IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 May 2024, accepted 10 July 2024, date of publication 15 July 2024, date of current version 1 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3427825

RESEARCH ARTICLE

TabCLR: Contrastive Learning Representation of Tabular Data Classification for Indoor-Outdoor Detection

MUHAMMAD BILAL AKRAM DASTAGIR, OMER TARIQ, AND DONGSOO HAN, (Senior Member, IEEE)

School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Yuseong, Daejeon 34141, South Korea

Corresponding author: Muhammad Bilal Akram Dastagir (bilal@kaist.ac.kr)

This work was supported by the Challengeable Future Defense Technology Research and Development Program through the Agency for Defense Development (ADD) funded by the Defense Acquisition Program Administration (DAPA), in 2024, under Grant 915027201.

ABSTRACT Indoor-outdoor detection (IOD) has gained prominence recently, particularly in positioning technology, leveraging smartphone-embedded sensors. It is pivotal in pedestrian localization, activity recognition, transportation mode classification, and power management of Internet of Things (IoT) devices. While several approaches have been explored for IOD, including threshold-based methods and machine learning-based models, challenges remain in addressing these models' temporal variations and computational complexities. Supervised learning approaches heavily rely on labeled datasets, which are costly and time-consuming to synthesize. We propose TabCLR, the first self-supervised learning (SSL) framework for IOD, to overcome these challenges. TabCLR utilizes contrastive learning representation tailored for tabular data classification using smartphone inertial sensors. It comprises data augmentation, a novel encoder network with self-attention, and an optimized contrastive loss function. Evaluation of TabCLR on multiple indoor-outdoor datasets demonstrates its superiority in both supervised and semi-supervised classification compared to existing methods. Notably, TabCLR outperforms SCARF by 6%-7%, indicating its effectiveness in capturing temporal feature representation patterns. Visualization analysis further illustrates TabCLR's distinctive clustering of feature embeddings compared to SCARF. TabCLR represents a significant advancement in SSL methodologies for indoor-outdoor detection classification. Its robust performance showcases its potential to enhance accuracy in indoor-outdoor integrated GPS systems, addressing critical challenges in IOD classification.

INDEX TERMS Contrasting learning representation, contrastive loss, IMU sensors, indoor-outdoor detection, self-supervise learning, tabular data.

I. INTRODUCTION

Indoor-outdoor detection (IOD) has become very popular in recent years due to its application in positioning technology, especially in environmental change detection using smartphone multimodal sensors. The deployment of location-based services on embedded systems leveraging low power consumption and on-device artificial intelligence offers significant economic and technical benefits in indoor-outdoor integrated GPS Systems [1], [2], [3] as shown in Figure 1. It has many applications in the fields of pedestrian localization and motion tracking [4], [5], [6], [7], [8], activity recognition [9], [10], transportation mode classification [11], [12], [13], power management and medical care [14] especially playing a key role in the smooth implementation of seamless positioning and navigation [15], [16], thus serving as a vital bridge between indoor and outdoor localization. IOD, along with other environmental conditions like time, weather, etc., provides personalized configurable services like adjusting the brightness of the smartphone screen and adjusting the volume on the device [17]. IOD models utilize smartphone sensor data to detect a user's environment.

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

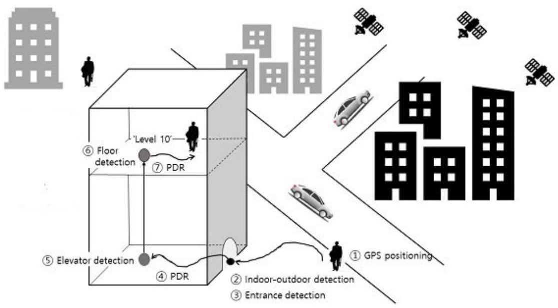

FIGURE 1. **Indoor-Outdoor-Integrated GPS System [2].**

Studies have been conducted that examine context-aware sensors showcasing distinct behaviors indoors and outdoors. Analyzing these attributes aids in predicting a user's behavior in a given setting. For instance, signals weaken when the user is passing through closed indoor spaces like a door, stairs, or elevator where lighting and magnetic fields fluctuate with indoor conditions, etc [18], [19].

Researchers have explored different approaches for IOD, for example, the threshold-based approach using sensor readings which make decisions on fixed pre-set values [20], [21], machine learning-based models adapting to diverse learning due to the detailed feature extraction from the data [17], [22], [23] but these suffer from limited exploration of temporal variations and computational complexities. Therefore, more complex models that integrate deep learning techniques for temporal aspects have been employed [24]. IOD is considered as a time series classification (TSC) problem in [25] while [26] considered IOD as multivariate time series classification which leveraging deep learning (DL) with self-attention mechanisms and spatial pyramids, achieving high accuracy but faces several challenges. For example, it relies on temporal relationships, and randomizing the dataset can disrupt these relationships, leading to degraded performance. Additionally, DL TSC data often has high dimensionality, as it is highly dependent upon the sequence of features, and it also requires more feature engineering to extract specific information for classification, which in results, impact the training time [27], [28].

All these DL approaches for IOD are based on supervised learning which is heavily dependent on the labeled datasets. However, labeling the datasets is a lengthy, time-consuming, and costly process. Therefore, this has led to the investigation of other methods like self-supervised learning. Self-supervised learning (SSL) is especially significant, as it enables deep learning progress by utilizing large amounts of unlabeled data to automatically learn features by learning important representations, eliminating the need for costly labeled datasets [29]. To address the challenge, semi-supervised learning approaches have been proposed for indoor and outdoor detection classification as mentioned in [30] and [31], but still, there is no framework available based on self-supervised learning for indoor-outdoor detection.

Herein lies the paramount importance of self-supervised learning approaches. Self-supervised learning, inspired by approaches like simCLR [32] initially designed for images followed by SCARF [33] for tabular classification, presents a promising direction. Adapting the contrastive learning representation for tabular data classification using inertial sensor data offers a groundbreaking opportunity for indoor-outdoor detection. Therefore, to our knowledge, we propose the first SSL framework for IOD called TabCLR, a contrastive learning representation learning of tabular data classification for indoor and outdoor detection using smartphone inertial sensors.

TabCLR encapsulates three important steps in contrasting learning representation. Firstly, in the augmentation phase, we randomly add corrupted features in the unlabeled dataset to generate positive pairs, fostering semantic similarity, while forming negative pairs by contrasting corrupted data with the remaining batch components. Secondly, a novel encoder model incorporates self-attention for temporal representation extraction and spatial features from the base encoder [33]. Thirdly, we encapsulate a modified form of NXENT contrastive loss function from simCLR [32], called spatial-context contrastive loss (SCCL), considering the dynamics of spatial context inherent in indoor-outdoor detection. Finally, we also present a fine-tuning phase, where labeled inertial sensor data is classified into indoor-outdoor classification using a lightweight DNN-based classifier after getting feature representation from the pre-trained novel encoder.

We conducted a comprehensive evaluation of multiple indoor-outdoor datasets, including our proposed mega indoor-outdoor detection (MIOD), deep indoor-outdoor detection (DIOD) and KAIST indoor-outdoor detection (KIOD), along with one public Cambridge indoor-outdoor detection (CIOD) [26]. The results demonstrate TabCLR's superiority in supervised and semi-supervised indoor-outdoor classification. Being the first SSL method for IOD, we also evaluated on both seen and unseen datasets and compared with SCARF [33], a popular tabular SSL method where results show that TabCLR outperforms it by $6\% - 7\%$, highlighting its effectiveness in capturing temporal feature representation patterns in inertial sensor data. Finally, we demonstrated using t-SNE [34] representation analysis that TabCLR exhibits a clearer 3D visualization of the clustering of feature embedding as compared to SCARF. In a nutshell, TabCLR is tailored specifically for indoor-outdoor detection classification, and its robust performance surpasses existing methods, showcasing its potential for enhancing indoor-outdoor detection accuracy in indoor-outdoor integrated GPS systems.

A. MAJOR CONTRIBUTIONS

The major contributions of our paper are summarized as follows:

1) We present TabCLR, a contrastive learning representation learning of tabular data classification for

indoor and outdoor detection that encapsulates the corruption mechanism and a novel encoder model that incorporates self-attention for temporal representation extraction spatial features, and we provide a pre-head training algorithm and fine-tuning mechanism.

2) We introduce a modified form of contrastive loss function known as spatial-context contrastive loss (SCCL), specifically tailored to enhance indoor-outdoor detection for the dynamics of spatial context.
3) We conducted a comprehensive evaluation, demonstrating that our framework outperforms the existing state-of-the-art methods for indoor and outdoor detection, including SCARF [33], a popular tabular SSL method.
4) We propose three new datasets, the mega indoor-outdoor detection (MIOD) dataset and deep indoor-outdoor detection (DIOD) encapsulating magnetometer, accelerometer, and gyroscope (MIOD), and the KAIST indoor-outdoor detection (KIOD) dataset encapsulating accelerometer and gyroscope.

B. PAPER ORGANIZATION

Our research work is structured in the following way. Section II highlights the state-of-the-art (S.O.T.A.) works in the IOD and SSL domains. Section III covers the detailed methodology section of the proposed framework. Section IV. Demonstrate and discuss the experiments, results, comprehensive analysis, and comparison, followed by a conclusion in Section V.

II. RELATED WORKS

The evolution of Indoor/Outdoor Detection (IOD) models utilizing smartphone sensor data represents a substantial leap in context-aware technology. Context-aware sensors, dissecting the nuanced behaviors exhibited in indoor and outdoor settings, serve as a cornerstone for predicting user behavior within a given environment. Notably, signals traversing through indoor materials tend to weaken, while dynamic fluctuations in lighting and magnetic fields characterize indoor conditions [18], [19]. Diverse methodologies have been explored in this domain, showcasing the evolution from threshold-based models using raw sensor readings [20], [21] to machine learning-based approaches adapting to varied data distributions [17], [22]. While machine learning models display superior performance compared to their threshold-based counterparts, inherent challenges persist, encompassing the limited exploration of temporal variations and the computational complexities tied to their implementation. Consequently, researchers have ventured into more sophisticated models integrating deep learning techniques to address these persisting challenges [17]. The adoption of deep learning models for temporal analysis signifies a significant stride, recognizing Indoor/Outdoor Detection as a time series classification problem [25]. Furthermore, studies like [26] have leveraged deep learning architectures infused with self-attention mechanisms and spatial pyramids. Despite achieving commendable accuracy rates, these approaches grapple with issues pertaining to imbalanced datasets and high memory usage, signifying potential avenues for further improvement. The trajectory of IOD models encapsulates a shift towards more sophisticated, data-driven paradigms, providing promising avenues to comprehend complex user behaviors in distinct settings. However, the persistence of challenges exists. For example, it relies on temporal relationships, and randomizing the dataset can disrupt these relationships, leading to degraded performance. Additionally, DL TSC data often has high dimensionality, as it is highly dependent upon the sequence of features, and it also requires more feature engineering to extract specific information for classification, which, as a result, impacts the training time [27], [28].

