Adversarial Contrastive Representation Learning for Passive WiFi Fingerprinting of Individuals

Varuna De Silva
Institute for Digital Technologies
Loughborough University
London, United Kingdom
V.D.De-Silva@lboro.ac.uk

Corentin Artaud
Institute for Digital Technologies
Loughborough University
London, United Kingdom
C.Artaud2@lboro.ac.uk

2024 14th International Conference on Pattern Recognition Systems (ICPRS) | 979-8-3503-7565-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPRS62101.2024.10677808

***Abstract*—Extracting valuable patterns from a dataset involves isolating the pattern of relevance from any other influential attributes in the data generation process. This becomes important when trying to repurpose existing datasets that were collected for a specific objective. When repurposing datasets, irrelevant influential attributes cannot necessarily be treated as noise. This paper aims to develop a representation learning scheme to identify individual users based on their radio-frequency footprint (WiFi Fingerprinting) within a collection of WiFi access points in a building. The question pertains to whether a user-dependent signature can be isolated from the Received signal strength indicator (RSSI) across a network of WiFi access points, which can then be used to identify individuals in a building without the need for line of sight sensors such as cameras. We repurposed a popular dataset called UJIndoorLoc, which was collected to localize individual users based on the RSSI. We propose a novel adversarial contrastive representation learning scheme based on a triplet loss to extract user-dependent patterns while minimizing the influence of location-dependent patterns inherent within this dataset. The results indicate a robust user-dependent pattern that can be isolated with an accuracy of 0.87. Such coarse-grained identification methods based on RSSI can be combined with fine-grained Channel State Information (CSI) for passively identifying individuals without requiring cameras.**

***Index Terms*—Contrastive Learning, User Identification, Triplet Loss, Passive Wireless Sensing**

I. Introduction

Identification of humans in enclosed settings without access to sensors such as cameras or audio is needed in many applications that preserve user privacy and covert intelligence-gathering operations. Recently, there has been a significant interest in electromagnetic imaging applications based on commercial wireless radio communication signals for applications in healthcare activity monitoring. Two essential techniques dominate current literature: active radar-based through wall sensing and passive WiFi-based activity monitoring, where a commercially available WiFi access point is listened to from outside the room to detect indoor activities. Using wireless radar to detect vital signs of humans is also an application that has gained significant interest [1]. Recent developments in data-driven AI have enabled significant gains in these domains, where multiple activities from a multi-user setting are identified with very high accuracy.

In contrast to activity recognition using WiFi signals, a relatively low emphasis has been placed on the personal identification of individuals from wireless signals. This is due mainly to many efforts being focused on datasets containing few individuals at most, and in most cases, the objective has been activity recognition. Identification of individuals is crucial in a multi-user scenario to enable tracking of the individuals. For example, in security applications such as responding to a hostage situation, one would need to know the perpetrator or understand the pattern of life in an organized crime scenario or even in civilian applications such as monitoring factory settings. However, identifying individuals from WiFi signals is challenging due to the nature of propagation and attenuation of wireless signals, which are very sensitive to environmental conditions.

Shi et al. proposed using WiFi signals to capture unique human physiological and behavioral characteristics inherited from their daily activities, including walking and stationary ones, and using a deep learning-based user authentication scheme to identify each user accurately [2]. WiWho demonstrated how step and walk analysis can be used to identify a person's walking gait from CSI and use the gait information to identify a person [3], achieving an identification accuracy of 92% to 80% for a group size of 2 to 6 people respectively, with only 2 - 3m of walking. WiID showcased a WiFi and gesture-based user identification system that can identify the user based on how they perform certain gestures [4]. WiGesID [5], a fine-grained gesture recognition-based system, works by identifying personalized spatiotemporal dynamic patterns from the gestures of different users. WiFi HuFu (WiHF) [6] identifies individuals based on distinct personalized motion change patterns prompted by arm gestures, encompassing rhythmic velocity fluctuation and characteristic pause distribution and maintaining uniformity across various domains. Observing that there are unique individual physiological characteristics underlying various body gestures of a person, a gesture-independent user identification scheme was proposed by Kong et al. [7]. RF-Capture is an active sensing system that operates by transmitting low-power RF signals (1/1000 the power of WiFi) and measuring and analyzing its reflections off different objects [8]. RF Capture demonstrated that the system could capture a representative human figure through walls and use it to distinguish between various users.

