İSTANBUL UNIVERSITY
CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2024; 24(1): 218-227

A Machine Learning Approach Based on Indoor Target Positioning by Using Sensor Data Fusion and Improved Cosine Similarity

Serpil Üstebay[1], Zeynep Turgut[1], Şafak Durukan Odabaşı[2], Muhammed Ali Aydın[2], Ahmet Sertbaş[2]

[1]Department of Computer Engineering, Istanbul Medeniyet University Faculty of Engineering and Natural Sciences, Istanbul, Turkiye
[2]Department of Computer Engineering, Istanbul University-Cerrahpaşa Faculty of Engineering, Istanbul, Turkiye

Cite this article as: S. Üstebay, Z. Turgut, Ş. Durukan Odabaşı, M. A. Aydın and A. Sertbaş, "A machine learning approach based on indoor target positioning by using sensor data fusion and improved cosine similarity," *Electrica*, 24(1), 218-227, 2024.

ABSTRACT

Indoor user positioning is a crucial problem in modern life. It has wide usage in health, security, smart homes, etc. Global positioning system (GPS) is used outdoors, and it does not work effectively in indoor areas since many things can degrade GPS positioning accuracy. All solutions for indoor areas aim to provide low-cost and high-accuracy positioning. In this study, a low-cost indoor positioning algorithm is developed. The fingerprint signal map of the building is measured with built-in digital sensors in smart devices. The measurements consist of Wi-Fi, bluetooth low energy, and magnetic field signals called data fusion. During the positioning phase, the proposed model, called improved cosine similarity, uses the cosine similarity and information gain method. Digital magnetometers measure magnetic fields with different approaches. In the proposed method, Kalman filter is used to reduce noise magnetic field signals since this variety can give rise to mistaken positioning. To compare the effectiveness of the proposed method, it was compared to K-nearest neighbor, support vector machines, linear discriminant analysis, artificial neural networks, decision trees, N-near neighbor, and binned neighbor algorithm. Based on the experimental data, it was concluded that the proposed architecture achieved higher accuracy rates by reducing distortion.

Index Terms—Kalman filter, multidimensional signal processing, multiple signal classification, sensor data fusion, simultaneous localization

I. INTRODUCTION

Positioning technologies, such as global positioning system (GPS), are commonly used for outdoor locations, but their accuracy can be affected by various factors. Indoor positioning systems (IPS) have been developed as solutions for indoor areas using technologies like Bluetooth, Wi-Fi, radiofrequency identification (RFID), ultrasound, ultra-wideband (UWB), and light signals. Each signal type has its own advantages and disadvantages, and there is no universal standard for IPS. These systems have diverse applications in various fields, such as airports, shopping malls, offices, and hospitals, providing benefits like guiding individuals and locating personnel/equipment. With the increasing popularity of location-based applications, the demand for accurate indoor positioning is growing. In this study, we developed a new machine learning algorithm-based positioning system that utilizes device-embedded sensors [Wi-Fi, Bluetooth low energy (BLE), magnetometer] to detect users' positions in indoor areas.

Positioning involves obtaining the geographic location information of a target using technologies like satellites, ultrasound, and UWB. Global positioning systems are commonly used for outdoor positioning, but their accuracy can be affected by factors like satellite geometry, atmospheric conditions, and receiver quality. Global positioning system devices calculate a target's position by measuring the distance from multiple GPS satellites. While GPS-enabled smartphones are generally accurate to within 4.9 m in open spaces, their accuracy may decrease near buildings, trees, and bridges. For indoor positioning, solutions based on Bluetooth, Wi-Fi, RFID, ultrasound, UWB, and light signals have been developed. Transmitter devices are installed in indoor spaces to cover the entire area, while receivers search for positions based on the received signal. Each signal type has its advantages and disadvantages, and there is no one-size-fits-all standard for IPS. Indoor positioning systems have a wide range of applications, from directing passengers and customers to the right locations at airports, shopping malls, and railway stations to finding or tracking personnel and equipment in offices, trade fairs, and

Corresponding author:
Şafak Durukan Odabaşı

E-mail:
safak.odabasi@iuc.edu.tr

Received: June 14, 2023

Revision Requested: July 14, 2023

Accepted: October 23, 2023

Publication Date: November 27, 2023

DOI: 10.5152/electrica.2023.23080

hospitals. With the rise of location-based applications in smart living, positioning systems will continue to evolve to provide more accurate results.

Indoor positioning techniques have become more important due to our growing dependence on smart devices and indoor activities. Recent studies have proposed various technologies, including Wi-Fi fingerprinting, device-based systems, heterogeneous approaches, and visible light communication (VLC)-based systems, each evaluated based on accuracy, complexity, energy efficiency, and cost. Researchers aim to enhance the performance and applicability of these technologies in diverse indoor environments through analysis and evaluation.

He and Chan [1] reviewed Wi-Fi-based indoor positioning technologies, highlighting Wi-Fi as a promising alternative to GPS. Xiao et al. [2] categorized IPS into device-based and device-free approaches. Yassin et al. [3] discussed indoor positioning techniques focusing on methodology and concepts, while Jang and Kim [4] introduced offline fingerprint-free indoor positioning technologies. Zafari et al. [5] explored user and device positioning techniques, and Guo Xiansheng et al. [6] surveyed fusion-based IPS, analyzing their characteristics.

The review [7] assessed indoor positioning technologies based on VLC, highlighting VLC's advantages in speed, latency, and security. It presented existing VLC-based positioning systems, including light-emitting diode and camera-based approaches, evaluating their accuracy and complexity. The study identified challenges like ambient light interference and proposed solutions. Evaluation criteria encompassed accuracy, complexity, reliability, and energy efficiency.

