GDOP and Reliability Aware Anchor Node Selection via Machine Learning for RSSI-Based Localization

Chao Sun
Electronic Engineering Department
Kwangwoon University
Seoul, South Korea
sunchao@kw.ac.kr

Youngok Kim*
Electronic Engineering Department
Kwangwoon University
Seoul, South Korea
kimyoungok@kw.ac.kr

***Abstract*— In RSSI-based localization systems, many conventional implementations assume that all available anchors within communication range participate in position estimation. While this simplifies system design, it can increase communication overhead, computational load, and susceptibility to poor geometry, especially when anchor density is high or the mobile node is in an unfavorable Geometric Dilution of Precision (GDOP) region. This paper addresses the problem of selecting both the number and the specific subset of anchors for each localization update, considering geometric configuration, link quality, and communication range constraints. We propose a learning-based anchor selection framework that predicts the most beneficial anchors using geometric and link-quality features, aiming to minimize GDOP and localization error while respecting resource constraints. Simulations demonstrate that the proposed approach, evaluated on a 10-anchor 2-D testbed, delivers 98.6% coverage of true classes with 44.4% Top-1 exact-subset accuracy over 65 active classes while consistently choosing low-GDOP, high-quality links and producing near-instant decisions without exhaustive search over 582 subsets.**

Keywords—anchor node selection, RSSI-based localization, machine learning, geometric dilution of precision, wireless sensor networks, positioning accuracy optimization

I. INTRODUCTION

Received Signal Strength Indicator (RSSI)–based localization has become an attractive solution for positioning in GPS-denied environments due to its low hardware cost, ease of deployment, and compatibility with low-power wireless networks [1], [2]. In such systems, the geometry of anchor node placement plays a critical role in determining localization accuracy through the Geometric Dilution of Precision (GDOP) [3]. High-GDOP configurations amplify measurement errors, making accurate position estimation challenging, especially in the presence of non-line-of-sight (NLOS) conditions and signal fluctuations [4].

In dynamic localization scenarios, the set of anchors available to the mobile node changes over time due to mobility and communication range limitations [5]. It is neither practical nor efficient to use all visible anchors at every localization update: some anchors may be too far to maintain a reliable link, and excessive anchor usage increases communication overhead and computational cost. Moreover, different anchor subsets yield different GDOP values, meaning that poor subset choices can significantly degrade localization accuracy even when measurement noise remains constant [6].

Early studies often treated anchor selection and placement in a rather ad-hoc manner. For example, Shang et al. [7] and Li et al. [8] selected anchors randomly within the network, with Shang et al. merely noting that a collinear set of anchors constitutes an "unlucky" choice without theoretical justification. More systematic approaches were later proposed. Doherty et al. [9] required anchors to be placed along the network boundaries, ideally at the corners, to ensure that unknown nodes remain within the convex hull of the anchors. Ash et al. [10] further provided analytical support for distributing anchors uniformly around the network perimeter, which minimizes localization errors under simple multilateration models. Karl and Willig [11], in their book, also emphasized the importance of perimeter anchor placement, reflecting the consensus that geometric configuration plays a decisive role in localization accuracy.

Beyond perimeter-based strategies, several works attempted to generalize anchor placement methods to more complex scenarios. Hara et al. [12] proposed a grid-based partitioning scheme, placing anchors at sub-rectangle centers to achieve target accuracy, although the method assumes a rectangular area and basic RSSI-based localization. Recognizing that poor anchor placement can significantly degrade performance in irregular or anisotropic networks. Cheng et al. [13] introduced the HyBloc algorithm, which combines multidimensional scaling (MDS) with proximity-distance mapping (PDM). By augmenting anchor geometry through artificial anchors in isotropic regions, HyBloc is able to mitigate the adverse effects of clustered or poorly distributed anchors. These studies collectively underscore the critical impact of anchor subset selection and placement strategies on localization robustness and accuracy.

Existing anchor selection strategies typically rely on either using all available anchors or applying simple heuristic rules such as selecting the nearest anchors or those with the widest angular spread. These methods do not explicitly account for the combined effects of measurement quality and geometric configuration, nor can they adapt dynamically to changing channel conditions. As a result, they may include anchors with poor link quality or unfavorable geometry, leading to suboptimal GDOP values and degraded positioning performance.