The current methods have two significant limitations for

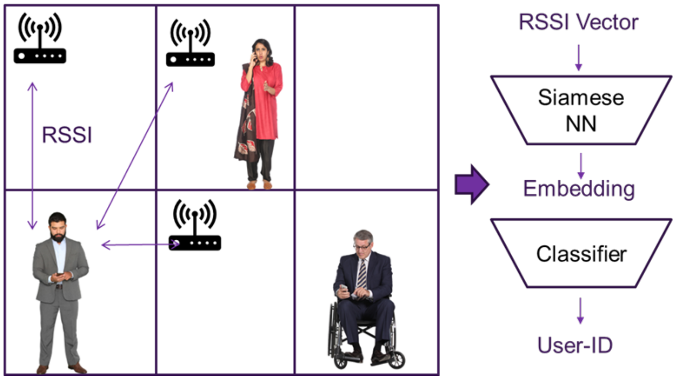

Fig. 1. The proposed method for user identification (WiFI fingerprinting) based on RSSI vector across a network of WiFi access points

ubiquitous personal identification through WiFi-based sensing: 1) they are tried and tested in a minimal number of environments (or domains as used in literature) [7] or 2) they are restricted to several individuals who are in the test set [6].

Most WiFi-based identification work relies on channel state information (CSI). Received signal strength indicator (RSSI), on the other hand, is coarse-grained information that measures the signal strength received by a user from a given access point. RSSI-based fingerprinting has been excessively used for localization activity, where RSSI at multiple wireless access points are utilized to locate the user [9], [10]. While CSI is very informative, accessing it is less straightforward than RSSI. RSSI-based localization activity has made significant domain adaptive results in the recent past.

An obvious question arises on the ability to do wireless sensing-based identification across domains in unknown room configurations and environments. A potential choice of methods for this is the fusion of fine-grained information from CSI and coarse-grained features from RSSI. While answering this question requires appropriate data collection, we test the feasibility of identifying individuals based on RSSI-based fingerprints in this paper. For this purpose, we use a dataset for indoor localization for location-independent user identification. We investigate a novel adversarial representation learning scheme to establish a user-dependent transformation. The proposed method, as illustrated in Figure 1, is based on contrastive learning on a Siamese Neural Network to learn the user-dependent patterns embedded within a vector RSSI captured through a network of wireless access points while minimizing the influence of location-dependent features such as the room in which the user was in.

The contributions of this paper are as follows:

1) A novel deep representation learning scheme based on contrastive and adversarial learning is proposed for low dimensional pattern recognition.
2) The proposed algorithm is utilized to develop a novel coarse-grained mechanism for identifying individuals through wireless signals.

The rest of this paper is organized as follows: a review of recent research into individual identification through wireless sensing and associated machine learning schemes is introduced in section II. The proposed technique for low-dimensional representation learning is presented in section III, followed by the results and discussion of how it is applied to a person identification task in section IV. We conclude the paper by highlighting potential future research directions in section V.

II. Background and Related Work

A. RSSI vs CSI for WiFi Fingerprinting

A typical channel attribute of WiFi utilized especially for localization is the RSSI that estimates the power level received from a wireless access point (AP) [11]. While RSSI diminishes as the receiver moves away from the AP in open spaces, this behavior alters in indoor spaces due to multipath and shadowing effects. Additionally, as RSSI is scalar, multiple AP links are needed to enhance fingerprint dimensionality, which is crucial for precise device localization.

Despite inherent limitations of WiFi signals with a lack of temporal and spatial resolution, advancements in technologies like MIMO and multicarrier have enhanced sensing resolution [5]. Among Wi-Fi signal parameters, CSI is widely utilized in sensing due to its ability to capture detailed features. Consequently, Wi-Fi sensing research now centers on extracting precise spatiotemporal features from CSI.

RSSI is an aggregated coarse-grained measure, whereas CSI measured by OFDM subcarriers is robust to environment noise, which can extract the factual information from the redundancy of the streams [12]. The main advantage of RSSI over CSI is that RSSI is available in all wireless devices, whereas CSI is still obtained only with specific NIC cards [12] in conjunction with particular tools [13].