Federated filtering is a vital technology for real-time indoor positioning, utilizing sensor data and machine learning algorithms to track individuals accurately by considering their unique characteristics. It offers privacy and security and finds applications in wayfinding, security, and facility management. Additionally, a proposed Voxel-Scale-Invariant Feature Transform(SIFT)-based algorithm [8] aims to enhance the efficiency of Light Detection and Ranging registration, providing high-precision indoor navigation. The method combines three sensors, federated filtering, and multiple algorithms to achieve improved indoor positioning accuracy for mobile robots.

[9] addresses indoor positioning challenges with Global Navigation Satellite System by using distributed sensors and a range-azimuth sensor. Their method combines primary and secondary data, improving position estimates with a recursive least squares framework. Simulation results align with the Cramer–Rao lower bound.

Fingerprint signal mapping is a two-phase mobile positioning technique for indoor environments [10]. In the offline stage, signal strength measurements from wireless access points (WAPs) are recorded at reference points and stored in a database. During the online phase, a smart device captures received signal strength (RSS) values, which are then used by a processing unit (PU) along with a fingerprinting signal map and machine learning techniques to predict the user's indoor location. The PU shares the predicted location with the target or building administrator.

Data fusion plays a critical role in enhancing IPS by integrating data from various sensors, including Wi-Fi, Bluetooth, magnetic field, and accelerometers [11, 12]. This approach improves the accuracy and robustness of positioning information in complex indoor environments, overcoming challenges like multipath and interference. Data fusion has broad applications, such as mobile robot navigation and augmented reality, and has the potential to transform indoor navigation and interaction.

Accurate positioning is the primary challenge in IPS [13]. Data fusion, as defined by the JDL Data Fusion Group, involves associating, correlating, and combining data from various sources to refine position and identity estimates [13]. Data fusion is categorized into three groups: measurement fusion, feature-level fusion, and decision-level fusion [14] (Fig. 1). Measurement fusion directly combines sensor data and performs feature extraction, while feature-level fusion converts sensor data into a single feature vector after extraction. Decision-level fusion utilizes different decision models for the final decision.

The JDL group categorizes data fusion processing for IPS into five levels: source pre-processing, object refinement, situation assessment, impact assessment, and process refinement [12]. In [15], researchers use a fusion model that combines radio signals and pedestrian dead reckoning (PDR) to reduce positioning errors in an office building from 2.25 - 4.75 m to 1.5 m.

[16] implemented a 2-dimensional PDR system using a smartphone and enhanced it with a map-matching algorithm and fusion of the

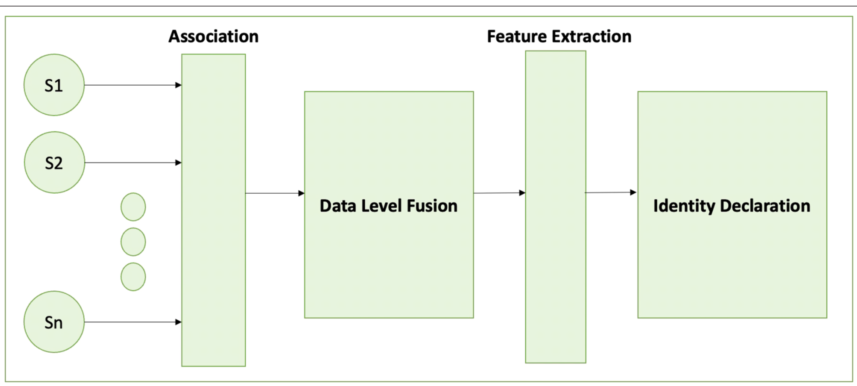


Fig. 1. Data fusion classification.

smartphone's digital barometer data, enabling a transition to a 3-dimensional PDR system.

[17] proposed a data fusion approach using motion and body activity information for indoor localization, achieving high accuracy in recognizing body activity and varying localization accuracy in different test scenarios.

[18] presented a fusion-based indoor positioning algorithm that incorporates Bluetooth, Wi-Fi, and RFID signals. Using cosine similarity and the adaptive weighted K-nearest neighbor (KNN) algorithm, they achieved improved accuracy compared to single-data-based solutions in a test environment with multiple wireless APs, RFID tags, and Wi-Fi beacons.

[19] employed decision-level fusion (Fig. 1), utilizing Kalman filtering to reduce Wi-Fi signal distortion. They clustered reference positions and trained random forest models for each group, which were then combined for mobile user localization.

Researchers strive for accurate indoor positioning of mobile users, preferring machine learning algorithms that offer high accuracy with limited data requirements. In studies such as [20] and [21], cosine similarity consistently outperformed other methods when comparing various similarity measurements for indoor positioning. Building on this, we propose an improved version of the cosine similarity method [22] to enhance the accuracy of indoor positioning models.

In this study, we have developed a new machine learning algorithm-based positioning system that detects users' positions in indoor areas by collecting signals with device-embedded sensors such as Wi-Fi, BLE, and magnetometer. To summarize, our contributions are as follows:

- We propose a new machine-learning algorithm to detect users' positions in indoor fields by collecting signals with device-embedded sensors like Wi-Fi, BLE, and magnetometer. We show that this algorithm can detect location with an accuracy of 93%.
- We prove that data fusion is more potent than using just one type of data. The performances of fusion data are analyzed in terms of predicting users'' location accuracy.
- We improve the cosine similarity method by using the information gain (IG) to weigh the impact of each data fusion. Our experiments indicate that the improved cosine similarity method can detect similarities more accurately than classic cosine similarity.
- We evaluate the performance of the proposed architecture using experimental data. Also, we use the Kalman filter to reduce distorted measurement. We find that the proposed architecture can detect location with an accuracy rate of 93% by reducing distortion, which is higher compared to the other methods.

The following sections of the paper will introduce the proposed model in Section II, followed by experimental results and system performance in Section III. We will then analyze the results and conclude in Section IV and Section V, respectively.

II. MATERIAL AND METHODS

This section contains comprehensive information about the fingerprinting method, the obtained signal map, and the test environment structure. Also, IG, cosine similarity, and Kalman filter methods are described. Besides, we detail the proposed model.

