

Hybrid Machine Learning Classifiers for Indoor User Localization Problem

Hamza Turabieh, Ahmad Alghamdi

Abstract: Wi-Fi technology is now everywhere either inside or outside buildings. Using Wi-Fi technology introduces an indoor localization service(s) (ILS). Determining indoor user location is a hard and complex problem. Several applications highlight the importance of indoor user localization such as disaster management, health care zones, Internet of Things applications (IoT), and public settlement planning. The measurements of Wi-Fi signal strength (i.e., Received Signal Strength Indicator (RSSI)) can be used to determine indoor user location. In this paper, we proposed a hybrid model between a wrapper feature selection algorithm and machine learning classifiers to determine indoor user location. We employed the Minimum Redundancy Maximum Relevance (mRMR) algorithm as a feature selection to select the most active access point (AP) based on RSSI values. Six different machine learning classifiers were used in this work (i.e., Decision Tree (DT), Support Vector Machine (SVM), k -nearest neighbors (kNN), Linear Discriminant Analysis (LDA), Ensemble-Bagged Tree (EBaT), and Ensemble Boosted Tree (EBoT)). We examined all classifiers on a public dataset obtained from UCI repository. The obtained results show that EBoT outperforms all other classifiers based on accuracy value/

Keywords: Machine learning, indoor user location, Classifications, Feature selection.

I. INTRODUCTION

Determining user location based on estimating the position of the mobile station (MS) is an essential task for many applications. Many wireless localization methods are available such as time of arrival (TOA), time difference of arrival (TDOA), angle of arrival (AOA), and RSSI [1]. Recently, indoor localization services attract a huge number of researchers and become a hot research topic [2]. Developing new applications that provide a set of services based on user localization is needed. Such applications for smart buildings, disaster management, the health sector, and smart cities [3]. Up to date, Global Positioning System (GPS) cannot navigate user location inside buildings. This shortage of GPS motivates researchers and developers to adopt different approaches to determine user location inside buildings. Traditional algorithms to determine indoor user location (i.e., fingerprinting) consists of two steps (i) Offline step, where the fingerprint database is created at an early stage, and (ii) Online step, where user position is determined based on RSSI. Comparing the current RSSI with stored RSSI signal to determine user location is a time-consuming

approach and not work well in-case of changing building infrastructure [4]. As a result, finding a robust fingerprint algorithm is needed to reduce the computational time based on machine learning methods by analyzing the RSSI database and not influence by changing the building infrastructure. Figure 1 demonstrates the process of generating the fingerprint database, collecting the RSSI values from different locations to build a machine learning model based on the collected data.

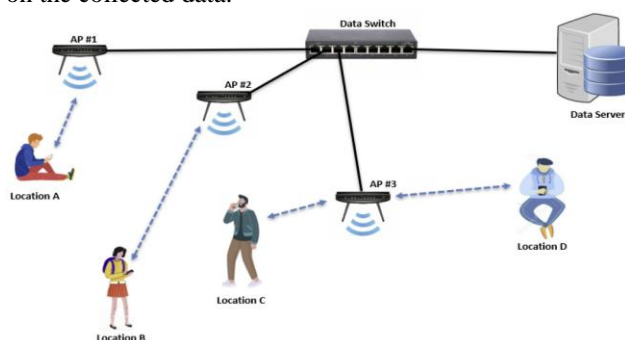


Fig. 1. Wi-Fi indoor localization system.

The main contribution of this work is to create a hybrid machine learning algorithm is proposed to determine indoor user location. A feature selection algorithm is employed to reduce the search space and model complexity. To perform a good analysis, we employed six different machine learning classification models and compare them based on average accuracy and p-value. The rest of this paper is organized as follows: Section 2. presents the related works of indoor localization problem. Section 3. presents the proposed approach. In Section 4. we explore the experimental dataset used in this paper. Section 5. shows the experimental results and analysis of the proposed approach. Section 6. presents the conclusion and future works.

