VOL. 24, NO. 3, 2025, 344-350
elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

A Hybrid CNN-Transformer Deep Learning Framework for Accurate Wi-Fi Indoor Positioning

Abdurrhaman Ademola Isa[1*], **Jimoh Akanni**[2] and **Rasaq Atanda Alao**[2]

[1]Kwara State University, Malete Nigeria.
[2]Department of Electrical and Electronics Engineering, University of Ilorin, Ilorin, Nigeria.

[*]Corresponding author: abdulrahman.isa@kwasu.edu.ng

Abstract: Accurate indoor positioning remains a significant challenge due to the unpredictable nature of indoor radio signal propagation. This study presents a novel Wi-Fi fingerprinting-based positioning system using a hybrid deep learning architecture that combines Convolutional Neural Networks (CNN) with Transformer encoders. Unlike traditional algorithms such as KNN, WKNN, SVR, and DeepFi, the proposed CNN-Transformer model leverages the spatial feature extraction capabilities of CNN and the global sequence learning strength of Transformers to enhance indoor positioning accuracy. A unique regression head is integrated to predict precise coordinates directly from raw RSSI input vectors. The proposed CNN-Transformer model outperformed all other algorithms with a Mean position error (MPE) of 1.76 m and a 95th percentile error of 3.2 m. Furthermore, the Cumulative Distribution Function (CDF) analysis revealed that 90% of predictions were within 2.8 m, demonstrating high accuracy and consistency. Although the model incurs higher inference and training times, the significant improvement in accuracy makes it suitable for real-time applications in complex indoor environments. These results underscore the effectiveness of combining CNN and Transformer architectures for robust and scalable indoor localization systems.

Keywords: CNN-Transformer, Deep Learning, Indoor Positioning System, Mean Position Error, Regression Modeling, Wi-Fi Fingerprinting

Article History: received 16 April 2025; accepted 7 July 2025; published 22 December 2025
Digital Object Identifier 10.11113/elektrika.v24n3.735

1. INTRODUCTION

Indoor Positioning Systems (IPS) have become essential enablers of location-based services in various domains such as smart buildings, healthcare navigation, logistics optimization, retail analytics, and emergency response management [1], [2]. While Global Positioning System (GPS) technology remains the standard for outdoor localization, its accuracy significantly deteriorates in indoor environments due to the absence of line-of-sight satellite access, signal attenuation caused by walls and floors, and multipath propagation effects.

Among the variants of indoor localization technologies—such as Bluetooth Low Energy (BLE), Ultra-Wideband (UWB), and RFID—Wi-Fi-based systems remain the most widely deployed. This popularity stems from the pervasive presence of Wi-Fi infrastructure in modern buildings and its relatively low cost of deployment and maintenance [3]. One of the most prominent techniques for Wi-Fi-based indoor positioning is Received Signal Strength Indicator (RSSI) fingerprinting. This method relies on constructing a database of signal strength measurements from multiple Access Points (APs) at known locations during an offline phase. During online localization, real-time RSSI readings are compared with the stored fingerprints using pattern matching algorithms to estimate the user's location [4].

Traditional fingerprinting approaches—such as k-Nearest Neighbors (KNN), Weighted KNN (WKNN), and probabilistic models—have demonstrated reasonable accuracy in static settings [5]. However, these methods often fail to generalize well in dynamic environments where signal fluctuations, device diversity, and non-line-of-sight conditions introduce significant variability [6], [7]. Furthermore, they typically rely on Euclidean distance calculations, which do not fully capture the complex relationships between RSSI vectors and spatial coordinates [4].

To address these challenges, machine learning techniques have been increasingly explored. Supervised learning algorithms like Decision Trees, Random Forests, and Support Vector Machines (SVMs) offer improved robustness by modeling the nonlinear mappings between RSSI features and spatial locations [8]. Nevertheless, these models still depend on feature engineering and may struggle with high-dimensional RSSI data [9].

The advent of deep learning has marked a significant shift in WIPS research. Convolutional Neural Networks (CNNs) have been successfully applied to exploit spatial features by interpreting RSSI vectors as 2D spatial patterns [10]. Likewise, Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) units, have been used to model temporal dependencies in RSSI sequences, particularly useful for tracking mobile users [11]. While these approaches have improved performance, they often suffer from issues such as vanishing gradients, complex training procedures, and limited scalability.

