Computers in Biology and Medicine 172 (2024) 108232

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Efficiently improving the Wi-Fi-based human activity recognition, using auditory features, autoencoders, and fine-tuning

Amir Rahdar [a], Mahnaz Chahoushi [a], Seyed Ali Ghorashi [b,*]

[a] AIFA, Dotin, Tehran, 1915718181, Iran
[b] Department of Computer Science & Digital Technologies, School of Architecture, Computing, and Engineering, University of East London, London, E16 2RD, UK

ARTICLE INFO

Keywords:
Autoencoder
Channel state information
Deep learning
Fine-tuning
Human activity recognition
Machine learning
Mel frequency cepstral coefficient

ABSTRACT

Human activity recognition (HAR) based on Wi-Fi signals has attracted significant attention due to its convenience and the availability of infrastructures and sensors. Channel State Information (CSI) measures how Wi-Fi signals propagate through the environment. However, many scenarios and applications have insufficient training data due to constraints such as cost, time, or resources. This poses a challenge for achieving high accuracy levels with machine learning techniques. In this study, multiple deep learning models for HAR were employed to achieve acceptable accuracy levels with much less training data than other methods. A pretrained encoder trained from a Multi-Input Multi-Output Autoencoder (MIMO AE) on Mel Frequency Cepstral Coefficients (MFCC) from a small subset of data samples was used for feature extraction. Then, fine-tuning was applied by adding the encoder as a fixed layer in the classifier, which was trained on a small fraction of the remaining data. The evaluation results (K-fold cross-validation and $K = 5$) showed that using only 30% of the training and validation data (equivalent to 24% of the total data), the accuracy was improved by 17.7% compared to the case where the encoder was not used (with an accuracy of 79.3% for the designed classifier, and an accuracy of 90.3% for the classifier with the fixed encoder). While by considering more calculational cost, achieving higher accuracy using the pretrained encoder as a trainable layer is possible (up to 2.4% improvement), this small gap demonstrated the effectiveness and efficiency of the proposed method for HAR using Wi-Fi signals.

1. Introduction

Human Activity Recognition (HAR) assumes a prominent role within the realm of research, propelling progress in Smart Homes and the Internet of Healthcare Things (IoHT). Its applications reverberate significantly within the healthcare sphere, encompassing functions such as Age and Gender Estimation, Monitoring of the Elderly and Individuals with Disabilities, and addressing ailments such as Alzheimer's, which necessitate ongoing care [1], as well as the identification of heart diseases [2]. HAR techniques are inherently classified into vision-based, sensor-based, and Wi-Fi-based paradigms. Vision-based techniques, despite their commendable accuracy, encounter constraints arising from environmental dependencies, encompassing factors like lighting conditions and background settings. A corollary concern pertains to the privacy encroachments linked to Vision-Based techniques [3]. Furthermore, while sensor-based HAR, exemplified by the utilization of deep recurrent neural networks and electroencephalogram EEG signals [4], showcases promising outcomes, it is reliant on sophisticated hardware for data recording. Conversely, sensor-based HAR involving more accessible devices such as smartphones emerges as a practical solution in real-world contexts, especially for data collection [5]. These endeavors have led to the conception of applications and frameworks where models are trained employing data derived from smartphones and even smartwatches [6]. However, the implementation of wearable sensors introduces a sense of discomfort and interruption during activity engagement, along with associated privacy trade-offs [7]. Within the domain of Wi-Fi signal attributes, Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) stand as pivotal markers. Owing to its user-friendly nature and reasonable precision, RSSI has found extensive application in HAR [8]. Nevertheless, CSI outperforms RSSI in terms of performance [9–11], thus meriting its selection as the preferred attribute in this study. CSI notably exhibits enhanced resilience in the face of obstacles and fluctuations in the distances between transmitters and receivers [12].

The domains of deep learning and machine learning have exerted a profound influence within the context of HAR, primarily focusing on the

* Corresponding author.
E-mail addresses: amir.rahdar@gmail.com (A. Rahdar), mahnazchahoushi@yahoo.com (M. Chahoushi), s.a.ghorashi@uel.ac.uk (S.A. Ghorashi).

https://doi.org/10.1016/j.compbiomed.2024.108232
Received 29 August 2023; Received in revised form 19 February 2024; Accepted 25 February 2024
Available online 27 February 2024

classification and prediction of activities. Notably, Convolutional Neural Networks (CNNs) have emerged as a pivotal catalyst in this arena [15]. A significant study [16] delves into the realm of deep learning methodologies, encompassing CNN, Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and attention techniques, all of which hold relevance for CSI-based HAR. These strategies aptly handle the intricacies of high-dimensional and time-series data. Pertaining to the efficacy of attention-based methodologies, recent investigations [17,18] underline the potential of the Convolution Block Attention Module (CBAM) in augmenting HAR models, encompassing both visual and time-series data. Furthermore, within [19], the proposition of a multi-resolution fusion convolution network (MRFC-Net) emerges as a means to enhance activity recognition accuracy. This study takes into consideration the challenge of HAR models encountering real-world data from previously unknown classes. To address this concern, the exploration of variational autoencoders, as potent tools for data augmentation, comes to the forefront.

Focusing on the surveillance of distinct groups, such as prisoners and the elderly, this study introduces a multiresolution fusion model rooted in convolutional neural networks. This innovative model facilitates the recognition of samples from unknown classes without imposing the constraint of prior classification. Meanwhile, in Ref. [20], a resource-efficient approach involving residual convolutional networks and a recurrent neural network (RCNN-BiGRU) surfaces, accompanied by an optimal feature set selection mechanism grounded in the Marine Predator Algorithm (MPA). This methodology showcases commendable performance, albeit with a relatively higher computational cost. Notably, the domain of data preprocessing stands as a cornerstone in the efficacy of classifiers. As a result, diverse preprocessing techniques have been meticulously scrutinized [21]. For instance, within [22], the conventional Mel Frequency Cepstral Coefficient (MFCC), commonly associated with audio signals, has been judiciously adapted to preprocess CSI time-series data. This adaptation, combined with the employment of Principal Component Analysis (PCA) and the transformation of data into one-dimensional time series, positions MFCC as a potent catalyst for feature extraction, resulting in exceptional classification capabilities. The preprocessing repertoire extends to encompass techniques such as Discrete Wavelet Transform (DWT) and Short-Time Fourier Transform (STFT). Subsequently, three distinct classifiers - CNN, LSTM, and Hidden Markov Model (HMM) - are harnessed for feature classification. The fusion of MFCC and CNN emerges as particularly promising [22]. Noteworthy is the approach adopted within [22], treating CSI streams akin to sound time series and leveraging MFCC for the extraction of auditory features. This not only streamlines training procedures but also optimizes computational resources. The choice of MFCC finds its rationale in the shared spectral attributes between the CSI time series and the human auditory spectrum. It is worth noting that while the utilization of MFCC underscores promising outcomes in HAR, alternative techniques such as wavelet-based feature extraction and preprocessing have also demonstrated excellent results in analogous tasks [23,24].

In various research endeavors, the incorporation of autoencoders has been explored to amplify the performance of classifiers. Autoencoders exhibit a range of variations, each tailored to specific objectives like denoising, feature extraction, generating synthetic data, and reducing dimensionality. The core principle underlying the deployment of autoencoders is to replicate input data at the output while preserving maximal resemblance. Comprising both an encoder and a decoder, once the encoder is trained, its integration into the classifier seamlessly enhances outcomes. Notably, in specific research studies, autoencoders are utilized to eliminate noise and enhance results [24].

Zou et al. [25] introduce a CSI-based technique named Autoencoder Long Term Recurrent Convolutional Network (AE-LRCN), which entails a convolutional neural network for feature extraction, a long short-term memory module to capture underlying temporal dependencies, and an autoencoder for noise removal. This approach yields high performance without necessitating specialized expertise and is characterized by efficient processing [25]. Guo et al. [26] propose an LSTM-based encoder and a CNN-based decoder to address the challenge of declining accuracy when the classifier is applied to different users. The method presented in Ref. [26] showcases robust classification performance, outperforming KNN, SVM, and RNN. A comprehensive study [27] encompasses an array of deep learning methodologies, encompassing autoencoders, RNN, LSTM, GRU, Multilayer Perceptron (MLP), and Random Forest (RF). Notably, autoencoders excel in the task of feature extraction. Amidst the backdrop of high-performing HAR techniques, the acquisition of ample data emerges as a challenge, often hindered by constraints or limited availability [28]. This constraint emanates from the labor-intensive process of data labeling, financial investments, and privacy considerations [29]. As a result, the pursuit of models capable of achieving commendable accuracy with fewer data instances becomes a pivotal endeavor. To address this challenge, diverse strategies have been proposed, including the utilization of Generative Adversarial Networks (GANs) and data augmentation to synthesize data instances. Nevertheless, these approaches encounter challenges in scenarios where data availability is limited, particularly in generating instances that accurately mimic real data [30]. Furthermore, certain studies advocate for the efficacy of more advanced autoencoder variants, such as variational autoencoders and adversarial autoencoders. These models exhibit proficiency in managing heterogeneous data from various devices, including diverse wearable sensors [31], and addressing data diversity stemming from different individuals [32].

Addressing the quandary of accessing adequate data, a recent study [33] introduced an innovative technique involving two autoencoders and two datasets. The first autoencoder underwent training on a substantial dataset, with its encoder subsequently applied to the second autoencoder. This facilitated the training of the second encoder using a smaller dataset. Furthermore, the output of the second autoencoder was labeled and employed to further train the model. Another study [13] tackled data scarcity by presenting a novel, based on Multi-Input Multi-Output Autoencoder (MIMO AE). Employing this autoencoder, three distinct methodologies were adopted to achieve optimal performance. The MIMO AE's encoder component extracted features based on their similarity, proving instrumental in the classifier's enhanced performance with fresh data. Alternatively, for computational efficiency and satisfactory outcomes, the encoder was integrated into the classifier without necessitating retraining. Additionally [13], investigated diverse proportions of training data to ascertain their influence. These endeavors collectively contribute to the resolution of data constraints in HAR, shedding light on techniques facilitating accurate activity recognition under restricted data scenarios. Furthermore, the combination of MIMO AE and MFCC for feature extraction in Automatic Speech Recognition (ASR) has been studied previously [34], further affirming the efficacy of this approach in optimizing feature extraction efficiency for speech recognition tasks.