II. RELATED WORKS

Selecting APs locations inside or outside buildings is not an easy task. APs positions determine their performances based on RSSI signals. Moreover, several applications depend on wireless sensors networks (WSNs) such as air pollution monitoring, smart homes and buildings, Internet of Things (IoT) applications and many other [5]. Several research papers highlight Wi-Fi sensors locations and their performances [6, 7]. So, Wi-Fi technology and APs applications have exponential growth in the future. Up to date, the Indoor user localization problem still gains great attention from researchers and the industrial world [2, 3].

Revised Manuscript Received on January 05, 2020.

* Correspondence Author

Hamza Turabieh*, Department of Information Technology, Taif University, Taif, Saudi Arabia. Email: h.turabieh@tu.edu.sa

Ahmad S. Alghamdi, Computer Engineering Department, Taif University, Taif, Saudi Arabia.

Hybrid Machine Learning Classifiers for Indoor User Localization Problem

Many applications provide services to the end-users based on their locations such as smart cities and buildings [8, 9, 10], health services [11], disaster management [12], and IoT applications [13, 14]. Up to our knowledge, no algorithm works reliably and accurately for ILS problem. The most suitable approach for indoor localization problem is the proliferation of smartphones.

Several research papers highlight ILS problem. For example, Rice and Harle [15] investigated a set of deterministic localization methods (i.e., Non-Linear Regression (NLR), Iterative Non-Linear Regression (INLR), Least Squares (LS), Random Sample Consensus (RANSAC) and Trilaterate on Minima (ToM)) to determine indoor user localization problem. The authors simulated their methods on real data collected from 550 m^2 environment. The obtained results show that NLR method outperforms other methods. Turabieh and Sheta [3] employed a cascaded layered recurrent neural network (L-RNN) to predict indoor user location. The authors adopted two public datasets in their works. The obtained results show the ability of L-RNN to determine indoor user location. Song et al. [16] employed the channel state information (CSI) to determine indoor user location. The authors employed multidimensional scaling (MDS) to find the Euclidean distance and time-reversal resonating strength between two positions (i.e., Actual and reference), then kNN algorithm is employed to determine the position. Wu et al. [17] employed CSI and NB classification model for passive indoor localization problems.

Sun et al. [18] proposed regression models based on Gaussian process to determine the spatial distribution of RSSI for indoor user localization problem. Haider et al. [19] employed a deep learning method to predict indoor user location based on RSSI values. The proposed approach is able to handle the missing data of RSSI. Khatib et al. [20] applied a deep extreme learning model with auto-encoder to predict the RSSI values. The authors noticed that increasing training samples with feature selection algorithm will improve the prediction accuracy.

Finding an accurate model for the indoor user localization problem is the main objective of previous research papers. The main problems of APs are their availability and interference, which makes this approach is more complex. As a result, finding robust classification models based on machine learning is needed.

III. PROPOSED APPROACH

The proposed method is depicted in Figure 2. In general, AP used to collect a huge amount of data. Finding the most important AP is needed. So the first step of this work is to employ a feature selection algorithm to select the most valuable AP. The (mRMR) algorithm is used as a feature selection algorithm. This step will reduce the search space and execution time for the classification model. We build a cascaded model to predict the building first, then the floor. Six different models are used in this work (i.e., DT, SVM, LDA, kNN, EBat, and EBoT). A cross-validation method is employed in this work with k-fold =10 to prevent overfitting problem.

A. mRMR algorithm

Feature selection algorithms try to extract the most valuable features that reflect the original dataset without

losing its value. In this work, we employed a well-known wrapper feature selection method called minimal redundancy maximal relevance criterion (*mRMR*) [21]. The main idea of *mRMR* method is to determine the most valuable features that maximize the mutual information between each feature and the proposed goal increase. The objective function of the mutual information is presented in Eq. (1).

