70.1	80.1	71.4
LOS+NLOS (averaged)	2000	60.6	73.55	74.5	87.7	90.7	90	80.35	75.85	82.05	79.2	80.1	78.85	89	93.7	89.55	93.8	90.15	94.25	93.5	94.45	93.7	95.2
	1000	41.1	49.95	57.1	74.3	83	77.7	69	66.65	76.3	73.4	79.05	75.85	84.2	84.7	84.35	86.6	84.05	89.55	85.05	89.7	85.8	89.8
	500	34.75	39.45	37.4	50.6	68.45	63.9	63.85	56.4	69.9	64.4	73.95	66.65	74.95	68.5	75.85	69.25	76.3	69.25	77.05	71.85	77.75	73.55
LOS+NLOS (single)	2000	60.2	69	74.7	87.8	90.4	88.8	80.1	74.5	81.2	78.6	79	74.5	88.5	94.3	89	94.3	88.3	94.6	93.5	94.5	94.2	94.9
	1000	42.1	52.9	57	74.6	81.7	77.2	68.1	65.6	75.2	71.8	77.4	72.5	82.1	87.8	82.2	89.2	82.7	89.5	84.2	89.6	85	89.3
	500	35.4	40.8	37.7	51.5	68.6	64.6	61.8	56.5	69.4	63.9	72.8	66	73.8	67.9	74.7	68.7	74.9	69.2	75.4	72.6	77.2	74.2

Fig. 5: Accuracy (%) results for the hierarchical approach (including separate LoS and NLoS results, their average) and the single multi-class classification approach. The color scale ranges from red (low accuracy) to green (high accuracy).

the wrong classifier, affecting final prediction accuracy. Furthermore, we note in passing that this result has a wider potential interest for other applications, beyond localization (e.g., beamforming, ranging, etc.).

We performed this evaluation using six models, including LR, KNN, RF, GBM, LightGBM, XGBoost, with the training data gradually increased from 5% to 80% in steps of 5%. Table III presents LoS / NLoS classifier (first-stage) results using AoA features estimated by MUSIC. All models performed perfectly across the evaluated metrics. For conciseness, the table reports the results at 80% of the training data.

We observed that when trained with sufficient data, the first-stage classifier maintains perfect accuracy, without negatively impacting the performance of the LoS and NLoS track classifiers in the second stage. With only 5% of the training data, all models achieved over 97% accuracy, demonstrating the strong discrimination ability of the AoA-based features. Remarkably, KNN and RF achieved 100% accuracy using just 5% of the training data across all tested window sizes $(500, 1000, 2000)$ for AoA estimation. Additionally, both models exhibited significantly lower training times than the others, highlighting their efficiency. However, KNN has a considerably longer inference time than LR, RF, or GBM. For real-time applications where inference speed is critical, LR and GBM provide ideal performance with inference times of $0.24 - 0.26$ ms and $0.78 - 0.82$ ms, respectively.

2) Performance of Track Classifiers: Since the LoS / NLoS classifier determines whether a given sample belongs to the LoS or NLoS region, it allows second-stage classifiers to focus on training and classifying tracks within their respective regions. This separation allows each classifier to be trained with more homogeneous data, improving overall performance. For training and evaluating the LoS track classifier, we used data from track 6, track 9, track 10, track 11, and track 12. For the NLoS track classifier, we used data from track 1, track 2, track 3, track 13, and track 20. Thus, both classifiers perform 5-class classification at this stage. To evaluate and compare the effectiveness of the proposed two-stage model, we also built a single-stage baseline classifier that performs direct track classification without the step of LoS / NLoS region separation. This single-stage model serves as the baseline performance.

TABLE III: Model performance (including accuracy, F1-score, and ROC AUC), training time in seconds (T (s)), and inference time in milliseconds (I (ms)) across different window sizes (W).