In [35], a novel Spectro-Temporal network (STnet) was introduced to extract temporal patterns and micro-Doppler features from radar signals. Comprising a spectro-stream and a temporal stream, STnet's efficacy rests upon STFT for classification, yielding commendable efficiency. It holds the potential to address sensor-based activity recognition, including Wi-Fi channel state information. It is pertinent to acknowledge, however, that STnet bears limitations with respect to location dependence or variations in distances between the radar and the activity performer. This drawback might impact result accuracy in settings with diverse distances or locations. Nevertheless, STnet stands as a significant contribution, enriching the arena of temporal pattern and micro-Doppler feature extraction. The study at hand concentrates on achieving robust performance with limited data through the fusion of MFCC and Multi-Input Multi-Output Autoencoder for feature extraction. Building upon the foundations established in Ref. [13], the central goal is to augment the performance of the model by employing the encoder within the classifier sans retraining. This endeavor aims to approximate outcomes closely aligned with those derived from the model boasting a retrained encoder in the classifier. The impetus behind enhancing the untrained encoder model lies in its superior efficiency, particularly concerning computational overhead, vis-à-vis other techniques proposed in Ref. [13].

In the preprocessing phase, the adoption of MFCC for time-series data significantly enhances classification accuracy. The extraction of features from CSI time series using MFCC culminates in deep learning methods showcasing heightened accuracy in activity classification. Consequently, the amalgamation of these two methodologies holds the promise of optimal performance and efficiency. Following preprocessing, the extracted features undergo classification through four approaches in this study. These approaches encompass the unaltered classifier, a classifier featuring a pretrained and untrainable encoder, a classifier integrating a pretrained and retrained encoder, and a classifier leveraging the encoder devoid of prior training.

2. The proposed method

This section elucidates the dataset, the preprocessing methodology, and the technique employed for feature extraction and classification. The initial elucidation pertains to the preprocessing technique, followed by the subsequent explication of four distinct approaches applied for model training utilizing the preprocessed data.

A. Selected Dataset

The study's efficacy is demonstrated through the utilization of the dataset sourced from Ref. [14]. This dataset comprises Wi-Fi-based CSI obtained via the Nexmon tool on a Raspberry Pi 4 GB. Specifically, data collection was executed using a Raspberry Pi and a Tp-link Archer C20 as an Access Point (AP) operating at a 20 MHz bandwidth. The data collection setup involved a Personal Computer (PC) generating traffic, achieved through activities such as pinging or streaming videos.

To ensure data quality, the collected CSI data underwent a noise reduction filtering process. The data acquisition settings encompassed Core 1, NSS mask 1, a sampling size of 4000, and a duration of 20 s. The dataset encapsulates seven distinct activities, each performed 20 times by three volunteers within a residential environment. Consequently, a total of 420 samples were generated. The dataset, along with corresponding labeled files, is accessible through the GitHub repository (https://github.com/parisafm/CSI-HAR-Dataset, Accessed on August 10, 2023).

The activities were executed while positioned between the transmitter and receiver, thereby inducing phenomena such as reflections, multipath fading, and scattering [36]. These actions led to variations in the CSI and the multipath transmission of Wi-Fi signals. The CSI data, offering insights into both signal amplitude and phase, facilitated the identification of these alterations [37]. Among published studies in the field of human activity recognition, there are some outstanding and recent publications that used this dataset as their main or secondary source of data [38–40]. These studies focused on using attention-based networks, advanced filtering or advanced feature fusion as their main approach, in order to improve the results of HAR models. Therefore, alongside some familiar studies, final results of these publication will be used a comparison reference in this manuscript, assuring to present a fair comparison, concerning this research.

B Channel State Information

Orthogonal Frequency-Division Multiplexing (OFDM) technologies find application in telecommunication networks for coherent information transmission using Wi-Fi signals over channels connecting transmitters and receivers. In the context of OFDM modulation, messages are encoded onto orthogonal subcarriers and transmitted. This enables the concurrent transmission of multiple signals with overlapping spectral ranges through a singular channel. In essence, instead of utilizing a solitary wideband channel frequency, OFDM modulation divides a single information stream among numerous closely spaced narrow-band subchannel frequencies.

In scenarios where obstacles are present, phenomena like reflection, scattering, and multipath fading come into play [36]. Consequently, when an individual conducts an activity between the transmitter and receiver, modifications occur in the propagation of the Wi-Fi network's multipath. CSI offers insights into both signal amplitude and phase, enabling the detection of changes in Wi-Fi signals. These changes encompass signal scattering, environmental attenuation, multipath fading, shadow fading, and power attenuation due to propagation distance in each transmission path [37]. CSI boasts numerous advantages over RSSI in this context, including heightened resilience, diminished susceptibility to environmental influences, and augmented information transmission [41].

Moreover, OFDM technology can be harnessed in Wi-Fi devices, wherein the IEEE 802.11 n/ac standard facilitates the division of bandwidth among orthogonal subcarriers. Concurrently, the utilization of Multiple Input Multiple Output (MIMO) antennas for both transmitters and receivers in Wi-Fi devices can amplify multiplexing benefits and reduce channel interference. The CSI data can be represented as a channel matrix:

$$\mathrm{CSI} = \begin{pmatrix} H_{1,1} & \dots & H_{1,r} \\ \vdots & \ddots & \vdots \\ H_{t,1} & \cdots & H_{t,r} \end{pmatrix} \tag{1}$$

where t is the number of transmitters, r is the number of receivers, and $H_{t,r}$ represents a vector that includes complex pairs of subcarriers. H can also be demonstrated as:

$$H_{t,r} = \left[h_{t,r,1}, \cdots, h_{t,r,k}\right] \tag{2}$$

where k represents the number of data subcarriers, and h is a complex number that incorporates the phase and amplitude of CSI. Therefore, each subcarrier can be expressed as:

$$h^{i}_{t,r} = A^{i}_{t,r} e^{j\theta^{i}_{t,r}}, i \in [1, \dots, k] \tag{3}$$

In the complex number h, A is the CSI amplitude, θ is the CSI phase, and i is the number of subcarriers in each channel. The number of available subcarriers can also vary according to the type of selected hardware or channel bandwidth. In 20MHZ bandwidth, Raspberry pi4 (Nexmon CSI Tool) can access 56 subcarriers.

Alterations in both phase and amplitude manifest as a consequence of activity engagement or environmental adjustments. Nonetheless, unsynchronized transmitters and receivers introduce haphazard phase offsets within the CSI, resulting in erratic transformations (Fig. (1)). Furthermore, the sampling frequency offset exerts an impact on the phase, whereas CSI generally maintains a relatively consistent range [42]. Consequently, CSI amplitude typically finds application as a more dependable metric.

In the context of human activity recognition, CSI plays a significant role. With the advancement of wireless technologies and sensing methodologies, wireless signals can sense human behaviors. The CSI data contains information about environmental changes, including the movement of humans in a specific environment. Therefore, it can be used to recognize human activities. Meanwhile, antenna selection can be done based on their sensitivity in accordance with different activities.

C MFCC and Representation of Data

The MFCC algorithm was employed for feature extraction in this study (refer to Fig. (2)). The shared attributes between CSI data and speech data rendered MFCC a fitting choice for feature extraction, as elucidated in Ref. [22]. Distinct shifts in overall spectral composition were evident in CSI samples across various activities, manifesting

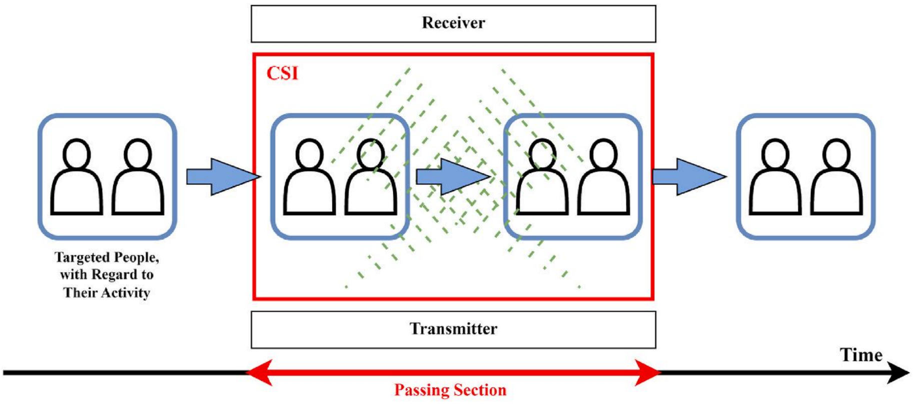

Fig. (1). Overview of CSI recording for the purpose of human activity recognition.

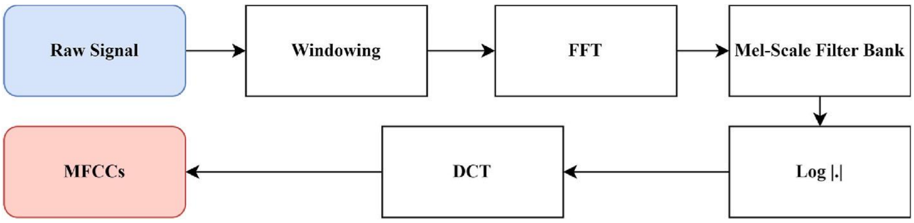

Fig. (2). The process of calculating MFCC values.

changes in frequency content during actions. Notably, all activities' occupied bandwidths fell within the auditory spectrum of human hearing, approximately 20–20 kHz [22]. CSI time series and sound signals exhibited a considerable overlap in a substantial portion of the human audible hearing spectrum, especially in the lower frequency range up to 1000 Hz, aligning seamlessly with the emphasis of MFCC [22]. Additionally, since sound and Wi-Fi signals adhered to comparable wave equations as electromagnetic waves within the measurement environment, they underwent analogous reflections, refractions, and diffractions from barriers and walls.

To process the original lengthy CSI data per record, a segmentation strategy was employed employing a sliding window with a span of 300 and a step of 30. This maneuver augmented the data accessible for subsequent procedures, resulting in 4545 samples for all classes. Each sample bore a structure of [300, 52], cumulating into a comprehensive data array of [4545, 300, 52]. Employing the MFCC algorithm, which demanded time-series data, was instrumental in representing each sample. Literature such as [22] advocated dimensionality reduction methods like PCA for employing the MFCC algorithm on the initial data. By applying PCA, the original data could be transformed into 4545 sample vectors, each possessing a length of 300. Alternatively, it was viable to utilize the MFCC algorithm on each of the 52 dimensions (thus calculating MFCCs for each of the vectors with the length of 300), culminating in a larger represented data array of [4545, 59, 13, 52]. Here, every sample had a form of [59, 13, 52]. Employing parameters selected through trial and error, a sample rate of 500 Hz (notably lower than the previously mentioned 1000 Hz), 13 cepstrum, a 512-size Fast Fourier Transform (FFT), and 26 filters in the filterbank were configured for the MFCC algorithm.