III. METHODOLOGY

Contrastive learning [32], a key aspect of SSL, has shown a major contribution to the image classification domain. However, the methods utilized in images cannot be directly applied to inertial sensor datasets for indoor-outdoor detection, which are commonly encountered in real-world location-based services due to the diverse nature of data representation. Therefore, a self-supervised learning framework is needed to address the IOD classification problem as shown in Figure 2. In this section, we present TabCLR, a contrastive learning representation of tabular data classification for indoor and outdoor detection. Unlike simCLR [32], it introduces a contrastive framework for indoor-outdoor detection using inertial sensors based on the augmentation technique [33]. First, we discuss the problem formulation and objective, random permutation feature corruption (RPFC), modified contrastive loss (SCCL), TabCLR framework, along with its components, and their interactions.

A. PROBLEM FORMULATION AND OBJECTIVE

We defined the problem of IOD as tabular data classification (TDC) by defining given a dataset of data collected from IMU Sensors for both indoor (I) and outdoor (O) environments, denoted as $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i represents the unlabeled inertial sensors data for the ith sample and y_i is the corresponding label indicating whether the sample is from an indoor or outdoor environment ($y_i \in \{I, O\}$), will used for fine-tuning after self-supervised learning. Our goal is to learn the inertial representation $f(x_i)$ that captures the spatio-temporal features crucial for indoor-outdoor classification. Our objective is to optimize the following:

$$\min_{\theta,\phi} \frac{1}{N} \sum_{i=1}^{N} \left[\log \frac{\exp(\text{sim}(f_\theta(x_i), f_\theta(x_j)))}{\sum_{k=1}^{N} \exp(\text{sim}(f_\theta(x_i), f_\theta(x_k)))} \right] \quad (1)$$

Here, f_θ denotes the encoding function parameterized by θ, and $\text{sim}(\cdot)$ represents the similarity measure between representations. The dataset is divided into an unlabelled

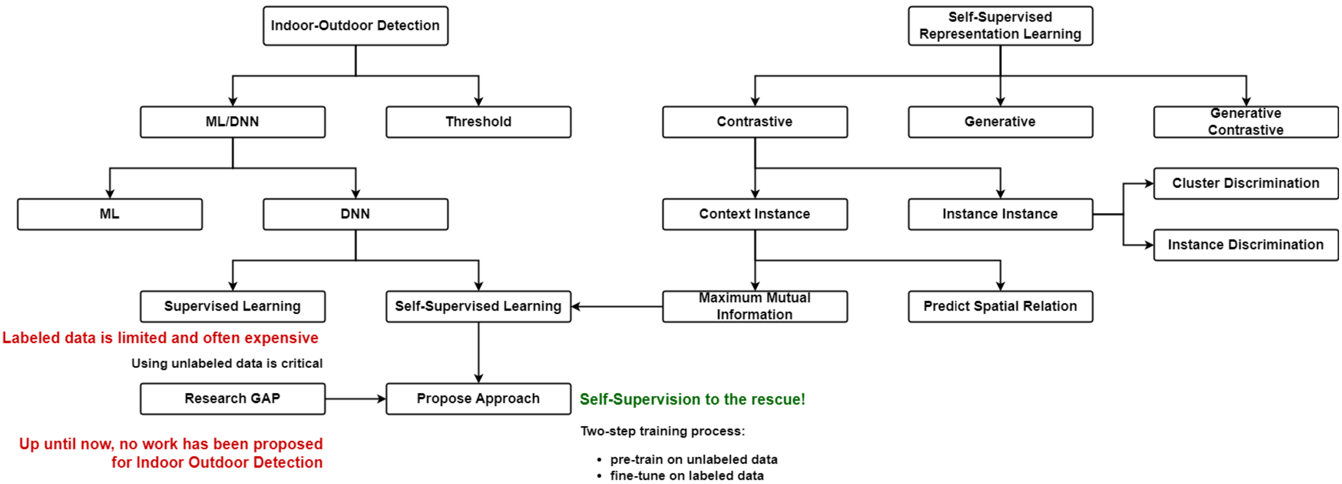

FIGURE 2. Indoor-outdoor detection in term of self-supervised representation learning.

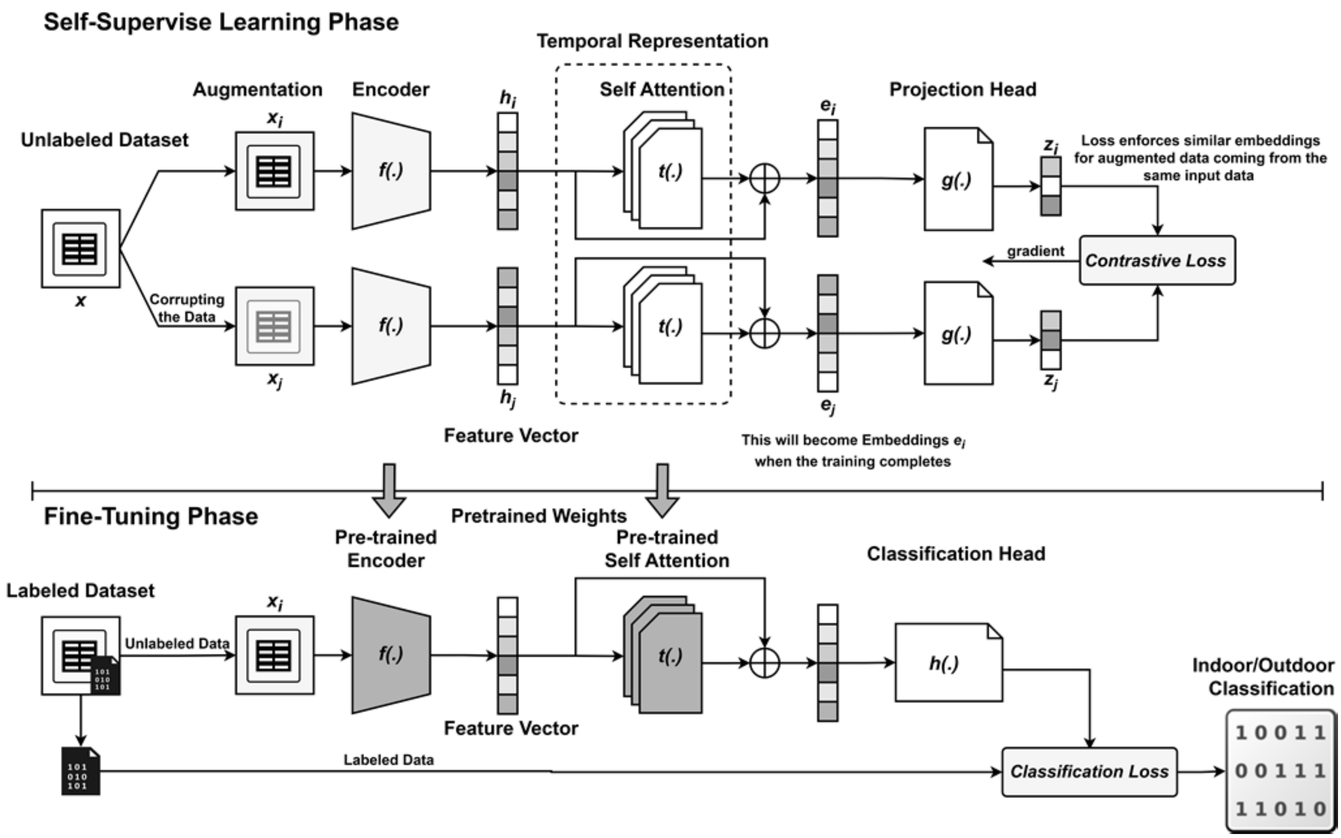

FIGURE 3. TabCLR: contrastive learning representation of tabular data for indoor-outdoor detection.

contrastive self-supervised learning phase and a labeled fine-tuning phase. During self-supervised learning, the objective is to minimize intra-cluster distances (d_{intra}) between semantically similar data points and maximize inter-cluster distances (d_{inter}) between dissimilar points:

$$\begin{aligned} d_{intra}(x_i, x_j) &= \|f(x_i) - f(x_j)\|_2 \quad \text{(Intra-cluster distance)} \\ d_{inter}(x_i, x_k) &= \|f(x_i) - f(x_k)\|_2 \quad \text{(Inter-cluster distance)} \end{aligned} \tag{2}$$

The model aims to minimize the ratio of intra-cluster distance to inter-cluster distance:

$$\text{Minimize } \frac{d_{intra}}{d_{inter}} \tag{3}$$

B. RANDOM PERMUTATION FEATURE CORRUPTION (RPFC)

TabCLR incorporates an important augmentation mechanism inspired by [33] for contrasting learning representation of inertial sensors data known as the random permutation feature corruption (RPFC), which is essential for enhancing the model's ability to distinguish between indoor and outdoor contexts. To create positive pairs, this mechanism corrupts a subset of features within the sensor data to generate positive pairs, while contrasting corrupted data within the remaining batch yields negative pairs. Then it utilizes empirical marginal distributions of features to guide the data augmentation process. Sampling from these distributions ensures the realism and representativeness of the augmented data, enhancing the quality of learned representations. The empirical sampling is mathematically represented as:

$$x_{\text{augmented}} \sim \text{Uniform}(\text{features\_low}, \text{features\_high}) \quad (4)$$

Let x be the input tensor of size $B \times M$ where B is the batch size, and M is the length of the input tensor. The corrupted tensor x_c is expressed as:

$$x_c = x \odot M_c + R \odot (\mathbf{1}_{B\times M} - M_c) \quad (5)$$

where $\odot$ represents element-wise multiplication, and $\mathbf{1}_{B\times M}$ is a tensor of ones with the same size as x. The algorithm steps are shown in Algorithm 1. which randomly corrupts a subset of features in each batch of the input tensor x and replaces the corrupted features with samples from the marginals. This corruption process is controlled by the parameters B, L, and the random permutation I.

