Dimensionality Reduction with Truncated Singular Value Decomposition and K-Nearest Neighbors Regression for Indoor Localization

Hang Duong Thi[1], Kha Hoang Manh[2*], Vu Trinh Anh[3], Trang Pham Thi Quynh[4], Tuyen Nguyen Viet[5]
Faculty of Electronics and Engineering, Hanoi University of Industry[1, 2, 4, 5]
Department of Electronics and Telecommunication, University of Engineering and Technology, Vietnam National University[1, 3]

***Abstract*—Indoor localization presents formidable challenges across diverse sectors, encompassing indoor navigation and asset tracking. In this study, we introduce an inventive indoor localization methodology that combines Truncated Singular Value Decomposition (Truncated SVD) for dimensionality reduction with the K-Nearest Neighbors Regressor (KNN Regression) for precise position prediction. The central objective of this proposed technique is to mitigate the complexity of high-dimensional input data while preserving critical information essential for achieving accurate localization outcomes. To validate the effectiveness of our approach, we conducted an extensive empirical evaluation employing a publicly accessible dataset. This dataset covers a wide spectrum of indoor environments, facilitating a comprehensive assessment. The performance evaluation metrics adopted encompass the Root Mean Squared Error (RMSE) and the Euclidean distance error (EDE)—widely embraced in the field of localization. Importantly, the simulated results demonstrated promising performance, yielding an RMSE of 1.96 meters and an average EDE of 2.23 meters. These results surpass the achievements of prevailing state-of-the-art techniques, which typically attain localization accuracies ranging from 2.5 meters to 2.7 meters using the same dataset. The enhanced accuracy in localization can be attributed to the synergy between Truncated SVD's dimensionality reduction and the proficiency of KNN Regression in capturing intricate spatial relationships among data points. Our proposed approach highlights its potential to deliver heightened precision in indoor localization outcomes, with immediate relevance to real-time scenarios. Future research endeavors involving comprehensive comparative analyses with advanced techniques hold promise in propelling the field of accurate indoor localization solutions forward.**

Keywords—Dimensionality Reduction; Indoor Positioning System; KNN regression; Truncated Singular Value Decomposition

I. INTRODUCTION

Indoor positioning has become a prominent research area in recent years, driven by the increasing demand for location-based services in various applications, such as indoor navigation, asset tracking, and context-aware services. Traditional positioning systems relying on Global Positioning System (GPS) are not always reliable indoors due to limited satellite signals penetration and multi-path effects, making them less accurate for indoor environments. This limitation has led to the emergence of alternative techniques, with machine learning proving to be a promising approach for indoor positioning tasks. Machine learning methods offer the ability to model complex relationships between Wi-Fi Received Signal Strength Indicator (RSSI) measurements and indoor locations, enabling accurate predictions in indoor settings [1]. Among the diverse machine learning algorithms, the K-Nearest Neighbors (KNN) algorithm has garnered significant attention and success in indoor positioning applications [2, 3]. KNN is a non-parametric and instance-based learning algorithm that classifies or predicts a target value based on the similarity of features from neighboring data points. Its simplicity and effectiveness have made it a popular choice for indoor positioning tasks. However, the conventional KNN algorithm can be sensitive to noise and imbalanced data, necessitating the exploration of specialized variants to improve performance.

One such variant is the Weighted K-nearest neighbors (WKNN), which assigns different weights to neighboring data points based on their distance or other factors [4]. This enables WKNN to give higher importance to closer points, leading to more accurate predictions and better handling of imbalanced data distributions. The authors in study [4] propose a method that utilizes Improved W-KNN to enhance indoor localization performance based on fingerprinting by leveraging the relationship between the nearest fingerprint and (K-1) auxiliary fingerprints to determine the position. In the quest for enhanced adaptability, Adaptive KNN adjusts the number of neighbors (K) based on the local density of data points, dynamically tailoring the algorithm to varying spatial distributions within the indoor environment [5]. The paper introduces an enhanced KNN algorithm featuring a variable K. The fundamental concept revolves around dynamically modifying the K value according to the discrepancies between measured signals and the corresponding values within the database. In this paper, adaptability contributes to improved performance across different regions with distinct data densities. Additionally, KNN Regression is used when predicting continuous target variables, such as indoor coordinates, making it particularly suitable for regression tasks in indoor positioning. Apart from KNN-based methods, deep learning techniques have also been explored for indoor positioning. Convolutional Neural Networks (CNN) [6, 7] and Long Short-Term Memory (LSTM) [8] networks are notable examples. CNN can effectively extract spatial features from Wi-Fi images, while LSTM can model temporal dependencies in time-series data, such as RSSI signals. In [7], the authors introduces an innovative method for converting Wi-Fi signatures into images, establishing a scalable fingerprinting framework utilizing convolutional neural networks (CNNs), These deep learning models have shown promise in achieving high accuracy in indoor positioning tasks, but they may require more extensive datasets and computation resources.

