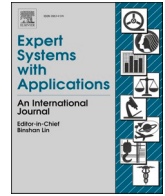


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Expert Systems With Applications

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

WISNet: A deep neural network based human activity recognition system

H. Sharen^{a,1}, L. Jani Anbarasi^{a,2}, P. Rukmani^{a,3}, Amir H. Gandomi^{b,c,*}, R. Neeraja^{a,5}, Modigari Narendra^{a,6}^a School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India^b Faculty of Engineering & IT, University of Technology Sydney, Ultimo, NSW 2007, Australia^c University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary

ARTICLE INFO

Keywords:

Human activity recognition
Deep Learning (DL)
Convolutional neural network
Smartphone
Sensors
Machine learning
LSTM

ABSTRACT

Nowadays, Human Activity Recognition (HAR) is a key research area with many ubiquitous innovative solutions, where both accelerometer and gyroscope data provide information about an observed person's physical activity. HAR offers a diverse variety of important applications, including healthcare, burglary detection, workplace monitoring, and emergency identification. Traditional recognition approaches rely on extracting handmade features from the obtained data to identify the type of human action. Additionally, the efficacy of these works is dependent upon the specific customized features that are chosen. One potential approach to tackle this issue is to utilize Convolutional Neural Networks (CNN) to automatically learn the relevant features. In this paper, we propose a deep learning model, WISNet, a custom 1D-CNN approach to recognize six complex human activities: Jogging, Walking Downstairs, Sitting, Standing, Walking and Climbing Upstairs. The model includes a Convolved Normalized Pooled (CNP_M) Block to generate significant features from the initial layers. An Identity and Basic (IDB_N) Block is incorporated to extract residual progressive features for capturing complex sequential data dependencies. Channel and Spatial attention (CAS_b) Block is integrated with the network to prioritize or minimize essential features based on relative weights. The proposed WISNet model achieved an enhanced accuracy and F1-score of 96.41 % and 0.95 for the HAR dataset by surpassing the existing transfer learning architectures such as Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and Recurrent Neural Network (SimpleRNN). By strategically integrating the CNP_M, IDB_N, and CAS_b blocks, this study aims to tackle distinct challenges encountered in the classification process by enhancing the discernment of features essential for the precise identification of multi-class human activity recognition. The seamless integration of these blocks within the model plays a pivotal role in elevating the overall performance of the WISNet architecture. The work also validates WISNet with similar open-source datasets (UCI-HAR and KU-HAR) and dissimilar open-source datasets (Sleep state detection, Fall detection, and ECG Heartbeat).

1. Introduction

Recognizing human activity is essential to analyze human interactions since it aids in the comprehension of a person's activity and movements. Intuitively, extracting the same information without human

intervention is challenging, because it requires understanding complicated elements such as their psychological state, physiological condition, and other factors. Furthermore, Human Activity Recognition (HAR) can assist surveillance systems in detecting illegal behavior. Although humans are seen to engage in a wide range of activities

* Corresponding author at: Faculty of Engineering & IT, University of Technology Sydney, Ultimo, NSW 2007, Australia.

E-mail addresses: sharen.h2020@vitstudent.ac.in (H. Sharen), janiabarasi.l@vit.ac.in (L. Jani Anbarasi), rukmani.p@vit.ac.in (P. Rukmani), gandomi@uts.edu.au (A.H. Gandomi), neeraja.r2020@vitstudent.ac.in (R. Neeraja), modigari.narendra@vit.ac.in (M. Narendra).¹ ORCID: <https://orcid.org/0000-0002-5969-3396>.² ORCID: <https://orcid.org/0000-0002-8904-2236>.³ ORCID: <https://orcid.org/0000-0002-9494-9422>.⁴ ORCID: <https://orcid.org/0000-0002-2798-0104>.⁵ ORCID: <https://orcid.org/0000-0002-6081-6269>.⁶ ORCID: <https://orcid.org/0000-0003-1852-2803>.<https://doi.org/10.1016/j.eswa.2024.124999>

Received 11 March 2024; Received in revised form 21 July 2024; Accepted 3 August 2024

Available online 22 August 2024

0957-4174/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

throughout their everyday lives, this research work only focuses on identifying a few essential human actions such as running, sitting, walking upstairs and downstairs, and walking. Due to technological improvements and its relevance in various fields such as home automation, telemedicine services, pervasive computing, robotics, computer engineering, physical sciences, health-related issues, natural sciences and industrial academic areas and so on, in recent years, HAR has become a popular area of research.