Algorithm 1 Random Permutation Feature Corruption

Input : Input tensor x, B (batch size), L (corruption length)
Output: Corrupted tensor x_c
$B, M \leftarrow$ size of x
$M_c \leftarrow$ tensor of zeros with the same size as x

for $i \leftarrow 1$ **to** B **do**
  $I \leftarrow$ random permutation of M elements with length L
  $M_c[i, I] \leftarrow$ True
end for
$R \leftarrow$ samples from marginals of size (B, M)
$x_c \leftarrow$ element-wise selection based on M_c

C. CONTRASTIVE LOSS

1) simCLR LOSS FUNCTION

The simCLR [32] loss function as shown in Eq. 6, originally designed for visual data, falls short in addressing the intricacies of inertial sensor data in the context of indoor-outdoor detection. Several limitations can be identified, especially since simCLR does not explicitly incorporate spatial context, a critical aspect in scenarios involving inertial sensors, and simCLR's focus on similarity without explicitly accounting for spatial dynamics makes it less sensitive to context-specific spatial patterns.

$$\mathcal{L}_{\text{C}} = -\frac{1}{2B}\sum_{i=1}^{2B}\log\left(\frac{e^{\text{sim}_{ii}/\tau}}{\sum_{j\neq i} e^{\text{sim}_{ij}/\tau}}\right) \quad (6)$$

2) PROPOSAL OF SCCL LOSS FUNCTION

To address these limitations, we propose spatial-context contrastive loss (SCCL), as shown in Eq. 7, specifically tailored for indoor-outdoor detection using inertial sensors. SCCL incorporates a spatial consistency term, enhancing its effectiveness in capturing spatial relationships.

$$\mathcal{L}_{\text{SCCL}} = -\frac{1}{2B}\sum_{i=1}^{2B}\log\left(\frac{e^{\text{sim}_{ii}/\tau}}{\sum_{j\neq i} e^{\text{sim}_{ij}/\tau}}\right) + \alpha \cdot \frac{1}{B}\sum_{k=1}^{B}\|\mathbf{z}_{ik} - \mathbf{z}_{jk}\|_2 \quad (7)$$

where sim_{ii} represents positive similarities, z_{ik} and z_{jk} are embeddings of anchor and positive samples, and α is the weight parameter for the spatial consistency term. Also, B is the batch size, τ is the temperature parameter, and α is the weight parameter for the spatial consistency term.

The first term aims to maximize the similarity between positive pairs (sim_{ii}) and minimize the similarity between negative pairs (sim_{ij}), effectively enhancing the clustering of similar samples. The second term enforces consistency in feature representations across augmentations, promoting robustness and discriminative feature learning. This design choice ensures that the self-supervised learning framework aligns with the particularities of inertial sensor signals, leading to more effective representation learning. This loss function encourages the model to learn meaningful representations conducive to indoor-outdoor detection classification.

D. TabCLR FRAMEWORK

In this section, we describe our proposed methodology for self-supervised contrastive learning with SCCL loss and self-attention. The algorithm is outlined in Algorithm 2, and Figure 3 provides a visual representation. For each mini-batch of examples from the unlabeled training data, a corrupted version $\tilde{x}(i)$ is generated for each example $x(i)$. The corruption process involves randomly sampling a fraction of features from each example uniformly at random. Random draws from the empirical marginal distribution of the respective feature and then replaces these selected features. Both the original $x(i)$ and corrupted $\tilde{x}(i)$ instances are passed through the encoder network (f), which incorporates self-attention (t) for temporal representation. The encoder network's output is then processed through the pre-train head network (g), which normalizes the outputs to lie on the unit hypersphere. This normalization is deemed crucial in practice. The resulting representations are denoted as $z(i)$ and $\tilde{z}(i)$ for the original and corrupted instances, respectively. The training objective involves optimizing the parameters of both the encoder (f) with self-attention (t) and the pre-train head (g) networks through stochastic gradient descent (SGD). The optimization process is guided by the SCCL loss function, as mentioned in Eq 7. To facilitate downstream tasks, a classifier is fine-tuned. The encoder network (f) with self-attention (t)

Algorithm 2 Pre-Head Training Algorithm

Input : Unlabeled training data $X \subseteq \mathbb{R}^M$, batch size B, temperature τ, alpha α, corruption rate c, encoder network f, self-attention t, pre-train head network g

Output: Encoder network with self-attention $\rightarrow e$

for *sampled tabular-batch* $x_B^{(i)} \subseteq X$ **do**

  for $i \in [B]$, *uniformly sample subset* I_i *from* $\{1, \ldots, M\}$ *of size* q **do**

    Define $\tilde{x}^{(i)} \in \mathbb{R}^M$ as follows: $\tilde{x}_j^{(i)} = x_j^{(i)}$ if $j \notin I_i$, otherwise $\tilde{x}_j^{(i)} = v$ where $v \sim X_j$.

  end for

  Initialize Q, K, V matrices with x by splitting x into Q, K, V

  $Q = x \cdot W_Q$, $K = x \cdot W_K$, $V = x \cdot W_V$ where W_Q, W_K, W_V are learned weights

  Compute scaled dot-product attention scores: $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_{\text{model}}}}\right) \cdot V$

  $t(x^{(i)}) \leftarrow \text{Attention}(Q, K, V)$

  $e(x^{(i)}) = f(x^{(i)}) \oplus t(x^{(i)})$

  $z^{(i)} = g(e(x^{(i)}))$, $\tilde{z}^{(i)} = g(e(\tilde{x}^{(i)}))$ for $i \in [B]$.

  $s_{i,j} = \frac{z^{(i)} \cdot z^{(j)}}{\|z^{(i)}\|_2 \cdot \|\tilde{z}^{(j)}\|_2}$ for $i, j \in [B]$.

  Define $L_{\text{SCCL}}(z_i, z_j) = -\frac{1}{2B} \sum_{k=1}^{2B} \log\left(\frac{e^{\text{sim}_{ij}^{(k)}/\tau}}{\sum_{l \neq k} e^{\text{sim}_{kl}^{(k)}/\tau}}\right) + \alpha \cdot \frac{1}{B} \sum_{m=1}^{B} \|z_i^{(m)} - z_j^{(m)}\|_2$

  Update encoder e and prehead g to minimize L_{SCCL} utilizing stochastic gradient descent.

end for

---

is retained, and a classification head (h) is attached to predict labels based on the output of the encoder. Cross-entropy classification loss is optimized, and the parameters of both the encoder and classifier are tuned. Pre-training is conducted for a pre-determined number of epochs. The specific number of epochs required depends on the model and dataset characteristics. To determine an appropriate stopping point, we propose using early stopping based on the validation of SCCL loss. The loss on a validation set generated by running the proposed method on unlabeled validation data is monitored during pre-training. Our updated proposed method leverages self-attention in the encoder for temporal representation, incorporating SCCL contrastive loss for self-supervised pre-training. The subsequent fine-tuning process includes both the pre-trained encoder and a classifier for downstream tasks. Early stopping on the validation SCCL loss is proposed to determine the optimal pre-training duration. The overall methodology aims to learn robust representations for diverse downstream tasks through self-supervised learning, especially for indoor-outdoor detection classification.

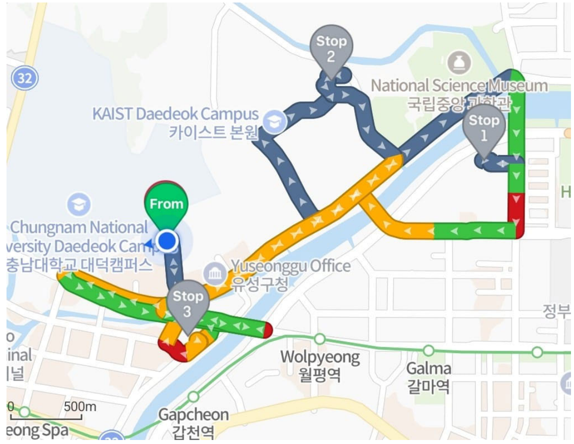

FIGURE 4. Propose MEGA indoor outdoor detection dataset (MIOD).

FIGURE 5. Propose KAIST indoor outdoor detection dataset (KIOD).

TABLE 1. Sensor contributions in public dataset CIOD [26] for indoor/outdoor detection.

Sensors	Contribution Highlights
Cellular Signal Strength	Signal variations indicating indoor-outdoor transitions
WiFi Signal Strength	Quality and intensity of nearby WiFi networks
Ambient Light Intensity	Differentiating environments based on light levels
Accelerometer	Detecting user motion patterns
Total Magnetic Intensity	Insights into magnetic field variations
Sound Intensity	Differentiating environments based on noise levels
Proximity Sensor	Reliability of ambient light data
Day/Night Label	Temporal context for diurnal variations

IV. EVALUATION OF OUR APPROACH

In this section, we conduct a comprehensive assessment of our methodology, comparing it with various other approaches in indoor-outdoor detection.

A. INDOOR-OUTDOOR DETECTION (IOD) DATASETS

1) PUBLIC CIOD DATASET

The Cambridge Indoor-Outdoor Detection (CIOD) dataset [26] was collected from various locations in close proximity,

TABLE 2. Sensor contributions in proposed datasets: MIOD (Seen), KIOD (Seen), and DIOD (Unseen) for indoor/outdoor detection.

Sensors	Contribution Highlights
Magnetometer	Provides valuable insights into magnetic signatures, helping to distinguish indoor and outdoor environments through analysis of magnetic field strength and direction variations.
Accelerometer	Captures movement patterns and velocity changes, aiding in differentiating between indoor and outdoor spaces by identifying acceleration patterns indicative of transitions.
Gyroscope	Offers information on rotational movements and orientation changes, assisting in identifying transitions between indoor and outdoor environments through analysis of rotational patterns associated with specific activities or environmental changes.

resulting in significant signal pattern variations across different environments. This dataset comprises feature sets collected over six months from two smartphones, the Redmi Note 9 and Huawei P30 lite. The dataset includes eight features: ambient light intensity, sound intensity, magnetic intensity, Reference Signal Received Quality (RSRQ), proximity sensor readings, and a binary variable indicating day and night. These features provide insights into daily activities such as work and travel. Data were recorded under various weather conditions, including rain and clear skies. Min-max normalization was applied to the sensor data at a sampling rate of 1 Hz, and significant outliers (top 1% of the sample distribution) were removed to ensure accurate data representation and avoid issues caused by malfunctioning sensors. Figure 6-a shows the dataset size, with over 1.4 million samples for training and over 14,000 samples for testing. The sensor contributions are detailed in Table 3.