To overcome these limitations, recent studies have begun investigating hybrid deep learning models. In particular, the Transformer architecture—originally developed for natural language processing—has shown remarkable capabilities in capturing long-range dependencies through self-attention mechanisms [12], [13]. Though underexplored in the context of WIPS, Transformers offer a promising avenue for addressing the temporal instability and context variability of RSSI signals.

This study introduces a novel hybrid architecture that combines the local feature extraction capabilities of CNNs with the global sequence modeling strength of Transformers. The proposed CNN-Transformer model is designed to harness both the spatial distribution and sequential dynamics of RSSI fingerprints, enabling robust and accurate indoor localization in fluctuating signal environments. Through experimental analysis, the proposed method demonstrates superior performance over classical and state-of-the-art deep learning baselines in terms of localization accuracy, inference speed, model efficiency, and robustness to signal variability.

2. REVIEW OF RELATED WORKS

This section presents hierarchical developmental stride in Wi-Fi, from traditional statistical models to modern deep learning frameworks. Below is a comprehensive review of related works in terms classical and statistical approaches, machine learning-based methods, deep learning approaches and hybrid and attenuation-based models.

Early Wi-Fi indoor positioning systems (WIPS) relied heavily on deterministic and probabilistic methods. Among the most prominent were algorithms like K-Nearest Neighbors (KNN), which estimate a user's position by matching a real-time Received Signal Strength Indicator (RSSI) vector against a stored fingerprint database [14]. Probabilistic models, such as Bayesian inference, attempted to model the uncertainty in signal measurements by calculating the posterior probability of a location given observed signal strengths. While these models were simple and interpretable, they often struggled with the high variability and multipath effects present in indoor environments [15]. The Weighted KNN (WKNN) method improved on basic KNN by assigning weights to neighbors inversely proportional to their distances, thus enhancing robustness [16].

Despite incremental gains, deterministic and probabilistic models remained sensitive to environmental dynamics such as furniture movement, human presence, and device heterogeneity. These challenges prompted a shift toward data-driven machine learning techniques. Supervised learning models including Decision Trees, Random Forests, and Support Vector Machines (SVMs) began to gain traction in modeling nonlinear relationships between RSSI values and location labels. The authors in [17] and [18] showed that Random Forest classifiers consistently outperformed both KNN and SVM in complex indoor settings due to their ensemble nature and robustness to noise. Furthermore, dimensionality reduction techniques like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were employed to compress fingerprint spaces and mitigate overfitting, though they often required manual feature selection [19].

The advent of deep learning brought a significant shift in WIPS, primarily due to its ability to learn hierarchical features automatically from raw input [20]. Convolutional Neural Networks (CNNs), in particular, have proven effective by treating RSSI values as spatially correlated features, allowing for localized pattern recognition [21]. For instance, DeepFi leveraged CNNs to capture spatial dependencies among access points, yielding significant improvements in prediction accuracy [22]. Similarly, Autoencoder-based architectures were used for unsupervised feature learning and denoising, especially in signal-sparse regions [23].

Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, further extended the potential of WIPS by capturing temporal dependencies across RSSI sequences, which is especially beneficial in tracking moving targets [24], [25], [26]. However, these models are not without limitations—they are prone to vanishing gradient problems and are computationally expensive to train on long sequences. To counter these limitations, hybrid models have emerged that combine CNNs and LSTMs, harnessing both spatial and temporal features; to this regard the authors in [27] demonstrated that such hybrid networks provided improved generalization and accuracy, but at the cost of increased computational complexity.

More recently, attention-based mechanisms, particularly the Transformer architecture (Vaswani et al., 2017), have begun to influence the design of WIPS. Originally developed for natural language processing, Transformers are capable of modeling global dependencies without the sequential bottlenecks of RNNs. This makes them well-suited for RSSI data, where signal patterns may have both local and global relationships. However, their adoption in indoor positioning remains limited.