B. Signal Processing Techniques for WiFi Passive Radar

WiFi passive radar literature obtains the Doppler-time plots by conventional FFT. FFT-based Doppler resolution cannot reveal and discriminate the rich and fast-varying Doppler/micro-Doppler signatures of human motions within a short (coherent integration time) CIT of 0.3s [15]. The classical superresolution algorithms, such as the minimum variance distortionless response (MVDR) and MUltiple SIgnal Classification (MUSIC) algorithms, require multiple signal snapshots to estimate the covariance matrix or perform eigen-analysis. MUSIC also needs to know the number of sources a priori.

However, in the Doppler processing of WiFi passive radar, only one temporal snapshot is available for each CIT. The number of the target's Doppler and micro-Doppler sources is also unknown. Thus, the classical superresolution algorithms cannot be applied. In [15], authors propose to use an Iterative Adaptive Approach (IAA) to replace the FFT in Doppler processing. IAA is a data-dependent, nonparametric, iterative adaptive algorithm based on the weighted least square (WLS) approach.

The received signals should be decomposed for multiple subjects to enable individual differences. Algorithms employ blind source separation (BSS) and variational mode decomposition (VMD) to separate various users, and heartbeat harmonics can separate the distinguishable respiration or heartbeat signal of different subjects [16]. However, these methods can not assign vital signs to each related subject, which needs localization and detection simultaneously [1].

Feature engineering through processing CSI signals is a popular option for developing cross-domain pattern recognition tasks. Widar3.0 [18] uses a feature called Body-coordinate Velocity Profile (BVP) extracted from the Doppler Frequency Shift (DFS) of raw CSI data to capture the relative velocity of arm movements. BVP is independent of user orientation, location, and environment, enabling cross-domain gesture recognition without additional data collection or model retraining. While BVP cannot preserve the personalized characteristics while performing arm gestures, in WiHF [6], authors derive a domain invariant feature from CSI that enables authentication and gesture recognition.

C. Machine Learning methods for identification of individuals

Bio-metrics enable authentication of individuals and can be either based on 1) physical characteristics such as fingerprints, face, iris, and hand print or 2) behavioral characteristics, such as gait, keystroke dynamics, and specific gestures [17]. Physical characteristics are more reliable; some may not change significantly, while behavioral traits can change over time.

WiWho developed their identification mechanism through supervised classification of the walking gait of known people on features extracted from CSI[3]. Spatio-temporal CSI features (captured of multiple nodes and across time) are processed through a dual task Convolutional Neural Network (CNN) based feature representation learning scheme for joint gesture recognition and identification in [5]. In [5], authors pre-train their network on tasks and identities to predict the similarity between the two data samples and test its performance on an entirely new set of 5 tasks and three identities based on a single data point as support.

Kong et al. [7] empirically observed that signals induced by body gestures can exhibit invariant individual uniqueness unrelated to specific body gestures. A gesture-independent user authentication system is developed in [7] by employing adversarial machine learning, where a GAN is trained to minimize the user authentication loss while maximizing the gesture recognition loss.

A human RF signatures synthesis system based on software-defined radio using conditional GANs is proposed in [19], with a successful generation of signatures of human occupancy. The GAN-generated spectrum simulates the wireless signals in an enclosed space occupied by a human subject using the baseline spectrums without human occupancy.

A key challenge in applying deep learning to RF datasets has been the greater difficulty, in comparison with computer vision, in acquiring datasets of sufficient sample size and statistical variance to train DNNs with high accuracy and generalization. The generation of synthetic data has been proposed through model-based synthesis and through generative deep networks [20]. Model-based approaches leverage video motion data to track the trajectories of humans and generate the expected RF micro-Doppler signature. While such methods are prone to errors introduced by backscattering from the entire human body, environmental sources of interference, sensor artifacts, or multipath, advanced forms of GANs have been shown to overcome such issues [20].

III. Methodology

A. The Dataset

We utilize the UJIndoorLoc dataset [9] for local independent user identification. It covers three buildings, each with four or more floors, and covers an area of almost 110m^2. It contains individual Wi-Fi fingerprints for around 20 users captured using Android devices. Each fingerprint contains information regarding the detected Wireless Access Point (WAP) through its corresponding Received Signal Strength Intensity (RSSI). The RSSI recorded values range from –104dBM (extremely poor signal) to 0dBM. Each fingerprint comprises 520 intensity values characterizing all the different WAPs across the three buildings. Each fingerprint additionally contains information regarding the coordinates (latitude, longitude, building-id, and floor), space (office, lab, etc.), and information about who recorded the RSSI (user), how it was recorded (android device and version of operating system) and when the capture was taken (timestep). The UJIndoorLoc dataset was presented as a classification problem (e.g., building or floor identification) and regression (e.g., longitude and latitude). We use this dataset as a classification problem for user identification.