Fig. 2. Test environment floor plan.

A. Database

In this paper, we use the fourth floor of a university building as a test environment. The floor plan is shown in Fig. 2. The test environment covers an area of 800 m^2, and its rooms are formed by using glass and concrete walls. Eight WAPs and four beacon devices are placed in different locations.

Firstly, reference points are determined in indoor fields. At this reference point, signal values are captured by using a smart device, and a fingerprinting signal map is obtained. The related signal map is used to train different machine-learning techniques for IPS. We captured RSS using four smart devices listed in Table I.

Received signal strength indicator (RSSI) and dBm are used for signal strength measurements and are different units of measurement that represent the same thing. While RSSI is a relative index, dBm is an absolute number representing power levels in mW. Received signal strength indicator is a term used to measure the relative quality of a received signal for a client device, but it does not have a definite value. Institute of Electrical and Electronics Engineers 802.11 standard specifies that RSSI can be on a scale from 0 to 255 and that each chipset manufacturer can define its own "RSSI_Max" value. In the test environment, Wi-Fi and Bluetooth signal measurements were obtained in dBm.

Beacons are Bluetooth radio-transmitter devices that operate with BLE specifications, have long-lasting battery consumption, and emit BLE signals at specific intervals. Each beacon sensor has its own coverage area. We captured signal strength (dBm) and the identification number of the device to create the signal map.

The Earth's magnetic field is a magnetic dipole field with an angular area of 11.5 degrees relative to the Earth's axis of rotation, as if it were a bar magnet placed at this angle in the center of the Earth.

TABLE I MOBILE DEVICE LIST

ID	Company/Model	Android Version
1	Sony Xperia	6.1
2	Xiaomi Mi 5 Prime	6.1
3	LG G4	6.0
4	LG G3	6.0

If electromotive force (EMF) is measured via an application from built-in digital sensors, X, Y, and Z values can be collected [23]. X represents northern intensity, Y represents eastern density, and Z represents vertical density.

The approaches used to measure the magnetic signals of the Earth are the Hall effect, giant magnetoresistance magnetometer sensors, magnetic tunnel junctions method, anisotropic magneto resistance, and Lorentz force sensor [24]. Each approach has its advantages and disadvantages. Digital sensors integrated into devices use one of these approaches according to brand and model. Therefore, signal measurements made at the same point may vary according to the approach used in the sensor.

B. Cosine Similarity

Cosine similarity tries to determine the relationship between two vectors in terms of the angle they form [25]. If the vectors are the same, the angle between them is 0, and if the vectors are different from each other, the angle between them becomes 0 [26]. If α and β are two n-dimensional vectors, the cosine of the angle between them is calculated by Eq. (1).

$$cosine\_sim(\alpha,\beta)=\sum_i(\alpha_i.\beta_i)/\left(\sqrt{\sum_i\alpha_i^2}\sqrt{\sum_i\beta_i^2}\right) \quad (1)$$

Since cos (0) = 1 and cos (180) = −1, the cosine value close to 1 represents the similarity of the two vectors.

C. Information Gain

The objective of the information-gaining method [27] is to find conditional variables that affect the decision variable in a data set consisting of many conditional variables. It is a measure of the use of multivariate analysis. An analysis is made according to entropy calculation. The information entropy of a random variable is used to measure the degree of its impairment. The entropy of class C is calculated as Eq. (2).

$$E(C)=\sum_{c\in C}p(c)log_2^{(p(c))} \quad (2)$$

$P(c)$ is the probability density function for the random variable C. The entropy of C conditioned on A is written as $E(C|A)$ and calculated as Eq. (3).

$$E(C\,|\,A=a)=\sum_{a\in A}p(C\,|\,A=a)E(C\,|\,A=a) \quad (3)$$

The IG value between the A feature and class C is calculated as Eq. (4). The feature with the highest IG value is the most powerful feature of the decision variable in the dataset. Also, a feature reduction method is used for selecting the best k feature.

$$IG(A)=E(C)-E(CA) \quad (4)$$

D. Kalman Filter

Kalman filter is the most important discovery of the 20th century. Although it is named a filter, we use it in linear systems to guess the next step. Its recursive structure (re-inputting the outputs into the filter) is the only filter that minimizes the estimation error in the existing filters. Kalman filter has two equations for estimation and correction [27-29] The estimation equation is shown in Eq. (5).

$$x_k=Ax_{k-1}+Bu_k+w_{k-1} \quad (5)$$

The measurement value of a signal (x_k) is obtained from the previous case x_{k-1}. The control signal is named u_k, and w_{k-1} is the noise of the previous measurement. A, B, and H indicate general representations of matrices. These values can be treated as numerical numbers. A matrix represents a state transition model, a B matrix represents a controlled model, and the H matrix represents the measurement model.

$$Z_k=Hx_k+v_k \quad (6)$$

The measurable value of a signal consists of a linear combination of the measured values of x_k, v_k, and w_{k-1} at Eq. (6).

III. EVALUATION

The constructed fingerprinting signal map is called TestDB, and it consists of 1664 signal measurements from seven rooms, which are separated from each other through glass or concrete walls. Missing values (NaN) are deleted from TestDB. Each vector consists of signals shown in Eq. (7), Eq. (8), and Eq. (9). A combination of vectors is presented in Eq. (10).

$$Wifi^{\rightarrow}=\left[RSS_1,...,RSS_n\right] \quad (7)$$

$$BLE^{\rightarrow}=\left[BLE_1,...,BLE_m\right] \quad (8)$$

$$EMF^{\rightarrow}=\left[X,Y,Z,G\right] \quad (9)$$

$$SM^{\rightarrow}=Wifi^{\rightarrow}\cup BLE^{\rightarrow}\cup EMF^{\rightarrow} \quad (10)$$