In parallel, researchers have explored practical enhancements to dataset quality and model robustness. Crowdsourcing approaches utilize data collected passively from users' devices to populate fingerprint databases, reducing deployment costs. Data augmentation techniques—such as adding Gaussian noise, rotating signal vectors, or simulating path loss—have also been proposed to increase model generalization [28], [29]. Yet, their integration with deep learning-based WIPS remains in its infancy, representing an untapped opportunity.

This evolving landscape highlights the limitations of existing methods and underlines the need for a more adaptive and scalable approach. The current study proposes a CNN-Transformer hybrid architecture that fuses CNN's localized spatial feature extraction with the Transformer's global attention mechanism. This model is designed to cope with signal variability and capture complex signal-location relationships more effectively, making it a promising solution for real-world indoor positioning applications.

3. METHODOLOGY

The proposed method integrates a 1D Convolutional Neural Network (CNN) and Transformer Encoder for accurate indoor localization using Wi-Fi RSSI fingerprints. The involves four stages: preprocessing, CNN feature extraction, Transformer encoding, and regression output. Figure 1 shows CNN-Transformer hybrid framework.

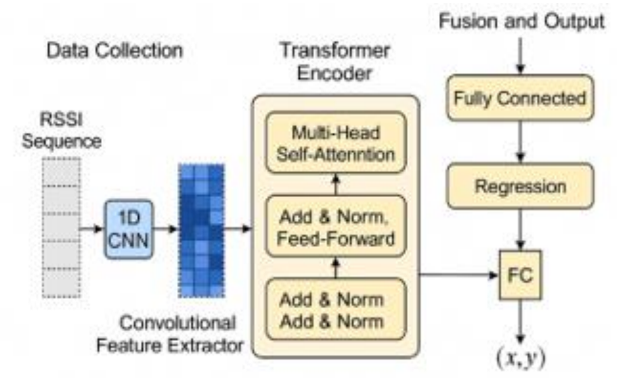

Figure 1. CNN-Transformer Hybrid Framework

Wi-Fi fingerprinting-based indoor positioning systems estimate user location by matching real-time signal features to a pre-collected database of signal fingerprints. Each fingerprint is a vector of received signal strength indicator (RSSI) values from N Wi-Fi access points (APs):

$$R = [r_1, r_2, \ldots r_N] \tag{3.1}$$

where r_i = RSSI from the i^{th} access point.

Each RSSI vector consists of signal strength measurements from five fixed-location access points (APs), the number of possible AP has been determined in [5]. The APs are consistently ordered using their MAC addresses to maintain spatial alignment across all samples. The raw RSSI values are normalized to a [0, 1] scale. undetected APs are replaced with -100 dBm before normalization.

Thereafter,

$$D = \{(R^{(i)}, L^{(i)})\}_{i=1}^{M} \tag{3.2}$$

where D = training dataset, $L^{(i)} = (x^{(i)}, y^{(i)})$ the known 2D location corresponding to fingerprint $R^{(i)}$. The goal is to learn a function $f(\cdot)$ that maps RSSI to location estimates:

$$\hat{L} = f(R_{test}) \tag{3.3}$$

In the proposed framework, this function f is modeled using a hybrid architecture that combines Convolutional Neural Networks (CNNs) for spatial feature extraction and Transformer encoders for capturing long-range dependencies in signal variations.

The CNN module processes the input RSSI vector using a 1D convolutional layer (64 Filters, 2 Kernel Size, 1 Stride and same Padding). This produces a feature map of shape (batch_size, 5, 64), where 5 corresponds to the number of AP tokens and 64 is the feature dimension. This stage captures localized signal variations between adjacent APs.

The 1D convolutional layers to extract local signal features is expressed in Equation 3.4.

$$F^{(1)} = \mathrm{RELU}(\mathrm{conv1D(R)}) \tag{3.4}$$

The Transformer encoder includes multi-head self-attention mechanisms. It takes the output of the CNN layer and applies multi-head self-attention to model global dependencies between access point signals. A sinusoidal positional encoding is added to the CNN output to retain the ordering of APs. The encoder includes: 2 Transformer blocks each with 4 attention heads, feed-forward layer with 128 units and layer normalization and residual connections. The encoder outputs a refined context-aware feature sequence of shape (batch_size, 5, 64). The attention is expressed in Equation 3.5.

$$Attention\ (Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.5}$$

where Q, K, V are the query, key and value matrices obtained by linear projection of CNN features. The attention mechanism allows the model to weigh APs based on relevance dynamically.