To address these limitations, this paper proposes a learning-based anchor node selection framework that jointly considers geometric and link-quality features to identify the most beneficial subset of anchors for trilateration. The proposed approach predicts both the number of anchors and the specific

subset to use at each localization update, aiming to minimize GDOP and localization error while reducing computational and communication load. The selected anchors are then used in conjunction with trilateration algorithm to estimate the mobile node's position. This design allows the system to maintain high positioning accuracy even under NLOS and high-GDOP conditions while achieving substantial reductions in computational cost, offering a scalable and robust solution for dynamic environments.

II. Model and Mathematical Formulation

A. System Setup

Assume that at time t, let the mobile node be located at $\hat{\mathbf{x}}_{t-1}$, estimated position from the previous step as shown in Fig. 1. Given a communication radius, the set of visible anchors is denoted by

$$\mathcal{A}_t = \{1,2,\dots,M\}, M = |\mathcal{A}_t| \tag{1}$$

The objective is to select a subset $S \subseteq \mathcal{A}_t$ with $|S| \geq 3$ that will be used for trilateration at the current step. Since trilateration requires at least three anchors, the minimum subset size is $K_{min} = 3$, while the maximum size can be up to all visible anchors M.

For any candidate subset S, the GDOP [3] is defined as

$$\text{GDOP}(S, \hat{\mathbf{x}}_{t-1}) = \sqrt{\text{tr}[(H_S^\top H_S)^{-1}]} \tag{2}$$

where

$$H_S = [\mathbf{u}_i^\top]_{i\in S}, \qquad \mathbf{u}_i = \frac{\mathbf{a}_i - \hat{\mathbf{x}}_{t-1}}{\|\mathbf{a}_i - \hat{\mathbf{x}}_{t-1}\|} \tag{3}$$

and $\mathbf{a}_i$ is the position of anchor i.

B. Objective Function

The overall objective function augments GDOP with a penalty term that accounts for link fluctuations or obstructions

$$J(S) = GDOP(S, \hat{\mathbf{x}}_{t-1}) + \lambda \sum_{i\in S} \omega_i \tag{4}$$

where $\lambda > 0$ balances geometric accuracy and link reliability penalties. ω_i is defined as a monotonically increasing function of the RSSI variance as

$$\omega_i = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \sigma_{ref}^2}, \qquad \sigma_{ref}^2 = median(\hat{\sigma}_j^2), j \in S \tag{5}$$

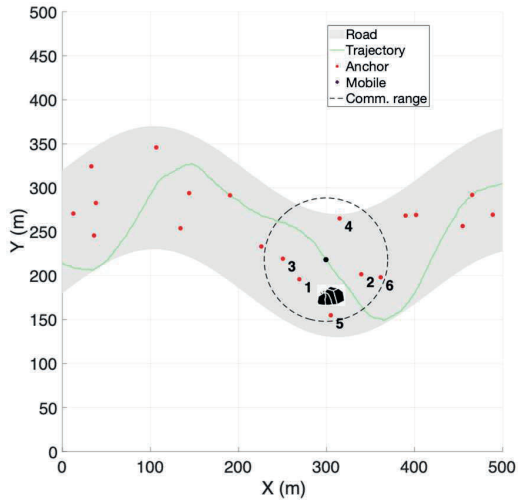


Fig. 1. Illustration of the localization scenario with 20 anchor nodes and a mobile node with a trajectory, showing a dashed communication range enclosing about six candidate anchors; an ML-based scheme selects a robust in-range subset for RSSI-based trilateration while explicitly accounting for obstacles that might degrade link quality.

where $\hat{\sigma}_i^2$ is the variance of the RSSI time series for link i, and σ_{ref}^2 denotes a robust scale reference that can be chosen as the median of the variances across all visible links.

The ground-truth subset is obtained by

$$S_t^\star = arg \min_{S\subseteq\mathcal{A}_t, |S|\geq 3} J(S) \tag{6}$$

which represents the optimal anchor subset that minimizes the objective function $J(S)$.