Model (W)	Acc.	F1	AUC	T (s)	I (ms)
LR (2000)	1	1	1	0.029	0.239
LR (1000)	1	1	1	0.123	0.238
LR (500)	1	1	1	0.315	0.264
KNN (2000)	1	1	1	0.002	72.657
KNN (1000)	1	1	1	0.002	77.847
KNN (500)	1	1	1	0.004	83.951
RF (2000)	1	1	1	0.377	12.485
RF (1000)	1	1	1	0.569	11.896
RF (500)	1	1	1	1.003	11.929
GBM (2000)	0.99	0.99	1	3.439	0.794
GBM (1000)	0.99	0.99	1	7.804	0.817
GBM (500)	0.99	0.99	1	17.426	0.824
LightGBM (2000)	1	1	1	19.498	11.020
LightGBM (1000)	1	1	1	27.386	11.485
LightGBM (500)	1	1	1	40.610	11.850
XGBoost (2000)	1	1	1	26.473	53.174
XGBoost (1000)	1	1	1	28.684	53.324
XGBoost (500)	1	1	1	28.540	52.469

The experimental results, presented in Fig. 5, show that the two-stage approach generally performs better or at least equally well when compared to the single-stage approach, particularly when using ESPRIT for AoA estimation and most individual models. Tree-based algorithms such as RF, GBM, XGBoost, and LightGBM achieve better results under the two-stage setup for both MUSIC and ESPRIT. Stacking models (choosing from top 2 to 6 individual classifiers) consistently show the highest classification performance, particularly when using ESPRIT.

In most cases (but not for the maximum accuracy achieved), ESPRIT outperforms MUSIC as the AoA estimation algorithm. Combining ESPRIT with a top-6 stacking model yields the overall best classification accuracy. However, this improvement comes with increased computational cost. Therefore, trade-offs between performance and computational efficiency must be considered. The corresponding training and inference times for ML models are shown in Table IV. Stacking the top-6 model has the highest results (98% with MUSIC, 95.4% with ESPRIT). Fig. 6 shows the accuracy comparison of the top-6 stacking model for LoS and NLoS tracks using MUSIC and

TABLE IV: The comparison of accuracy, training time, and inference time across ML models for MUSIC and ESPRIT (window size = 2000), sorted by accuracy.

Models	Accuracy (%)	Training (s)	Inference (s)
MUSIC			
Stacking Top-6	97.98	1023.9869	0.0657
RF	94.30	0.5239	0.0110
GBM	89.46	10.6841	0.0018
XGBoost	89.10	117.3329	0.0212
LightGBM	85.98	103.7215	0.0063
KNN	76.62	0.0020	0.0383
LR	71.09	0.6774	0.0003
ESPRIT			
Stacking Top-6	95.38	1260.2022	0.0693
RF	92.61	0.7599	0.0108
LR	86.73	0.7763	0.0002
KNN	85.06	0.0020	0.0374
XGBoost	84.69	120.0175	0.0211
GBM	83.41	28.4534	0.0017
LightGBM	82.32	137.4263	0.0044

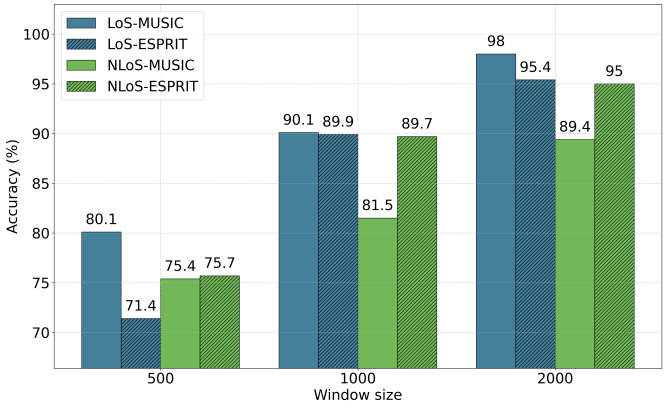

Fig. 6: Accuracy comparison of the top-6 stacking ensemble between LoS and NLoS across different window sizes using MUSIC and ESPRIT.

ESPRIT. Still, the training time exceeds 1000 seconds, and its inference time (above 60 ms) is significantly higher than that of other algorithms. Conversely, lightweight models such as LR and KNN provide fast inference but result in lower accuracy, achieving 71.1% and 76.6% with MUSIC, respectively. RF delivers an excellent balance, achieving high accuracy (94.3% with MUSIC, 92.6% with ESPRIT) with low training time ($0.52 - 0.76$ s) and low inference (approximately 11 ms).