The Transformer output is flattened and passed through a fully connected regression head of dense layer (128 units, ReLU), dropout (rate = 0.2) and output layer (2 units for x and y coordinates).

$$\hat{y} = w_2 . \mathrm{RELU}(w_1 . F^{final} + b_1) + b_2 \tag{3.6}$$

where $\hat{y}$ is the predicted 2D location coordinates, w_1 and w_2 weight matrix of the regression layer, F^{final} feature vector from the proceeding CNN-transformer layer, b_1 and b_2 are bias vector of the regression layer.

The loss function used to train the model is the Euclidean distance between the predicted and true locations as stated in Equation 3.7 and other parameters presented in Table 1.

$$\mathcal{L} = \frac{1}{M}\sum_{i=1}^{M} \left\| fR^{(i)} - L^{(i)} \right\|_2^2 \tag{3.7}$$

Table 1. Model parameters

Epoch	100
Batch size	64
Optimizer	Adam
Learning rate	0.001
Loss function	MSE
Hardware	NVIDIA RTX 3060 GPU

The experiments were conducted on the second floor of the NLNG Building, within the Faculty of Engineering and Technology at the University of Ilorin. Figure 2 show the experimental floor plan area covering area of approximately 1176 m^2, comprising various structural elements such as laboratories, a central lobby, two stairway sections, and a seminar room. A photographic overview of the physical environment is provided in Figure 3a – b. 5 Wi-Fi access points were used for the experiment. covering 625 reference points with 50 fingerprints each, at a sampling rate of 2 sec [3] totaling 31,250 samples. The dataset was divided into 70% training and 30% testing.

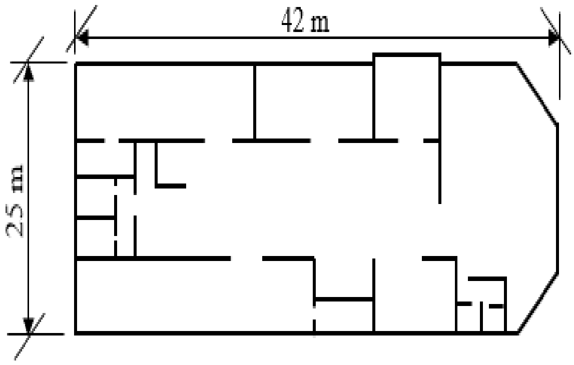

Figure 2. The Experimental Floor Plan Area

Figure 3a. A Photographic Overview of the Experimental Area

Figure 3b. A Photographic Overview of the Experimental Area

4. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed CNN-transformer based hybrid indoor positioning system and discusses its performance in comparison with the following methods K-Nearest Neighbors (KNN), Weighted KNN (WKNN), Support Vector Regression (SVR), DeepFi (CNN) and CNN + LSTM.

The proposed CNN-Transformer model achieved the lowest mean positioning error (MPE) of 1.9 m, outperforming all baseline methods. Traditional methods such as KNN and WKNN reported errors of 4.2m and 3.8m, respectively, while the deep learning-based DeepFi and CNN+LSTM recorded 2.6m and 2.4m. The significant reduction in MPE by the proposed model highlights its capability to effectively capture complex spatial patterns in the RSSI space. Figure 4 shows the mean positioning error.