2) PROPOSED MIOD AND KIOD (SEEN) DATASETS

The Mega Indoor-Outdoor Detection (MIOD) and KAIST Indoor-Outdoor Detection (KIOD) datasets are meticulously collected for indoor and outdoor environment classification. MIOD uses data from the Samsung Galaxy Fold, including accelerometer, magnetometer, and gyroscope sensors, sampled uniformly at 10 Hz within a 15 km radius around KAIST University (Figure 6-b). The trajectory forms a closed-loop path with three stops representing different activity contexts, with 10% allocated to starting and ending points, 25% to indoor-outdoor transitions, and 20% to predominantly indoor activities. Similarly, KIOD covers 1.2 kilometers from the KAIST N1 Building to the KAIST Subway (Figure 6-c), focusing on accelerometer and gyroscope sensors with sampling rates and trajectory characteristics identical to MIOD. Labeling follows established conventions, distinguishing between "Indoor" and "Outdoor" classes. Both datasets support the development of classification algorithms by offering precise geographical routes, strategic stops, and careful labeling, contributing to the advancement of indoor-outdoor detection using inertial sensors. MIOD and KIOD datasets have over 1 million and 40,000 samples for training, and over 14,000 and 400 samples for testing, respectively. Sensor contributions are detailed in Table 4.

TABLE 3. Public dataset CIOD [26] sensor contributions in indoor/outdoor detection.

Sensors	Contribution Highlights
Cellular Signal Strength	Signal variations indicating indoor-outdoor transitions
WiFi Signal Strength	Quality and intensity of nearby WiFi networks
Ambient Light Intensity	Distinguishing environments based on light levels
Accelerometer	Detecting user motion patterns
Total Magnetic Intensity	Insights into magnetic field variations
Sound Intensity	Differentiating environments based on noise levels
Proximity Sensor	Reliability of ambient light data
Day/Night Label	Temporal context for diurnal variations

3) PROPOSED DIOD (UNSEEN) DATASET

We also propose the Deep Indoor-Outdoor Detection (DIOD) dataset for comprehensive evaluation after training on the MIOD dataset. Like MIOD, DIOD is collected using the Samsung Galaxy Fold, sampled uniformly at 10 Hz at six locations within KAIST: Indoor Corridor, Indoor Hallway, Indoor Stairs, Outdoor Campus, Parking Lots, and Outdoor Roads. The dataset comprises 50% indoor samples (20% for Indoor Corridor and Indoor Hallway each, 10% for Indoor Stairs) and 50% outdoor samples (25% for Outdoor Campus, 10% for Parking Lot, 15% for Outdoor Roads). Figure 6-d shows the dataset size, with over 1.4 million samples for training and over 14,000 samples for testing.

V. EVALUATION AND DISCUSSION

In this section, we investigate and evaluate our approach through comprehensive evaluation and compare the results with various other approaches in indoor-outdoor detection.

A. INDOOR-OUTDOOR DETECTION (IOD) DATASETS

1) PUBLIC CIOD DATASET

Cambridge indoor-outdoor detection [26] (CIOD) was obtained by collecting data from various locations in close proximity to each other. This resulted in significant variations in signal patterns in different environments. The dataset includes feature sets that were collected over a period of six months from smartphones, specifically two Redmi Note 9 and Huawei P30 lite devices. The dataset consists of 8 features, namely ambient light intensity, sound intensity, magnetic intensity, quality of Reference Signal Received Quality (RSRQ), proximity sensor readings, and a binary variable indicating day and night. These features provide information about daily activities such as work and travel. The data was recorded under different weather conditions, including rain and clear sky. To normalize the sensor data, min and max normalization operations were applied with a sampling rate of 1 Hz. Prior to normalization, any significant

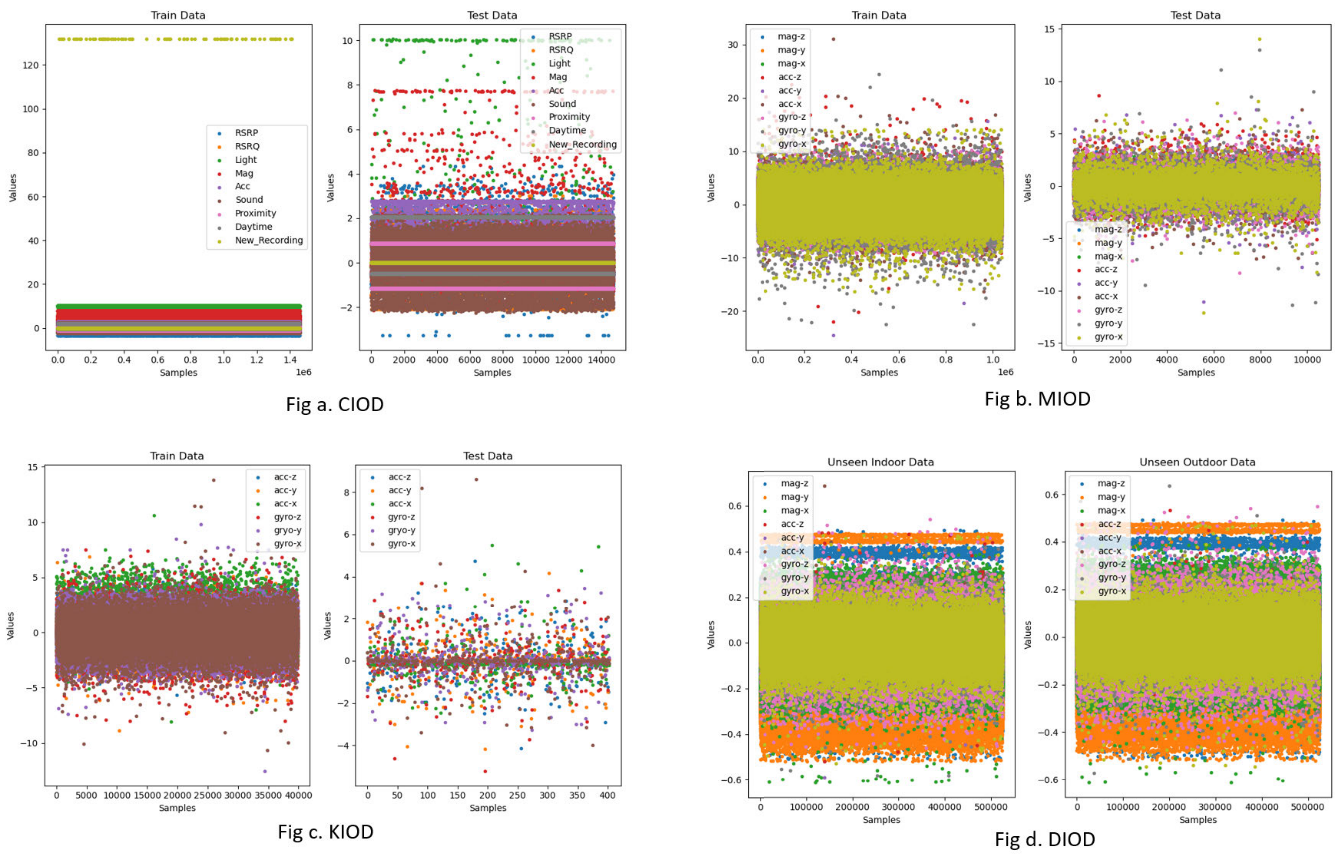

FIGURE 6. Size of the public CIOD [26], proposed MIOD, KIOD and DIOD Datasets for IOD.

TABLE 4. Propose dataset MIOD(Seen), KIOD(Seen) and DIOD(Unseen) sensor contributions in indoor/outdoor detection.

Sensors	Contribution Highlights
Magnetometer	Valuable for detecting changes in orientation and direction, magnetometer data in IO detection offer insights into magnetic signatures, aiding in distinguishing indoor and outdoor environments by analyzing variations in magnetic field strength and direction.
Accelerometer	Essential for capturing movement patterns and velocity changes, accelerometer data in IO detection help differentiate between indoor and outdoor spaces by identifying acceleration patterns indicative of transitions, such as sudden velocity changes or shifts in gravitational orientation.
Gyroscope	Gyroscopic data in IO detection provide information about rotational movements and orientation changes, aiding in identifying transitions between indoor and outdoor environments by analyzing rotational patterns associated with specific activities or environmental changes.

outliers, identified as the top 1% of the sample distribution, were removed to ensure accurate data representation and to prevent any issues caused by malfunctioning smartphone sensors. The size of the dataset is shown in Figure 6-a, with more than 1.4 million samples for the training while more than 14000 samples are used for the testing. Sensors used are mentioned in Table 3 along with each sensors contribution.

2) PROPOSE MIOD AND KIOD (SEEN) DATASETS

The mega indoor-outdoor detection (MIOD) and KAIST indoor-outdoor detection (KIOD) datasets are carefully collected for the classification of indoor and outdoor environments. MIOD uses data from the Samsung Galaxy Fold, including accelerometer, magnetometer, and gyroscope sensors, sampled uniformly at 10 Hz within a 15 km radius around KAIST University as shown in Figure 4. Its trajectory forms a closed-loop circular path with three stops, each representing different activity contexts. Stop proportions vary, with 10% allocated to starting and ending points, 25% to the indoor-outdoor transitions and 20% to predominantly indoor activities. Similarly, KIOD covers 1.2 kilometers from the KAIST N1 Building to the KAIST Subway as shown in Figure 5, focusing on accelerometer and gyroscope sensors with sampling rates and trajectory characteristics identical to those of MIOD. Its labeling follows established conventions, distinguishing between "Indoor" and "Outdoor" classes. Both datasets support the development of classification algorithms by offering precise geographical routes, strategic stops, and careful labeling, thereby contributing to the advancement of indoor-outdoor detection classification problems using only inertial sensors. The size of the datasets MIOD and KIOD are shown in Figure 6-b and Figure 6-c, with more than 1 million samples and 40000 samples for the training while more than 14000 samples and 400 samples are used for the testing respectively. Sensors used are mentioned in Table 4 along with each sensors contribution.

3) PROPOSE DIOD DATASETS (UNSEEN) DATASET

We also proposed Deep Indoor Outdoor Detection (DIOD) as unseen dataset for a comprehensive evaluation after training in the MIOD dataset. Like MIOD, DIOD is collected using Samsung Galaxy Fold including sampled uniformly at 10 Hz at various six locations within KAIST like Indoor Corridor, Indoor Hallway, Indoor Stairs, Outdoor Campus, Parking lots, Outdoor Roads. 50% of the samples are collected for indoor environment e.g 20% for Indoor Corridor and Indoor Hallway each, while 10% for Indoor Stairs. While remaining 50% of the samples are collected for outdoor environment e.g 25% for Outdoor Campus, 10% for the Parking lot and 15% for the Outdoor Roads. The size of the dataset is shown in Figure 6-d, with more than 1.4 million samples for the training while more than 14000 samples are used for the testing.