Wifi$^{\rightarrow}$ represents received Wi-Fi signal strength, BLE$^{\rightarrow}$ represents received Bluetooth signal strength from 4 beacon devices, and EMF is signal values of digital sensors integrated into mobile phones. EMF$^{\rightarrow}$ values are precise and can easily be affected by the environment. We get different EMF$^{\rightarrow}$ values with two different devices at the same reference points under the same conditions. We assume that variation is caused by the digital sensor's measurement approaches. SM$^{\rightarrow}$ shown in Eq. (10), is a fusion data vector that combines Wifi$^{\rightarrow}$, BLE$^{\rightarrow}$, and EMF$^{\rightarrow}$ signals at the same reference point. TestDB, at Eq. (11), encompasses 1664 fusion data vector measurements.

$$TestDB=\left[SM^{\rightarrow}{}_1 SM^{\rightarrow}{}_2 \vdots SM^{\rightarrow}{}_t\right]=\left[RSS_1\ldots G_1 \vdots \cdot \vdots RSS_t\ldots G_t\right] \quad (11)$$

The entropy of each feature (Eq. (12)) is calculated to gather the contribution of distinct data to indoor positioning. Entropy values help us to find IG, which is broadly used for feature reduction. Entropy is a measure of the randomness or disorder of a system. High entropy is denotative that a related feature has a low impact on classification, or else it means high impact.

$$E(T_x)=-\sum_i^c p_i * log_2(p_i) \quad (12)$$

$P=[p_1, p_2, \ldots p_c]$ is the probability distribution of the TestDB matrix, whose formula is shown in Eq.(13), and it is calculated according to class labels (c). Each column of TestDB is subdivided as T_1, T_2, ... T_{16} to calculate the conditional entropy value of each feature. So that means each T_x is a feature of TestDB.

$$E(TestDB\,|\,T_x)=\sum_i\frac{T_{xi}}{T_x}E(T_{xi}) \quad (13)$$

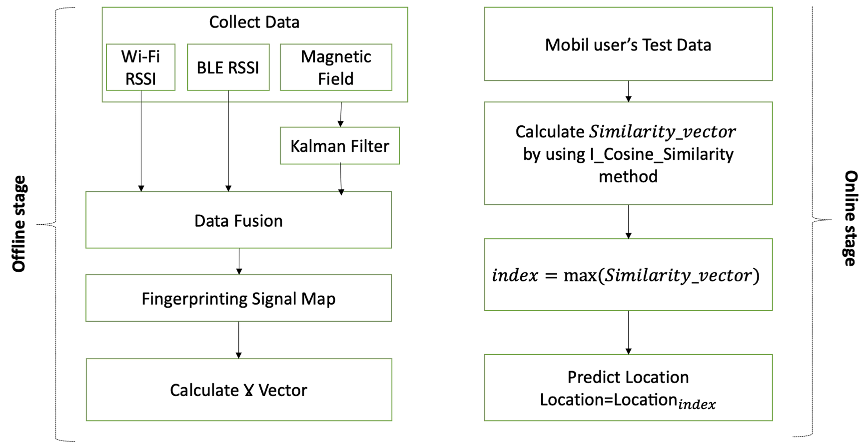


Fig. 3. Flow chart of the developed indoor localization system.

$$IG\left(TestDB, T_x\right) = E\left(T_x\right) - E(TestDB \mid T_x) \tag{14}$$

$$\gamma = \left[IG_1, IG_2, \ldots, IG_k\right] \tag{15}$$

The impact of T_x feature on the classification is called IG, which is calculated as shown in Eq. (14). The γ vector [Eq. (15)] stores TestDB's IG for each feature. The γ values range from 0 to 1. 1 indicates that the feature has a high contribution to classifying data. Conversely, 0 means that the feature has a low contribution.

In this study, we adopt the IG method to the cosine similarity. We name this new method the improved cosine similarity method (*I* – Cos), which is calculated using Eq. (16).

$$I - Cos\left(\propto, \beta, \gamma\right) = \frac{\sum \alpha_i, \beta_i, \gamma_i}{\sqrt{\sum_i \alpha_i^2 \gamma_i}\sqrt{\sum_i \beta_i^2 \gamma_i}} \tag{16}$$

Fig. 3 shows the flow chart of the proposed positioning model. The model starts in the offline phase. At that phase, RSS values are captured at the reference points. We applied the Kalman filter to reduce the distortion of EFM signals. The output of that stage is the fingerprint fusion signal map. Target positioning is done at the online stage. Target's smart device captures received signal values. These are sent to a positioning server. The positioning server predicts the target's position using the I-Cos similarity method.

A. Results

We divide TesDB into two parts. We use the first part to train the model, and the second part is used to test the model. We ensure that both parts include all reference point measurements. Non-error rate (NER) for the model's performance metric is used at the bottom.

$$NER = \frac{TP}{TP + FP} \tag{17}$$

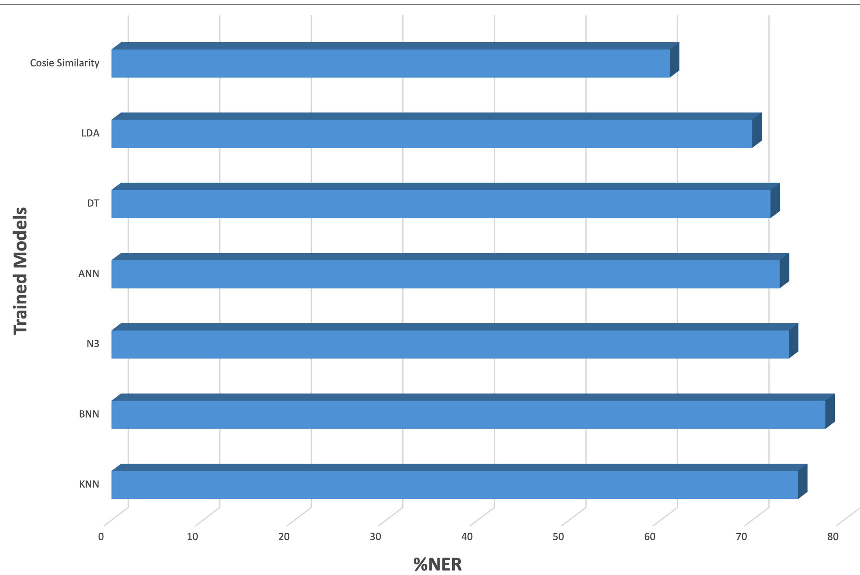


Fig. 4. Comparison of accuracy rates obtained from positioning models created with TestDB Wi-Fi signals.