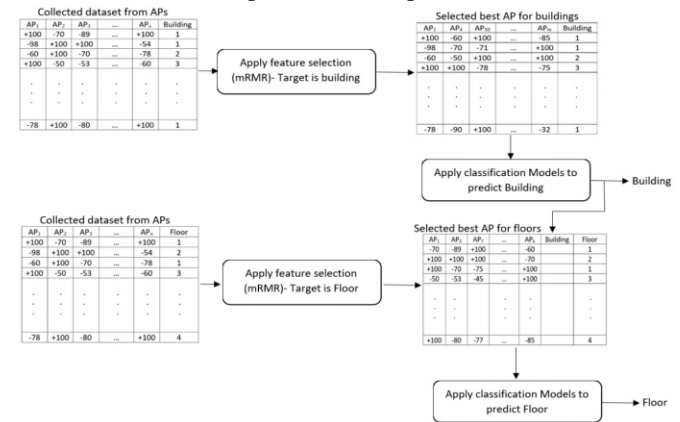


Fig. 2. Proposed hybrid model.

$$\max M(S, t) = \frac{1}{|S|} \sum_{F_i \in S} I(F_i, t) \quad (1)$$

where S is a set of features with features F_i , t presents the target, $M(S, t)$ presents the average of the mutual information between each feature and t , $I(i)$ refers to the measure of dependency the density of feature, and x_i presents the density of the t . In simple, if two features have a power separability on the t , and both features are highly interrelated, it is not acceptable to select both of them. Eq.(2) presents the main concept of minimum redundancy is to select a set of features that are mutually different. The minimization of minimum redundancy between two features.

$$\min R(S) = \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j) \quad (2)$$

where $R(S, t)$ denotes the average of the mutual information between two selected features. The *mRMR* method tries to maximize the average of the respective function information between each feature F_i and target t , and minimize of the mutual information between two selected features $R(S)$. Interested readers about *mRMR* can read [21].

IV. DATASET

In this work, we adopted a public dataset obtained called UJIIndoorLoc that was introduced in 2014 by Arnau et al. [22]. The dataset tries to determine the user location inside campus consists of a set of buildings and floors. The input variables present the RSSI values obtained from 520 access points (AP). The dataset has three building and each building has a set of floors between four to five. The weak RSSI single is encoded as +100 dBm. The dataset consists of 21048 records. Table I explores samples of the dataset. Figure 3 demonstrates the RSSI data pattern for 520 APs. It is clear that each AP works for a short period due to its location inside or outside buildings and floors. Interested users can download the dataset from UCI Machine Learning Repository [23].

Table-I. Sample data for UJIIndoorLoc dataset.

AP ₁	AP ₂	AP ₃	...	AP ₅₂₀	Floor	Building
+100	-54	-48		-20	4	3
-85	+100	+100		-63	1	0
-60	-74	-99		-66	0	1

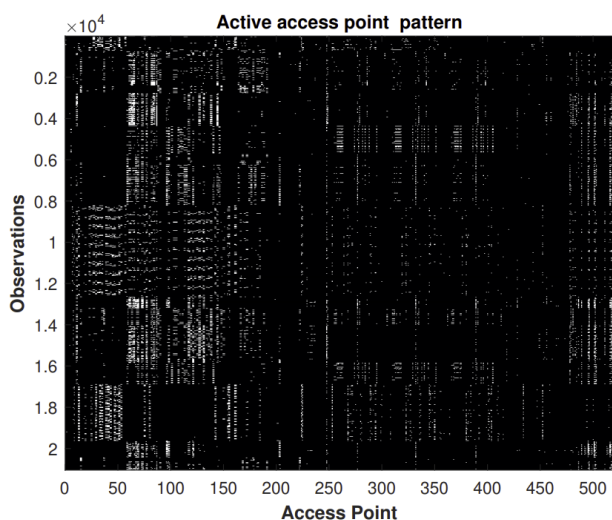


Fig. 3. Data pattern of RSSI values for all APs.