B. DATA PRE-PROCESSING

In the pre-processing phase, data are transformed into numerical representations using a one-hot encoding scheme. This encoding method facilitates the representation of categorical variables as binary vectors, thereby enabling their integration into machine learning algorithms. Subsequently, the encoded categorical features undergo careful analysis and aim to evaluate the robustness of the model to perturbations and uncertainties in the data representations. Addressing the issue of missing data, a systematic approach is adopted to handle absent values in the dataset. Initially, feature columns that consistently lack data across observations are identified and subsequently excluded from further analyses. Following this step, imputation strategies are employed to address missing values within categorical and numerical attributes. Specifically, for categorical features, missing entries are imputed using the mode, representing the most frequently occurring category observed across the entirety of the dataset. In contrast, numerical features with missing values are imputed utilizing the mean of the available data. By implementing these pre-processing techniques, the dataset is prepared for subsequent analyses, ensuring the integrity and reliability of the data for modeling purposes.

C. MODEL ARCHITECTURE AND TRAINING

We use the following settings across all experiments. As delineated previously, the neural network architecture is partitioned into distinct components: an encoder f augmented with attention t, denoted collectively as e, a pre-training head g, and a classification head h. The inputs to both g and h are derived from the outputs of e. For uniformity, all three constituent models are instantiated as ReLU networks with shared hidden dimensions of 256. The encoder f comprises four layers, while both g and h consist of two layers each. In the experimental setup, both the proposed approach and SCARF framework, along with the autoencoder baselines, utilize the pre-training head g. The optimization process for all models and their components is conducted using the Adam optimizer with a default learning rate of 0.0001. A batch size of 256 is employed for both pre-training and fine-tuning phases. Unsupervised pre-training methods incorporate early stopping with a patience parameter of 10 on the validation loss. Similarly, supervised fine-tuning adheres to this early stopping criterion, utilizing classification error as the validation metric, which has shown marginally superior performance. Fine-tuning is constrained to a maximum of 100 epochs, while pre-training epochs are capped at 5000. A static validation set is constructed over 10 epochs. The corruption rate c is fixed at 0.8, with α set to 1 and a temperature parameter τ set to 1, and also different variations were tested. Each experiment is repeated 50 times with distinct train/validation/test splits to ensure the robustness of the results. The model architecture is implemented using PyTorch, with CUDA serving as the backend for neural network training and inference. Experimentation is conducted on a hardware setup comprising a PC equipped with an Intel i9 CPU operating at 3.20 GHz, 32GB of DDR5 random access memory, and an RTX 4090 GPU. The proposed neural network framework is trained under a contrastive self-supervised learning paradigm, followed by fine-tuning with labeled data for indoor-outdoor detection.

D. COMPARISON WITH RELATED IOD WORKS

Our investigation involved a detailed comparison of various indoor-outdoor detection (IOD) models, as presented in Table 5. This comparison involves a supervised approach using ML-based classification approaches, DL models, and threshold-based methodologies. Although ML-based classification approaches, such as Random Forest [30], [35] and Multi-Layer Perceptron [35], [36], have been widely used for classification tasks, including indoor-outdoor detection, they may struggle to capture complex temporal patterns present in sensor data, particularly in dynamic and noisy environments. Moreover, feature engineering and selection are crucial for the effectiveness of these algorithms, which can be labor intensive and require domain expertise. In DL approaches, we specifically investigated DenseNet-LSTM [24] and CAP-ALSTM [26] models designed for IOD classification. Although these leverage deep neural networks to automatically learn features from raw sensor data, making it a promising approach for indoor-outdoor detection, DL TSC comes with its own set of challenges. One significant limitation is the loss of connectivity when using techniques such as randomization to the dataset for efficient and more efficient training. Time-series data are inherently based on temporal relationships, and randomizing the dataset can disrupt these relationships, leading to degraded performance. Furthermore, DL TSC data often have high dimensionality,

TABLE 5. Indoor outdoor detection accuracy comparison with related works.

Method	CIOD Dataset									Approach
	Accuracy	F1-Score	Precision	Recall	MD	TT	SPT	RM	MP	
SenseIO [21]	67.1 ± 5.80	77.8 ± 5	77.6 ± 3.0	77.4 ± 5.3	5	-	-	-	-	TH
IODetector [20]	68.10 ± 8.47	77.7 ± 6.79	77.9 ± 3.5	77.8 ± 5.37	9	-	-	-	-	TH
RF [30], [35]	85.59 ± 8.42	87.75 ± 7.6	85.2 ± 4.5	87.75 ± 6.2	2	**123**	7.1	2.9	-	SL
MLP [35], [36]	86.98 ± 6.50	89.14 ± 5.8	84.3 ± 3.7	89.14 ± 5.0	2	312	6.3	16.4	176K	SL
Dense-LSTM [24]	88.05 ± 6.42	89.84 ± 5.56	87.5 ± 2.6	89.84 ± 4.0	4	68440	17.1	500.1	146M	SL
CAP-ALSTM [26]	89.36 ± 5.28	90.97 ± 5.06	89.8 ± 3.1	90.97 ± 4.5	3	4458	4.9	5.4	74K	SL
MB-SVM-HMM [23]	$92.17, \pm 2.23$	92.35 ± 2.46	91.7 ± 1.9	92.35 ± 2.1	5	-	-	-	-	SL
SSDL [31]	89.77 ± 6.72	88.03 ± 8.14	88.1 ± 4.2	88.03 ± 5.3	30	-	-	-	-	Semi-SL
Proposed Approach	**92.61±2.17**	**93.03±2.28**	**93.51±2.85**	**93.14±2.17**	**0.01**	882	**0.177**	**0.13**	32K	**Self-SL**

MD = Median Delay (s), *TT* = Training Time (s),
SPT = Single Prediction Time (msec),
RM = Required Memory (Mb),
MP = Model Parameters
TH = Threshold, *SL* = Supervise Learning

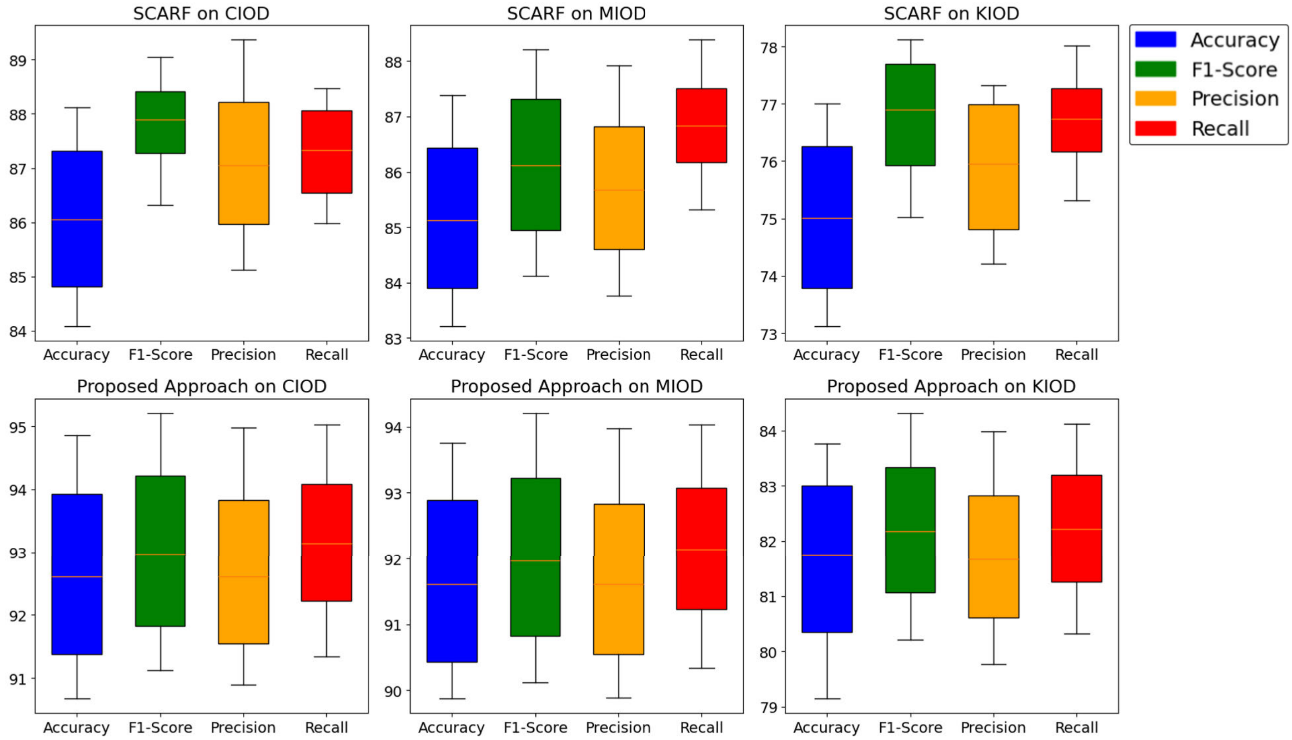

FIGURE 7. Comprehensive evaluation of the accuracy, F1-Score, Precision and Recall with the comparison of the proposed approach with the baseline SCARF on the indoor outdoor datasets (CIOD, MIOD, and KIOD) repeated 100 times.

as they are highly dependent on the sequence of features, and it also requires more feature engineering to extract specific information for classification, which, as a result, impacts training time [27], [28].

Furthermore, for a well-rounded comparison, our evaluation includes two threshold-based models, IODetector [20], and SenseIO [21]. These approaches rely on predefined thresholds to classify indoor and outdoor environments based on sensor readings. Although simple and easy to implement, threshold-based methods often lack adaptability and robustness, as they may fail to generalize well to different environmental conditions and sensor variations. Additionally, setting appropriate thresholds can be challenging and may require manual tuning, making these approaches less scalable and prone to errors. We also compare a semi-supervised learning approach [31] on partially labeled radio data obtained inside the network and from 3GPP signal measurements. Due to the heavy energy consumption of the signal collection, it is not suitable for IoT devices, which require low power support. Another limitation of this model is that data collection occurred every 15 seconds, assuming that a user cannot change their environment twice within a 30-second timeframe. This assumption diminishes its suitability for real-time implementation and timely detection. Similarly, the latest work, MB-SVM-HMM approach [23], while achieving impressive accuracy (94.22%) with a 9-second delay, makes it less suitable for practical implementation. This highlights the need for alternative methods that balance accuracy with resource efficiency for broader deployment. On the other hand, our proposed method using contrastive learning with attention has achieved similar accuracy regardless of the temporal length, leading to lower power consumption, faster processing, and improved real-time performance, making it more suitable for deployment on smartphones and similar devices. We evaluated each model using a diverse set of metrics to assess their strengths and weaknesses. Accuracy, F1-score, precision and recall measure the overall classification performance while considering potential imbalances in the data. Delay statistics were analyzed to understand how each model behaves during environmental changes. Computational efficiency factors like parameter counts, memory requirements, prediction time, and training time were evaluated to assess resource consumption and practicality and compare them with related works. Our self-supervised approach outperformed other works in several key aspects like accuracy, F1-score, precision and recall. Our approach achieved the highest accuracy, F1-score, precision and recall, demonstrating its effectiveness in correctly classifying indoor and outdoor environments. In addition, our model boasts significantly faster single prediction times compared to other methods, making it suitable for real-time applications where quick response is crucial. Finally, our approach requires less memory compared to several other methods, potentially reducing hardware demands and improving resource efficiency while Random Forest and Multi-Layer Perceptron exhibited faster training times, but at the cost of accuracy, they necessitate a substantial and diverse volume of training data to achieve satisfactory performance [31].