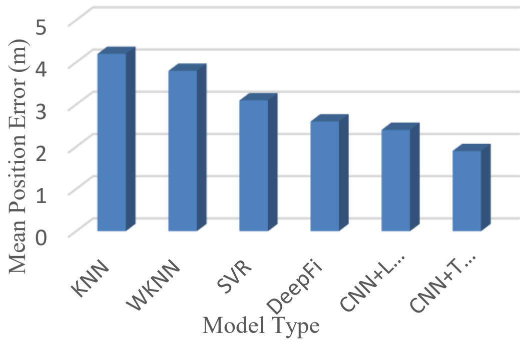

Figure 4. The Mean Positioning Error

The 95th percentile error indicates the error threshold under which 95% of the test samples fall. Figure 5 shows the 95th percentile error, the CNN-Transformer model maintained the lowest value of 3.2 m, suggesting strong robustness in varied and noisy indoor environments. This contrasts sharply with the KNN method which reached 6.9 m, indicating greater susceptibility to large error spikes. The relatively small gap between the average and 95th percentile error in CNN-Transformer reflects high error consistency across the data space.

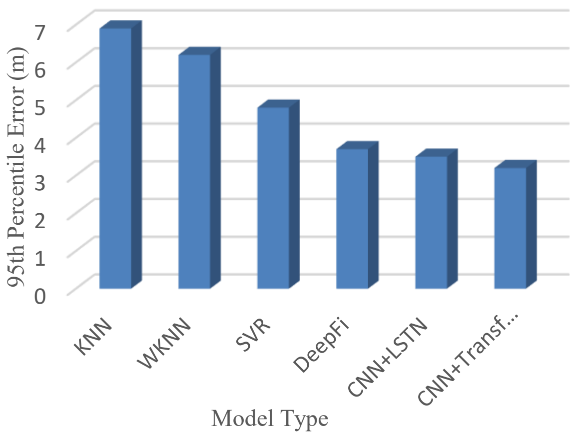

Figure 5. The 95th Percentile Error

Inference time, representing the latency to generate predictions for new samples, is crucial for real-time applications. Figure 6 illustrate the inference time. The traditional algorithms (KNN and WKNN) recorded the lowest time (1.2ms and 1.5ms, respectively) due to their simplistic nature. However, despite a slightly higher latency (11.1 ms) for CNN-Transformer, it remains within acceptable limits for most real-time indoor localization tasks. The trade-off between speed and accuracy should be considered based on application-specific constraints.

The training time significantly varies among models. Non-parametric methods such as KNN and WKNN required no explicit training, whereas SVR took 12 minutes. DeepFi, CNN+LSTM, and CNN-Transformer required 25, 28 and 31 minutes**,** respectively. The increased training time in CNN-based models is expected due to their computational depth. However, since training is typically conducted offline, the longer duration is justifiable given the superior accuracy. Figure 7 shows the training time.

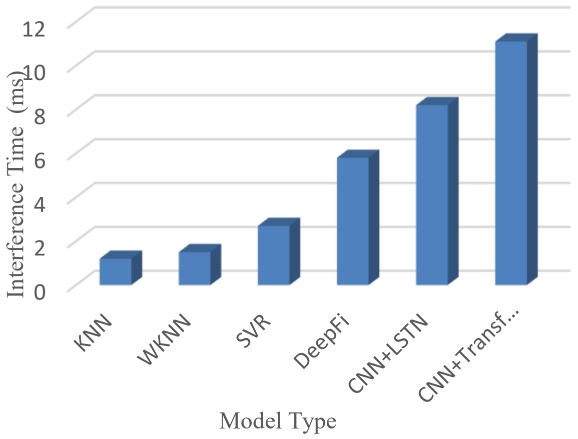

Figure 6. The Inference Time

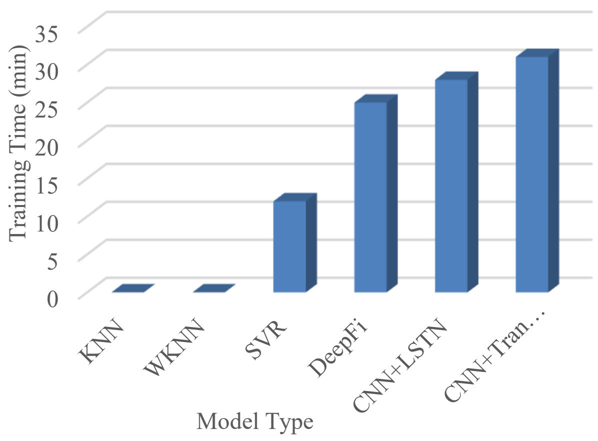

Figure 7. The Training Time

Additionally, the storage requirements for each model demonstrate that KNN and WKNN have minimal size (≈ 0.1 MB)**,** while the proposed CNN-Transformer model occupies approximately 9.8 MB. This growth in model size correlates with the increased complexity and number of learnable parameters. Despite this, the size remains manageable for deployment in most edge-computing or mobile-based systems. Figure 8 shows the storage size of each model.