B. Data Pre-processing

The RSSI values recorded contain a positive value of 100 to denote a not-detected WAP because the WAPs from one building are not detected from another. To standardize our data, we convert all the undetected values to fall within the initial negative range found within the detected WAPs. Since an inferior signal value is lower than or equal to –104dBM, we convert the undetected WAPs to –105dBM to keep our range constant and not have the undetected WAPs considered outliers with their initial value of 100.

Performing this conversion enables us to scale and transform the dataset. We perform both min-max normalization and value inversion. Min-max normalization allows us to scale our values to the range of [0,1] such that our minimum recording RSSI values are 1. However, we require that the minimum held represents the worst signal values in our case. Therefore, we perform value inversion such that the shallow RSSI values are close to 0, where all undetected WAPs would be represented as 0 and all detected and high signal WAPs would be close to 1.

In addition, through statistical analysis, we found that some samples within the dataset never triggered any WAPs. Hence, we removed each of these samples, reducing the number of WAPs from 520 to 465. These 465 WAPs cover the three buildings found within the dataset. Therefore, we concentrated only on building 2 (as it contains the most available samples) and further reduced the dataset to include only the WAPs triggered at any time within this building. This left us with 203 available WAPs across three floors representing one building.

C. Adversarial Contrastive Learning for Extraction of Individual Fingerprint

Contrastive learning is a representation learning technique that learns an embedding space for input vectors. Similar pairs of inputs are positioned closer together in the embedding space, and dissimilar ones are placed far from each other. A basic form of contrastive learning involves pairs of either similar inputs or dissimilar inputs as given in (1):

$$\begin{aligned}\mathcal{L}_{\text{cont}}(x_i, x_j, \theta) = \mathbb{1}[y_i \neq y_j] \max(0, \epsilon - \|f_\theta(x_i) - f_\theta(x_j)\|_2^2) \\ + \mathbb{1}[y_i = y_j] \|f_\theta(x_i) - f_\theta(x_j)\|_2^2\end{aligned} \tag{1}$$

We utilize this learning technique to cluster each user for identification. The architecture we employ for contrastive learning resolves around a Siamese neural network (also called a twin network). It uses the same underlying parameters working in tandem with different input vectors to compute their comparisons. In our case, these would be both similar and dissimilar vectors. We employ a multi-layered perceptron (MLP) composed of 4 layers with ReLU activation functions for our use case. We aim to learn a compressed embedding dimension of 16 from an input dimension of 203.

Our loss function is based on the triplet loss, where an anchor (also known as the reference) is compared by a Euclidean distance metric to some positive (matching input) and negative (non-matching input) where the aim is to minimize the distance between the anchor and positive and maximize that between the anchor and the negative. The triplet loss works directly on the embedded distances. In other words, it performs a distance measure on the compressed representations provided by the model. For our specific use case, we use two correlated triplet losses.

To perform these two triplet losses, we divide the dataset into two components u and l and apply the same loss function $\mathcal{L}_k$ on the respective data anchor, positive input and negative input for the associated data component k given as follows.

$$\begin{aligned}\mathcal{L}_k(\mathbf{x}_k, \mathbf{x}_k^+, \mathbf{x}_k^-) = \sum_{\mathbf{x}_k \in \mathcal{X}} \max(0, \|f(\mathbf{x}_k) - f(\mathbf{x}_k^+)\|_2^2 \\ - \|f(\mathbf{x}_k) - f(\mathbf{x}_k^-)\|_2^2 + \epsilon_k)\end{aligned} \tag{2}$$

The first component u relies on finding anchors, positive inputs, and negative inputs based solely on the *user-id*. Where $\mathbf{x}_u$ represents the anchor, $\mathbf{x}_u^+$ represents the positive input and $\mathbf{x}_u^-$ the negative input. We use an Euclidean distance as our metric and a soft margin of 2 for ϵ.