IV. Conclusions and Future Perspectives

In this work, we proposed a hierarchical ML-based framework for track identification, leveraging robust AoA features estimated using high-resolution algorithms (MUSIC and ESPRIT). Extensive experiments have shown that the proposed framework achieves perfect classification between LoS and NLoS regions. Furthermore, an accuracy of 98% was achieved for LoS track identification, while an accuracy of 95% was achieved for NLoS track identification. For future work, combining both azimuth and elevation angles will be used to further increase accuracy. Another promising direction is the application of convolutional or graph-based neural networks to capture spatio-temporal information embedded in AoA distributions. Furthermore, we can extend region-based localization toward fine-grained coordinate-level positioning for more precise localization. This work can serve as a foundation for location-based trust evaluation as we proposed in [15]. Additionally, the model's outputs can be incorporated in location-based physical layer authentication as proposed in [4].

Acknowledgment

B. Trinh-Nguyen has been supported by the IPAL project CONNECTING. A. Chorti and S. Berri have been partially supported by the EC through the Horizon Europe/JU SNS project ROBUST-6G (Grant Agreement no. 101139068), the EU HORIZON MSCA-SE TRACE-V2X project (Grant No. 101131204), the ANR-PEPR 5G Future Networks projects, and the CYU INEX-PHEBE and NESLI (grant "Investissements d'Avenir" ANR-16-IDEX-0008) projects. The authors would like to thank S. Wesemann, G. Kaltbeitzel, D. Wiegner, M. Kinzler, S. Merk and S. Woerner from Nokia for realizing the channel measurements and sharing the data.

References

[1] M. Mitev, M. Shakiba-Herfeh, A. Chorti, M. Reed, and S. Baghaee, "A Physical Layer, Zero-Round-Trip-Time, Multifactor Authentication Protocol," *IEEE Access*, vol. 10, 2022.

[2] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[3] B. Yang and E. Yang, "A Survey on Radio Frequency Based Precise Localisation Technology for UAV in GPS-Denied Environment," *Journal of Intelligent & Robotic Systems*, vol. 103, no. 3, 2021.

[4] T. M. Pham, L. Senigagliesi, M. Baldi, G. P. Fettweis, and A. Chorti, "Machine Learning-Based Robust Physical Layer Authentication Using Angle of Arrival Estimation," in *IEEE GLOBECOM 2023*, 2023.

[5] D. Burghal, A. T. Ravi, V. Rao, A. A. Alghafis, and A. F. Molisch, "A Comprehensive Survey of Machine Learning Based Localization With Wireless Signals," *arXiv preprint arXiv:2012.11171*, 2020.

[6] A. Sobehy, É. Renault, and P. Mühlethaler, "CSI-MIMO: K-Nearest Neighbor Applied to Indoor Localization," in *IEEE ICC 2020*, 2020.

[7] T. Yang, A. Cabani, and H. Chafouk, "A Survey of Recent Indoor Localization Scenarios and Methodologies," *Sensors*, vol. 21, 2021.

[8] F. Bao, S. Mazokha, and J. O. Hallstrom, "Mobintel: Passive Outdoor Localization via RSSI and Machine Learning," in *2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2021, pp. 247–252.

[9] S. Chen, J. Fan, X. Luo, and Y. Zhang, "Multipath-Based CSI Fingerprinting Localization With a Machine Learning Approach," in *2018 Wireless Advanced (WiAd)*, 2018.

[10] M. K. Shehzad, L. Rose, S. Wesemann, and M. Assaad, "ML-Based Massive MIMO Channel Prediction: Does It Work on Real-World Data?" *IEEE Wireless Communications Letters*, vol. 11, no. 4, 2022.

[11] D. H. Wolpert, "Stacked Generalization," *Neural Networks*, 1992.

[12] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[13] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. Acoust., Speech, Signal. Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[14] N. P. Waweru, D. B. O. Konditi, and P. K. Langat, "Performance Analysis of MUSIC, Root-MUSIC and ESPRIT DOA Estimation Algorithm," *Int. J. Electrical Computer Energetic Electronic and Communication Engineering*, vol. 8, no. 01, pp. 209–216, 2014.

[15] B. Trinh-Nguyen, S. Berri, S. G. Teo, T. Truong-Huu, and A. Chorti, "A Framework for Global Trust and Reputation Management in 6G Networks," in *7th International Conference on Machine Learning for Networking (MLN 2024)*, Reims, France, 2024.