The authors in [9, 10] conducted an overview study on several data dimensionality reduction methods and their effectiveness in reducing computation time while preserving information. Data preprocessing is of paramount importance in indoor positioning tasks to enhance model performance and reduce computational complexity. High-dimensional Wi-Fi RSSI data can be computationally intensive and challenging to handle. To address this issue, dimensionality reduction techniques are applied to retain critical information while significantly reducing the number of features. Principal Component Analysis (PCA) is a popular method for dimensionality reduction, but it may not be ideal for all scenarios due to its requirement for the data to be centered and scaled. As an alternative, Truncated Singular Value Decomposition (Truncated SVD) [11] is employed, which is a variant of PCA that can efficiently handle large datasets and does not require data centering. In this paper, we propose an approach for indoor positioning that combines Truncated SVD for dimensionality reduction of Wi-Fi RSSI data with KNN regression for accurate indoor location estimation. The solutions mentioned above all aim to reduce errors in location estimation. However, a challenge arises in studies using Wi-Fi signals, where the use of high-dimensional data complicates the training process due to the time required for both training and prediction. Therefore, reducing data dimensionality is considered an effective solution to reduce model complexity and simultaneously enhance data processing flexibility. Combining Truncated SVD and KNN Regression yields a flexible method applicable to various indoor positioning scenarios, not constrained by specific data structures or characteristics. The proposed approach will be extensively evaluated and compared with other state-of-the-art methods to demonstrate its effectiveness and applicability.

In the subsequent sections, we will delve into the details which are as follows: Section II is about the related works, Section III deals with the proposed approach. Section IV gives results, and discussion, and concludes with implications for future research in Section V, thereby contributing to the advancement of indoor positioning technology and its real-time applications.

II. Related Works

A. Challenges in Indoor Localization using Wi-Fi Signals

Addressing the problem of enhancing accuracy in indoor positioning still encounters numerous difficulties due to the persistence of challenges such as limited GPS signals. Unlike outdoor environments where Global Positioning System (GPS) signals are readily available, indoor spaces often lack direct access to GPS signals due to signal attenuation caused by walls, ceilings, and other structural elements. This limitation hampers the effectiveness of traditional GPS-based localization techniques. Multipath effects further complicate indoor positioning. Indoor environments introduce multipath effects, where wireless signals bounce off surfaces and create multiple signal paths. This leads to signal interference, phase shifts, and fluctuations, making signal strength-based localization less accurate and reliable. Moreover, signal propagation variability within indoor spaces can vary significantly due to factors such as furniture placement, architectural elements, and interference from electronic devices. This variability challenges the establishment of consistent and reliable signal patterns for accurate localization. Non-line-of-sight (NLOS) conditions due to obstructions like walls and obstacles can block the direct line between the transmitter and receiver, introducing additional complexities in signal propagation and affecting accuracy. The phenomenon of multipath fading, where signals arriving via different paths interfere constructively or destructively, contributes to signal fluctuations and inaccuracies in distance estimation. Dealing with high-dimensional data, such as Wi-Fi Received Signal Strength Indicator (RSSI) readings from multiple access points, is a computational challenge in indoor localization. The presence of interference and noise from electronic devices further impacts the accuracy of localization algorithms. Different indoor environments with unique layouts and architectural features add complexity to localization algorithms, as a one-size-fits-all approach may not be effective. Privacy concerns arising from collecting and analyzing personal data in indoor localization require careful consideration of data handling and user consent. Furthermore, many indoor localization applications demand real-time accuracy for guiding users or tracking assets, posing a challenge in achieving both precision and speed. Addressing these challenges necessitates innovative techniques that consider the intricacies of indoor environments. The proposed methodology aims to tackle these obstacles by combining dimensionality reduction and regression techniques for accurate indoor localization. In the context of employing methods such as K-Nearest Neighbors (KNN), Weighted K-Nearest Neighbors (WKNN), Adaptive K-Nearest Neighbors (Adaptive KNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) for indoor position prediction, various limitations become apparent. For K-Nearest Neighbors (KNN), its sensitivity to the choice of K neighbors introduces computational complexities as K increases. Additionally, KNN is sensitive to noise and imbalanced data. Weighted K-Nearest Neighbors (WKNN) presents challenges in effectively setting weight parameters to enhance prediction performance. Adaptive K-Nearest Neighbors (Adaptive KNN) involves uncertainty in dynamically determining the optimal number of neighbors (K) for individual cases. Convolutional Neural Networks (CNN) demand substantial training data and computational resources, especially in real-time scenarios. Long Short-Term Memory (LSTM) requires longer time-series data for complex pattern recognition. Overall, while these methods offer unique strengths to tackle indoor positioning challenges, they also exhibit limitations, necessitating careful customization and consideration of the specific environment for an effective solution.

To address these challenges comprehensively, innovative techniques that account for the intricacies of indoor environments are indispensable. The proposed methodology aims to overcome these hurdles by synergistically employing dimensionality reduction techniques and regression methodologies for accurate indoor localization. The ensuing sections delve into the details of this approach, experimental evaluation, and results, underscoring its efficacy in tackling these persistent challenges and enhancing indoor positioning accuracy.