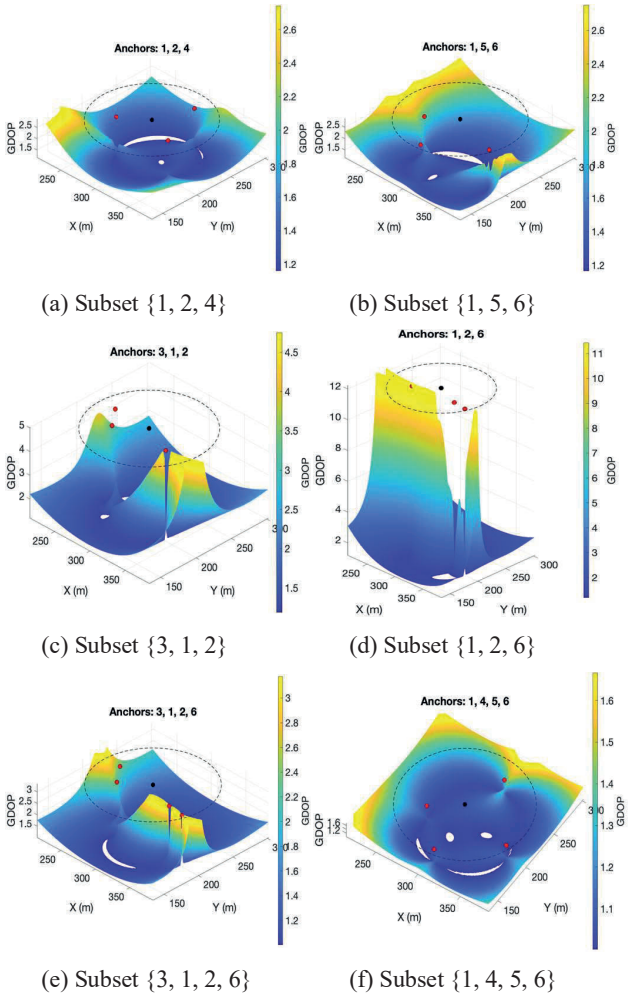
Since the anchor indices are unordered in subset formation, the cardinality of the candidate subset space is determined by the binomial coefficient summation

$$|\mathcal{S}_t| = \sum_{k=K_{min}}^{M} \binom{M}{k} \tag{7}$$

For the case of six visible anchors $M = 6$ as in our setup, this expands to

$$|\mathcal{S}_t| = \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 20 + 15 + 6 + 1 = 42 \tag{8}$$

meaning all 42 candidate subsets. This exhaustive evaluation against the objective $J(S)$ is computationally feasible when M is small, but scales exponentially for larger M, motivating the use of learning-based subset selection.



(a) Subset {1, 2, 4} (b) Subset {1, 5, 6}

(c) Subset {3, 1, 2} (d) Subset {1, 2, 6}

(e) Subset {3, 1, 2, 6} (f) Subset {1, 4, 5, 6}

Fig. 2. Comparative analysis of GDOP for anchor-node subsets within a mobile node's communication range.

C. *Illustrative GDOP Example*

To provide an intuitive understanding of how anchor-node placement affects GDOP, Fig. 2 presents the GDOP distributions for six representative anchor subsets within the mobile node's communication range. Each surface plot visualizes the spatial variation of GDOP values with different anchor configurations. As observed, subsets with more favorable geometric diversity (e.g., well-separated anchors such as {1, 2, 4} or {1, 4, 5, 6}) result in lower GDOP across the region, indicating better positioning accuracy. In contrast, subsets with colinear anchors (e.g., {1, 2, 6}) exhibit larger GDOP values, implying poor geometry and degraded localization performance. This example highlights the necessity of subset selection, as the choice of anchors has a direct impact on the positioning accuracy.

III. Machine Learning-Based Anchor Subset Selection

In this section, we formulate the problem of anchor subset selection as a supervised classification task, where the goal is to identify the most reliable subset of anchors for trilateration-based localization. The classification relies on a set of engineered features extracted from the communication links, including (GDOP), RSSI statistics, and link quality indicator (LQI) statistics.