The second component l relies on finding the same inputs but concerning both the *floor* and *space-id*. Where $\mathbf{x}_l$ represents the anchor, $\mathbf{x}_l^+$ represents the positive input and $\mathbf{x}_l^-$ the negative input in relation to the location identifiers.

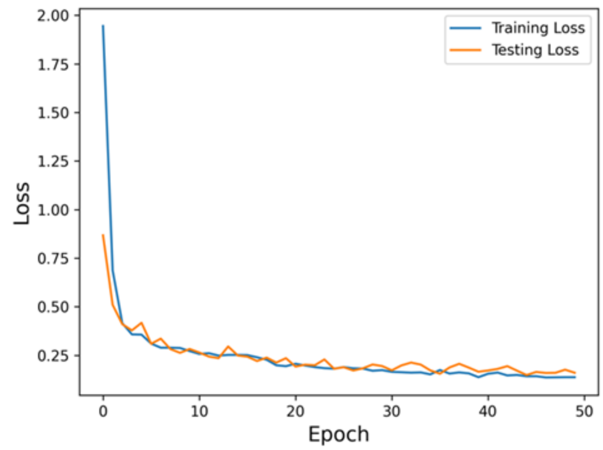

Fig. 2. Cross-Validation Loss Curves

Our objective is to minimize the distance of representations for the same user while maximizing the distance for different users. On the other hand, we need to remove the influence of location-based signatures to identify the unique signature of a given user. We penalize the representations learned that contain a location-based signature to achieve this objective. We formalize these conflicting objectives using a combination of the above two losses:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{user}}(\mathbf{x}_u, \mathbf{x}_u^+, \mathbf{x}_u^-) - \beta \mathcal{L}_{\text{loc}}(\mathbf{x}_l, \mathbf{x}_l^+, \mathbf{x}_l^-) \tag{3}$$

IV. Results and Discussions

A. Training of Adversarial Contrastive Learning

After our data pre-processing, we were left with around 9000 samples within building 2. We perform an 80/20 train/test split and train our model for 50 epochs. Figure 2 demonstrates the training and testing loss curves achieved.

After training, we utilize the testing set to validate the representational embeddings learned by our model concerning WAPs RSSI inputs relative to specific users. We perform both a PCA, illustrated in 3 and TSNE, illustrated in 4, embedding to demonstrate the learned latent representation. We compress the learned latent representation from 16 to 2 components using PCA and TSNE decompositions.

The decompositions retrieved from both PCA and TSNE demonstrate the capabilities of contrastive learning in separating users into their clusters. Except for a few outliers in both approaches, we can conclude that the underlying representation learned simply from RSSI data enables a solid separation. Interestingly, we can extract several interesting aspects relative to how this dataset was constructed. For example, examining Users 5, 11, and 14 demonstrates that the registered WAPs and the users' locations were very close together. We can see an apparent clustering of groups of users, which goes along with how the dataset was gathered; some users stayed on the same floors, and some moved around more

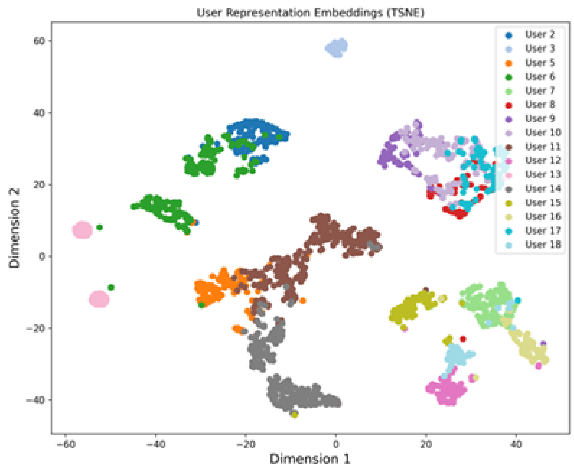

Fig. 3. TSNE Embedding against Testing set

than others (e.g., User 13). It is also clear that within these samples, some participants triggered the same type of WAPs when the recordings occurred, meaning they followed a similar path around the building when recording these RSSI values.