V. EXPERIMENTAL RESULTS

In this paper, we employed two types of experiments. The first experiment is to predict the building, while the second one is to predict the floor inside the building. All experiments were performed using MATLAB-R2019b. Each classification model is evaluated based on average accuracy and standard deviation (std) for the accuracy that was evaluated using the cross-validation method with k-fold=10.

Table II shows the number of selected APs from mRMR algorithm. mRMR algorithm selects 324 valuable APs for buildings and 227 APs for floors. It clear that mRMR algorithm is able to reduce the search space of this problem. Figure 4 demonstrates the selected APs for buildings and floors, respectively.

Table-II. Results of mRMR algorithm.

	Buildings	Floors
Number of selected APs	324	227
Percentage	62%	43.60%

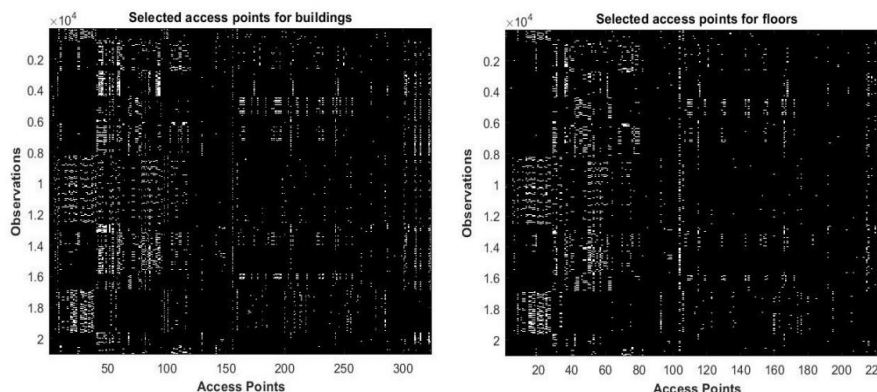


Fig. 4. Active access points for buildings (left) and floors (right).

Table III explores the obtained results for building prediction phase and floors prediction phase. For building prediction results, It is clear that mRMR-EBot model outperforms all the classification models with average accuracy equals to 95.8. While the LDA is the worst classification model with average accuracy equals to 87.1. The results of floors prediction results, mRMR-EBot algorithm gains higher accuracy results compared to other algorithms. For more statistical analysis, we employed a Wilcoxon statistical evaluation method (pvalue) between all classification models with a significance level of 0.05. From Table IV, It can be noticed, that the performances of all algorithms are not similar due to all p-values are less than 0.05.

Figure V demonstrates the box-plot diagrams for buildings and floors prediction. We can see that the performance of EBoT is stable and gain a robust performance. Moreover, the performance of the DT model is good for building prediction, while the performance of DT is the worst for floors prediction.

Table 5 demonstrates a comparison between our best results with the literature. It is noticed that our proposed approach outperforms other results in the literature. Moreover, selecting the most valuable APs (i.e., features) enhances the performance of classification models.

VI. CONCLUSION AND FUTURE WORKS

In this work, we employed six different machine learning classifiers to predict user location based on RSSI signals. We proposed a hybrid model between a wrapper feature selection (i.e., mRMR algorithm) and machine learning classifiers. The proposed approach consists of two phases (i.e., building prediction phase, and floor prediction phase). A deep analysis was performed on a public dataset obtained from the UCI repository. The obtained results show that mRMR is able to reduce the search space and enhance the performance of the classification model. The performance of mRMR-EBot model outperforms other models with an average accuracy 96.7% for buildings and 96.3% for floors. The future work will investigate a neural network either standard learning or deep learning models, with different wrapper feature selection algorithms.

Table-III. Obtained results for buildings and floors for each classifier.