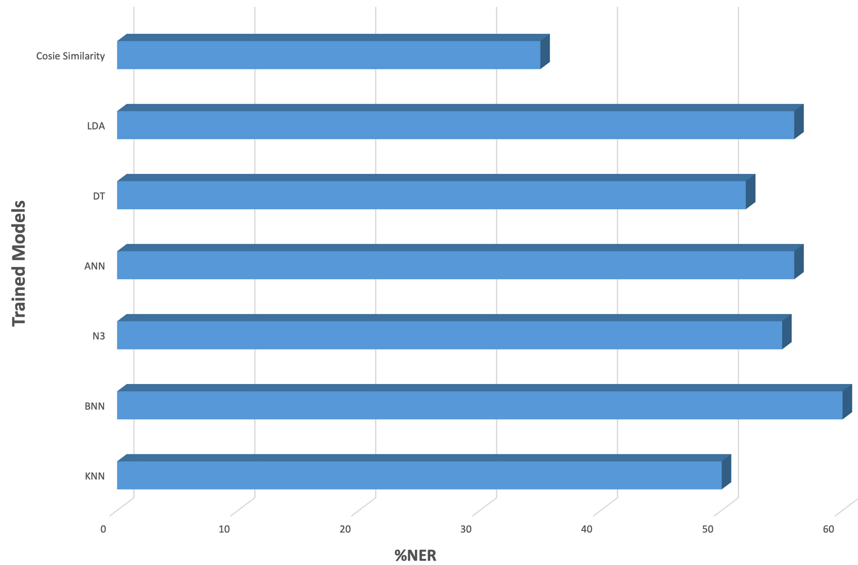


Fig. 5. Comparison of accuracy rates obtained from positioning models created with TestDB BLE signals.

Here, TP is the number of true predicted samples, and FP is the number of false predicted ones. In the first step of the study, we want to demonstrate the power of the fusion data set. Therefore, TestDB signal sources are used separately. In the fusion data set, only Wi-Fi signals, BLE signals, and EFM values are used. We trained all groups using KNN, binned neighbor algorithm (BNN) [29], N3 [30], N-near neighbor, decision trees, support vector machines (SVM), linear discriminant analysis (LDA), and cosine similarity methods, respectively. The ratio of the training data set is 80%, and the testing data set is 20%. Models are coded by using Python 2.7 programming language.

Positioning test results tested by using Wi-Fi signals are shown in Fig. 4. According to the different tests, the best accuracies are obtained with these parameters: Neighbor number is 4 for KNN, BNN alpha value is 2.50, and N3 alpha value is 1.25. Artificial neural network (ANN) hidden layer size is chosen as 16. Gaussian kernel function in the SVM algorithm and linear kernel function in the LDA algorithm are used. The highest accuracy is obtained by using the BNN algorithm. The same models are trained using BLE signals, and the results are shown in Fig. 5. The highest accuracy is obtained with the BNN algorithm (59% NER), and the lowest accuracy value is obtained by the cosine similarity method. Positioning models by using EFM signal test results are shown in Fig. 6. Similarly, the BNN algorithm retains its success with 46.94% NER. Cosine similarity is the worst algorithm by 27% NER. It has been determined that relying on a single data source is insufficient when attempting to pinpoint a target within an indoor setting. Hence, the indoor positioning models should be developed using fusion data instead of sole data.

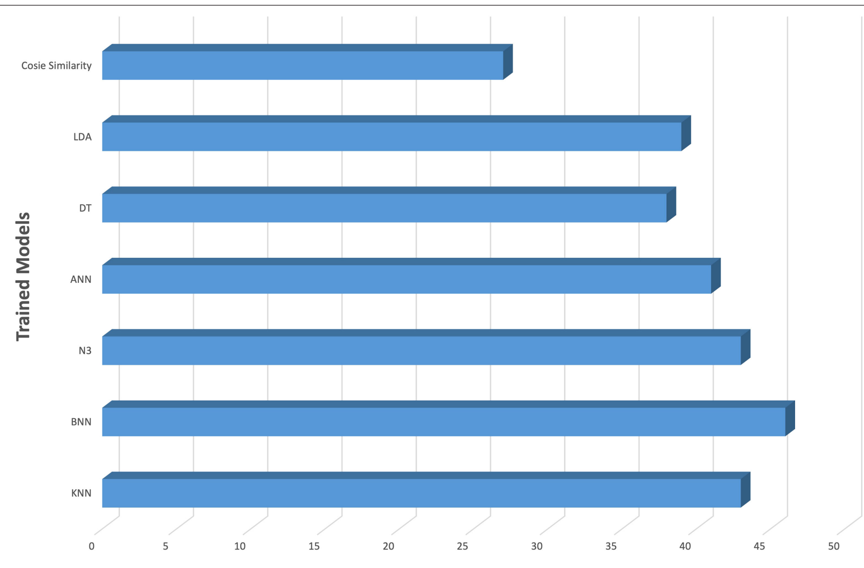


Fig. 6. Comparison of accuracy rates obtained from positioning models created with TestDB magnetic field signals.