B. Dimensionality Reduction Techniques

In the realm of indoor positioning, addressing the challenges posed by high-dimensional data is crucial for achieving accurate and efficient results. This section provides an overview of various dimensionality reduction techniques that have been employed to tackle the complexity of indoor positioning datasets. Dimensionality reduction aims to extract essential information from the data while reducing its dimensionality, thus enhancing the efficiency of subsequent analysis and prediction processes.

One commonly used technique is Principal Component Analysis (PCA) [12-14], which projects the original data onto a new orthogonal coordinate system defined by its principal components. By retaining the most significant dimensions and discarding less informative ones, PCA simplifies the data representation while preserving as much variance as possible. Another approach, [15] Truncated SVD, is a variant of PCA that efficiently approximates the original data matrix by retaining only the top singular values and corresponding singular vectors. This method is particularly suitable for large datasets and offers advantages in terms of computational efficiency.

In addition to these techniques, various other methods can also play a role in dimensionality reduction. However, the choice of method depends on the characteristics of the data and the specific requirements of the indoor positioning task. By effectively reducing the dimensionality of the input data, these techniques contribute to enhancing the performance of subsequent algorithms and models for accurate indoor positioning. The following sections will delve into the details of how these techniques are applied and their impact on the proposed methodology.

1) Truncated singular value decomposition (Truncated SDV) Method: The Truncated Singular Value Decomposition (Truncated SVD) method is a dimensionality reduction technique commonly employed to mitigate the challenges associated with high-dimensional data in various applications, including indoor positioning. This approach builds upon the concept of Singular Value Decomposition (SVD), which decomposes a data matrix into three separate matrices representing its singular values and corresponding left and right singular vectors. In the context of indoor positioning, Truncated SVD involves retaining only the top singular values and their corresponding singular vectors, effectively reducing the dimensionality of the data while preserving its essential information. This process is particularly beneficial for managing large datasets, as it significantly decreases the computational burden and enhances the efficiency of subsequent analysis. The core idea of Truncated SVD is to approximate the original data matrix using a lower-dimensional representation that captures the most significant patterns and relationships within the data. By selecting a specific number of singular values to retain, this method allows researchers and practitioners to balance between dimensionality reduction and preserving relevant information.

Truncated SVD finds applications in various fields, including image processing, natural language processing, and data compression. In the context of indoor positioning, it offers a valuable tool for preprocessing Wi-Fi Received Signal Strength Indicator (RSSI) data, effectively reducing its dimensionality while maintaining its inherent structure. The reduced-dimension representation obtained through Truncated SVD can then be used as input for subsequent algorithms, such as K-Nearest Neighbors (KNN) regression, to enhance the accuracy and efficiency of indoor positioning predictions. The goal of Truncated SVD is to reduce the dimensionality of the data by retaining a limited number of important singular values and vectors. This helps to simplify the complexity of the original data and create a reduced version that can be used in various tasks such as classification, prediction, and indoor positioning. Algorithm 1 presents Truncated SVD Algorithm.

Algorithm 1: Truncated SVD Algorithm

Input: The initial data is a matrix A with dimensions m x n, where m is the number of samples and n is the dimensionality of the data.
Output: The matrix A_reduced represents the reduced data and includes the most important components from the original data.
Step 1: Compute Singular Value Decomposition (SVD): Perform the Singular Value Decomposition on the data matrix **A**: **A = U * Σ * Vt**
Where:

- **U** is the matrix containing the left singular vectors (columns) of A.
- **Σ** is the diagonal matrix containing the singular values of A.
- **Vt** is the matrix containing the right singular vectors (rows) of A.

Step 2: Select Number of Components: Choose the number of components (singular values and vectors) that you want to retain after dimensionality reduction. This is an important parameter to adjust the level of dimensionality reduction.
Step 3: Truncate Singular Values and Vectors: Keep only the singular values and vectors corresponding to the number of components selected in the previous step. Create the truncated matrices **U** reduced and **Vt** reduced.
Step 4: Reconstruct Reduced Data: Generate a new data matrix (reduced data) by multiplying the truncated matrix **U** reduced, the diagonal matrix **Σ** reduced (containing singular values), and the transpose of the matrix Vt_reduced: **A**_reduced = **U**_reduced * **Σ**_reduced * **Vt**_reduced

---

2) K nearest neighbors regression algorithm: KNN regression is a non-parametric algorithm that relies on the similarity of feature vectors to make predictions. It assumes that similar instances will have similar target values. The algorithm doesn't involve model training like some other regression algorithms; instead, it stores the entire dataset and calculates predictions based on the K nearest neighbors of the query instance. The choice of K is crucial, as a small K might lead to noisy predictions, while a large K might lead to overly smoothed predictions. KNN regression can be sensitive to outliers and irrelevant features, so preprocessing the data and feature selection can impact its performance. Fig. 1 describes the KNN regression for improving the accuracy of indoor localization.