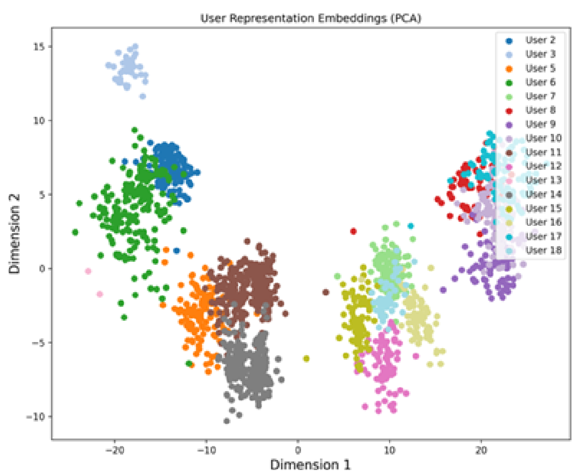

Fig. 4. PCA Embedding against Testing Set

B. Individual Identification Performance

a) Performance based on a downstream Classification task: To demonstrate the capabilities of our embedder with respect to further identifying users based on RSSI data, we first feed our WAP RSSI data into our frozen contrastive learning model; we then feed the embedded outputs, represented as a latent dimension of 16 and first test a Neural network classifier using five estimators. Through cross-validation, the random forest classifier reached a mean cross-validation score of 89% with a test accuracy of 87%. This showcases that our embedder can properly separate users based solely on RSSI data and can further classify said users based solely on these representations. Table I showcases the classification report achieved for the random forest classifier. Similarly, we have tested it with SVM, Random Forest, and Logistic Regression to achieve similar results. The similarity could have been because each model's feature selection process was identical.

TABLE I
CLASSIFICATION REPORT FOR NEURAL NETWORK-BASED CLASSIFIER

	Precision	*Recall*	*f1-score*
Accuracy			0.87
Macro Avg.	0.86	0.84	0.84
Weighted Avg.	0.87	0.87	0.87

For our demonstration, however, these similar results showcase the robustness and reliability of our approach. They also show that the embeddings are not dependent on the idiosyncrasies of a single algorithm. Additionally, they enable flexibility in this approach, with performances across multiple classifiers. The selection of the model based on criteria such as ease of implementation, computational efficiency, or available resources is more flexibly approachable.

b) Ablation experiment to evaluate the performance on location identification: To justify the triplet loss used within our contrastive learning model and to further expand on the capabilities of the classifiers, we perform ablations to test if the locality of each user is indeed separated from the RSSI values within our representation. We examine the capabilities of a feed-forward network (NN) to classify the room identifications within which a user could be depending on the RSSI values.

Given the representations from the contrastive model, we train an NN model to classify the room identification. Table II demonstrates the classification report achieved for this approach. We can see that the model cannot classify the room-id from the representations themselves; showcasing the learned representation does remove any locality information and further justifies our triplet loss function. Table III showcases the capabilities of the NN directly from the RSSI values (that is, what is used to train the contrastive model). We showcase adequate results in predicting which room the user is in.

TABLE II
CLASSIFICATION REPORT OF NN FOR ROOM-ID BASED ON REPRESENTATIONS

	Precision	*Recall*	*f1-score*
Accuracy			0.04
Macro Avg.	0.01	0.02	0.01
Weighted Avg.	0.02	0.04	0.01

As we can conclude from Table II, classifying the room identifications directly from the representations confirms that the representations separate the locality from the user identification with a mediocre accuracy of 4%. Furthermore, when classifying the room identifications directly from the RSSI WAP signals, represented in Table 8, we can achieve a satisfactory classification of locality with an unstable accuracy of 70% (see the variations in precision and recall in Table III).

TABLE III
CLASSIFICATION REPORT OF NN FOR ROOM-ID BASED ON REPRESENTATIONS

	Precision	***Recall***	***f1-score***
Accuracy			0.70
Macro Avg.	0.59	0.58	0.55
Weighted Avg.	0.72	0.70	0.68

The results presented here justify our approach for the triplet loss and further enhance our hypothesis that our loss function design separates the locality information contained within the RSSI WAP values for user identification.

C. Discussions

This study aims to investigate the effectiveness and presence of patterns that may exist in a multi-building, multi-floor, multi-room dataset of WiFi fingerprinting based on RSSI values.

The individual identification experiment used contrastive learning to learn a feature representation to distinguish between different users, independent of the domain features such as the floor, room, and the location within the room. A Siamese neural network was trained with multiple loss functions: contrastive loss, label-aware contrastive loss, and triplet loss. We measure the effectiveness of representation learning through training a downstream classifier (based on the transformed features) and through visual analysis of the features.