TABLE I. Extracted Per-Link Features

Feature Category	Feature Description	Mathematical Expression
Signal Strength	RSSI Mean	$\mu_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}$
	RSSI Variance	$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T} (r_{i,t} - \mu_i)^2$
	RSSI Max	$r_i^{\max} = \max_{\tau} r_{i,\tau}$
	RSSI Min	$r_i^{\min} = \min_{\tau} r_{i,\tau}$
Link Quality	LQI Mean	$\bar{q}_i = \frac{1}{T}\sum_{t=1}^{T} q_{i,t}$
	LQI Variance	$\hat{\sigma}_{q,i}^2 = \frac{1}{T-1}\sum_{t=1}^{T} (q_{i,t} - \bar{q}_i)^2$
	LQI Max	$q_i^{\max} = \max_{\tau} q_{i,\tau}$
	LQI Min	$q_i^{\min} = \min_{\tau} q_{i,\tau}$
Geometry	Relative Angle	$\theta_i = \arctan\frac{y_i - \hat{y}_{t-1}}{x_i - \hat{x}_{t-1}}$
	Estimated Distance	$\hat{d}_i = \|\mathbf{a}_i - \hat{\mathbf{x}}_{t-1}\|$

A. *Feature Design*

To enable robust subset classification, features are extracted at two levels: (i) per-link features that characterize individual anchor–mobile links in TABLE , and (ii) subset-level descriptors that aggregate link information and geometric measures in TABLE . The per-link features capture signal statistics and link quality, while subset-level descriptors provide geometric diversity and overall stability of the candidate subset.

The key geometric metric is the GDOP, which quantifies spatial geometry. In addition, the average distance and minimum angular separation relative to the mobile node describe anchor placement diversity. Statistical descriptors such as RSSI and LQI mean, variance, maximum, and minimum capture the stability and reliability of the selected links.

By combining per-link features with subset-level descriptors, the feature space integrates both signal reliability and geometric diversity. This allows the machine learning classifier to effectively discriminate between high-quality and low-quality anchor subsets.

TABLE II. Subset-Level Descriptors

Feature Category	Feature Description	Mathematical Expression
Geometric Feature	GDOP	$\mathrm{GDOP}(S, \hat{\mathbf{x}}_{t-1}) = \sqrt{\mathrm{tr}[(H_S^{\mathrm{T}} H_S)^{-1}]}$
	Avg. Distance	$\bar{d}(S_t) = \frac{1}{\lvert S_t \rvert} \sum_{i \in S_t} d_i$
	Min Angular Separation	$\Delta\theta_{\min}(S_t) = \min_{i \neq j \in S_t} \min(\lvert\theta_i - \theta_j\rvert, 2\pi - \lvert\theta_i - \theta_j\rvert)$
Statistical Reliability	RSSI Mean*	$\mathrm{RSS\_mean}(S_t) = \frac{1}{\lvert S_t \rvert} \sum_{i \in S_t} \mu_i^{(r)}$
	RSSI Variance*	$\mathrm{RSS\_var}(S_t) = \frac{1}{\lvert S_t \rvert} \sum_{i \in S_t} \sigma_i^{2(r)}$
	RSSI Max*	$\mathrm{RSS\_max}(S_t) = \max_{i \in S_t} r_i^{\max}$
	RSSI Min*	$\mathrm{RSS\_min}(S_t) = \min_{i \in S_t} r_i^{\min}$
Link Quality	LQI Mean*	$\mathrm{LQI\_mean}(S_t) = \frac{1}{\lvert S_t \rvert} \sum_{i \in S_t} \mu_i^{(q)}$
	LQI Variance*	$\mathrm{LQI\_var}(S_t) = \frac{1}{\lvert S_t \rvert} \sum_{i \in S_t} \sigma_i^{2(q)}$
	LQI Max*	$\mathrm{LQI\_max}(S_t) = \max_{i \in S_t} q_i^{\max}$
	LQI Min*	$\mathrm{LQI\_min}(S_t) = \min_{i \in S_t} q_i^{\min}$

* Averaged within the selected subset.