Concurrent and interleaving goal and activity recognition, as described by [Hu and Yang \(2008\)](#) can be achieved through the utilization of various wireless and sensor networks, either addressing multiple goals concurrently or focusing on a single-goal recognition approach. Similarly, [Tapia et al. \(2004\)](#) introduced the recognition of activities within home settings using a collection of small and simple state-change sensors, aiming to identify activities such as toileting, bathing, and grooming, with detection accuracies ranging from 25 % to 89 %. [Arif and Jalal \(2021\)](#) assessed the estimation and detection of different human body actions across various video and image scenes, utilizing factors such as the position of body portions (head, torso, arms, and legs), as well as size and orientation within the scene, to recognize the actions.

Many of these methods use data from several sensors to detect a single-user's single-activity at a time. Multiple-user multiple-activity, however, exists as a simultaneous and interleaving activity within a single series of events ([Jalal et al., 2012](#)). Due to the high availability of sensors and accelerometers, their low cost and low power consumption, live data streaming, and advancements in computer vision, Machine Learning (ML), Artificial Intelligence (AI), and Internet of Things (IoT), HAR has become one of the most popular study fields.

The purpose of this activity recognition is to use sensor data to recognize and detect simple and complicated behaviors in real-world situations. This is a difficult process, since the data generated by the sensors is sometimes confusing in terms of the activity that is happening. One of the benefits of HAR is that it provides information about a person's behavior, allowing computers to help people with their daily duties more effectively. Recognizing actions, on the other hand, is a challenging endeavor, due to the complexity and uncertainty of human activities, and as a result, only a few approaches ([Kim et al., 2009](#)) deal with sophisticated activity recognition. ML methods can solve activity recognition problems that require user input, a large amount of training data, sequential data, and a complex network. Deep learning, on the other hand, overcomes challenges like scale invariance caused due to different paced frequencies and local dependencies of the nearby signals faced by ML algorithms, resulting in various improvements in these applications ([Thapa et al., 2020](#); [Wang et al., 2017](#)). The aim of HAR is to identify a user's high-level activity, using a variety of sensors and actions. Either logical-based methods or probabilistic approaches can be chosen in this scenario ([Kautz, 1987](#)). Wearable sensor-based activity recognition became popular in the century, following the extensive success of machine ([Ponce et al., 2016](#)) and DL ([Ordóñez & Roggen, 2016](#)) techniques. Data from wearable sensors or body-worn sensors are analyzed using CNN ([Bevilacqua et al., 2018](#)), recurrent neural networks ([Singh et al., 2017](#)), or other DL algorithms to distinguish single, or basic activities.

[Sharma et al. \(2008\)](#) used artificial neural networks (ANN), whereas [Khan \(2013\)](#) included decision trees to identify human activities. Later, k-nearest neighbors (kNN) algorithm was identified as one of the finest classifiers, although it is still unsuccessful to distinguish activities that were highly similar ([Wu et al., 2012](#)). [Ronao and Cho \(2016\)](#) proposed a deep convolutional neural network named convent to perform efficient and effective HAR using smartphone sensors' data. This model performed automatic and data-adaptive techniques to extract elite features from raw data. The convnet generates more relevant and complex features in each subsequent layer, reducing the complexity levels. [Mekruksavanich and Jitpattanukul \(2021\)](#) proposed a framework that utilises wearable devices and DL algorithms to achieve multi-class user

identification. The wearable devices' tri-axial gyroscopes and tri-axial accelerometers were utilised to obtain more detailed information about users throughout various activities. The CNN model had a peak accuracy of 91.77 %, while the LSTM model achieved a higher accuracy of 92.43 %.

[Jia et al. \(2021\)](#) suggested a DL-based technique for HAR, using stepped frequency continuous wave (SFCW) radar. This method also used multi-frequency spectrograms and a ring that includes several parallel convolutional layers, together with a sparse autoencoder to identify and integrate numerous human activity feature maps.

Non-wearable sensors pose intrusion concerns, while vision-based technologies like webcams raise privacy issues and are costly, despite their effectiveness in HAR. Installing such technology in homes raises portability challenges. Therefore, leveraging an AI-based sensor HAR system can offer quick and beneficial services. With ample training data, CNNs have shown significant success in tasks like object recognition and data classification, making them suitable for classifying human activity movements using data from devices equipped with accelerometers and gyroscopes. This study examines the WISDM (Wireless Sensor Data Mining) Smartphone and Smartwatch Activity and Biometrics dataset ([Weiss, 2019](#)), comprising six classes: Jogging, Walking Downstairs, Sitting, Standing, Walking, and Climbing Upstairs, totaling 109,816 instances.