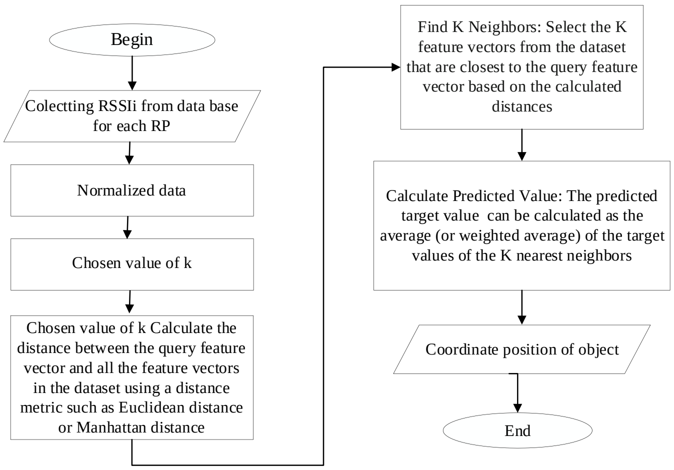

Fig. 1. The algorithm of WKNN for indoor positioning system.

III. Proposed Approach

This section presents the proposed solution that combines dimensionality reduction using Truncated Singular Value Decomposition (Truncated SVD) with KNN regression for indoor position prediction.

A. Proposal Model Block Diagram

Fig. 2 describes the block diagram illustrates the proposed approach that integrates two main components: Truncated Singular Value Decomposition (Truncated SVD) and K Nearest Neighbors (KNN) regression. The process begins with the collection of Wi-Fi RSSI data, which is the initial step for indoor position prediction.

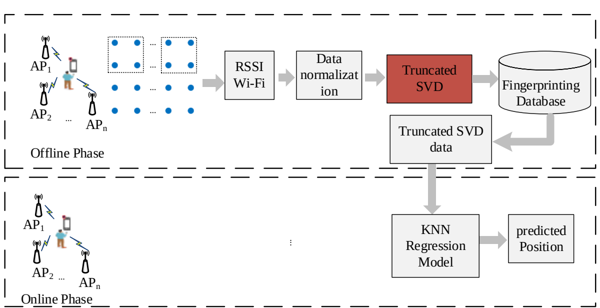

Fig. 2. The structure of proposed approach based on KNN regression.

Data collection and preprocessing:

- Wi-Fi RSSI data is gathered from multiple access points within the indoor environment.
- The collected data is preprocessed to remove noise, handle missing values, and normalize the features.

Truncated Singular Value Decomposition (Truncated SVD):

- The preprocessed Wi-Fi RSSI data undergoes Truncated SVD, a dimensionality reduction technique.
- Truncated SVD reduces the dimensionality of the data while retaining the most significant features that capture the underlying patterns.
- The transformed data is then ready for further processing.

K Nearest Neighbors (KNN) regression:

- The transformed data from Truncated SVD serves as input to the KNN regression model.
- KNN regression aims to predict the indoor position based on the similarity of the transformed data points.
- The model identifies the K nearest neighbors to the input data point and uses their positions to estimate the target position.

Indoor position prediction:

- The combination of Truncated SVD and KNN regression results in an accurate indoor position prediction.
- The estimated position is output as the final result of the model.

The block diagram demonstrates how the proposed method utilizes Truncated SVD for dimensionality reduction to handle the high-dimensional Wi-Fi RSSI data effectively. The reduced-dimensional data is then fed into the KNN regression model, which leverages the spatial relationships between data points to predict the indoor position accurately. This integrated approach aims to overcome the challenges of noise, signal variability, and dimensionality while providing enhanced precision in indoor position prediction.

B. Indoor Positioning Dataset

In this study, we assessed our proposed solution using an online dataset [16], previously standardized for indoor positioning research. This dataset, employed in previous work including [17] aimed to enhance indoor positioning precision by LSTM algorithms. The study in [6] achieved a positioning error are range 2.5 meters to 2.7 meters on the public dataset [16]. The paper [16] verified the dataset's normalization and reliability for indoor localization research. The dataset covered a library space of over 300 square meters on the 3rd and 5th floors, collected over 15 months, and comprising 60,000 measurements. It included object positions, Wi-Fi AP access point RSSI values, execution time, and identification data. With 448 Wi-Fi AP access points at around 2.65 meters above the ground on both floors, a fingerprint database was created from multiple locations and directions, including front, back, left, and right measurements. Offline training involved known reference points and a Samsung Galaxy S3 phone equipped with an application to capture RSSI data. The dataset was divided into training (16,704 fingerprints from 24 reference points) and test sets (46,800 fingerprints from 106 reference points), each containing 448 RSSI indicators from access points. Following the approach of [6], our experiments on the normalized public dataset utilized the Minmaxscale() function and Formula (1).

$$X_{scaled} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X is the initial value of the feature, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values in the feature.