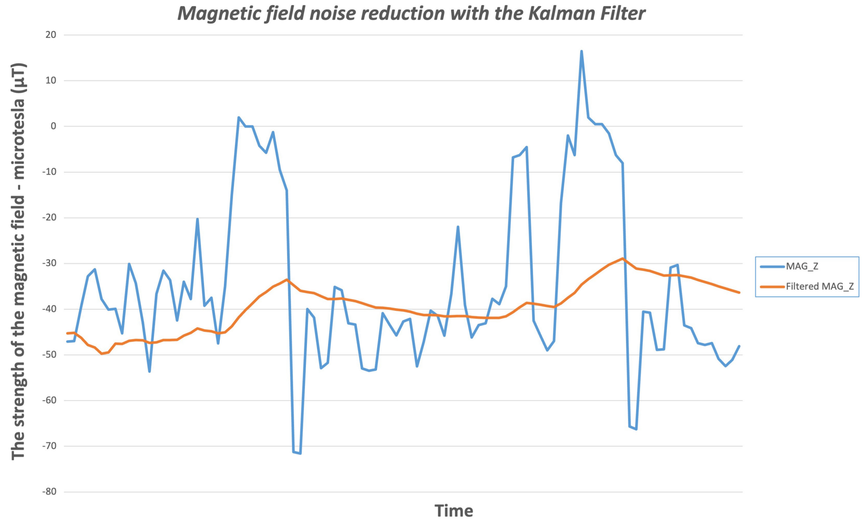


Fig. 7. Magnetic field noise reduction with the Kalman filter.

TABLE II COSINE METHOD RESULTS DEVELOPED WITH TESTDB

Training Partition	Cosine Similarity	I-Cos Similarity	Proposed Method
30%	76.469	78.617	82.626
40%	79.165	81.466	85.815
50%	81.639	83.54	88.223
60%	83.416	85.248	89.67
70%	85.521	87.152	90.547
80%	86.171	88.115	91.968
90%	87.865	89.764	93.253

Similar results are obtained among machine learning algorithms independent of the data source. The highest accuracy is obtained with the BNN algorithm, and the lowest accuracy is obtained with SVM and LDA algorithms. BNN, N3, and KNN algorithms are based on Euclidean distance. But BNN and N3 methods need more processing time than KNN. We do not recommend it for IPS.

In all three cases, the cosine similarity presents the lowest accuracy rates. We aimed to increase the accuracy rate of the cosine similarity method. For this purpose, we have designed a three-step model. In the first step, the data fusion technique was selected for indoor mapping. Secondly, the Kalman filter is chosen to reduce the noises caused by measurement approaches of digital sensors. A and H matrices are treated as numerical values, and the values are assigned as $A=1$ and $H=1$. We ignore the B value by assigning a 0 value to it. The process noise is set to 1e-5, and the measurement noise is set to 1e-2. A sample Kalman filtering result is shown in Fig. 7.

In the last step, we boost the cosine similarity method with IG. All results obtained are compared in Table II. Thus, a higher accuracy rate is obtained when the I-Cos method is applied. The results are shown in Fig. 8. Comparative test results are listed in Table II.

Cosine similarity and improved cosine similarity process time (microseconds). Fig. 9 shows the processing time for determining the position of the target. The horizontal axis indicates the ratio of training data to TestDB. The vertical axis represents the processing time in microseconds to locate a target. In the improved cosine similarity method, the gain vector used as the weight causes an increase in process time. Table III shows the NER of different positioning models. The test results

Fig. 8. Accuracy results with cosine similarity, improved cosine similarity, and Kalman filter-improved cosine similarity.

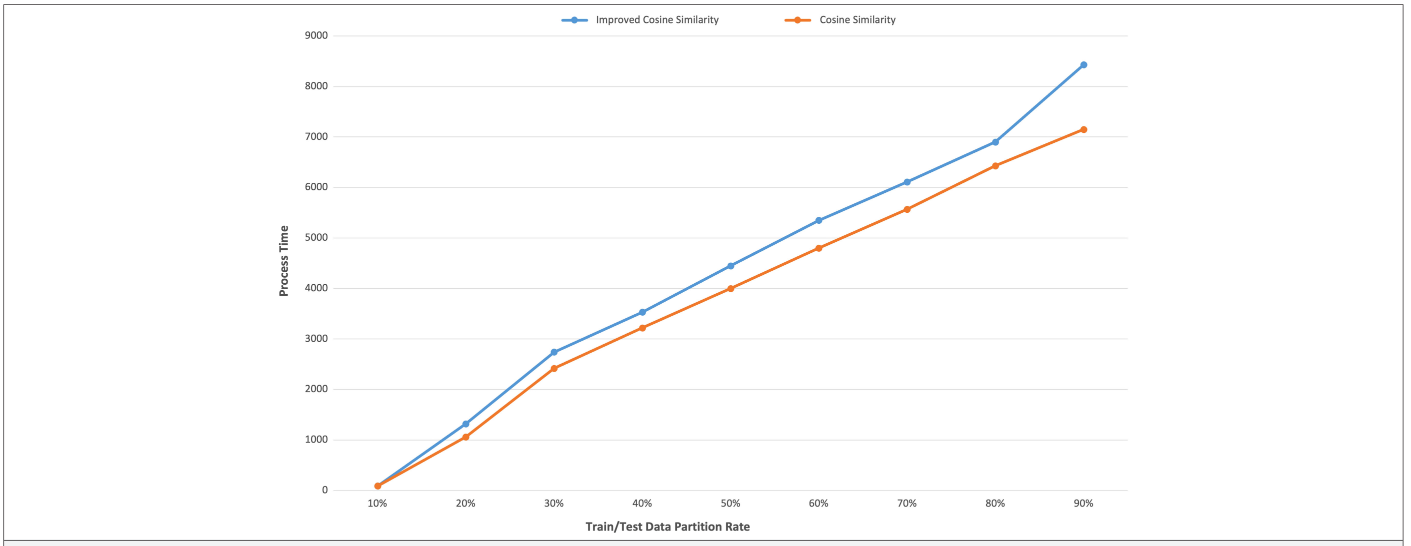


Fig. 9. Cosine similarity and Improved cosine similarity process time (microseconds).

show that data fusion provides higher positioning accuracy than a single data structure. The cosine similarity method with data fusion achieves 87% NER. The lowest accuracy value was obtained with ANN by 80%. Artificial neural network needs much more data for the model training. Creating a Fingerprinting signal map is a low-cost method according to the device, but it requires too much time.