D Autoencoder

An autoencoder represents a variant of artificial neural networks designed to autonomously learn the compression and subsequent reconstruction of data. This construct comprises an input layer, an output layer, and one or more concealed layers. The process of converting input data into a condensed representation of lower dimensions is termed encoding, while the converse process of restoring the original data from this encoded form is referred to as decoding. Positioned within the latent space, this encoded representation harbors distilled features that encapsulate the most pivotal information inherent in the input data.

In this study, an autoencoder is harnessed to extract analogous features from instances belonging to identical classes. This endeavor aims to enhance classifier performance by leveraging these acquired features. Notably, the encoder within the autoencoder is educated and preserved for subsequent utilization within the classifier. The decoder, entrusted with the task of restoring the input data, remains unutilized within the classifier's context in this investigation. The function that is used to make a nonlinear mapping of the x input at the encoder is as follows:

$$d_i = \sigma(wx_i + b) \tag{4}$$

where σ is a nonlinear activation function, d_i is encoded features, and w and b are weight and bias, respectively. The decoder function used to reconstruct the input data is as follows:

$$\widehat{d}_i = \sigma(\widehat{w} d_i + \widehat{b}) \tag{5}$$

where $\widehat{d}_i$ is the output of the decoder, which is designed to be exactly similar to the original input, while $\widehat{w}$ and $\widehat{b}$ are, respectively, the weight and bias of the decoder.

E. Multi-Input Multi-Output Autoencoder

This investigation introduces a pioneering amalgamation of the MFCC algorithm and MIMO AE for the extraction of features. The MIMO AE, characterized by two inputs and two outputs, extracts shared information from its initial inputs and retains them within the output of the encoder subnetwork. Subsequently, the decoder subnetwork reconstructs the inputs using these derived features. During the classification phase, the classifier network discerns the type of activity, disregarding extraneous attributes like personal identity, gender, or age. The impact of these non-essential attributes on the input can be mitigated through the utilization of the MIMO AE encoder, potentially bolstering the efficacy of the trained classifier (Refer to Fig. (3)). The MIMO AE training necessitates pairs of samples that share the same activity type but differ in non-targeted attributes. Any dissimilarity related to non-targeted aspects suffices, even if the activity is replicated.

The characteristics and behaviors of the MIMO AE are clarified through the utilization of RGB color images, as demonstrated in Fig. (4). One image portrays a purple shade, while the other exhibits a green hue. Both of these images are utilized as inputs for the two-input two-output AE architecture. Within this configuration, the encoder component distills the color blue as a salient feature. Blue serves as a representation of the commonalities between the two images. Deviations in color, symbolized by the hues red and green, are retained as weights and biases within the decoder during the feature extraction process. Ultimately, these divergences merge with the shared color (blue), culminating in the reconstruction of the input images at the output. It is important to emphasize that achieving an exact replication of the original inputs remains an elusive endeavor. In the context of training the MIMO AE with CSI data, studies such as [43] attest to the enhanced performance achieved in comparison to prior methodologies for various tasks.

The MIMO AE is architected as depicted in Fig. (5), with inputs and outputs conforming to the contours of the extracted MFCCs. The deliberate similarity in feature shape to the original inputs facilitates equitable comparisons in future studies. A mere 1.5% of all samples are selected for MIMO AE training, ensuring minimal overlap between the data utilized for MIMO AE training and that earmarked for classifier training, validation, and evaluation. Escalating the number of pairs fails to considerably diminish the validation loss of the MIMO AE. Given that faithful input reconstruction is not a foremost objective, this characteristic does not encumber subsequent investigations. All layers employ the ReLU function as their activation function, except for the last layer which remains devoid of activation. The training process adheres to a batch size of 64, spans 300 epochs, features a learning rate of 0.0001, incorporates a kernel size of 5, and employs Mean Square Error (MSE) as the loss function, orchestrated by the Adam learning algorithm. Notably, minor alterations in the hyperparameters of the devised MIMO AE, such as learning rate and dropout, do not elicit substantial alterations in the results.

F. Fine-Tuning

Transfer Learning (TL) is a machine learning strategy that heightens the efficacy of a successive model by incorporating a preexisting model (with predetermined weights and biases) as a foundational framework for the subsequent analogous model. This technique has garnered notable currency in the realm of HAR challenges in contemporary times. Hernandez et al. [43] furnish an exhaustive examination of the manifold applications of TL within HAR research. Another avenue within the transfer learning paradigm is fine-tuning. Additionally, Ray et al. [44] have explored the deployment of TL in HAR, particularly in the context of visual data. Further, Pavliuk et al. [45] conducted an investigation into the synergy between TL and wavelet transform for feature extraction, introducing a novel amalgamation of methodologies that yields discernible outcomes.

In the framework of transfer learning, an established model is integrated into a distinct yet conceptually linked model. The discerning choice of the constituent from the primary model for integration into the secondary model carries substantive import, with the aim of augmenting the ultimate outcomes of the latter through harnessing the outputs of the former. To elucidate, as depicted in Fig. (6), subsequent to the training of the inaugural network on dataset 1, model 1 is primed for assimilation into an alternate network. Following this, network 2, encompassing the trained model 1, undergoes training with dataset 2. Importantly, model 1 within network 2 can be subjected to subsequent rounds of training. In the proposed methodology, only the encoder component of the autoencoder is harnessed within the classifier.

Delineating the fine-tuning method and TL reveals a disparity in the selection of training datasets for the initial and subsequent models. TL entails the utilization of an entirely distinct dataset for training the initial and second models. Conversely, in fine-tuning, a modest portion of the dataset is allocated for training the initial model, and the residual dataset is allocated for training the second model. In both paradigms, following the training of the initial model and the determination of its weights and biases, the ensuing alterations in the second model are subtler compared to the scenario where the second model is trained from the ground up. In this study, fine-tuning is adopted due to the limited dataset allocation for training the autoencoder model. A fraction of the remaining data is arbitrarily chosen for the training of the classifier. Notably, the data deployed for autoencoder training remains detached from the training process of the classifier.

G. Designed Classifier

The classifier's architecture is presented in Fig. (7). The initial stratum of this configuration corresponds to the encoder subnetwork of the MIMO AE, which can be optionally omitted in the fundamental rendition of this classifier. In three distinct approaches, this encoder can function as a pretrained and untrainable layer, a pretrained and trainable layer, or an untrained layer. This gives rise to a total of four potential approaches, encompassing the option of not integrating any form of encoder. The comprehensive array of conceivable scenarios is visually

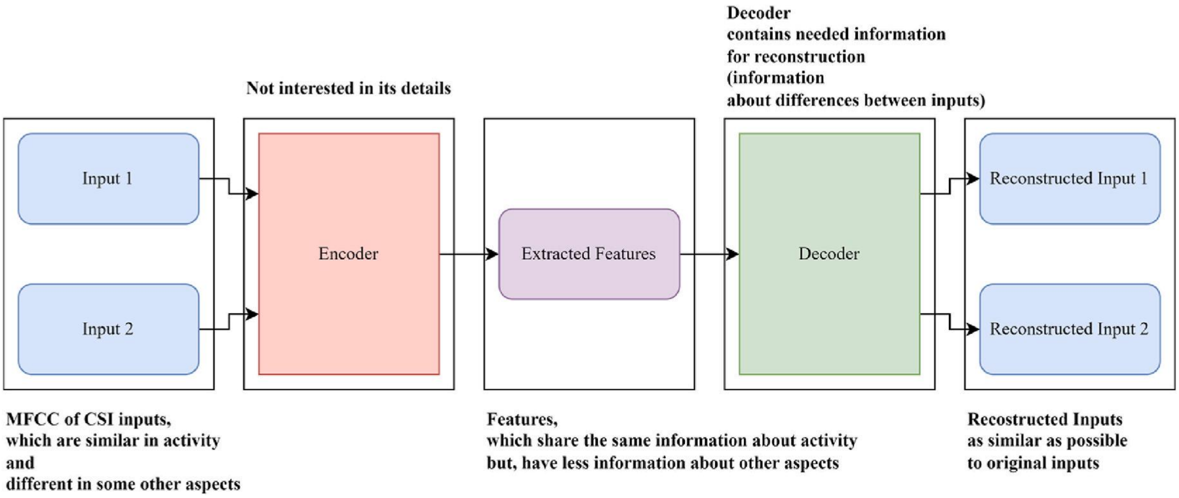
Not interested in its details
Decoder
contains needed information
for reconstruction
(information
about differences between inputs)
Input 1
Input 2
Encoder
Extracted Features
Decoder
Reconstructed Input 1
Reconstructed Input 2
MFCC of CSI inputs,
which are similar in activity
and
different in some other aspects
Features,
which share the same information about activity
but, have less information about other aspects
Recostructed Inputs
as similar as possible
to original inputs

Fig. (3). Details, regarding the subnetworks of the MIMO AE.

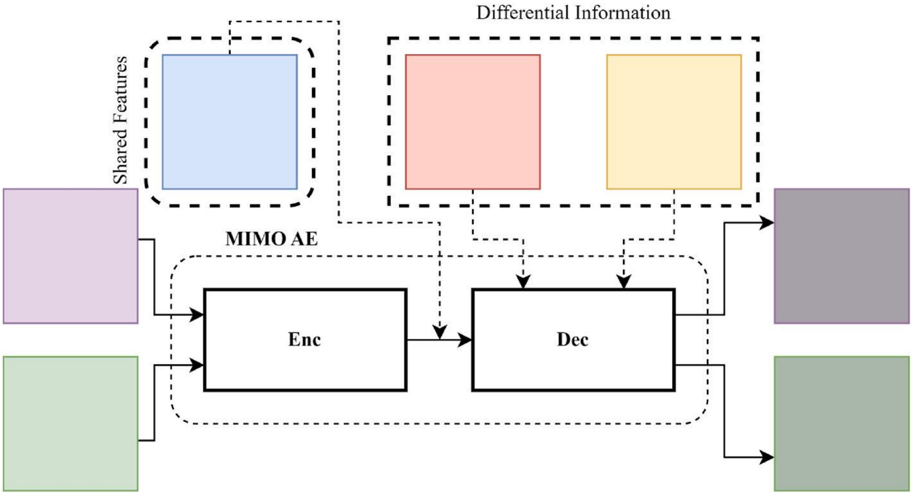

Fig. (4). Simple explanation of trained MIMO-AE and relation of inputs and extracted features, using colors.