Our approach offers a compelling balance between training speed and overall performance. Our model achieved the highest mean accuracy (92.61%) independent of temporal lengths, surpassing previous work MB-SVM-HMM [23]. This demonstrates its consistent performance regardless of the data window size. Compared to DenseNet-LSTM [24] and CAP-ALSTM [26], our approach leverages a less complex architecture for high-level feature extraction, potentially reducing computational overhead and being advantageous in scenarios where resource constraints are a concern. Our self-supervised learning approach eliminates the need for large labeled datasets, requiring only a small number of labels for fine-tuning. This makes it more data-efficient compared to traditional supervised learning methods. While most IOD models achieve high accuracy on static datasets, our approach excels in recognizing environment transitions regardless of the temporal length of the data window. This makes it particularly well-suited for applications where frequent environmental changes occur. Overall, our self-supervised approach demonstrates superior performance and efficiency in indoor-outdoor detection, particularly for capturing the transition relationship between temporal environment states.

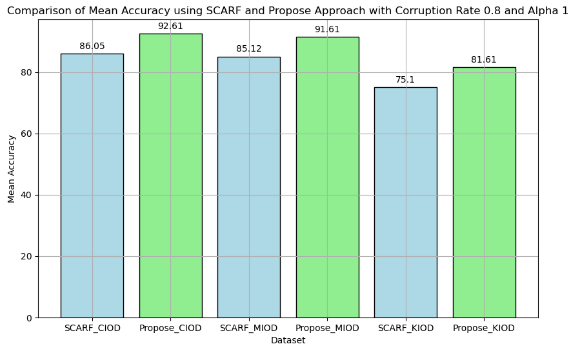

FIGURE 8. Comparison of mean accuracy using SCARF and propose approach with corruption rate 0.8 and alpha 1.

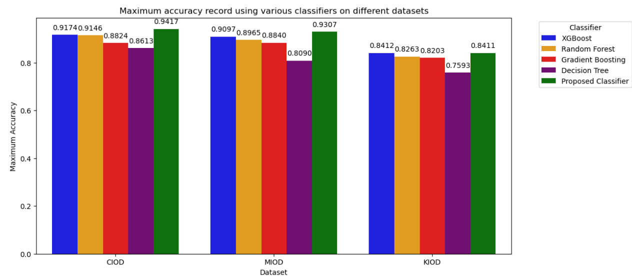

FIGURE 9. Comparison of maximum accuracy record using various classifiers on different datasets.

This makes it a promising solution for real-world applications requiring accurate and efficient indoor-outdoor detection and also removes the need for large dataset labeling, which is a time-consuming and uneconomical solution.

E. COMPREHENSIVE EVALUATION

We also conduct a comprehensive study to evaluate the effectiveness of TabCLR's proposed components, such as the spatial-context contrastive loss and the novel encoder model, in improving indoor-outdoor classification performance. By selectively modifying these components and analyzing the resulting changes, the study provides insights into the critical factors influencing TabCLR's effectiveness.

1) COMPARISON WITH BASELINE SSL PAPER ON IOD DATASETS

In our experimental setup, TabCLR, being the first SSL framework for indoor outdoor detection classification, we compare TabCLR with SCARF [33], a popular tabular SSL method. We demonstrate the superiority of our proposed approach against SCARF. We implemented comprehensive experimentation on three distinct indoor-outdoor datasets: CIOD, MIOD, and KIOD, each subjected to varying corruption rates and alpha parameters repeated 100 times. The outcomes of these experiments are illustrated through in Figure 7, while the corresponding mean accuracies with a corruption rate of 0.8 and α of 1 are presented in Figure 8. Notably, our approach consistently outperforms all

TABLE 6. Comprehensive evaluation with different performance metrics of indoor outdoor detection for various datasets (seen and unseen datasets).

Dataset	Accuracy		F1-Score		Precision		Recall	
	SL	PL	SL	PL	SL	PL	SL	PL
Evaluation on Seen Dataset using SCARF [33] Approach								
CIOD	84.67 ± 2.50	86.11 ± 2.91	84.94 ± 2.41	86.20 ± 2.13	84.35 ± 2.54	86.69 ± 2.66	84.40 ± 2.18	86.27 ± 2.02
MIOD	83.55 ± 2.47	84.86 ± 2.89	83.71 ± 2.37	85.08 ± 2.10	83.09 ± 2.51	85.46 ± 2.62	83.29 ± 2.16	84.05 ± 2.99
KIOD	74.60 ± 2.20	75.14 ± 2.67	74.58 ± 2.11	75.44 ± 2.86	74.27 ± 2.22	75.52 ± 2.34	74.31 ± 2.92	75.54 ± 2.77
Evaluation on Unseen Indoor Dataset DIOD using SCARF [33] Approach trained on MIOD								
Indoor Corridor	70.17 ± 2.14	71.94 ± 2.62	70.65 ± 2.99	71.73 ± 2.76	70.81 ± 2.93	71.94 ± 2.76	70.19 ± 2.84	71.53 ± 2.66
Indoor Hallway	70.23 ± 2.99	71.16 ± 2.58	70.99 ± 2.81	71.49 ± 2.55	70.71 ± 2.67	71.89 ± 2.64	70.46 ± 2.68	71.39 ± 2.55
Indoor Stairs	71.04 ± 2.75	72.19 ± 2.57	71.04 ± 2.58	72.32 ± 2.71	71.04 ± 2.53	72.23 ± 2.71	71.04 ± 2.55	72.41 ± 2.60
Evaluation on Unseen Outdoor Dataset DIOD using SCARF [33] Approach trained on MIOD								
Outdoor Campus	71.86 ± 2.31	72.95 ± 2.14	71.86 ± 2.22	72.95 ± 2.17	71.86 ± 2.29	72.95 ± 2.19	71.86 ± 2.27	72.95 ± 2.18
Parking lot	71.71 ± 2.35	72.37 ± 2.18	71.71 ± 2.26	72.37 ± 2.19	71.71 ± 2.33	72.37 ± 2.24	71.71 ± 2.31	72.37 ± 2.22
Outdoor Roads	71.95 ± 2.40	72.62 ± 2.25	71.95 ± 2.33	72.62 ± 2.27	71.95 ± 2.36	72.62 ± 2.32	71.95 ± 2.38	72.62 ± 2.29
Evaluation on Seen Dataset using Proposed Approach								
CIOD	91.45 ± 2.68	**92.61 ± 2.17**	91.89 ± 2.59	**93.03 ± 2.28**	92.11 ± 2.71	**93.51 ± 2.85**	92.08 ± 2.33	**93.14 ± 2.17**
MIOD	90.15 ± 2.82	**91.47 ± 2.38**	90.19 ± 2.35	**92.18 ± 2.53**	91.24 ± 2.63	**92.13 ± 2.59**	91.12 ± 2.27	**92.27 ± 2.31**
KIOD	80.31 ± 2.79	**81.71 ± 2.74**	80.49 ± 2.51	**82.18 ± 2.48**	81.35 ± 2.74	**82.25 ± 2.83**	81.37 ± 2.62	**82.43 ± 2.58**
Evaluation on Unseen Indoor Dataset DIOD using Proposed Approach trained on MIOD								
Indoor Corridor	80.43 ± 2.08	**82.02 ± 2.15**	81.52 ± 1.95	**82.13 ± 1.80**	81.61 ± 1.91	**82.27 ± 1.94**	81.75 ± 1.98	**82.34 ± 1.86**
Indoor Hallway	80.22 ± 2.34	**81.75 ± 1.88**	81.36 ± 2.12	**81.58 ± 1.72**	81.47 ± 2.01	**81.69 ± 1.86**	81.63 ± 1.94	**81.89 ± 1.88**
Indoor Stairs	81.79 ± 2.62	**83.04 ± 1.76**	82.01 ± 2.18	**83.20 ± 1.90**	81.82 ± 2.06	**83.38 ± 1.97**	81.99 ± 1.99	**83.16 ± 1.85**
Evaluation on Unseen Outdoor Dataset DIOD using Proposed Approach trained on MIOD								
Outdoor Campus	80.07 ± 2.46	**81.51 ± 2.21**	80.16 ± 2.18	**81.62 ± 2.28**	80.22 ± 2.32	**81.72 ± 2.25**	80.29 ± 2.36	**81.83 ± 2.16**
Parking lot	80.13 ± 2.29	**82.65 ± 2.14**	81.34 ± 2.06	**82.71 ± 2.17**	80.42 ± 2.24	**82.82 ± 2.10**	81.55 ± 2.29	**82.92 ± 2.13**
Outdoor Roads	80.06 ± 2.13	**81.90 ± 2.17**	81.12 ± 2.10	**81.94 ± 2.04**	80.28 ± 2.05	**81.77 ± 2.08**	80.36 ± 2.09	**81.83 ± 2.07**

SL = SimCLR Loss Function [32], *PL* = Proposed Loss Function

three datasets, exhibiting an improvement ranging between 6.0% to 7.0%. We also evaluated different machine learning classifiers like XGBoost, Random Forest, Gradient Boosting, Decision Tree etc and compared them with a very lightweight proposed DNN-based classifier, as shown in Figure 9. It can be observed if we increase the complexity of ML, it can achieve the same level of accuracy at the cost of higher model parameters, but our simple proposed lightweight classifier achieves the same accuracy but in very few model parameters. SCARF [33] incorporate XGBoost classifier, but our classifier outperforms on two datasets(CIOD and MIOD) while almost the same on KIOD but in very few model parameters. Therefore, this significant performance enhancement underscores the efficacy of our proposed methodology across diverse datasets, affirming its robustness and superiority in indoor-outdoor detection scenarios.