These results demonstrate that the dataset contains patterns suitable for identifying individuals, with varying accuracies depending on how identification is set up. However, combined with explainable AI methods such as LIME, such identification can be helpful for different applications.

When a Random Forest classification was performed on the original dataset without transformation, the system achieved a classification accuracy of 0.75 (Due to the space limitation, we have not provided this result). This result demonstrates that the RSSI vector across the WAPs can be a valuable feature for the given dataset to distinguish between the users. However, it should be noted that the RF classifier captures details about where the specific user was through the WAPs activated. Since not all the users in the dataset collected information at all the locations tested, the classification result should be taken cautiously as a possibility to identify individuals. However, it could be argued that specific users will most likely be present only in a limited set of locations in a practical setting such as an office building.

D. Implications for Future work

While most of the existing identification methods of individuals from WiFi signals depend on CSI-based fine-grained feature analysis, most methods have been demonstrated only within a limited number of rooms with few individuals. The methods proposed in this paper, based on RSSI fingerprinting across different floors/rooms and a large number of individuals, could complement this capability. Thus, a fusion technique could help the generalisability of WiFi-based identification techniques.

Practical implementation of Authentication: The current method relies on RSSI values measured on an individual's mobile device. RSSI of a given device can also be estimated at the WAP. Thus, we can conceive one practical implementation of authentication through cooperative sensing across the WAPs for identification. Classification-based methods can be used as an authentication scheme when data can be collected for known users. If users cannot be classified confidently, that can indicate a new (unknown) user, or if the user has been in a room/location, they don't usually have.

V. CONCLUSIONS

This paper aims to utilize the popular UJIndoorLoc dataset used for localizing individuals in a building to extract unique signatures (fingerprints) left by individuals in this dataset. We propose a contrastive learning algorithm to learn an embedding space that minimizes the embedding distance between the same individual while maximizing the embedding distance between different users. While learning this embedding space, we also reduce the influence of location-based patterns by adversarially weighting the location-dependent triplet loss against the user-dependent triplet loss. We evaluate the performance of the representation learning scheme by a downstream classification method, which achieves an accuracy of 0.87 for user identification, with an f1-score of 0.86. A downstream classifier on the location shows a drop of accuracy from 0.7 to 0.04, demonstrating that the embedding space learned has removed the influence of location-based patterns.

The RSSI-based user identification method considered in this paper identifies a user who always carries a phone, thus limiting its application to such situations. RSSI-based distributed sensing can be a low-complex alternative to supplement and scale CSI-based identification to adapt to different domains. There is a need to collect dedicated datasets for improved representation learning for the identification of individuals, especially for extending the use of the method to identify individuals who do not carry mobile phones.

ACKNOWLEDGMENT

This work is funded by the Engineering and Physical Sciences Research Council under grant agreement number EP/X028631/1: A Trustworthy Robotic Autonomous System to Support Casualty Triage.

REFERENCES

[1] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Multi-Resident Non-Contact Vital Sign Monitoring Using Radar: A Review," IEEE Sensors Journal, vol. 21, no. 4, pp. 4061–4084, Feb. 2021.

[2] C. Shi, J. Liu, H. Liu, and Y. Chen, “Smart User Authentication through Actuation of Daily Activities Leveraging WiFi-enabled IoT,” in Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, in Mobihoc ’17. New York, NY, USA: Association for Computing Machinery, Jul. 2017, pp. 1–10. doi: 10.1145/3084041.3084061.

[3] Y. Zeng, P. H. Pathak, and P. Mohapatra, “WiWho: WiFi-Based Person Identification in Smart Spaces,” in 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Apr. 2016, pp. 1–12. doi: 10.1109/IPSN.2016.7460727.

[4] M. Shahzad and S. Zhang, “Augmenting User Identification with WiFi Based Gesture Recognition,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2, no. 3, p. 134:1-134:27, Sep. 2018, doi: 10.1145/3264944.

[5] R. Zhang, C. Jiang, S. Wu, Q. Zhou, X. Jing, and J. Mu, “Wi-Fi Sensing for Joint Gesture Recognition and Human Identification From Few Samples in Human-Computer Interaction,” IEEE Journal on Selected Areas in Communications, vol. 40, no. 7, pp. 2193–2205, Jul. 2022, doi: 10.1109/JSAC.2022.3155526.