B. *Machine Learning Model Selection*

We frame anchor-subset selection as a multi-class classification problem where each class corresponds to one candidate subset S with sizes 3–5 and in total 582 classes for 10 anchors. The ground-truth label for a mobile location is the subset that minimizes $J(S)$ in Eq. (4). We adopt XGBoost because it handles tabular inputs with heterogeneous, non-linear interactions and provides calibrated class probabilities with fast inference. Given a training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^{M}$, where $\mathbf{x}_i$ is the feature vector and y_i is the ground-truth class label corresponding to the optimal anchor subset, XGBoost builds an ensemble of decision trees.

The prediction function is defined as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \qquad f_k \in \mathcal{F} \tag{9}$$

where $\mathcal{F}$ is the space of regression trees and K is the number of trees. The training objective is

$$\mathcal{L} = \sum_{i=1}^{M} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{10}$$

with $\ell(y_i, \hat{y}_i)$ is differentiable convex loss, e.g., softmax cross-entropy for classification. $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ is regularization term penalizing the complexity of each tree, where T is the number of leaves and w are the leaf weights.

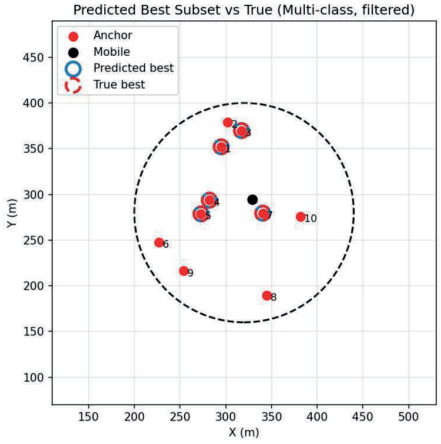


Fig. 3. Overlay of the predicted best subset and the ground-truth best subset at a representative location.

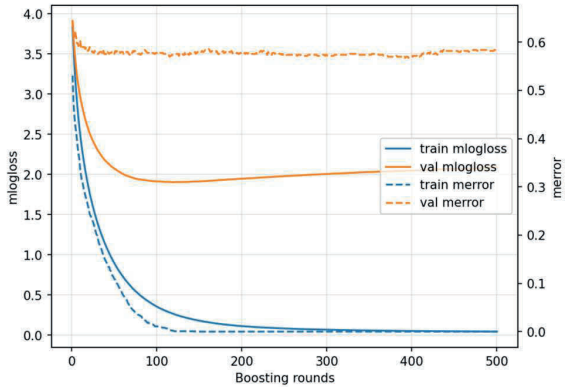


Fig. 4. Multiclass XGBoost training curves showing mlogloss and classification error versus boosting rounds.

C. *Model Training*

Synthetic datasets are generated by uniformly sampling mobile positions inside the communication region as shown in Fig. 3. For each position, we compute the label by exhaustive minimization of $J(S)$ over all subsets and simulate per-link RSSI via a log-distance model with anchor-specific noise levels. To avoid degenerate rare classes and enable stratified splits, we retain only classes with at least two training samples; this leaves 65 active classes out of 582. We train an XGBoost multi-class model with 500 trees, depth 6, learning rate 0.05, softmax cross-entropy, using a stratified 75/25 train/validation split. Fig. 4 plots the training curves that shows loss decreases steadily and validation error stabilizes, indicating convergence without instability. On a held-out test set of 500 positions, the model covers 98.6% of true classes (i.e., they were seen during training) and achieves 44.4% Top-1 exact-subset accuracy on the covered portion.