2) COMPARISON ON VARIOUS DATASETS (SEEN AND UNSEEN DATASETS)

The results presented in the Table 6 demonstrate the performance of both the SCARF approach and a proposed algorithm using the simCLR loss function [32] and the proposed loss function in various datasets for indoor-outdoor detection. We evaluated the robustness of our approach using both seen and unseen datasets. The proposed algorithm consistently outperforms SCARF [33] in terms of accuracy, F1-score, precision, and recall as shown in Table 6. For example, on the CIOD dataset, the proposed algorithm achieves an accuracy of 91.45% compared to SCARF's 84.67%, indicating a substantial improvement. This trend persists in other seen datasets, highlighting the effectiveness of the proposed algorithm in accurately detecting indoor-outdoor scenes. Furthermore, when considering unseen datasets, the superiority of the proposed algorithm becomes even more apparent. In both indoor and outdoor unseen datasets, the proposed algorithm consistently achieves higher accuracy, F1-score, precision, and recall compared to SCARF. For example, on the Indoor Corridor dataset, the proposed algorithm achieves a mean accuracy of 80.43±2.08% and 82.02±2.15% using simCLR loss function [32] and the proposed loss function compared to SCARF's 70.17±2.14% and 71.94±2.62%, indicating significant improvements. These results underscore the efficacy of the proposed algorithm in indoor-outdoor detection tasks. The superior performance of the proposed algorithm is attributed to its temporal representation in self-supervised learning using the self-attention approach, which incorporates advanced feature extraction and using the proposed contrastive loss function tailed for indoor-outdoor detection. Furthermore, the proposed algorithm demonstrates superior performance with more extensive datasets with various sensors such as the CIOD public dataset [26] along with the three proposed datasets (two seen and unseen) employing IMU Sensors, allowing it to learn richer representations of indoor outdoor detection. Overall, these results highlight our algorithm's robustness and generalizability to new and unseen environments, making it a compelling choice for real-world applications. We rigorously validate the performance of our algorithm through cross-validation techniques and tests on diverse datasets that represent a wide range of indoor

TABLE 7. The average relative gain in accuracy achieved by adding pre-training methods to reference methods across different variations in the indoor-outdoor detection setting.

Variations	Relative Gain in Accuracy					
	SCARF	SCARFAE	nonoiseAE	add.noiseAE	SCARFdisc	Proposed Approach
100% labeled training						
control	2.352	2.244	1.107	1.559	0.574	**8.789**
dropout	1.609	1.196	0.623	1.228	-1.312	**7.285**
mixup	1.720	1.183	-0.377	0.971	-0.307	**7.469**
labelsmooth	1.522	0.711	-0.002	1.040	-0.894	**7.234**
distill	2.392	2.186	0.823	1.431	-0.394	**8.981**
25% labeled training						
dropout	2.212	1.848	2.013	1.155	-0.322	**8.197**
mixup	2.809	0.730	0.106	0.439	0.466	**8.001**
labelsmooth	2.303	0.705	-0.564	0.196	-0.206	**8.876**
distill	3.609	2.441	1.969	2.263	1.795	**9.378**
self-train	3.839	2.753	1.672	2.839	2.559	**9.463**
tri-train	3.549	2.706	1.455	2.526	1.920	**9.158**
30% label noise						
control	2.261	1.988	0.914	1.612	-1.408	**8.554**
dropout	2.004	2.058	0.900	1.471	-2.540	**8.632**
mixup	2.739	1.723	0.116	1.409	0.189	**8.374**
labelsmooth	2.558	1.474	0.703	1.395	-1.337	**8.444**
distill	2.881	2.296	-0.239	1.659	-0.226	**8.703**
deepknn	2.001	1.281	0.814	1.348	0.088	**8.822**
bitempered	2.680	2.915	0.435	1.387	-1.147	**8.987**

scenarios along with outdoor scenarios. By evaluating the accuracy and robustness of the algorithm of the proposed approach in different buildings and sensor combinations, we demonstrated and ensured reliable performance in real-world applications.

The Table 7 presents results obtained from fully labeled training data, 25% labeled training data, and full training data subjected to 30% label noise, evaluating the relative gain in accuracy when integrating pre-training methods with reference methods. The rows represent different reference methods, while the columns depict various pre-training methods. The key findings from the table indicate that both the SCARF approach and the proposed approach consistently outperform alternative methods across different scenarios. Notably, they excel not only in enhancing control methods but also in improving methods specifically tailored for the indoor-outdoor detection (IOD) setting. In the case of fully labeled training data, the proposed approach demonstrates substantial relative gains in accuracy across all considered variations compared to SCARF and other methods. Particularly, it achieves remarkable improvements in accuracy, indicating its effectiveness in leveraging pre-training techniques for enhancing indoor-outdoor detection performance. When only 25% of the training data is labeled, the proposed approach continues to outperform alternative methods, showcasing its robustness and adaptability in scenarios with limited labeled data. Despite the reduced training data, the proposed approach maintains significant gains in accuracy, highlighting its potential for efficient utilization of limited resources. Moreover, in scenarios with 30% label noise, the proposed approach again exhibits superior performance, indicating its resilience to label noise and its ability to learn robust representations for indoor-outdoor detection tasks. Overall, the results underscore the effectiveness of both the SCARF approach and the proposed approach in improving indoor-outdoor detection accuracy, even in challenging conditions such as limited labeled data and label noise. These findings highlight the promising potential of the proposed approach for advancing indoor-outdoor detection systems and addressing real-world challenges in diverse application scenarios.

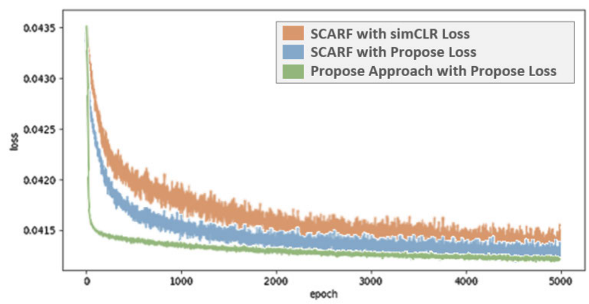

FIGURE 10. Comparison of loss function with proposed loss function and propose approach.

3) COMPARISON OF CONTRASTIVE LOSS FUNCTION

In our investigation, we examined the significance of our proposed loss function, the spatial-context contrastive loss (SCCL), and conducted a comparative analysis of its effectiveness against the well-established simCLR loss function [32]. SCCL, optimized for indoor-outdoor detection employing inertial sensors, is devised to address the limitations of simCLR in capturing spatial dynamics. Our

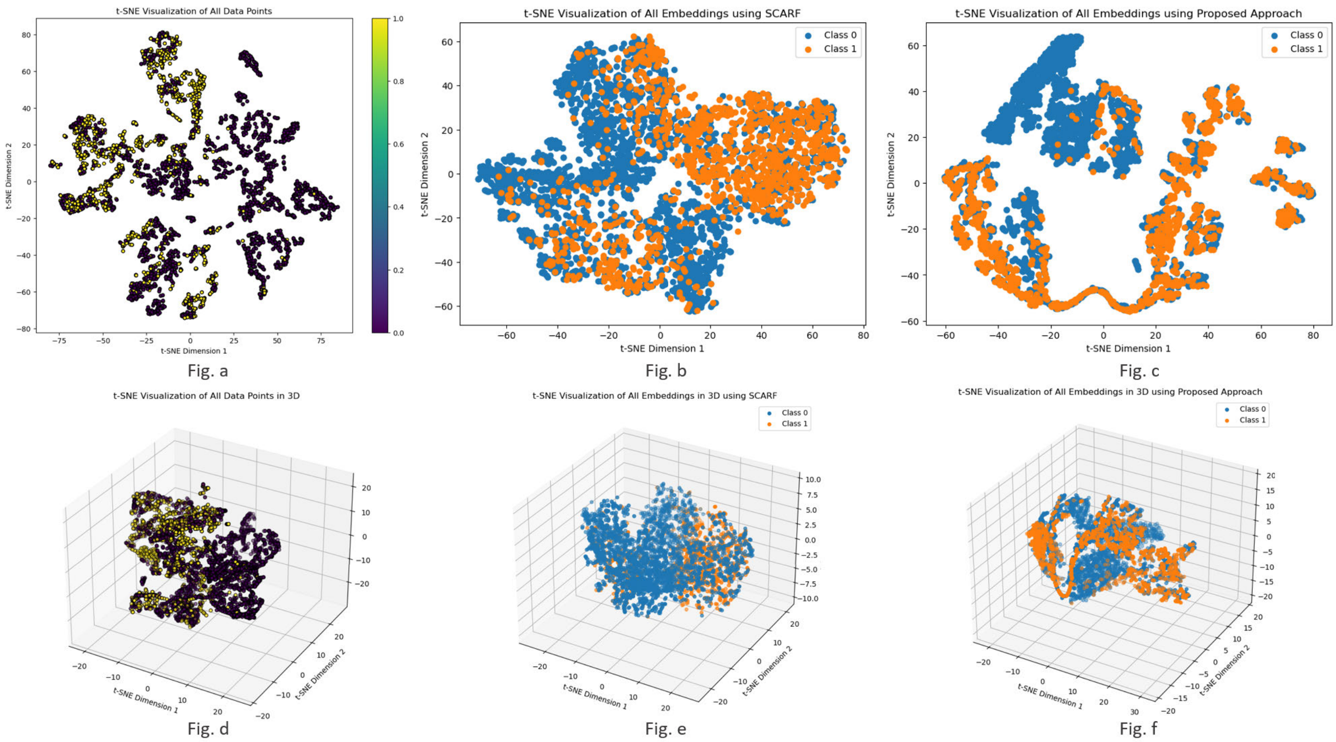

FIGURE 11. Comparison of TSNE visualization using scarf and propose approach.

comprehensive analysis demonstrated that SCCL outperforms simCLR in our experimental setup, showcasing its superiority in modeling spatial context and augmenting sensitivity to context-specific spatial patterns, as illustrated in Figure 10. Our proposed loss function, SCCL, along with our methodological approach, aims to tackle several observed constraints in SCARF [33] when employing the simCLR loss function. Particularly, SCCL mitigates the tendency of the loss function to exhibit rapid descent at the onset of training, followed by a gradual decline until reaching the early stopping point. Furthermore, SCCL strives to alleviate the noisy and high variance nature inherent in the training loss arising from the stochastic nature of our corruption method, as depicted in Figure 10 with orange color. By integrating spatial context modeling and introducing a spatial consistency term, SCCL enhances the resilience and stability of the training process, thereby improving the performance of SCARF with the simCLR loss function, as demonstrated in Figure 10 with blue color. Our approach, coupled with the SCCL loss function, yields optimized pre-training and validation loss curves, as depicted in Figure 10 with green color.