C. Error Estimation Criteria

We employed four machine learning-based evaluation criteria to assess the proposed solution's effectiveness. Initially, we used the mean absolute error (MAE) to gauge the average absolute error within the prediction dataset. Further evaluation employed mean squared error (MSE) or RMSE, widely utilized in machine learning regression problems, to quantify the squared error between predicted and actual values. Additionally, the determination coefficient R2 was employed as a measure of the model's predictive capability. R2 assesses how well the model predicts the dependent variable based on independent variables, showcasing the goodness-of-fit. Higher R2 values signify better model fit, with a range from 0 to 1. Negative values can emerge if the model performs worse than a constant model predicting the mean. Despite its usefulness, R2 has assumptions and limitations that warrant consideration. This coefficient's range lies between 0 and 1. A value closer to 1 indicates strong model fit and accurate prediction of position.

Furthermore, we introduced the EDE as an additional metric to measure prediction accuracy. The EDE calculates the direct geometric distance between predicted and actual positions, offering a straightforward measure of how far the predictions deviate from the true positions. This distance was calculated using the Euclidean distance formula, providing valuable insight into the spatial accuracy of the proposed solution.

$$MAE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left| Pos_{itrue} - Pos_{iest} \right| \tag{2}$$

$$MSE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left(Pos_{itrue} - Pos_{iest} \right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left(Pos_{itrue} - Pos_{iest} \right)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} \left(Pos_{itrue} - Pos_{iest} \right)^2}{\sum_{i=1}^{N_{test}} \left(Pos_{itrue} - avg(y) \right)^2} \qquad avg(y) = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} Pos_{itrue} \tag{5}$$

$$EDE = \frac{1}{N_{test}} \sum_{1}^{n} \sqrt{\left(Pos_{itrue} - Pos_{iest} \right)^2} \tag{6}$$

Where Pos_{itrue} is the $i-th$ observed position, Pos_{itest} is the $i-th$ estimation position. $i-th$ is the number of samples in the test dataset.

IV. Results and Discussion

In this study, we conducted a thorough investigation to determine the optimal number of principal components (components) to retain when applying the truncated SVD technique. The objective was to minimize the EDE, a critical metric for assessing the accuracy of our proposed indoor positioning solution.

The process of choosing the appropriate value for components began with a systematic survey across a range of values, assessing the performance of the solution at each step. We employed the EDE as the primary evaluation criterion, aiming to identify the component's value that yielded the lowest error. Our experimentation revealed that at the number of components = 35, the solution achieved the minimal EDE. This observation was consistent with our goal of minimizing error while ensuring computational efficiency, as retaining 35 principal components struck an optimal balance between accuracy and resource consumption. By retaining 35 principal components, we effectively reduced the dimensionality of the data while preserving the critical information necessary for accurate indoor localization. This choice optimized the model's ability to capture relevant spatial relationships among data points, resulting in a significant reduction in the EDE. In conclusion, our survey and experimentation led us to select the number of components = 35 as the optimal configuration for the truncated SVD algorithm. This choice aligns with our objective of achieving the lowest EDE while maintaining efficiency, demonstrating the effectiveness of this approach in improving indoor positioning accuracy.

The analysis of Fig. 3, which depicts the relationship between the number of components in truncated SVD (n) and the EDE, reveals several key insights:

Firstly, at n = 35, we observe the lowest EDE, indicating that this configuration results in the most accurate indoor positioning predictions. This point aligns with our earlier discussion, highlighting n = 35 as the optimal choice for retaining principal components. When n is smaller than 35, the reduction in dimensionality becomes excessive, leading to a loss of critical information required for accurate localization.

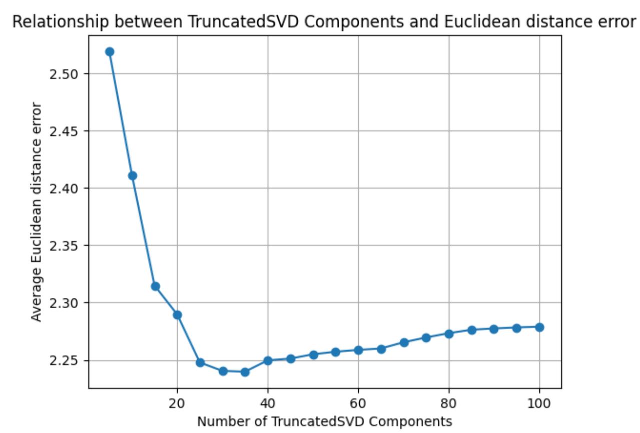

Fig. 3. The relationship between the number of truncated SVD components and euclidean distance error.

Conversely, when the numbers of components (n) are greater than 35, the model becomes overly complex, potentially introducing noise and diminishing its predictive capabilities. The fact that the EDE consistently rises for values of n smaller or larger than 35 emphasizes the importance of careful parameter tuning in the truncated SVD technique. It highlights the delicate balance between dimensionality reduction and information preservation.