Classification model	Buildings		Floors	
	Accuracy	Std.	Accuracy	Std.
mRMR-DT	92.6	0.02	74.6	0.17
mRMR-SVM	89.3	0.04	82.40	0.12
mRMR-kNN	88.6	0.17	84.12	0.22
mRMR-LDA	87.1	0.02	79.62	0.42
mRMR-EBat	94.6	0.12	90.13	0.17
mRMR-EBot	96.7	0.06	96.3	0.12

Table-IV. P-value results between all classification algorithms.

Compared algorithms	Building	Floor	Compared algorithms	Building	Floor
DT vs SVM	0.004	0.03	SVM vs kNN	0.006	0.030
DT vs LDA	0.001	0.02	SVM vs NB	0.002	0.010
DT vs kNN	0.004	0.05	SVM vs Ensemble-Bagged Tree	0.003	0.007
DT vs NB	0.003	0.001	SVM vs Ensemble-Boosted Tree	0.004	0.003
DT vs Ensemble-Bagged Tree	0.002	0.006	kNN vs Ensemble-Bagged Tree	0.003	0.040
DT vs Ensemble-Boosted Tree	0.003	0.008	kNN vs Ensemble-Boosted Tree	0.002	0.001
SVM vs LDA	0.007	0.03	Ensemble-Bagged Tree vs Ensemble-Boosted Tree	0.004	0.003

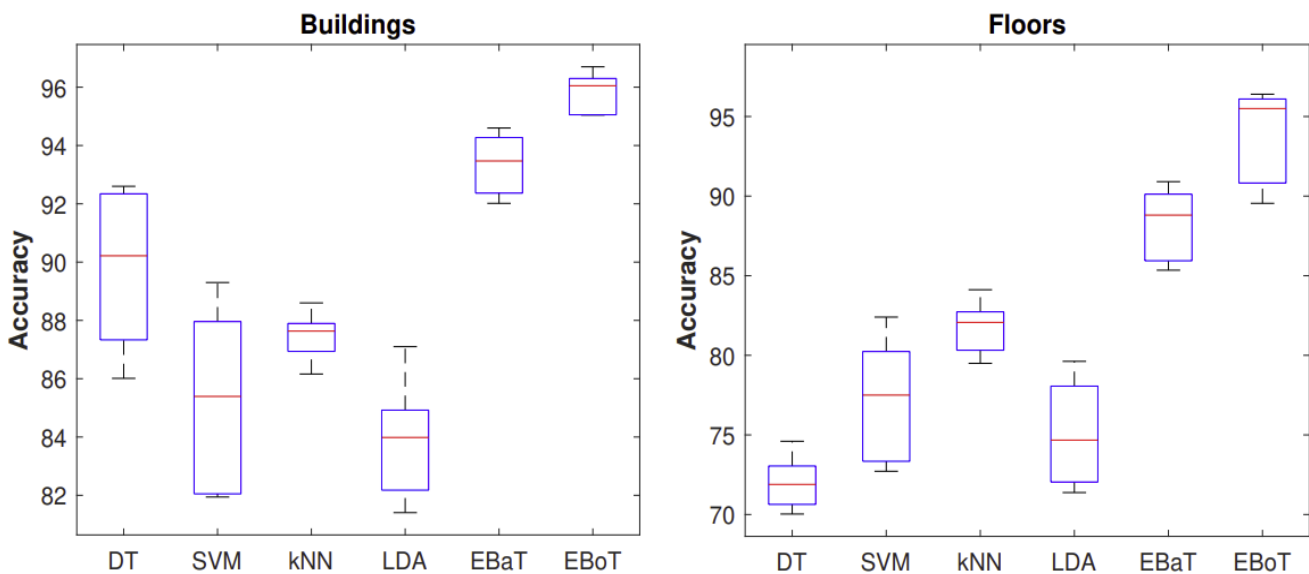


Fig. 5. Boxplot diagrams for all classifiers.

Table-V. Comparison with the state-of-the-art methods based on the average accuracy values.