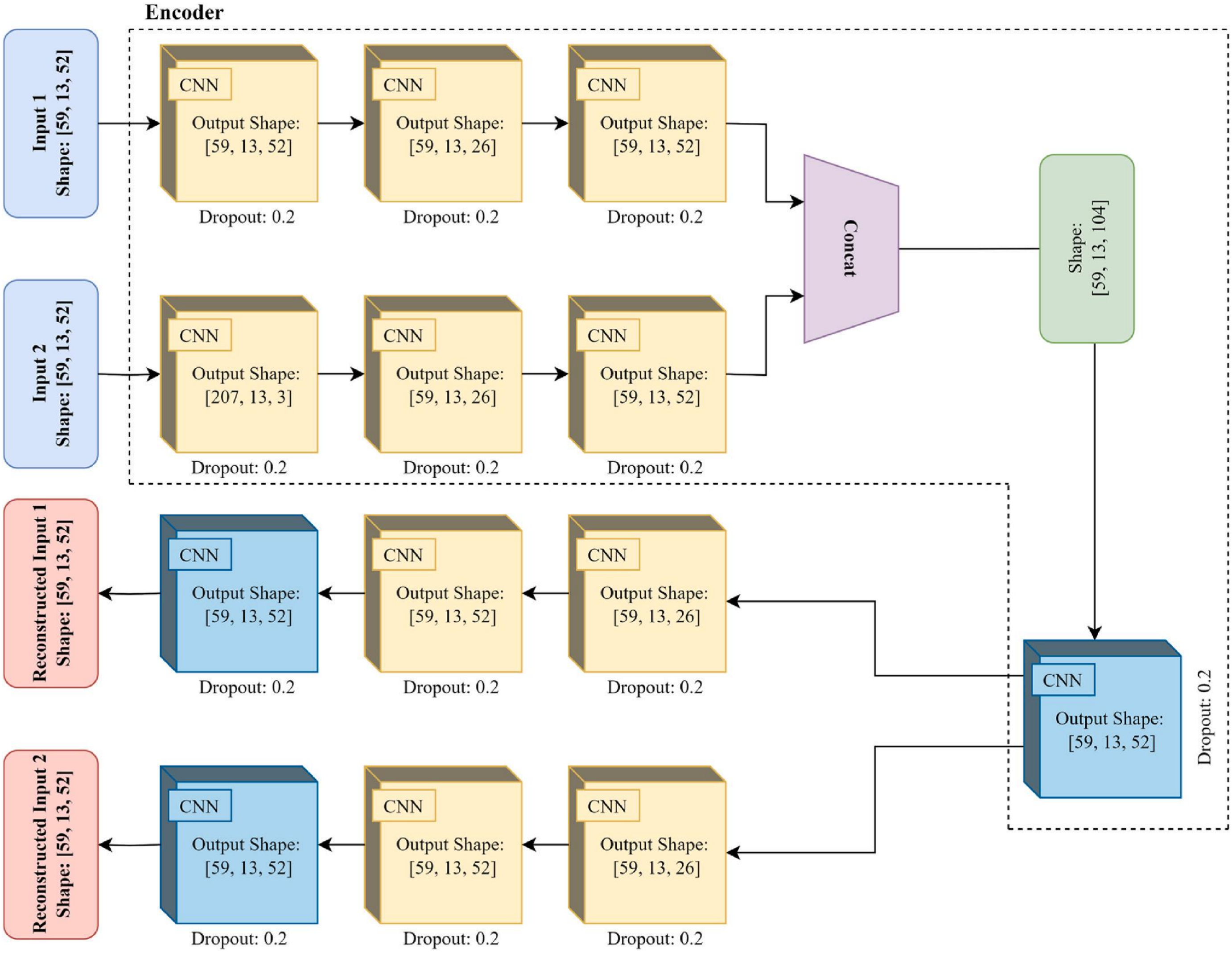

Fig. (5). The designed MIMO AE.

depicted in Fig. (8). Additionally, a range of distinct hyperparameters is systematically examined to meticulously fine-tune the classifier, thereby ensuring its resilience and efficacy.

Concerning the hyperparameters governing the devised classifier, the Relu function is engaged as the activation function across all strata, with the exception of the output layer, where Softmax serves as the

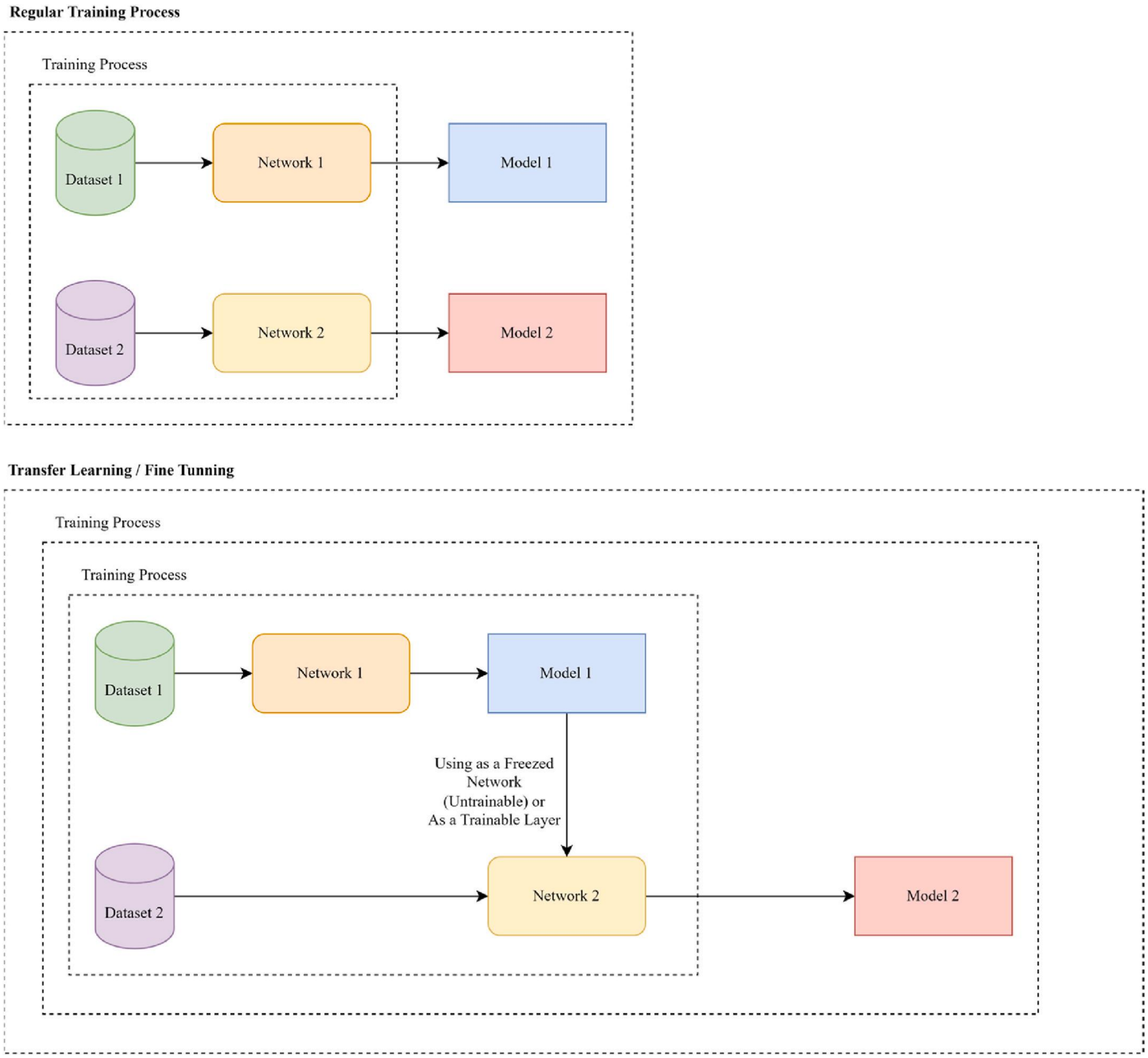

Fig. (6). General concept of transfer learning/fine-tuning, based on using pretrained models in the training process of a secondary model.

designated activation function. Furthermore, all convolutional strata adopt a kernel size of 7. In relation to the remaining hyperparameters, a batch size of 16, 150 epochs, Categorical Cross Entropy (selected for its applicability in multi-class classification tasks) as the designated loss function, and the Adam algorithm for the learning process, alongside an initial learning rate of 0.0001, are determined through iterative refinement. The adaptive nature of the learning rate is managed by the Adam algorithm throughout the training procedure.

3. Results of simulations

This section proceeds to showcase the outcomes achieved through the utilization of the proposed MIMO AE in three distinct methodologies. Initially, 20% of the available dataset is reserved exclusively for the assessment phase, thereby ensuring a rigorous and comprehensive evaluation of the model's performance. Subsequently, leveraging the remaining dataset, experiments are meticulously conducted across a range of proportions, spanning from 10% to 50% (equivalent to 8%–48% of the total available dataset) for both the training and validation phases. Within these two phases, 80% of the available data is devoted to training endeavors, while the remaining 20% is exclusively allocated for validation purposes. It is pertinent to emphasize that a minor fraction of the original dataset is specifically allocated and employed to train the MIMO AE prior to initiating this process. This methodology effectively empowers the pretrained encoder to aptly distill and capture pivotal features inherent within the data. Through the systematic assessment of the classifier's performance across varying data fractions, the overarching objective is to showcase its adeptness in attaining notable levels of accuracy despite the constraints imposed by limited training data.