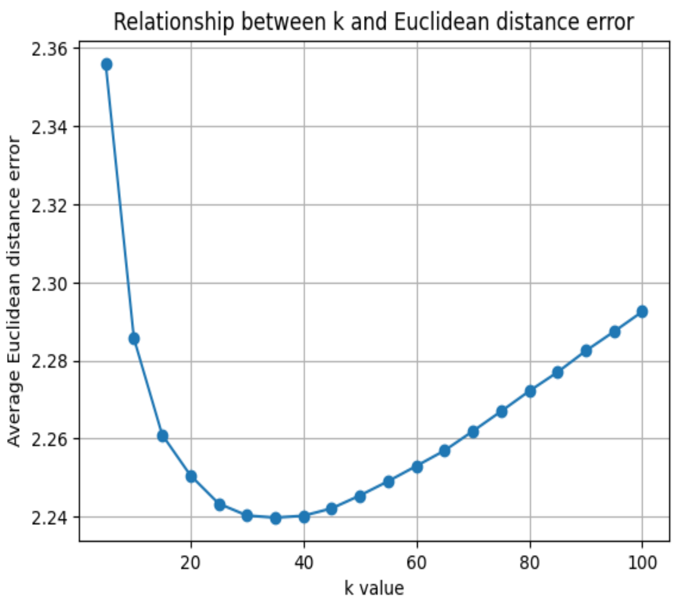

Fig. 4. The relationship between k and euclidean distance error.

Fig. 4, which illustrates the relationship between the number of nearest neighbors (k) and the EDE, provides crucial insights into our study. The plot demonstrates a clear trend that supports the idea that selecting k = 30 is an optimal choice for our KNN regression-based indoor positioning solution. This specific k-value results in the lowest EDE, indicating the highest precision in predicting indoor positions. When k deviate from this optimal k-value, either by choosing k values smaller or larger than 30, we consistently observe an increase in EDE. This pattern emphasizes the sensitivity of our model's performance to the choice of k. If k is smaller than 30, the model may not adequately capture essential spatial relationships, leading to less accurate predictions. Conversely, when k exceeds 35, the model might over smooth the data, potentially losing critical local information, which results in increased prediction errors. The observation that EDE increases for k values smaller or larger than 30 underscores the importance of selecting the appropriate parameter for KNN regression. It reinforces the idea that finding the right balance in the number of neighbors considered is vital for achieving accurate indoor localization. Fig. 4 highlights the significance of choosing k = 30 as the optimal parameter for our KNN regression-based indoor positioning solution. This choice leads to the lowest EDE value, indicating the highest accuracy in position prediction. Deviating from this value consistently results in increased EDE, affirming the effectiveness of our proposed approach in enhancing indoor localization accuracy.

Fig. 5 provides a visual representation of 100 randomly selected real positions (depicted as blue dots) and their corresponding predicted positions (depicted as red dots). This graph serves as a valuable illustration of the performance of our indoor positioning model. The meanings of the parameters on Fig. 5 are described below:

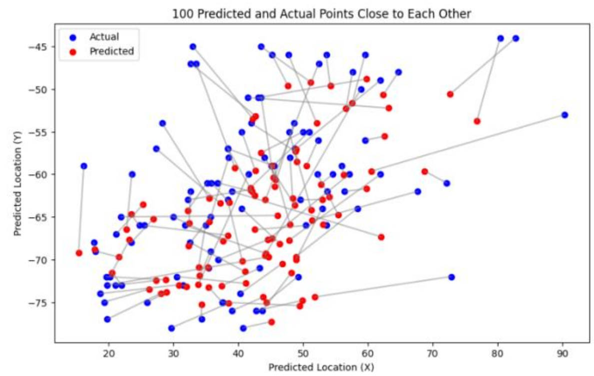

Fig. 5. 100 predicted and real positions.

1) Blue dots - real positions: The blue dots represent the actual positions of objects in the indoor environment, providing a reference for the ground truth. These positions are based on the collected dataset.

2) Red dots - predicted positions: The red dots, on the other hand, signify the positions predicted by our proposed indoor positioning model. These predictions are generated using the combination of Truncated SVD for dimensionality reduction and KNN regression for position estimation.

3) Visual comparison: By visually comparing the red and blue dots, it's evident that our model's predictions are generally very close to the actual positions. This alignment between the predicted and actual positions highlights the accuracy and effectiveness of our proposed solution.

4) Scattered distribution: The distribution of both red and blue dots across the graph demonstrates that the model is capable of predicting positions in various locations throughout the indoor environment. This showcases the versatility and applicability of our approach across different scenarios.

5) Few outliers: While most of the red dots closely match the blue dots, there may be a few outliers where the predicted positions slightly deviate from the actual positions. These outliers could be attributed to factors such as signal interference or complex spatial relationships in the indoor environment.

Fig. 5 visually reinforces the accuracy and reliability of our proposed indoor positioning solution. The close alignment between the predicted and actual positions across a range of locations underscores the model's effectiveness in accurately estimating indoor positions, thus contributing to enhanced indoor localization.

Fig. 6 shows the CDF of our proposed solution compared to the research [15]. The results in Fig. 6 clearly indicate that our solution performs significantly better in predicting locations when compared to the reference paper on the same public dataset and similar situations. Fig. 6 also demonstrates that, for distance errors of two meters or less, our solution achieves an accuracy rate of over 50%, while the research [15] reaches around 30%. These results provide valuable insights into the performance distribution of the solution concerning distance errors. Observing this relationship, it is evident that:

For distance errors less than or equal to 1 meter, the probability of achieving such accuracy is approximately 20%.