IV. DISCUSSION

The test results show that the proposed method model achieves 93% accuracy, while cosine similarity achieves 89% accuracy. We increased the accuracy of the model by 7%. Both have the same complexity. However, extra multiplications caused an increase in processing time. Due to this, target positioning is determined later than the original method. This delay can be minimized through the use of parallel programming techniques, such as OpenMP and Cuda. An indoor positioning algorithm has been tackled in [17], which fuses Bluetooth, Wi-Fi, and RFID data. In this paper, in addition to Wi-Fi and RFID, we used EMF values. In the literature, the cosine similarity algorithm is used as equipment dependent [19]. They use Kalman filter on Wi-Fi signals. However, we propose a solution for indoor positioning by using fusion at the decision level. We prefer to use Kalman filter on EMF signals because we don't want to lose the original EMF signal value of the environment because of environmental effects. Additionally, we did not test the proposed method with deep learning methods due to the size of our dataset. It can be considered the main limitation of the study.

TABLE III COMPARATIVE NER TEST RESULTS OF PROPOSED METHOD AND OTHER ML METHODS

	Training/Testing Partition						
Algorithm	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
KNN	72	75	77	79	80	82	83
BNN	73	76	77	78	80	81	82
N3	72	74	76	77	79	79	81
ANN	72	74	76	78	78	79	80
DT	77	79	81	83	85	85	86
SVM	77	80	82	83	84	85	86
LDA	68	69	69	69	69	69	70
Cosine similarity	76	79	81	83	85	86	87
I-Cos similarity	78	81	83	85	87	88	89
Proposed method	82	85	88	89	90	91	93

Abbreviations: KNN, K-Nearest Neighbors Algorithm; BNN, Binned Neighbor Algorithm; N3, N-Near Neighbor; ANN, Artificial Neural Networks; DT, Decision Trees; SVM, Support Vector Machines; LDA, Linear Discriminant Analysis.

V. CONCLUSION

In this study, a new classification model is proposed for IPS. This model is designed as a fusion of different data types. Wi-Fi, BLE, and EMF data were applied to the developed I-Cos method, which consisted of a combination of IG and cosine similarity methods. Kalman filter is also used to clean noise from digital sensors that measure with different approaches in mobile devices. In the test environment created for the developed positioning system, the positions of the rooms are determined correctly with a 93% accuracy rate. As the signal map obtained from the test environment grows, the accuracy of the developed model increases. As a future study, a more extensive dataset is planned to be created to provide more accurate positioning on a centimeter basis.

Peer-review: Externally peer-reviewed.

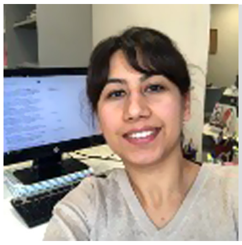
Author Contributions: Concept – S.Ü., Z.T., Ş.D.O., M.A.A., A.S.; Design – S.Ü., Z.T., Ş.D.O., M.A.A., A.S.; Supervision – S.Ü., M.A.A., A.S.; Funding – S.Ü.; Materials – S.Ü., Z.T.; Data Collection and/or Processing – S.Ü., Z.T.; Analysis and/or Interpretation – S.Ü., Z.T.; Literature Review – S.Ü., Ş.D.O.; Writing – S.Ü., Ş.D.O.; Critical Review – S.Ü., Ş.D.O.

Declaration of Interests: The authors have no conflict of interest to declare.

Funding: The authors declared that this study has received no financial support.