Rank	Approach	Average accuracy (%)
1	mRMR-EBot (our approach)	96.70
2	CNN [24]	95.41
3	Cascaded L-RNN [3]	93.55
4	Scalable DNN [25]	92.89
5	SAE+ classifier [26]	91.10

REFERENCES

1. Y. Yue, L. Cao, J. Hu, S. Cai, B. Hang, and H. Wu, "A novel hybrid location algorithm based on chaotic particle swarm optimization for mobile position estimation," *IEEE Access*, vol. 7, pp. 58 541–58 552, 2019.
2. M. Kwak, Y. Park, J. Kim, J. Han, and T. Kwon, "An energy-efficient and lightweight indoor localization system for internet-of-things (iot) environments," *Proc. ACM Interact. Mob. Wearable Ubiquitous*

Technol., vol. 2, no. 1, pp. 17:1–17:28, Mar. 2018. [Online]. Available: <http://doi.acm.org/10.1145/3191749>

3. H. Turabieh and A. Sheta, "Cascaded layered recurrent neural network for indoor localization in wireless sensor networks," in *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, Oct 2019, pp. 1–6.

4. P. Jiang, Y. Zhang, W. Fu, H. Liu, and X. Su, "Indoor mobile localization based on wi-fi fingerprint's important access point," *International Journal of Distributed Sensor Networks*, vol. 11, no. 4, p. 429104, 2015. [Online]. Available: <https://doi.org/10.1155/2015/429104>
5. S. R. J. Ramson and D. J. Moni, "Applications of wireless sensor networks — a survey," in *2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT)*, Feb 2017, pp. 325–329.
6. L. Zhang, J. Liu, H. Jiang, and Y. Guan, "Senstrack: Energy-efficient location tracking with smartphone sensors," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3775–3784, Oct 2013.
7. J. Koo and H. Cha, "Localizing wifi access points using signal strength," *IEEE Communications Letters*, vol. 15, no. 2, pp. 187–189, February 2011.
8. E. Curry, S. Dustdar, Q. Z. Sheng, and A. Sheth, "Smart cities – enabling services and applications," *Journal of Internet Services and Applications*, vol. 7, no. 1, p. 6, Jun 2016. [Online]. Available: <https://doi.org/10.1186/s13174-016-0048-6>
9. A. Filippopolitis and E. Gelenbe, "An emergency response system for intelligent buildings," in *Sustainability in Energy and Buildings*, N. M'Sirdi, A. Namaane, R. J. Howlett, and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 265–274.
10. M. Dziubany, R. Machamer, H. Laux, A. Schmeink, K. Gollmer, G. Burger, and G. Dartmann, "Machine learning based indoor localization using a representative k-nearest-neighbor classifier on a low-cost iot hardware," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 2050–2054.
11. S. Doeweling, T. Tahiri, P. Sowinski, B. Schmidt, and M. Khalilbeigi, "Support for collaborative situation analysis and planning in crisis management teams using interactive tabletops," in *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS '13. New York, NY, USA: ACM, 2013, pp. 273–282. [Online]. Available: <http://doi.acm.org/10.1145/2512349.2512823>
12. M. Ibrahim, M. Torki, and M. ElNainay, "Cnn based indoor localization using rss time-series," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, June 2018, pp. 01 044–01 049.
13. S. K. Pandey and M. A. Zaveri, "Localization for collaborative processing in the internet of things framework," in *Proceedings of the Second International Conference on IoT in Urban Space*, ser. Urb-IoT '16. New York, NY, USA: ACM, 2016, pp. 108–110. [Online]. Available: <http://doi.acm.org/10.1145/2962735.2962752>
14. Z. Zhao, M. Zhang, G. Xu, D. Zhang, and G. Q. Huang, "Logistics sustainability practices: an iot-enabled smart indoor parking system for industrial hazardous chemical vehicles," *International Journal of Production Research*, vol. 0, no. 0, pp. 1–17, 2020. [Online]. Available: <https://doi.org/10.1080/00207543.2020.1720928>
15. A. Rice and R. Harle, "Evaluating lateration-based positioning algorithms for fine-grained tracking," in *Proceedings of the 2005 Joint Workshop on Foundations of Mobile Computing*, ser. DIALM-POMC '05. New York, NY, USA: ACM, 2005, pp. 54–61. [Online]. Available: <http://doi.acm.org/10.1145/1080810.1080820>
16. Q. Song, S. Guo, X. Liu, and Y. Yang, "Csi amplitude fingerprinting-based nb-iot indoor localization," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1494–1504, June 2018.
17. Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, "Passive indoor localization based on csi and naive bayes classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1566–1577, Sep. 2018.
18. W. Sun, M. Xue, H. Yu, H. Tang, and A. Lin, "Augmentation of fingerprints for indoor wifi localization based on gaussian process regression," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 896–10 905, Nov 2018.
19. A. Haider, Y. Wei, S. Liu, and S.-H. Hwang, "Pre- and post-processing algorithms with deep learning classifier for wi-fi fingerprint-based indoor positioning," *Electronics*, vol. 8, no. 2, 2019. [Online]. Available: <http://www.mdpi.com/2079-9292/8/2/195>
20. Z. E. Khatab, A. Hajihoseini, and S. A. Ghorashi, "A fingerprint method for indoor localization using autoencoder based deep extreme learning machine," *IEEE Sensors Letters*, vol. 2, no. 1, pp. 1–4, March 2018.
21. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of maxdependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.
22. J. Torres-Sospedra, R. Montoliu, A. Mart'inez-Uso, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2014, pp. 261–270.
23. D. Dua and C. Graff, "UCI machine learning repository," 2019. [Online]. Available: <http://archive.ics.uci.edu/ml>
24. J. Jang and S. Hong, "Indoor localization with wifi fingerprinting using convolutional neural network," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, July 2018, pp. 753–758.
25. K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on wi-fi fingerprinting," *Big Data Analytics*, vol. 3, no. 1, p. 4, Apr 2018. [Online]. Available: <https://doi.org/10.1186/s41044-018-0031-2>
26. M. Nowicki and J. Wietrzykowski, "Low-effort place recognition with wifi fingerprints using deep learning," in *Automation 2017*, R. Szewczyk, C. Zielinski, and M. Kaliczynska, Eds. Cham: Springer International Publishing, 2017, pp. 575–584.