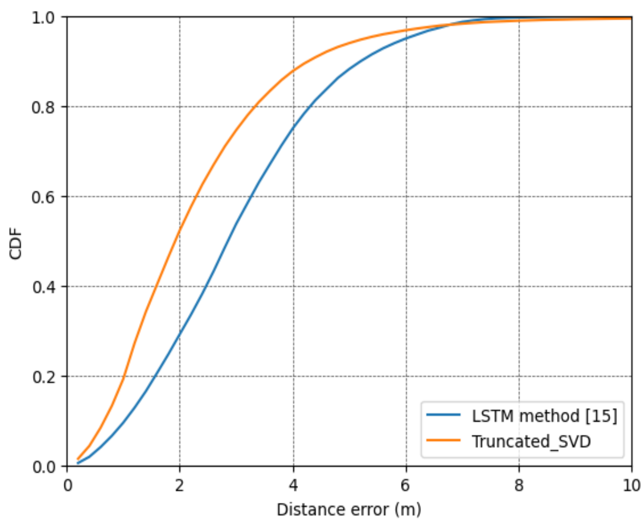

Fig. 6. CDF of the positioning error distance.

Expanding the acceptable error threshold to two meters significantly increases the probability of success to around 52%.

This visualization is further reinforced by the visual representation in Fig. 7, which vividly illustrates the proportion of predictions falling within different error ranges. It is evident that the proposed solution demonstrates a notable capability in achieving sub-2 meters accuracy, making it suitable for a range of indoor positioning applications. In accordance with the evaluation criteria outlined in Section III(C), the performance metrics for our model are as follows: a Root Mean Square Error (RMSE) stand is 1.97 meters, EDE is 2.23 meters, Mean Squared Error (MSE) is calculated at 3.91 meters, and R-squared (R^2) demonstrates a value of 0.69. Additionally, our model shows promising results in terms of error percentages, with 19.41% of errors falling within a 1-meter range and 52.36% within a 2-meter range. The RMSE value is approximately 1.97 meters, indicating a relatively small average deviation between predictions and actual values. This reflects the model's accuracy in estimating positions. The average Euclidean distance is around 2.23 meters, signifying the average difference between predictions and actual values. This is a critical evaluation criterion widely used in location-related tasks. Mean Squared Error (MSE) value is approximately 3.91 meters, representing the average squared error between predictions and actual values. This value indicates the variability of errors and can be used for model comparison. R-squared (R2): The R2 value is approximately 0.69, indicating the model's accuracy in explaining data variance. A high R2 value close to 1 suggests that the model is reasonably good at explaining the data. Percentage of errors within 1 meter: Around 19.41% of predictions have errors within 1 meter, demonstrating that the model achieves a relatively good level of accuracy in predicting positions with errors less than 1 meter. Percentage of errors within 2 meters: Approximately 52.36% of predictions have errors within 2 meters, which is an acceptable threshold for many real-time applications.

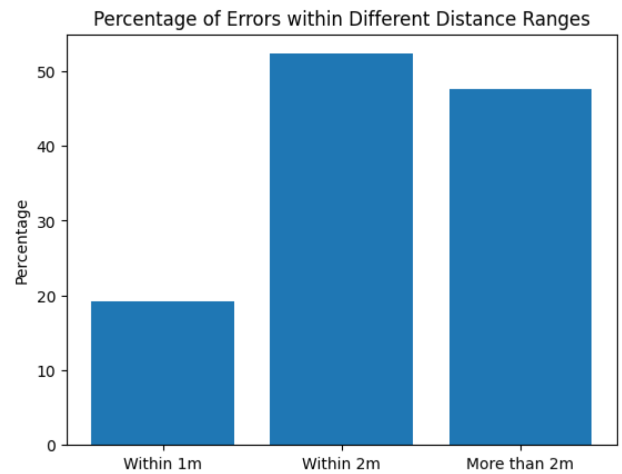

Fig 7. The chart of prediction errors with varying error distances.

In [17], the authors applied deep learning models, specifically RNN/LSTM, to predict indoor positions, achieving accuracy with distance errors ranging from 2.5 meters to 2.7 meters. A significant contribution of our study lies in the substantial enhancement of position prediction accuracy offered by our proposed solution. More precisely, we reduced the error margin from the range of 2.5 meters - 2.7 meters to a mere 2.23 meters, marking a remarkable improvement ranging from 10.8% to 17.4%. The simulation results demonstrate the effectiveness of the Truncated SVD and KNN Regression combination in significantly bolstering position prediction capabilities, particularly in error reduction. Furthermore, our solution also made substantial strides in prediction time efficiency when compared to the findings in reference [17]. As indicated in [17], the training time using the RNN model was 564.1396 seconds, with 10.0848 seconds required for testing. When LSTM was employed, the training process took 581.3599 seconds, and testing consumed 10.1721 seconds. In contrast, our solution demonstrated remarkable efficiency, taking only 4.5523 seconds to complete the same experimental scenario. These results underscore not only the enhancement in accuracy achieved by our approach but also its superior efficiency in prediction time. Additionally, our method simplifies the intricate training process commonly associated with traditional machine learning and deep learning methodologies.