A. Numerical Results of K-Fold Cross-Validation

This section offers a comprehensive exposition of the simulation outcomes pertaining to four distinct models: the designed classifier, the

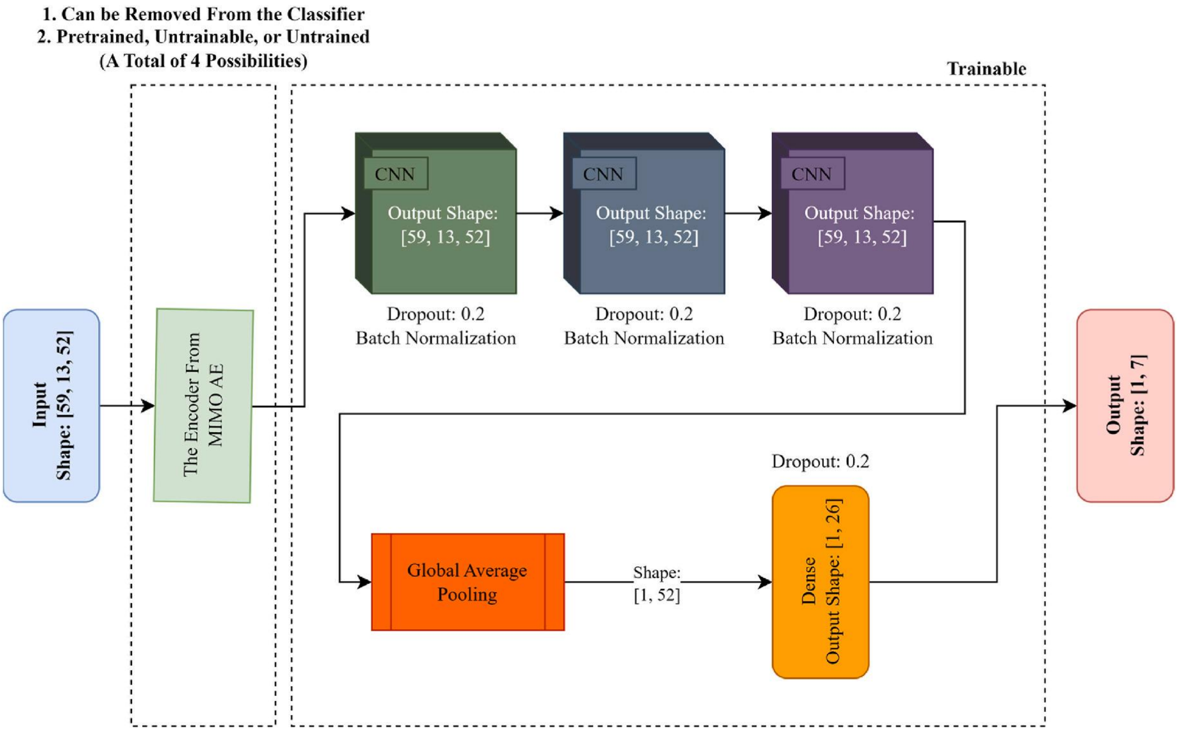

Fig. (7). The designed classifier, including the three possible approaches of utilizing the encoder from the MIMO AE.

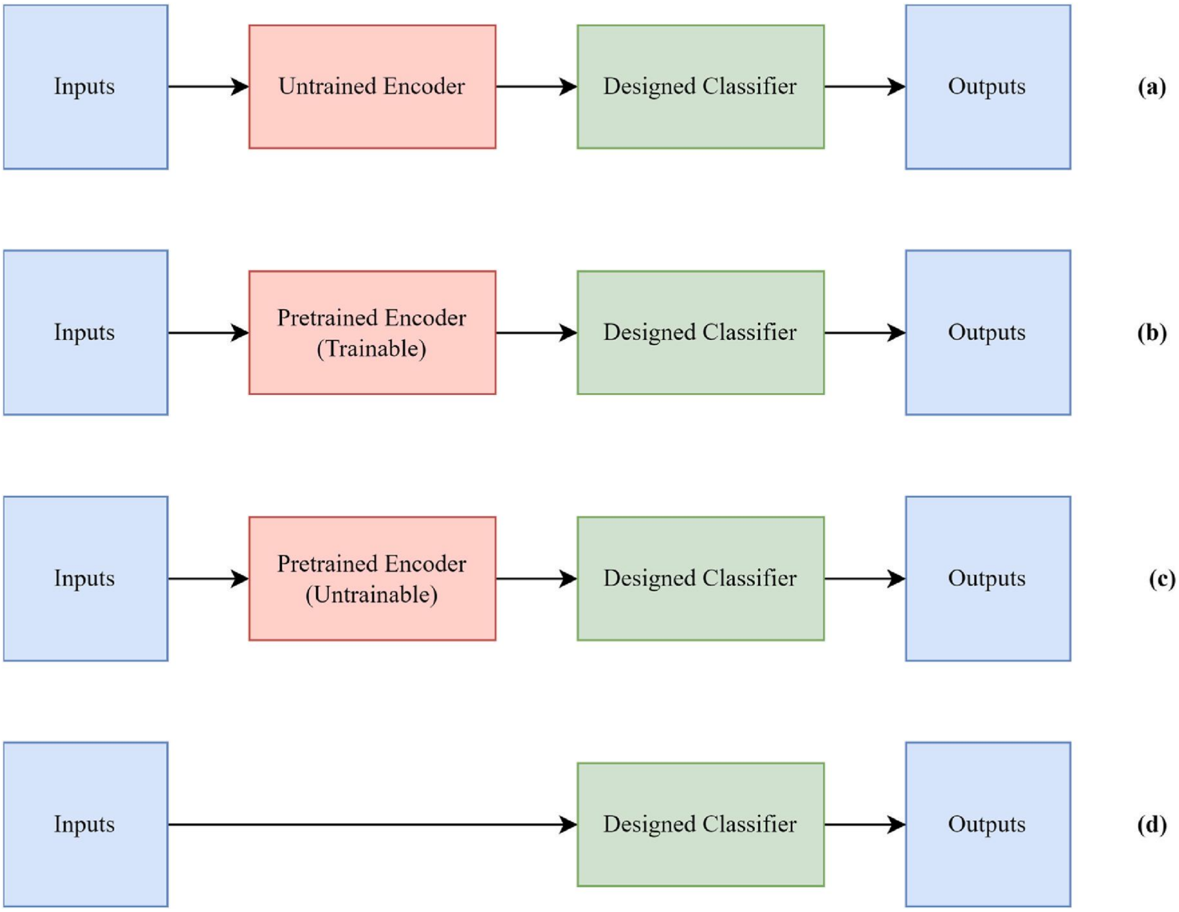

Fig. (8). All possible scenarios a) Designed Classifier and Untrainable Encoder b) Designed Classifier and Trainable Encoder c) Designed Classifier and Untrainable Encoder and d) Designed Classifier, regarding the utilizing the encoder from MIMO AE in the designed classifier.

designed classifier with the pretrained and untrainable encoder, the designed classifier with the pretrained and trainable encoder, and the designed classifier with the untrained encoder. The primary objective underlying these model comparisons is twofold. Firstly, with respect to the designed classifier paired with the pretrained and untrainable encoder, the intention is to underscore the merits of employing this particular encoder variant by juxtaposing its performance against that of the designed classifier. Secondly, in the case of the designed classifier accompanied by the pretrained and trainable encoder, this comparative analysis serves to ascertain that any potential enhancement is not solely attributed to the use of a larger network. This is validated by comparing it with an extended iteration of the designed classifier, namely the version incorporating the untrained encoder. Through these systematic comparisons, substantive insights can be derived concerning the influence of diverse encoder configurations on the classifier's performance.

In the context of the data segmentation process presented, the utilization of K-fold cross-validation with $K = 5$ is employed to ensure the consistency and reliability of the results. Given the diverse data segments involved in this study's training process, a judicious application of the cross-validation methodology is paramount. To this end, the evaluation data is initially segregated from the training and validation data via K-fold cross-validation. Subsequently, segments of varying magnitudes are harnessed for subsequent training and validation phases. This approach facilitates a comprehensive and stringent appraisal of the model's performance while accommodating the varied dataset segmentations. The holistic concept is visually illustrated in Fig. (9). The conclusive evaluation outcomes, encompassing the utilization of up to 50% of the training and validation data, are meticulously presented in Table (1). It is imperative to note that when deploying more extensive datasets for training and validation (using 60% of available data or more), all four models attain near-impeccable results, with accuracy surpassing the 97% threshold. However, it is noteworthy that this study is explicitly tailored to address scenarios characterized by limited training data volume. As such, these high-performance instances are intentionally omitted from this discourse. Instead, the paper accentuates the efficacy and enhancements engendered by the proposed approach in contexts characterized by constrained data resources.

Table (1)
The recognition results, regarding all four possible approaches and different data sizes (the training and validation data).

The Classifier	10% of the Data	20% of the Data	30% of the Data	40% of the Data	50% of the Data
Designed Classifier	69.2%	76.7%	79.3%	85.5%	93%
Designed Classifier with Untrainable Encoder	80.1%	86,1%	90.3%	93.7%	96.2%
Designed Classifier with Untrained Encoder	64%	68.3%	79.8%	80.8%	82.9%
Designed Classifier with Trainable Encoder	71.1%	82.4%	93.7%	96.3%	99.1%

B. Confusion Matrices of Best Results, Regarding Each Model

The accuracy acquired via the K-fold cross-validation method stands as a reliable benchmark for upcoming research endeavors. Given the intricate nature of this challenge as a seven-class classification, it holds paramount significance to ensure that recognition rates for each distinct class maintain parity or close alignment with the overarching attained accuracy. To effectively address this concern, we present the confusion matrices linked to one specific fold within the set of five folds, as facilitated by the K-fold cross-validation technique. These matrices are visually elucidated in Fig. (10), specifically considering the utilization of

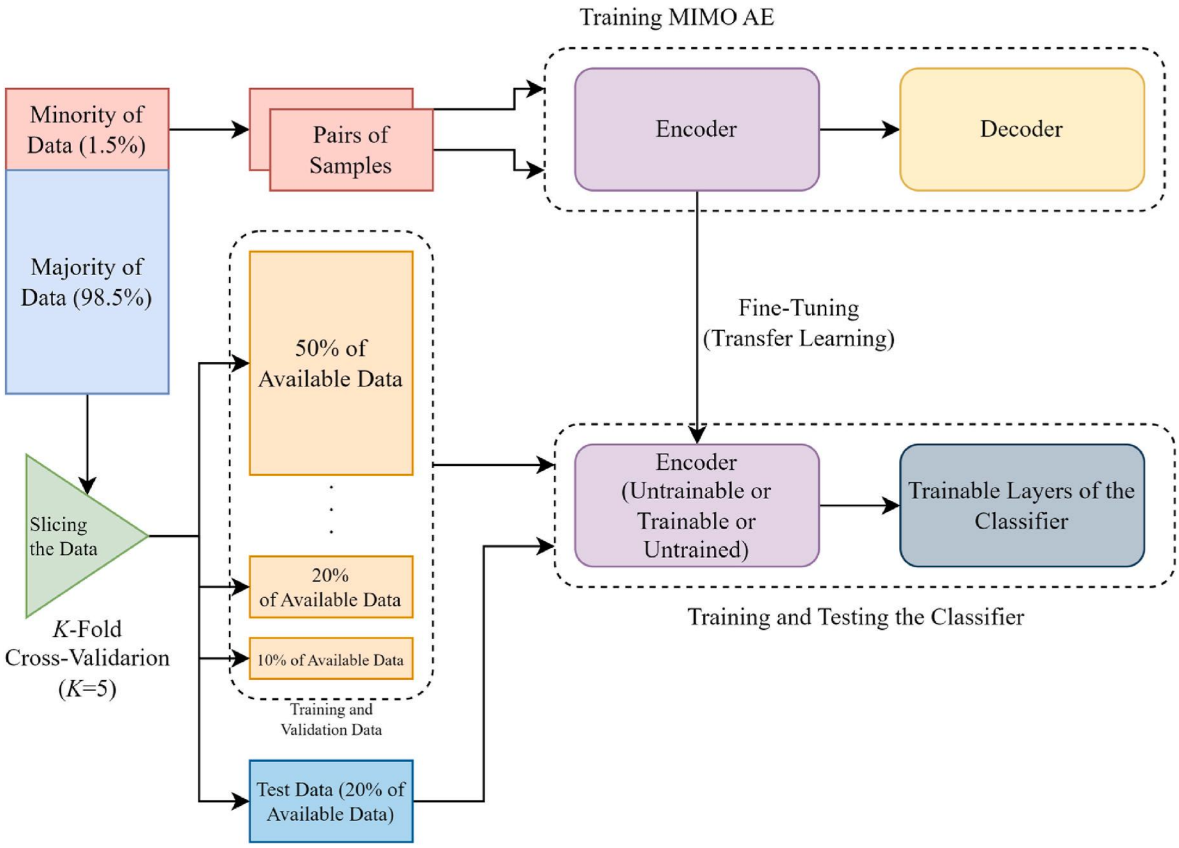
Training MIMO AE
Minority of Data (1.5%)
Majority of Data (98.5%)
Pairs of Samples
Encoder
Decoder
Fine-Tuning (Transfer Learning)
50% of Available Data
20% of Available Data
10% of Available Data
Training and Validation Data
Test Data (20% of Available Data)
Slicing the Data
K-Fold Cross-Validarion (K=5)
Encoder (Untrainable or Trainable or Untrained)
Trainable Layers of the Classifier
Training and Testing the Classifier