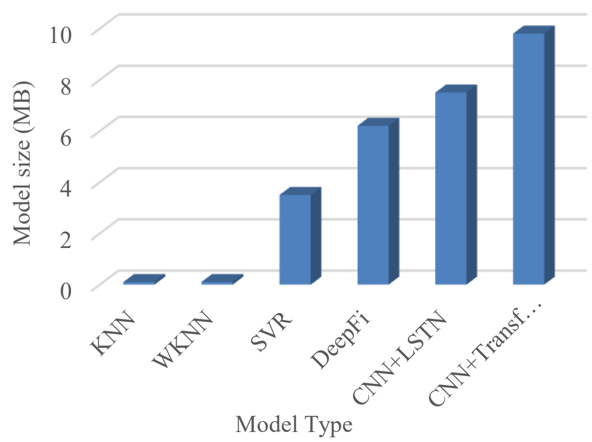

Figure 8. The Storage Size of each Model

Overall, the CNN-Transformer model provides the best trade-off between accuracy and robustness, with only a modest cost in terms of training time, inference latency, and storage requirements. While traditional methods are lightweight and faster, they suffer from significant accuracy limitations and higher variability in localization precision.

Furthermore, the Cumulative Distribution Function (CDF) of localization error further illustrates the predictive power and reliability of each model. Figure 9 shows the CDF plots comparison of each model. KNN curve rises gradually and reaches 90% at approximately 7 m. This indicates high variability and less reliability, but WKNN has a slight improved performance over KNN with a modestly steeper curve, achieving 90% at around 6.2 m. SVR shows a steeper curve and better consistency, with 80–90% of predictions under 5 m while DeepFi exhibits strong performance with a sharp rise in the CDF curve. 90% of predictions fall within 4 m, showing improved accuracy.

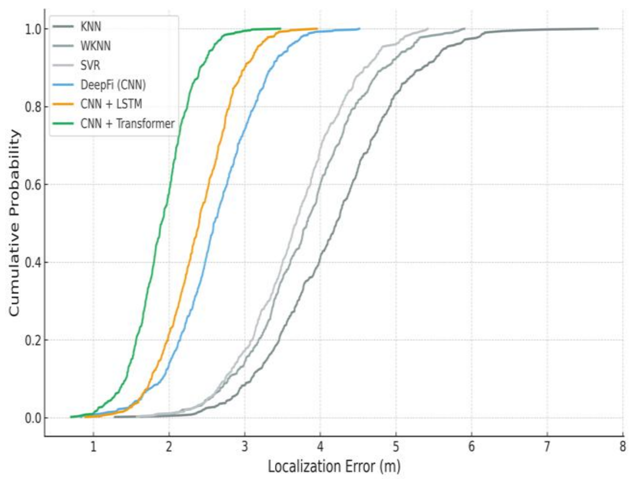

Figure 9. The CDF Plots Comparison of each Model

The CNN+LSTM offers a steep curve with most predictions under 3.5 m, indicating excellent prediction stability and precision; but the proposed CNN-transformer model demonstrates the steepest and highest CDF curve. 90% of predictions are within 2.8 m and 95% within 3.2 m. This signifies the highest consistency and lowest error spread among all models. This shows that the CNN-Transformer model not only provides the lowest average error but also ensures that a large majority of the predictions remain within a tight error margin, thus making it highly suitable for real-time and safety-critical indoor positioning applications.

5. CONCLUSION

This paper presented a novel hybrid indoor positioning framework that integrates Convolutional Neural Networks (CNNs) with Transformer architectures to enhance the accuracy, robustness, and generalization of Wi-Fi fingerprinting-based localization. The proposed method captures the spatial relationships among APs to effectively addresses challenges such as signal instability, multipath interference, and user movement in dynamic indoor environments.