V. Conclusions

In conclusion, our study highlights the significant effectiveness of the proposed solution, which combines Truncated Singular Value Decomposition (Truncated SVD) with K-Nearest Neighbors (KNN) regression for indoor positioning. This innovative approach brings about a substantial improvement in prediction accuracy while meeting real-time requirements. One of the notable advantages lies in the simplicity and efficiency of KNN regression. Unlike traditional machine learning and deep learning solutions that require complex pre-training processes, our method does not burden the user with such complexities. This streamlines the implementation and makes it an attractive choice for various indoor localization scenarios. By integrating Truncated SVD as a dimensionality reduction technique, we enhance the model's robustness and precision. Through rigorous experimentation, we determined that setting Truncated SVD's number of components to 35 minimizes EDEs in predictions, further showcasing the effectiveness of this hybrid approach. This combined methodology not only advances indoor positioning accuracy but also ensures that the solution is practical for real-time applications. In summary, our work presents a powerful and efficient solution for indoor positioning, opening doors to improved location-based services and applications.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

This work was supported by the Hanoi University of Industry (HaUI) under Grant No. 24-2023-RD/HĐ-ĐHCN.

REFERENCES

[1] X. Feng, K. A. Nguyen, Z. J. J. o. I. Luo, and Telecommunication, "A survey of deep learning approaches for WiFi-based indoor positioning," vol. 6, no. 2, pp. 163-216, 2022.

[2] S. A. Zibaei and R. A. Abbaspour, "Evaluation of Improved K-Nearest Neighbors for Indoor Positioning System in Real Complex Buildings," in 2023 9th International Conference on Web Research (ICWR), 2023, pp. 12-19.

[3] X. Zheng, R. Cheng, and Y. Wang, "RSSI-KNN: A RSSI Indoor Localization Approach with KNN," in 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), 2023, pp. 600-604.

[4] Z. Liu, X. Luo, and T. He, "Indoor positioning system based on the improved W-KNN algorithm," in 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2017, pp. 1355-1359.

[5] M. F. Mosleh, R. A. Abd-Alhameed, and O. A. Qasim, "Indoor Positioning Using Adaptive KNN Algorithm Based Fingerprint Technique," in Broadband Communications, Networks, and Systems: 9th International EAI Conference, Broadnets 2018, Faro, Portugal, September 19–20, 2018, Proceedings 9, 2019, pp. 13-21: Springer.

[6] A. Alitaleshi, H. Jazayeriy, and J. J. E. A. o. A. I. Kazemitabar, "EA-CNN: A smart indoor 3D positioning scheme based on Wi-Fi fingerprinting and deep learning," vol. 117, p. 105509, 2023.

[7] S. Tiku, A. Mittal, and S. Pasricha, "A Scalable Framework for Indoor Localization Using Convolutional Neural Networks," in Machine Learning for Indoor Localization and Navigation: Springer, 2023, pp. 159-176.

[8] N. Singh, S. Choe, and R. J. I. A. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," vol. 9, pp. 127150-127174, 2021.

[9] M. Ashraf et al., "A Survey on Dimensionality Reduction Techniques for Time-series Data," 2023.

[10] G. E. Hinton and R. R. J. s. Salakhutdinov, "Reducing the dimensionality of data with neural networks," vol. 313, no. 5786, pp. 504-507, 2006.

[11] W.-C. Lu, Y.-C. Cheng, and S.-H. Fang, "A study of singular value decomposition for wireless LAN location fingerprinting," in 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), 2016, pp. 466-470: IEEE.

[12] A. Li, J. Fu, H. Shen, and S. J. I. I. o. T. J. Sun, "A cluster-principal-component-analysis-based indoor positioning algorithm," vol. 8, no. 1, pp. 187-196, 2020.

[13] S.-H. Fang and T. J. I. T. o. M. C. Lin, "Principal component localization in indoor WLAN environments," vol. 11, no. 1, pp. 100-110, 2011.

[14] H. Li, W. Quan, G. Ji, and Z. Qian, "Wireless indoor positioning algorithm based on PCA," in 2015 International Conference on Artificial Intelligence and Industrial Engineering, 2015, pp. 8-9: Atlantis Press.

[15] W. C. Lu, Y. C. Cheng, and S. H. Fang, "A Study of Singular Value Decomposition for Wireless LAN Location Fingerprinting," in 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), 2016, pp. 466-470.

[16] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. J. D. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," vol. 3, no. 1, p. 3, 2018.

[17] H.-Y. Hsieh, S. W. Prakosa, and J.-S. Leu, "Towards the implementation of recurrent neural network schemes for WiFi fingerprint-based indoor positioning," in 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), 2018, pp. 1-5: IEEE.