Fig. (9). The process of slicing data for different aspects of the proposed study.

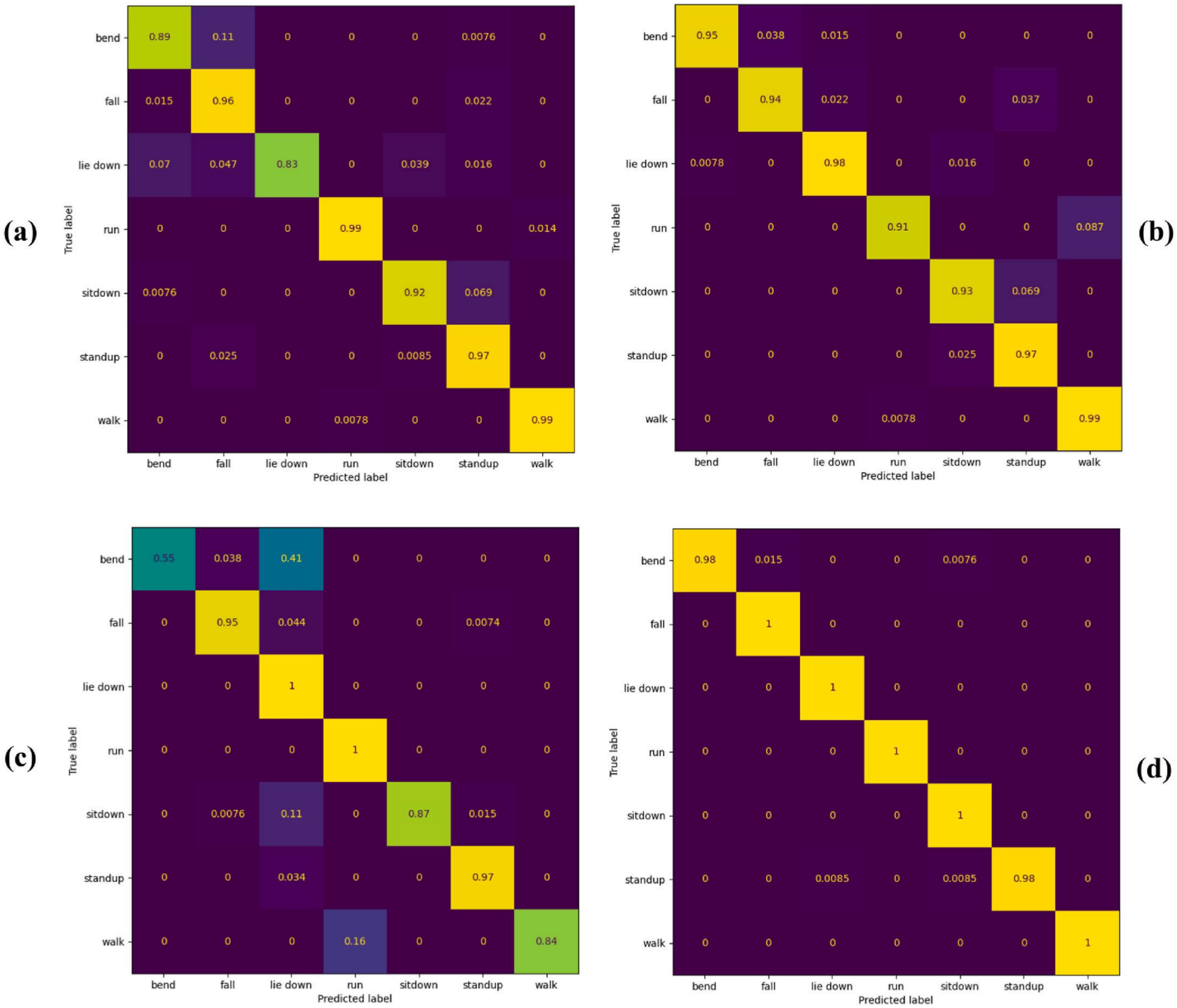

Fig. (10). Confusion matrix, regarding the a) Designed Classifier b) Designed Classifier and Untrainable Encoder c) Designed Classifier and Untrained Encoder, and d) Designed Classifier and Trainable Encoder, using 50% of available data for the training and validation phases.

50% of the available dataset. The elected fold epitomizes the most comprehensive and inclusive outcomes among the spectrum of available alternatives. Essentially, it circumvents any anomalies or isolated outcomes that may manifest within the majority of alternative folds. A comprehensive examination of these graphical representations unveils the accomplishment of commendable recognition rates by each meticulously designed model across the entirety of sample classes. This observation underscores the profound effectiveness and unwavering resilience embedded within the proposed models, thereby faithfully delineating activities within every distinct class.

C. Comparison of the Results, Obtained from the Introduced Models with Themselves

When examining data with limited proportions (e.g., 10% and 20% of available data), as indicated in Table (1), the classifier designed with an untrainable encoder showcases superior performance among the various models. Despite the relatively meager quantity of training samples, a noteworthy observation arises from the comparison of accuracy between the classifier designed and the classifier with the untrainable encoder. For instance, when utilizing 10% and 20% of the available data for training and validation (resulting in respective accuracy values of 69.2% and 76.7% versus 80.1% and 86.1%), a meaningful discrepancy is apparent. This differentiation is intriguing, particularly given that these two networks share an identical count of trainable

Table (2)
The total number of trainable and untrainable parameters per model.

Type of Parameter	Designed Classifier	Designed Classifier with Untrainable Encoder	Designed Classifier with Untrained Encoder	Designed Classifier with Trainable Encoder
Total Parameters	399,835	805,747	805,747	805,747
Trainable Parameters	399,523	399,523	805,435	805,435
Untrainable Parameters	312	406,224	312	312

parameters, as outlined in Table (2). The presence of prior knowledge embedded in the pretrained encoder evidently influences the evaluation outcomes of these models. Notably, even when affording more data for the training of both models, the classifier with the untrainable encoder consistently outperforms its more fundamental counterpart, i.e., the standard classifier. This observation underscores the compensatory role of efficient and effective feature extraction in mitigating the relatively weaker training outcomes in the absence of a sizable dataset.

Considering the utilization of a pretrained encoder as a trainable layer, as depicted in Table (1), the potential for attaining higher recognition rates compared to the aforementioned approaches is apparent. However, this heightened performance comes at the expense of heightened computational demands, stemming from an increased tally of trainable parameters. This, in turn, translates to an extended training process duration, as revealed in Tables (2) and (3). While the results portrayed in Table (1) depict the classifier designed with the trainable encoder as showcasing the best general performance, a comprehensive perspective necessitates a closer examination. When evaluating the duration of the training process and the number of trainable parameters, alongside a performance improvement of nearly 3% over the classifier with the untrainable encoder, the classifier with the untrainable encoder emerges as the most valuable model presented in this study. It is noteworthy that the discernibly enhanced performance of the classifier with the trainable encoder, in comparison to its counterpart with the untrained encoder, despite a comparable count of trainable parameters, underscores the previously discussed assertion. Namely, the use of a pretrained encoder in both trainable and untrainable capacities lead to a marked improvement, particularly when confronted with limited training data availability.

D. Comparison of the Results, Obtained from the Introduced Models with Other Studies

In order to establish an equitable benchmark for comparison against analogous studies, two specific references, denoted as [13,14], elected to employ an identical dataset with similar objectives. Notably [13], conducted a repetition of simulations following the model introduced in Ref. [14], encompassing a spectrum spanning from 10% to 50% of the available data (equivalent to 80% of the total dataset) for both training and validation purposes. Consequently, the findings derived from these two investigations offer a relevant comparative foundation to the results of the present study. The accuracies attained across the presented models and the aforementioned antecedents are diligently compiled in Table (4).

Upon a thorough examination of Table (4), it becomes evident that the designed classifier coupled with the untrainable encoder has outperformed the preceding models described in Refs. [13,14]. Notably, this observation is particularly striking considering that one of the models under scrutiny encompasses a pretrained and trainable encoder, underlining the tangible influence of employing MFCC data representation throughout the entire process. Furthermore, the distinct contrast in the tally of trainable parameters between the designed classifier featuring the trainable encoder (805,435) and its predecessor (2,227,107) [13], coupled with the fact that the former surpasses the latter, underscores the pivotal role of a proficient and efficacious feature extraction methodology.

Table (3)
The processing time (sec) of proposed methods.

Percentage of Used Dataset for Training and Validation	Designed Classifier (Train/ Test)	Designed Classifier and Untrainable Encoder (Train/Test)	Designed Classifier and Untrained Encoder (Train/Test)	Designed Classifier and Trainable Encoder (Train/Test)
10%	71.06, 0.35	110.17, 0.67	137.26, 0.93	133.46, 0.99
20%	94.68, 0.34	144.77, 0.67	190.04, 0.73	182.27, 0.77
30%	123.99, 0.58	179.74, 0.64	247.27, 0.68	235.51, 0.72
40%	146.04, 0.35	215.62, 0.74	301.23, 0.75	283.83, 0.71
50%	159.64, 0.36	234.6, 0.68	341.83, 0.72	331.42, 0.7

Table (4)
Comparing the evaluation accuracy of proposed methods with the results of [13, 14].