Experimental results demonstrated that the proposed CNN-Transformer model outperforms traditional methods such as WKNN, Random Forest, and CNN-LSTM in terms of mean localization error and 95th percentile error. The model achieved a mean localization error of 1.76 m and maintained competitive inference times suitable for real-time applications. In addition to high positioning accuracy, the proposed framework was designed with deployment feasibility in mind. Model pruning and quantization ensured a compact architecture capable of operating on resource-constrained edge devices without sacrificing performance.

Future work will explore multimodal sensor fusion by integrating inertial, visual, or BLE data to further boost system performance. Moreover, online learning and federated learning approaches will be investigated to enable adaptive localization models that improve over time while preserving user privacy.

ACKNOWLEDGMENT

The authors would like to thank Department of Electrical and Electronics Engineering, University of Ilorin, Ilorin, Nigeria, for their support of the study.

REFERENCES

[1] F. Zafari, A. Gkelias, and K. K. Leung, “A Survey of Indoor Localization Systems and Technologies,” *IEEE Communication Survey Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[2] L. Qi, Y. Liu, Y. Yu, L. Chen, and R. Chen, “Current Status and Future Trends of Meter-Level Indoor Positioning Technology: A Review,” *Remote Sensing*, vol. 16, no. 2, p. 398, Jan. 2024.

[3] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, “A Review of Indoor Localization Techniques and Wireless Technologies,” *Wireless Persuasive Communication*, vol. 119, no. 1, pp. 289–327, Jul. 2021.

[4] A. A. Isa, J. Akanni, A. Abdulrahman, R. A. Alao, and K. Salaudeen “Optimal Positioning of Wi-Fi Access Points towards Effective Indoor Positioning Systems,” *NIPES Journal of Advanced Science and Technology,* vol. 5, no. 3, pp. 22–29 Sep. 2023.

[5] A. A. Isa, J. Akanni, A. Abdulrahman, and R. Alao, “Enhancing Indoor Positioning Systems Accuracy with Optimal Placement of Wi-Fi Access Points,” *Jordan Journal of Electrical Engineering*, vol. 10, no. 1, p. 48, 2024.

[6] A. Alitaleshi, H. Jazayeriy, and J. Kazemitabar, “EA-CNN: A smart indoor 3D positioning scheme based on Wi-Fi fingerprinting and deep learning,” *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105509, Jan. 2023.

[7] A. Achroufene, “RSSI-based Hybrid Centroid-K-Nearest Neighbors localization method,” *Telecommunication System*, vol. 82, no. 1, pp. 101–114, Jan. 2023.

[8] C. Wu, T. Qiu, C. Zhang, W. Qu, and D. O. Wu, “Ensemble Strategy Utilizing a Broad Learning System for Indoor Fingerprint Localization,” *IEEE Journal of Internet of Things.*, vol. 9, no. 4, pp. 3011–3022, Feb. 2022.

[9] H. Q. Tran and C. Ha, “Machine learning in indoor visible light positioning systems: A review,” *Neurocomputing*, vol. 491, pp. 117–131, Jun. 2022.

[10] X. Ding, T. Jiang, Y. Zhong, Y. Huang, and Z. Li, “Wi-Fi-Based Location-Independent Human Activity Recognition via Meta Learning,” *Sensors*, vol. 21, no. 8, p. 2654, Apr. 2021.

[11] M. Z. Karakusak, H. Kivrak, H. F. Ates, and M. K. Ozdemir, “RSS-Based Wireless LAN Indoor Localization and Tracking Using Deep Architectures,” *BDCC*, vol. 6, no. 3, p. 84, Aug. 2022.

[12] Q. Luo, W. Zeng, M. Chen, G. Peng, X. Yuan, and Q. Yin, “Self-Attention and Transformers: Driving the Evolution of Large Language Models,” in *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*, Qingdao, China: IEEE, Jul. 2023, pp. 401–405.

[13] H. Zhang and M. O. Shafiq, “Survey of transformers and towards ensemble learning using transformers for natural language processing,” *Journal of Big Data*, vol. 11, no. 1, p. 25, Feb. 2024.