F. t-SNE REPRESENTATION ANALYSIS

To illustrate the representation in the proposed method, we employ t-SNE (t-Distributed Stochastic Neighbor Embedding) [34] to reduce the dimensionality of the extracted high-dimensional features from the model for visualization purposes. We visualized samples from the indoor-outdoor dataset. Subsequently, we visualize the features extracted from the same signals by both the pre-trained models of SCARF and the proposed approach. Figure 11 illustrates the results of dimensionality reduction for the original samples and the extracted features of both approaches. Each color corresponds to a distinct category of indoor and outdoor classes. Distinguishing the original time-domain signals is challenging. However, following pretraining, the majority of the features from the samples are grouped by category. Notably, the pretraining phase does not include any category information, indicating the model's capability to acquire proficient representation learning through self-supervised pretraining. The bottom subplots of Figure 11 demonstrates clearly the clustering of features in 3D visualization. In summary, the proposed method exhibits a progressive development of feature extraction capabilities compared to the baseline SCARF approach [33].

VI. CONCLUSION

This research explores the application of self-supervised learning (SSL) to indoor-outdoor detection (IOD), introducing TabCLR, the first SSL framework explicitly designed for IOD leveraging smartphone inertial sensor data. Our approach utilizes a novel encoder with self-attention, contrastive learning representation for tabular data, and a modified contrastive loss function. Extensive evaluation of benchmark and proposed indoor-outdoor inertial datasets demonstrates the mean classification accuracy of 92.61%, outperforming S.O.T.A. in indoor-outdoor detection. The TabCLR shows outstanding performance evaluation in terms of a single prediction time of $177\mu s$, memory usage of $0.13MB$, and total parameters of $32K$. The fast execution time with the low memory demands of TabCLR makes it an ideal neural network model for resource-constrained IoT-enabled IOD systems and smartphones. The practical significance of TabCLR extends across diverse applications, including location-based services, smart home automation, health and fitness tracking, environmental monitoring, security and surveillance, retail analytics, and emergency response. By accurately distinguishing between indoor and outdoor environments, TabCLR enables enhanced functionality and contextual awareness in various real-world scenarios. Overall, the efficient execution and low resource requirements of TabCLR position it as a promising solution for advancing indoor-outdoor detection capabilities in IoT-enabled systems and smartphone applications, paving the way for innovative and reliable context-aware services in the digital era.

REFERENCES

[1] S. Li, Z. Qin, H. Song, C. Si, B. Sun, X. Yang, and R. Zhang, "A lightweight and aggregated system for indoor/outdoor detection using smart devices," *Future Gener. Comput. Syst.*, vol. 107, pp. 988–997, Jun. 2020.

[2] KAIST News. *An AI-Based, Indoor/Outdoor-Integrated (IOI) GPS System To Bring Seismic Waves in the Terrains of Positioning Technology*. Accessed: Jul. 22, 2022. [Online]. Available: https://news.kaist.ac.kr/newsen/html/news/?mode=V&mng_no=22070&skey=category&sval=research&list_s_date=&list_e_date=&GotoPage=1

[3] Qualcomm. *First Mobile Platform Designed With Generative AI in Mind*. Accessed: Jul. 22, 2022. [Online]. Available: https://www.qualcomm.com/products/mobile/snapdragon/smartphones/mobile-ai

[4] W. Shao, F. Zhao, C. Wang, H. Luo, T. M. Zahid, Q. Wang, and D. Li, "Location fingerprint extraction for magnetic field magnitude based indoor positioning," *J. Sensors*, vol. 2016, no. 1, pp. 1–16, 2016.

[5] F. Zhao, H. Luo, X. Zhao, Z. Pang, and H. Park, "HYFI: Hybrid floor identification based on wireless fingerprinting and barometric pressure," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 330–341, Feb. 2017.

[6] H. Luo, F. Zhao, M. Jiang, H. Ma, and Y. Zhang, "Constructing an indoor floor plan using crowdsourcing based on magnetic fingerprinting," *Sensors*, vol. 17, no. 11, p. 2678, Nov. 2017.

[7] Q. Wang, H. Luo, A. Men, F. Zhao, and Y. Huang, "An infrastructure-free indoor localization algorithm for smartphones," *Sensors*, vol. 18, no. 10, p. 3317, Oct. 2018.

[8] Q. Wang, H. Luo, A. Men, F. Zhao, X. Gao, J. Wei, Y. Zhang, and Y. Huang, "Light positioning: A high-accuracy visible light indoor positioning system based on attitude identification and propagation model," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 2, Feb. 2018, Art. no. 155014771875826.

[9] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 235–249, Jun. 2014.

[10] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.

[11] A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2406–2417, Oct. 2015.

[12] D. Shin, D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, and G. Schmitt, "Urban sensing: Using smartphones for transportation mode classification," *Comput., Environ. Urban Syst.*, vol. 53, pp. 76–86, Sep. 2015.

[13] Y. Qin, H. Luo, F. Zhao, Z. Zhao, and M. Jiang, "A traffic pattern detection algorithm based on multimodal sensing," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 10, Oct. 2018, Art. no. 155014771880783.

[14] M. I. Gilmour, M. S. Jaakkola, S. J. London, A. E. Nel, and C. A. Rogers, "How exposure to environmental tobacco smoke, outdoor air pollutants, and increased pollen burdens influences the incidence of asthma," *Environ. Health Perspect.*, vol. 114, no. 4, pp. 627–633, Apr. 2006.

[15] J. Cheng, L. Yang, Y. Li, and W. Zhang, "Seamless outdoor/indoor navigation with WiFi/GPS aided low cost inertial navigation system," *Phys. Commun.*, vol. 13, pp. 31–43, Dec. 2014.

[16] M. Jia, Y. Yang, L. Kuang, W. Xu, T. Chu, and H. Song, "An indoor and outdoor seamless positioning system based on Android platform," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Aug. 2016, pp. 1114–1120.

[17] Y. Zhu, H. Luo, Q. Wang, F. Zhao, B. Ning, Q. Ke, and C. Zhang, "A fast indoor/outdoor transition detection algorithm based on machine learning," *Sensors*, vol. 19, no. 4, p. 786, 2019.

[18] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications*. New York, NY, USA: Springer, 2001.

[19] T.-H. Yi, H.-N. Li, and M. Gu, "Effect of different construction materials on propagation of GPS monitoring signals," *Measurement*, vol. 45, no. 5, pp. 1126–1139, Jun. 2012.

[20] P. Zhou, Y. Zheng, Z. Li, M. Li, and G. Shen, "IODetector: A generic service for indoor outdoor detection," in *Proc. 10th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2012, pp. 113–126.

[21] M. Ali, T. ElBatt, and M. Youssef, "SenseIO: Realistic ubiquitous indoor outdoor detection system using smartphones," *IEEE Sensors J.*, vol. 18, no. 9, pp. 3684–3693, May 2018.

[22] I. Saffar, M. L. A. Morel, K. D. Singh, and C. Viho, "Machine learning with partially labeled data for indoor outdoor detection," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–8.

[23] F. Zhu, W. Chen, F. Guo, and X. Zhang, "Combining context connectivity and behavior association to develop an indoor/outdoor context detection model with smartphone multisensor fusion," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 2883–2898, Jan. 2024.

[24] Y. Zhu, H. Luo, F. Zhao, and R. Chen, "Indoor/outdoor switching detection using multisensor DenseNet and LSTM," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1544–1556, Feb. 2021.

[25] S. A. Hamideche, M. L. A. Morel, K. Singh, and C. Viho, "Indoor-outdoor detection using time series classification and user behavioral cognition," in *Proc. 14th IFIP Wireless Mobile Netw. Conf. (WMNC)*, Sousse, Tunisia, Oct. 2022, pp. 7–14.

[26] S. Bakirtzis, K. Qiu, I. Wassell, M. Fiore, and J. Zhang, "Deep-learning-based multivariate time-series classification for indoor/outdoor detection," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24529–24540, Dec. 2022.

[27] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Müller, "Deep learning for time series classification: A review," 2018, *arXiv:1809.04356*.

[28] N. Mohammadi Foumani, L. Miller, C. Wei Tan, G. I. Webb, G. Forestier, and M. Salehi, "Deep learning for time series classification and extrinsic regression: A current survey," 2023, *arXiv:2302.02515*.

[29] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 1, p. 2, Dec. 2020.

[30] V. Radu, P. Katsikouli, R. Sarkar, and M. K. Marina, "A semi-supervised learning approach for robust indoor-outdoor detection with smartphones," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2014, pp. 280–294.

[31] I. Saffar, M. L. A. Morel, K. D. Singh, and C. Viho, "Semi-supervised deep learning-based methods for indoor outdoor detection," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[32] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020, *arXiv:2002.05709*.

[33] D. Bahri, H. Jiang, Y. Tay, and D. Metzler, "SCARF: Self-supervised contrastive learning using random feature corruption," 2021, *arXiv:2106.15147*.

[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[35] W. Wang, Q. Chang, Q. Li, Z. Shi, and W. Chen, "Indoor-outdoor detection using a smart phone sensor," *Sensors*, vol. 16, no. 10, p. 1563, Sep. 2016.

[36] L. Wang, L. Sommer, T. Riedel, M. Beigl, Y. Zhou, and Y. Huang, "NeuralIO: Indoor outdoor detection via multimodal sensor data fusion on smartphones," in *Proc. Int. Summit Smart City 360°*, 2019, pp. 127–138.

MUHAMMAD BILAL AKRAM DASTAGIR received the master's degree from the University of Leicester, U.K. He is currently pursuing the Doctor of Philosophy (Ph.D.) degree with the School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He was a Senior Design Engineer with the National Space Agency, SUPARCO, for seven years. His research interests include the Internet of Things, deep neural networks (DNN), self-supervised learning, multi-variant time series data classification, indoor-outdoor detection, floor change detection, and deep learning for indoor localization solutions.

OMER TARIQ received the B.S. degree in electrical engineering from UET, Pakistan, in 2014. He is currently pursuing the Ph.D. degree with the School of Computing, Korea Advanced Institute of Science and Technology. He was a Team Lead of FPGA design and verification with the National Space Agency and the Centres of Excellence in Science and Applied Technologies (CESAT), Pakistan, for seven years. His research interests include the Artificial Intelligence of Things (AIoT) and multimodal sensor fusion algorithms for localization, mapping, navigation, and perception.

DONGSOO HAN (Senior Member, IEEE) received the Ph.D. degree in information science from Kyoto University. He is currently a Professor of computer science with Korea Advanced Institute of Science and Technology (KAIST), where he is also the Director of the Indoor Positioning Research Center. His research interests include indoor positioning, pervasive computing, and location-based mobile applications.