Percentage of Used Dataset for Training and Validation	Best Results of [14]	Best Results of [13]	Designed Classifier with Untrainable Encoder	Designed Classifier with Trainable Encoder
10%	64.76%	37%	80.1%	71.1%
20%	68.27%	68.2%	86,1%	82.4%
30%	78.26%	87.27%	90.3%	93.7%
40%	74.64%	93.21%	93.7%	96.3%
50%	76.72%	94.49%	96.2%	99.1%

While there are two studies available for results comparison using different ratios of training data [13,14], three other studies (recently published) are also selected for a fairer comparison. These studies [38–40] did not try to find or evaluate the optimized size of training data, and they used the complete dataset for all three processes of training, validation, and evaluation. Based on the result comparison, it is safe to say that this study manages to introduce models that are more efficient and effective than the models introduced in these three studies.

Concerning a more detailed comparison [38], managed to achieve 99.6% accuracy using K-fold cross-validation for $K = 5$ (similar to this study). While they achieved outstanding accuracy, the results presented in this manuscript (Table (4)) demonstrate an accuracy of 99.1% using only half the training data. Furthermore, our studies showed that the most promising model of the research (Designed Classifier with Trainable Encoder) is capable of achieving an accuracy of 99.8% in the same condition (using full training data). It is worth mentioning that using more than 50% of the training data is not an aim of this study; this extra step has been done only to provide a more precise comparison.

Moreover, in the case of [39], the researchers managed to achieve an accuracy of 97.6% using K-fold cross-validation for $K = 5$ (similar to this study), where, similar to Ref. [38], it lacks the efficiency and effectiveness of the proposed methods in this manuscript. In the case of [40], the researcher decided to focus on the recognition of only sitting and standing, without using K-fold cross-validation. In the recognition of sitting and standing, authors managed to reach accuracies of 94.51% and 96.04%, respectively, where in both cases, their models lag behind the models proposed in this manuscript. In case of average accuracy of 99.8% (using full training data) in this manuscript, accuracies of 99.76% and 98.38% are achieved, in case of recognizing sitting and standing, respectively.

E Discussion

The outcomes presented in the study underscore the superior performance achieved through the amalgamation of the designed classifier with the pretrained and trainable encoder. It is of significance to note that the designed classifier accompanied by the pretrained and untrainable encoder exhibits notably fewer trainable parameters (equivalent to the designed classifier) and still manages to attain a mere 3% decrease in accuracy when compared to the optimal outcome. Consequently, this particular approach emerges as the preeminent model among the four alternatives proposed, adeptly harmonizing computational costs with recognition precision. In other words, if the study's limitations are constrained solely by the number of samples, employing a classifier with a pretrained and trainable encoder is the optimal approach. On the other hand, if restricted access to more robust hardware poses an additional limitation, utilizing a classifier with a pretrained and untrainable encoder not only incurs a significantly lower computational cost but also results in only a marginal decrease of nearly 3% in accuracy compared to the highest achievable recognition rate. Howe, considering the fact that accessing robust hardware is not a challenge to the most of recent studies, using classifier with a pretrained and trainable encoder is a more promising approach.

Furthermore, it is pivotal to acknowledge that the discernible 16% variance in performance between the designed classifier, incorporating the pretrained and trainable encoder, and its counterpart with the untrained encoder, can be solely attributed to the prior knowledge imbibed within the pretrained encoder. This discernment, congruous with prior scholarly works such as [13,14], accentuates the potency of fine-tuning and transfer learning in enhancing the efficacy of comparatively compact classifiers while mitigating computational overheads. Additionally, upon meticulous analysis of the presented confusion matrices, all suggested models distinctly attain a commendably consistent class-wise accuracy, correlating well with the overall precision of each model.

A previous study [13] had already demonstrated the efficacy of employing MIMO AE to secure satisfactory results in HAR, even when confronted with limited training data. However, previous outcomes unveiled a substantial disparity of approximately 13% (pertaining to 50% of accessible data for training and validation) between the performance of the designed classifier partnered with the pretrained and untrainable encoder and its equivalent furnished with the pretrained and trainable encoder [13]. In the current investigation, given the modest 3% divergence within the outcomes of these classifiers, it is rational to contend that the designed classifier integrating the pretrained and untrainable encoder surfaces is the most promising approach introduced herein. It merits mention that all outcomes, inclusive of those emanating from the designed classifier, notably exceed the comparative results elucidated in Ref. [13]. This marked enhancement is a direct outcome of the deployment of the MFCC algorithm as the preprocessing methodology, a facet hitherto unexplored in the preceding study.

4. Conclusion

The study explores the efficacy of amalgamating a formulated classifier with various iterations of a pretrained MIMO AE encoder, employing MFCC derived from CSI data for HAR with limited training data. The outcomes illustrate that adopting the devised classifier in tandem with the pretrained and untrainable encoder yields an exceptional recognition rate. This is accomplished while retaining a relatively modest count of trainable parameters, thus establishing a harmonious equilibrium between computational expense and recognition precision. Consequently, this model emerges as the most valuable among the four proposed strategies. While human activity recognition models face various challenges, such as lack of sufficient data, accuracy, optimization and security, the suggested models reached a better level, regarding of such aims, in two manners, i.e., lack of sufficient data and accuracy.

Additionally, the investigation underscores the significance of fine-tuning and transfer learning with the pretrained encoder, culminating in a noteworthy enhancement in the performance of the designed classifier paired with the pretrained and trainable encoder, as opposed to its untrained counterpart. This observation aligns harmoniously with prior scholarly inquiry, validating the effectiveness of leveraging such methodologies to heighten the proficiency of compact classifiers sans inordinate computational strain. A juxtaposition of these findings with outcomes from a previous study, which concentrated on utilizing MIMO AE for HAR within a restricted dataset, reveals a striking progression. The introduction of the formulated classifier with the pretrained and untrainable encoder substantially diminishes the performance discrepancy when compared to the model equipped with the pretrained and trainable encoder. This narrowing culminates in a modest 3% distinction, attributable to the utilization of the MFCC algorithm for feature extraction.

In summation, the amalgamation of MFCC, MIMO AE, and fine-tuning techniques herald new avenues for HAR exploration, yielding auspicious outcomes and paving the trajectory for the development of more efficient and precise activity recognition systems in the times ahead.

Summary

The study delves into Wi-Fi-based Human Activity Recognition (HAR) employing Channel State Information (CSI) and deep learning models. Challenges arise in acquiring sufficient training data across diverse scenarios. Thus, a unique strategy is adopted: the utilization of a pretrained Multi-Input Multi-Output Autoencoder (MIMO AE) coupled with Mel Frequency Cepstral Coefficients (MFCC) for feature extraction. This innovative approach yields remarkable outcomes, leveraging notably fewer training data in contrast to conventional machine learning methods.

Among the array of approaches explored, the application of the designed classifier with the pretrained and untrainable encoder emerges as the most efficacious. It showcases a noteworthy recognition rate while maintaining a relatively modest count of trainable parameters. This equilibrium successfully navigates the trade-off between computational cost and recognition accuracy. The study underscores the significance of fine-tuning and transfer learning with the pretrained encoder. This practice leads to substantial enhancements in performance for the designed classifier with the pretrained and trainable encoder, outperforming the version with the untrained encoder. This correlation aligns with earlier research, affirming the efficacy of such techniques in bolstering small classifiers without overwhelming computational demands.

Upon comparing the present findings with previous research concentrated on MIMO AE for HAR with limited data, the proposed approach manifests a significant stride. The incorporation of the designed classifier with the pretrained and untrainable encoder effectively narrows the performance disparity, resulting in a mere 3% deviation. This distinction can be exclusively attributed to the application of the MFCC algorithm for feature extraction. In essence, the study introduces a pioneering avenue, harmonizing MFCC, MIMO AE, and finely tuned methodologies, promising heightened efficiency and precision in the realm of HAR research.

The funding statement

The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability

The data, used in this study is publicly available on the web, and accessible at https://github.com/parisafm/CSI-HAR-Dataset (Accessed on 7 March, 2024).

CRediT authorship contribution statement

Amir Rahdar: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Mahnaz Chahoushi:** Writing – original draft, Software, Investigation. **Seyed Ali Ghorashi:** Writing – review & editing, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGBT Service in order to enhance the language quality of the manuscript, by editing the original text, written by the authors. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Park, N. Kim, G.H. Lee, J.K. Choi, MultiCNN-FilterLSTM: resource-efficient sensor-based human activity recognition in IoT applications, Future Generat. Comput. Syst. 139 (Feb. 2023) 196–209, https://doi.org/10.1016/j.future.2022.09.024.

[2] A. Maity, A. Pathak, G. Saha, Transfer learning based heart valve disease classification from Phonocardiogram signal, Biomed. Signal Process Control 85 (Aug. 2023) 104805, https://doi.org/10.1016/j.bspc.2023.104805.

[3] A. Sarkar, S.K.S. Hossain, R. Sarkar, Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm, Neural Comput. Appl. 35 (7) (Mar. 2023) 5165–5191, https://doi.org/10.1007/s00521-022-07911-0.

[4] P. Khan, Y. Khan, S. Kumar, M.S. Khan, A.H. Gandomi, HVD-LSTM based recognition of epileptic seizures and normal human activity, Comput. Biol. Med. 136 (2021) 104684, https://doi.org/10.1016/j.compbiomed.2021.104684. Sep.

[5] I.M. Pires, F. Hussain, G. Marques, N.M. Garcia, Comparison of machine learning techniques for the identification of human activities from inertial sensors available in a mobile device after the application of data imputation techniques, Comput. Biol. Med. 135 (2021) 104638, https://doi.org/10.1016/j.compbiomed.2021.104638. Aug.

[6] L. Köping, K. Shirahama, M. Grzegorzek, A general framework for sensor-based human activity recognition, Comput. Biol. Med. 95 (Apr. 2018) 248–260, https://doi.org/10.1016/j.compbiomed.2017.12.025.

[7] H. Abedi, A. Ansariyan, P.P. Morita, A. Wong, J. Boger, G. Shaker, AI-powered noncontact in-home gait monitoring and activity recognition system based on mm-wave FMCW radar and cloud computing, IEEE Internet Things J. 10 (11) (Jun. 2023) 9465–9481, https://doi.org/10.1109/JIOT.2023.3235268.