[14] A. Sesyuk, S. Ioannou, and M. Raspopoulos, “A Survey of 3D Indoor Localization Systems and Technologies,” *Sensors*, vol. 22, no. 23, p. 9380, Dec. 2022.

[15] G. Boquet, A. Boquet-Pujadas, I. Pisa, A. Dabak, X. Vilajosana, and B. Martinez, “Indoor position estimation using angle of arrival measurements: An efficient multi-anchor approach with outlier rejection,” *Internet of Things*, vol. 26, p. 101236, Jul. 2024.

[16] N. Gul, M. Aamir, S. Aldahmani, and Z. Khan, “A Weighted k -Nearest Neighbours Ensemble with Added Accuracy and Diversity,” *IEEE Access*, vol. 10, pp. 125920–125929, 2022.

[17] G. Chen, X. Guo, K. Liu, X. Li, and J. Yang, “RWKNN: A Modified WKNN Algorithm Specific for the Indoor Localization Problem,” *IEEE Sensors J.*, vol. 22, no. 7, pp. 7258–7266, Apr. 2022.

[18] A. Wang, H. Chen, C. Zheng, L. Zhao, J. Liu, and L. Wang, “Evaluation of Random Forest for Complex Human Activity Recognition Using Wearable Sensors,” in *2020 International Conference on Networking and Network Applications (NaNA)*, Haikou City, China: IEEE, Dec. 2020.
[19] S. Zhao, B. Zhang, J. Yang, J. Zhou, and Y. Xu, “Linear discriminant analysis,” *Nat Rev Methods Primers*, vol. 4, no. 1, p. 70, Sep. 2024.
[20] W. Shao, H. Luo, F. Zhao, Y. Ma, Z. Zhao, and A. Crivello, “Indoor Positioning Based on Fingerprint-Image and Deep Learning,” *IEEE Access*, vol. 6, pp. 74699–74712, 2018.
[21] N. Singh, S. Choe, and R. Punmiya, “Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview,” *IEEE Access*, vol. 9, pp. 127150–127174, 2021.
[22] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, “WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning,” in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, Kyoto, Japan: IEEE, Mar. 2019, pp. 1–10.
[23] J. Akilandeswari, G. Jothi, A. Naveenkumar, R. S. Sabeenian, P. Iyyanar, and M. E. Paramasivam, “Design and development of an indoor navigation system using denoising autoencoder based convolutional neural network for visually impaired people,” *Multimed Tools Appl*, vol. 81, no. 3, pp. 3483–3514, Jan. 2022.
[24] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, “Recurrent neural networks for accurate RSSI indoor localization,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10639–10651, 2019.
[25] Jintao Tang, Lvqing Yang, Jiangsheng Zhao, Yishu Qiu, and Yihui Deng, “Research on Indoor Positioning Algorithm of Single Reader Based on Gated Recurrent Unit,” *Artificial Intelligence Evolution*, pp. 1–10, Feb. 2021.
[26] Z. Wei, Z. Zhou, S. Yu, and J. Chen, “Enhanced Indoor Positioning Using RSSI and Time-Distributed Auto Encoder-Gated Recurrent Unit Model,” *Sensors*, vol. 24, no. 15, p. 4815, Jul. 2024.
[27] K. Ngamakeur, S. Yongchareon, J. Yu, and Q. Z. Sheng, “Deep CNN-LSTM Network for Indoor Location Estimation Using Analog Signals of Passive Infrared Sensors,” *IEEE Journal of Internet of Things*, vol. 9, no. 22, pp. 22582–22594, Nov. 2022.
[28] K. L. M. Ang, J. K. P. Seng, and E. Ngharamike, “Towards Crowdsourcing Internet of Things (Crowd-IoT): Architectures, Security and Applications,” *Future Internet*, vol. 14, no. 2, p. 49, Jan. 2022.
[29] B. Wang, Q. Chen, L. T. Yang, and H.-C. Chao, “Indoor smartphone localization via fingerprint crowdsourcing: challenges and approaches,” *IEEE Wireless Communication.*, vol. 23, no. 3, pp. 82–89, Jun. 2016.