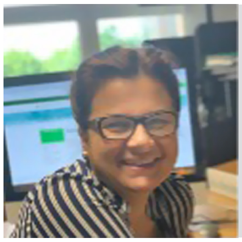
REFERENCES

1. S. He, and S.-H. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surv. Tutor.*, vol. 18, no. 1, pp. 466–490, 2015. [CrossRef]
2. J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv. (CSUR)*, vol. 49, no. 2, pp. 1–31, 2017. [CrossRef]
3. A. Yassin *et al.*, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Commun. Surv. Tutor.*, vol. 19, no. 2, pp. 1327–1346, 2016. [CrossRef]
4. H. Jang, "Kim, Indoor positioning technologies without offline fingerprinting map: A survey," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 1, pp. 508–525, 2018.
5. F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 3, pp. 2568–2599, 2019. [CrossRef]
6. X. Guo, N. Ansari, F. Hu, Y. Shao, N. R. Elikplim, and L. Li, "A survey on fusion-based indoor positioning," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 1, pp. 566–594, 2019. [CrossRef]
7. A. Motroni, A. Buffi, and P. Nepa, "A survey on indoor vehicle localization through RFID technology," *IEEE Access*, vol. 9, pp. 17921–17942, 2021. [CrossRef]
8. H. Li, L. Ao, H. Guo, and X. Yan, "Indoor multi-sensor fusion positioning based on federated filtering," *Measurement*, vol. 154, p. 107506, 2020. [CrossRef]
9. B. Pardhasaradhi, G. Srinath, J. Raghu, and P. Srihari, "Position estimation in indoor using networked GNSS sensors and a range-azimuth sensor," *Inf. Fusion*, vol. 89, pp. 189–197, 2023. [CrossRef]
10. P. Bahl, and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," Proc. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 2, pp. 775–784, 2000. [CrossRef]
11. H. Mehrabian, and R. Ravanmehr, "Sensor fusion for indoor positioning system through improved RSSI and PDR methods," *Future Gener. Comput. Syst.*, vol. 138, pp. 254–269, 2023. [CrossRef]
12. H. B. Mitchell, *Multi-sensor Data Fusion: An Introduction*. Springer Science & Business Media, 2007.
13. E. Franklin, *Data Fusion Lexicon, Joint Directors of Laboratories, Technical Panel for C3, Data Fusion Sub-Panel*, San Diego: Naval Ocean Systems Center, 1987.
14. P. Juntama, White Jr., *Multi-sensor Data Fusion Techno Briefing, Data & Analytics, Technology*, 2014.
15. F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, p. 704504, 2013. [CrossRef]
16. Y. Li, Y. Zhuang, P. Zhang, H. Lan, X. Niu, and N. El-Sheimy, "An improved inertial/wifi/magnetic fusion structure for indoor navigation," *Inf. Fusion*, vol. 34, pp. 101–119, 2017. [CrossRef]
17. R. Faragher, and R. Harle, "SmartSLAM-an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing," In *Ion GNSS*, vol. 13, pp. 1–14, 2013.
18. Y. Li, Z. He, J. Nielsen, and G. Lachapelle, "Using wi-fi/magnetometers for indoor location and personal navigation," International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE Publications, 2015. [CrossRef]
19. X. Guo, W. Shao, F. Zhao, Q. Wang, D. Li, and H. Luo, "WiMag: Multimode fusion localization system based on Magnetic/WiFi/PDR," International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE Publications, 2016. [CrossRef]
20. C. C. Chiu, J. C. Hsu, and J. S. Leu, "Implementation and analysis of Hybrid Wireless Indoor Positioning with iBeacon and wi-fi," 8th International Congress on Ultra Modern Telecommunications and Control. Systems and Workshops (ICUMT). IEEE Publications, 2016, pp. 80–84. [CrossRef]
21. L. Kanaris, A. Kokkinis, A. Liotta, and S. Stavrou, "Fusing Bluetooth Beacon Data with wi-fi Radio maps for Improved Indoor Localization," *Sensors (Basel)*, vol. 17, no. 4, p. 812, 2017. [CrossRef]
22. G. De Blasio, A. Quesada-Arencibia, C. R. García, R. Moreno-Díaz, and J. C. Rodríguez-Rodríguez, "Analysis of distance and similarity metrics in indoor positioning based on Bluetooth low energy," International Conference on Ubiquitous Computing and Ambient Intelligence. Cham: Springer, 2017, pp. 213–224.
23. J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, 2015. [CrossRef]
24. Y. Cai, Y. Zhao, X. Ding, and J. Fennelly, "Magnetometer basics for mobile phone applications," *Electron. Prod.*, vol. 54, no. 2, 2012.
25. J. Li, J. Fu, A. Li, W. Bao, and Z. Gao, *An Improved WKNN Indoor Fingerprinting Positioning Algorithm Based on Adaptive Hierarchical Clustering, Advanced Computational Methods in Life System Modeling and Simulation*. Singapore: Springer, 2017, pp. 253–262.
26. J. Yang, Y. Li, W. Cheng, Y. Liu, and C. Liu, "EKF–GPR-based fingerprint renovation for subset-based indoor localization with adjusted cosine similarity," *Sensors (Basel)*, vol. 18, no. 1, p. 318, 2018. [CrossRef]
27. F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic Cosine Similarity," The 7th International Student Conference on Advanced Science and Technology ICAST. Seoul, South Korea, 2012.
28. G. Bishop, and G. Welch, "An introduction to the Kalman filter," *Proc of SIGGRAPH, Course 8*, 2001.
29. C. Lee, and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Inf. Process. Manag.*, vol. 42, no. 1, pp. 155–165, 2006. [CrossRef]
30. R. Todeschini, D. Ballabio, M. Cassotti, and V. Consonni, "N3 and BNN: Two new similarity-based classification methods in comparison with other classifiers," *J. Chem. Inf. Model.*, vol. 55, no. 11, pp. 2365–2374, 2015. [CrossRef]



Serpil Üstebay was born in Türkiye. She received Ph.D. degree from İstanbul University, in 2018. She currently works at Computer Engineering Department, İstanbul Medeniyet University as an Assistant Professor. Her research interests include machine learning, deep neural network, and cyber security.

Zeynep Turgut is currently an Assistant Professor at the Department of Computer Engineering, Istanbul Medeniyet University, Türkiye. She received her Ph.D. degree from the Department of Computer Engineering, Istanbul University, Türkiye, in 2018. Her primary research interests are computer networks, indoor localization, and intrusion detection.



Şafak Durukan Odabaşi received the B.S., M.S., and Ph.D. degrees in computer engineering from the Istanbul University, Istanbul, Türkiye in 2005, 2008, and 2013. She worked as a Research Assistant and Assistant Professor of Computer Engineering in Istanbul University and Istanbul University-Cerrahpaşa between 2005-2019. She was as a Visiting Assistant Professor with Illinois Wesleyan University, USA from 2019 to 2021. She is currently working as an Assistant Professor at Istanbul University – Cerrahpaşa. Her research areas are next generation networks, IoT and cybersecurity.

Muhammed Ali Aydin received the Ph.D. degree in computer engineering from Istanbul University. He was a Postdoctoral Researcher with the Department of Computer Science, Telecom SudParis. He is currently an Associate Professor with the Department of Computer Engineering, Istanbul University-Cerrahpaşa. His research interests include cyber security, cryptography, network security, and communication-network protocols.

Ahmet Sertbaş was born in Istanbul, Türkiye in 1965. He received the B.Sc. and M.Sc. degrees in electronic engineering from Istanbul Technical University, Istanbul, in 1986 and 1990, respectively, and the Ph.D. degree in electric-electronic engineering from Istanbul University, Istanbul, in 1997. Since 2000, he has been an Assistant Professor, an Associate Professor, and a Professor with the Computer Engineering Department, İstanbul University, and a Professor with the Computer Engineering Department, Istanbul University-Cerrahpaşa, since 2018. His research interests include image processing, artificial intelligence, computer arithmetic, and hardware security. He has 25 articles in indexed SCI-SCIE journals and many journal articles not indexed SCI-SCIE and international conference papers.