[8] J. Liu, H. Liu, Y. Chen, Y. Wang, C. Wang, Wireless sensing for human activity: a survey, IEEE Communications Surveys & Tutorials 22 (3) (2020) 1629–1645, https://doi.org/10.1109/COMST.2019.2934489.

[9] W. Cui, B. Li, L. Zhang, Z. Chen, Device-free single-user activity recognition using diversified deep ensemble learning, Appl. Soft Comput. 102 (Apr. 2021) 107066, https://doi.org/10.1016/j.asoc.2020.107066.

[10] X. Wang, C. Yang, S. Mao, ResBeat: resilient breathing beats monitoring with realtime bimodal CSI data, in: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Dec. 2017, pp. 1–6, https://doi.org/10.1109/GLOCOM.2017.8255021.

[11] Y. Xu, W. Yang, M. Chen, S. Chen, L. Huang, Attention-based gait recognition and walking direction estimation in Wi-Fi networks, IEEE Trans. Mobile Comput. 21 (2) (Feb. 2022) 465–479, https://doi.org/10.1109/TMC.2020.3012784.

[12] J. Yang, Y. Liu, Z. Liu, Y. Wu, T. Li, Y. Yang, A framework for human activity recognition based on WiFi CSI signal enhancement, Int. J. Antenn. Propag. 2021 (2021) e6654752, https://doi.org/10.1155/2021/6654752. Feb.

[13] M. Chahoushi, M. Nabati, R. Asvadi, S.A. Ghorashi, CSI-based human activity recognition using multi-input multi-output autoencoder and fine-tuning, Sensors 23 (7) (Jan. 2023), https://doi.org/10.3390/s23073591. Art. no. 7.

[14] P. Fard Moshiri, R. Shahbazian, M. Nabati, S.A. Ghorashi, A CSI-based human activity recognition using deep learning, Sensors 21 (21) (Jan. 2021), https://doi.org/10.3390/s21217225. Art. no. 21.

[15] Md M. Islam, S. Nooruddin, F. Karray, G. Muhammad, Human activity recognition using tools of convolutional neural networks: a state of the art review, data sets, challenges, and future prospects, Comput. Biol. Med. 149 (2022) 106060, https://doi.org/10.1016/j.compbiomed.2022.106060. Oct.

[16] E. Shalaby, N. ElShennawy, A. Sarhan, Utilizing deep learning models in CSI-based human activity recognition, Neural Comput. Appl. 34 (8) (Apr. 2022) 5993–6010, https://doi.org/10.1007/s00521-021-06787-w.

[17] T.R. Mim, et al., GRU-INC: an inception-attention based approach using GRU for human activity recognition, Expert Syst. Appl. 216 (Apr. 2023) 119419, https://doi.org/10.1016/j.eswa.2022.119419.

[18] Md M. Islam, S. Nooruddin, F. Karray, G. Muhammad, Multi-level feature fusion for multimodal human activity recognition in Internet of Healthcare Things, Inf. Fusion 94 (Jun. 2023) 17–31, https://doi.org/10.1016/j.inffus.2023.01.015.

[19] J. Li, H. Xu, Y. Wang, Multiresolution fusion convolutional network for open set human activity recognition, IEEE Internet Things J. 10 (13) (Jul. 2023) 11369–11382, https://doi.org/10.1109/JIOT.2023.3243476.

[20] A.M. Helmi, M.A.A. Al-qaness, A. Dahou, M. Abd Elaziz, Human activity recognition using marine predators algorithm with deep learning, Future Generat. Comput. Syst. 142 (May 2023) 340–350, https://doi.org/10.1016/j.future.2023.01.006.

[21] S. Yousefi, H. Narui, S. Dayal, S. Ermon, S. Valaee, A survey on behavior recognition using WiFi Channel State information, IEEE Commun. Mag. 55 (10) (Oct. 2017) 98–104, https://doi.org/10.1109/MCOM.2017.1700082.

[22] T. Tegou, A. Papadopoulos, I. Kalamaras, K. Votis, D. Tzovaras, Using auditory features for WiFi Channel State information activity recognition, SN COMPUT. SCI. 1 (1) (2019) 3, https://doi.org/10.1007/s42979-019-0003-2. Jun.

[23] I.A. Showmik, T.F. Sanam, H. Imtiaz, Human activity recognition from wi-fi CSI data using principal component-based wavelet CNN, Digit. Signal Process. 138 (Jun. 2023) 104056, https://doi.org/10.1016/j.dsp.2023.104056.

[24] N. Dua, S.N. Singh, S.K. Challa, V.B. Semwal, M.L.S. Sai Kumar, A survey on human activity recognition using deep learning techniques and wearable sensor data, in: N. Khare, D.S. Tomar, M.K. Ahirwal, V.B. Semwal, V. Soni (Eds.), Machine Learning, Image Processing, Network Security and Data Sciences, In Communications in Computer and Information Science, Springer Nature Switzerland, Cham, 2022, pp. 52–71, https://doi.org/10.1007/978-3-031-24352-3_5.

[25] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, C.J. Spanos, DeepSense: device-free human activity recognition via autoencoder long-term recurrent convolutional network, in: IEEE International Conference on Communications (ICC), 2018, pp. 1–6, https://doi.org/10.1109/ICC.2018.8422895. May 2018.

[26] L. Guo, et al., Towards CSI-based diversity activity recognition via LSTM-CNN encoder-decoder neural network, Neurocomputing 444 (Jul. 2021) 260–273, https://doi.org/10.1016/j.neucom.2020.02.137.

[27] A. Dahou, M.A.A. Al-qaness, M.A. Elaziz, A.M. Helmi, MLCNNwav: multi-level convolutional neural network with wavelet transformations for sensor-based human activity recognition, IEEE Internet Things J. (–1) (2023) 1, https://doi.org/10.1109/JIOT.2023.3286378.

[28] A. Mihoub, A deep learning-based framework for human activity recognition in Smart Homes, Mobile Inf. Syst. 2021 (2021) e6961343, https://doi.org/10.1155/2021/6961343. Sep.

[29] R. Vrskova, P. Kamencay, R. Hudec, P. Sykora, A new deep-learning method for human activity recognition, Sensors 23 (5) (Jan. 2023) 5, https://doi.org/10.3390/s23052816.

[30] M. Nabati, H. Navidan, R. Shahbazian, S.A. Ghorashi, D. Windridge, Using synthetic data to enhance the accuracy of fingerprint-based localization: a deep learning approach, IEEE Sensors Letters 4 (4) (2020) 1–4, https://doi.org/10.1109/LSENS.2020.2971555. Apr.

[31] S. Mitra, P. Kanungoe, Smartphone based human activity recognition using CNNs and autoencoder features, in: International Conference on Trends in Electronics and Informatics, ICOEI), 2023, pp. 811–819, https://doi.org/10.1109/ICOEI56765.2023.10126051. Apr. 2023.

[32] K. Thapa, Y. Seo, S.-H. Yang, K. Kim, Semi-supervised adversarial auto-encoder to expedite human activity recognition, Sensors 23 (2) (Jan. 2023), https://doi.org/10.3390/s23020683. Art. no. 2.

[33] A.G. Prabono, B.N. Yahya, S.-L. Lee, Atypical sample regularizer autoencoder for cross-domain human activity recognition, Inf. Syst. Front 23 (1) (Feb. 2021) 71–80, https://doi.org/10.1007/s10796-020-09992-5.

[34] A. Rahdar, D. Gharavian, W. Jęśko, Serial weakening of human-based attributes regarding their effect on content-based speech recognition, IEEE Access 11 (2023) 24394–24406, https://doi.org/10.1109/ACCESS.2023.3255982.

[35] F. Luo, E. Bodanese, S. Khan, K. Wu, Spectro-temporal modeling for human activity recognition using a radar sensor network, IEEE Trans. Geosci. Rem. Sens. 61 (2023) 1–13, https://doi.org/10.1109/TGRS.2023.3270365.

[36] X. Cheng, B. Huang, J. Zong, Device-Free human activity recognition based on GMM-HMM using Channel State information, IEEE Access 9 (2021) 76592–76601, https://doi.org/10.1109/ACCESS.2021.3082627.

[37] Y. Fang, F. Xiao, B. Sheng, L. Sha, L. Sun, Cross-scene passive human activity recognition using commodity WiFi, Front. Comput. Sci. 16 (1) (2021) 161502, https://doi.org/10.1007/s11704-021-0407-8. Oct.

[38] S. Mekruksavanich, W. Phaphan, N. Hnoohom, A. Jitpattanakul, Attention-based hybrid deep learning network for human activity recognition using WiFi Channel State information, Appl. Sci. 13 (15) (Jan. 2023), https://doi.org/10.3390/app13158884. Art. no. 15.

[39] G. Lim, B. Oh, D. Kim, K.-A. Toh, Human activity recognition via score level fusion of Wi-Fi CSI signals, Sensors 23 (16) (Jan. 2023), https://doi.org/10.3390/s23167292. Art. no. 16.

[40] O. Custance, S. Khan, S. Parkinson, Classifying participant standing and sitting postures using Channel State information, Electronics 12 (Jan. 2023) 21, https://doi.org/10.3390/electronics12214500. Art. no. 21.

[41] J. Su, Z. Liao, Z. Sheng, A.X. Liu, D. Singh, H.-N. Lee, Human activity recognition using self-powered sensors based on multilayer Bi-directional long short-term memory networks, IEEE Sensor. J. (–1) (2022) 1, https://doi.org/10.1109/JSEN.2022.3195274.

[42] M.H. Kabir, M.H. Rahman, W. Shin, Csi-Ianet, An inception attention network for human-human interaction recognition based on CSI signal,", IEEE Access 9 (2021) 166624–166638, https://doi.org/10.1109/ACCESS.2021.3134794.

[43] N. Hernandez, J. Lundström, J. Favela, I. McChesney, B. Arnrich, Literature review on transfer learning for human activity recognition using mobile and wearable devices with environmental technology, SN COMPUT. SCI. 1 (2) (2020) 66, https://doi.org/10.1007/s42979-020-0070-4. Feb.

[44] A. Ray, M.H. Kolekar, R. Balasubramanian, A. Hafiane, Transfer learning enhanced vision-based human activity recognition: a decade-long analysis, International Journal of Information Management Data Insights 3 (1) (Apr. 2023) 100142, https://doi.org/10.1016/j.jjimei.2022.100142.

[45] O. Pavliuk, M. Mishchuk, C. Strauss, Transfer learning approach for human activity recognition based on continuous wavelet transform, Algorithms 16 (2) (Feb. 2023), https://doi.org/10.3390/a16020077. Art. no. 2.