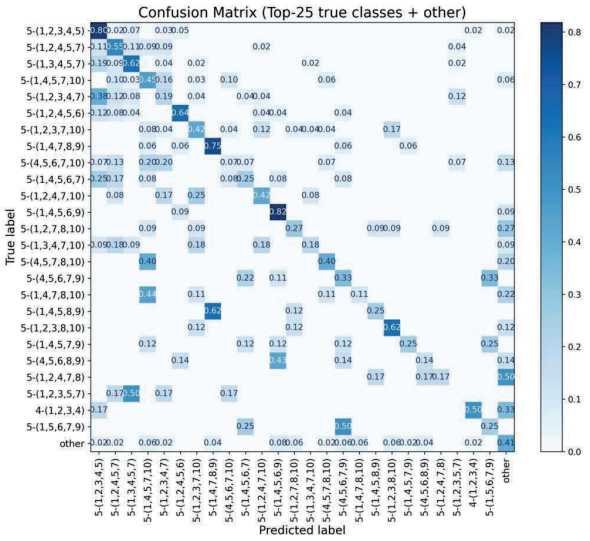


Fig. 5. Row-normalized confusion matrix for the 25 most frequent true classes with the remaining classes.

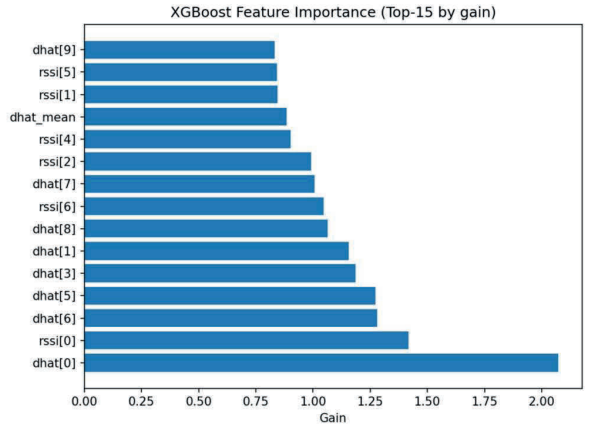


Fig. 6. XGBoost feature importance (top 15 features ranked by gain) highlighting the most influential per-link and aggregate features.

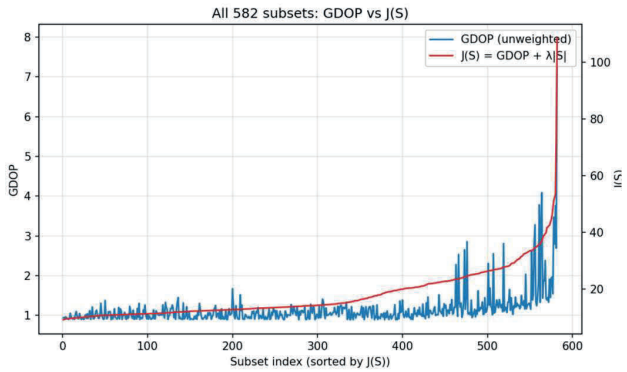


Fig. 7. All 582 candidate subsets, sorted by GDOP and $J(S)$ at a representative position.

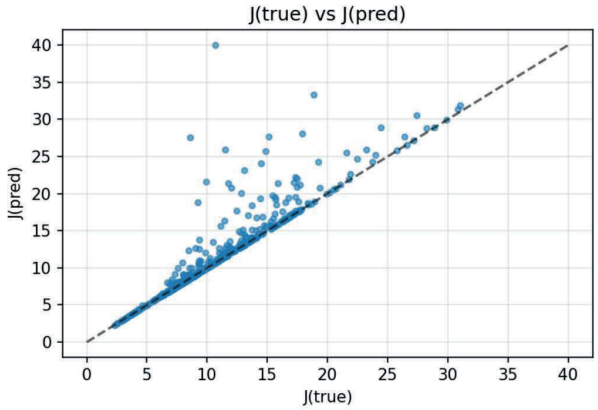


Fig. 8. Scatter of $J(\text{true})$ versus $J(\text{pred})$ across test positions with the $y = x$ line indicating near-optimality of predictions.