[6] C. Li, M. Liu, and Z. Cao, “WiHF: Gesture and User Recognition With WiFi,” IEEE Transactions on Mobile Computing, vol. 21, no. 2, pp. 757–768, Feb. 2022, doi: 10.1109/TMC.2020.3009561.

[7] H. Kong et al., “Push the Limit of WiFi-based User Authentication towards Undefined Gestures,” in IEEE INFOCOM 2022 - IEEE Conference on Computer Communications, May 2022, pp. 410–419. doi: 10.1109/INFOCOM48880.2022.9796740.

[8] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, “Capturing the human figure through a wall,” ACM Trans. Graph., vol. 34, no. 6, pp. 1–13, Nov. 2015, doi: 10.1145/2816795.2818072.

[9] J. Torres-Sospedra et al., “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. doi: 10.1109/IPIN.2014.7275492.

[10] J. Bi, Y. Wang, B. Yu, H. Cao, T. Shi, and L. Huang, “Supplementary open dataset for WiFi indoor localization based on received signal strength,” Satellite Navigation, vol. 3, no. 1, p. 25, Nov. 2022, doi: 10.1186/s43020-022-00086-y.

[11] A. Gassner, C. Musat, A. Rusu, and A. Burg, “OpenCSI: An Open-Source Dataset for Indoor Localization Using CSI-Based Fingerprinting.” arXiv, Apr. 14, 2022. Accessed: Aug. 28, 2023. [Online]. Available: http://arxiv.org/abs/2104.07963

[12] F. Li, M. A. A. Al-qaness, Y. Zhang, B. Zhao, and X. Luan, “A Robust and Device-Free System for the Recognition and Classification of Elderly Activities,” Sensors (Basel), vol. 16, no. 12, p. 2043, Dec. 2016, doi: 10.3390/s16122043.

[13] E. Salomon, L. H. Botler, K. Diwold, C. A. Boano, and K. Romer, “Comparison of Channel State Information driven and RSSI-based WiFi Distance Estimation”.

[14] D. K. Pin Tan et al., “Integrated Sensing and Communication in 6G: Motivations, Use Cases, Requirements, Challenges and Future Directions,” in 2021 1st IEEE International Online Symposium on Joint Communications Sensing , Feb. 2021, pp. 1–6. doi: 10.1109/JCS52304.2021.9376324.

[15] H. Sun, L. G. Chia, and S. G. Razul, “Through-Wall Human Sensing With WiFi Passive Radar,” IEEE Transactions on Aerospace and Electronic Systems, vol. 57, no. 4, pp. 2135–2148, Aug. 2021, doi: 10.1109/TAES.2021.3069767.

[16] Z. Li, T. Jin, Y. Dai, and Y. Song, “Through-Wall Multi-Subject Localization and Vital Signs Monitoring Using UWB MIMO Imaging Radar,” Remote Sensing, vol. 13, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/rs13152905.

[17] Y. Zhang et al., “Widar3.0: Zero-Effort Cross-Domain Gesture Recognition With Wi-Fi,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 8671–8688, Nov. 2022, doi: 10.1109/TPAMI.2021.3105387.

[18] S. D. Regani, Q. Xu, B. Wang, M. Wu, and K. J. R. Liu, “Driver Authentication for Smart Car Using Wireless Sensing,” IEEE Internet of Things Journal, vol. 7, no. 3, pp. 2235–2246, Mar. 2020, doi: 10.1109/JIOT.2019.2958692.

[19] J. Liu, R. Ewing, E. Blasch, and J. Li, “Synthesis of Passive Human Radio Frequency Signatures via Generative Adversarial Network,” in 2021 IEEE Aerospace Conference (50100), Mar. 2021, pp. 1–9. doi: 10.1109/AERO50100.2021.9438476.

[20] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, “Physics-Aware Design of Multi-Branch GAN for Human RF Micro-Doppler Signature Synthesis,” in 2021 IEEE Radar Conference (RadarConf21), May 2021, pp. 1–6. doi: 10.1109/RadarConf2147009.2021.9455194.

[21] M. J. Bocus et al., “OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors,” Sci Data, vol. 9, no. 1, Art. no. 1, Aug. 2022, doi: 10.1038/s41597-022-01573-2.