AUTHORS PROFILE



Hamza Turabieh: is an Associate professor at the Information Technology Department- Faculty of Science and Information Technology- Taif University. Dr. Turabieh received his B.A., M.Sc. degrees in Computer Science from Balqa Applied University in 2004 and 2006 respectively, in Jordan. Dr. Turabieh obtained his Ph.D. from National University of Malaysia (UKM) in 2010,

his research interests and activities lie at the interface of Computer Science and Operational Research. Intelligent decision support systems, search and optimization (combinatorial optimization, constraint optimization, multi-modal optimization, and multi-objective optimization) using heuristics, local search, hyper-heuristics, meta heuristics (in particular memetic algorithms, particle swarm optimization), hybrid approaches and their theoretical foundations.



Ahmed S. AlGhamdi was born in Jeddah, Saudi Arabia, in 1975. He received the B.E. degree in electronic engineering from the technical college, Riyadh, Saudi Arabia, in 1997, and the Master degree and Ph.D. degrees in computer and electrical engineering from the Curtin University Perth, Western Australia, Australia, in 2008 and 2014, respectively. In 1998, he joined the

Department of instrumentation, at KEMYA Company. Since September 1990, he has been with the Department of Computer Engineering, faculty of Computer Information Technology, Taif University, where he was an Assistant Professor. His current research interests include Fuzzy logic, transportation system, communication, 3D printing.