D. *Model Evaluation*

The confusion matrix over the Top-25 most frequent test classes in Fig. 5 is diagonally dominant, with off-diagonal mass mostly between subsets of the same cardinality or with similar geometry. This pattern reflects the intrinsic difficulty of distinguishing near-optimal subsets when RSSI noise is comparable across anchors. XGBoost gain in Fig. 6 ranks range-related features $\hat{d}$ and RSSI of individual anchors, plus global statistics as most informative, followed by a few per-link quality indicators. This confirms that the model primarily leverages geometry and link reliability, which aligns with the design of $J(S)$. The ordered curves in Fig. 7 show GDOP and $J(S)$ for all 582 subsets at the demo position. The sharp rise of $J(S)$ in the tail indicates many subsets are clearly sub-optimal; our classifier avoids exploring them and focuses probability mass on the low-$J(S)$ region. Fig. 8 compares $J(\text{true})$ against $J(\text{pred})$ across test positions. Points cluster near the diagonal, showing that—even when the exact subset differs—the predicted set is typically near-optimal in objective value.

IV. Conclusion

We presented a learning-based anchor selection framework that predicts, for each localization update, the subset of anchors minimizing a GDOP-aware objective under link-quality variability. On a 10-anchor testbed with 3–5-anchor candidates in 582 classes, the model covers 98.6% of true classes and achieves 44.4% Top-1 exact-subset accuracy over the 65 active classes, while producing fast decisions without online combinatorial search.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2024-00461079, Development of Adaptive On-Device Software Technology for Environmental Adaptation in Unmanned Vehicle Surveillance Equipment).

REFERENCES

[1] J. Luomala and I. Hakala, "Analysis and evaluation of adaptive RSSI-based ranging in outdoor wireless sensor networks," *Ad Hoc Netw.*, vol. 87, pp. 100–112, May 2019, doi: 10.1016/j.adhoc.2018.10.004.

[2] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Commun. Surv. Tutor.*, vol. 11, no. 1, pp. 13–32, 2009, doi: 10.1109/SURV.2009.090103.

[3] S.-H. Bach and S.-Y. Yi, "Constrained Least-Squares Trilateration for Indoor Positioning System Under High GDOP Condition," *IEEE Trans. Ind. Inform.*, vol. 20, no. 3, pp. 4550–4558, Mar. 2024, doi: 10.1109/TII.2023.3326535.

[4] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Comput. Netw.*, vol. 43, no. 4, pp. 499–518, Nov. 2003, doi: 10.1016/S1389-1286(03)00356-6.

[5] L. L. de Oliveira, G. H. Eisenkraemer, E. A. Carara, J. B. Martins, and J. Monteiro, "Mobile Localization Techniques for Wireless Sensor Networks: Survey and Recommendations," *ACM Trans Sen Netw*, vol. 19, no. 2, p. 36:1-36:39, 2023, doi: 10.1145/3561512.

[6] T. Kunz and B. Tatham, "Localization in Wireless Sensor Networks and Anchor Placement," *J. Sens. Actuator Netw.*, vol. 1, no. 1, pp. 36–58, June 2012, doi: 10.3390/jsan1010036.

[7] Y. Shang, W. Rumi, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 11, pp. 961–974, Nov. 2004, doi: 10.1109/TPDS.2004.67.

[8] L. Li and T. Kunz, "Cooperative node localization using nonlinear data projection," *ACM Trans Sen Netw*, vol. 5, no. 1, p. 1:1-1:26, 2009, doi: 10.1145/1464420.1464421.

[9] L. Doherty, K. S. J. pister, and L. El Ghaoui, "Convex position estimation in wireless sensor networks," in *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213)*, Apr. 2001, pp. 1655–1663 vol.3. doi: 10.1109/INFCOM.2001.916662.

[10] J. N. Ash and R. L. Moses, "On optimal anchor node placement in sensor localization by optimization of subspace principal angles," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008, pp. 2289–2292. doi: 10.1109/ICASSP.2008.4518103.

[11] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005.

[12] S. Hara and T. Fukumura, "Determination of the placement of anchor nodes satisfying a required localization accuracy," in *2008 IEEE International Symposium on Wireless Communication Systems*, Oct. 2008, pp. 128–132. doi: 10.1109/ISWCS.2008.4726032.

[13] K.-Y. Cheng, K.-S. Lui, and V. Tam, "HyBloc: Localization in Sensor Networks with Adverse Anchor Placement," *Sensors*, vol. 9, no. 1, pp. 253–280, Jan. 2009, doi: 10.3390/s90100253.