The major contribution of the proposed WISNet architecture is as follows:

- The incorporated identity and basic blocks extract both global and locally optimized progressive features, which are subsequently employed to generate the refined features.
- Integrating attention-based spatial and channel module enables the acquisition and prioritization of salient features at diverse hierarchical levels, with learned attention coefficients assigning greater importance to significant traits for more accurate classification in subsequent layers.
- To acquire precise classification features, the CNN design architecture incorporates CNP_M block by enhancing the weights and learning procedure through batch normalization leading to a reduction in the number of features throughout the depth layers. Consequently, this facilitates accurate and efficient classification of human activities.

The rest of the paper is organized as follows: [Section 2](#) summarizes work related to the field of AI detailing 1D-CNN for Signal Processing, LSTM, GRU, and SimpleRNN, [Section 3](#) presents the proposed WISNet architecture in detail, [Section 4](#) elaborates the experimental analysis and evaluation methods, results obtained from the proposed model for the six class HAR using WISDM dataset, and in the end, [Section 5](#) concludes the work with the future scope.

2. Related work

[Dahou et al. \(2022\)](#) introduced a novel HAR system that combines the Binary Arithmetic Optimization Algorithm (BAOA) with CNN. The CNN is responsible for learning and extracting features from the input data, while the BAOA is utilized to generate the most optimal features. The selected feature was categorized based on distinct activities using the support vector machine (SVM). The HAR model was assessed using three distinct public datasets, namely UCI-HAR, WISDM-HAR, and KU-HAR datasets. The results obtained validate the efficacy of the proposed model, as it achieves competitive performance metrics of 95.23 %, 99.5 %, and 96.8 % for the UCI-HAR, WISDM-HAR, and KU-HAR datasets, respectively. [Xiao et al. \(2021\)](#) developed a learning system known as HARFLS, which allows individual users to effectively and collaboratively manage their activity recognition task while ensuring safety. In this study, a Perceptive Extraction Network (PEN) is employed as the feature extractor for each user. The PEN is designed to identify and extract local characteristics from the HAR data. Additionally, a

Relation Network is utilized, which combines LSTM and an attention mechanism. The primary objective of the Relation Network is to uncover and analyze the global relationships that are concealed within the data. The performance evaluation involves the utilization of four commonly employed datasets, including WISDM, UCI-HAR 2012, OPPORTUNITY, and PAMAP2. In this examination, it is observed that PEN exhibits superior performance compared to the existing methods for HAR.

Athota and Sumathi (2022) introduced a Hybrid Learning Algorithms (HLA) approach for constructing robust classification methods in the field of HAR using data collected from wearable sensors. Convolution Gated Fusion Algorithm (CGFA) and Convolution Memory Fusion Algorithm (CMFA) are used in this method to efficiently capture local features as well as long-term and gated-term dependencies in sequential data. The Amalgam Learning Model was implemented on the WISDM dataset, resulting in the attainment of accuracy rates of 97.76 % and 94.98 % for smartwatch and smartphone devices, respectively, using the CMFA approach. Additionally, the CGFA approach yielded accuracy rates of 96.91 % and 84.35 % for smartwatch and smartphone devices, respectively.

Gao et al. (2021) introduced a novel dual attention technique known as DanHAR. This method combines channel and temporal attention within residual networks, aiming to enhance the capability of feature representation for sensor-based HAR tasks. Extensive experiments were performed on four publicly available HAR datasets, along with a weakly labeled HAR dataset. The results showed relative improvements of 2.02 %, 4.20 %, 1.95 %, 5.22 %, and 5.00 % respectively over regular Convolutional Neural Networks (ConvNets) on the OPPORTUNITY dataset, PAMAP2 dataset, UNIMIB SHAR dataset, WISDM dataset, and the weakly labeled HAR dataset. In their study, Panja et al. (2023) introduced a novel approach to tackle the instance selection problem in smartphone sensing-based human activity recognition (HAR). Their suggested approach uses a hybrid selection and training pipeline that combines evolutionary computing and the closest neighbor principle. The authors of this study have presented a clustering technique, which is afterward followed by an instance selection strategy based on a Genetic technique. The WISDM dataset and UCI-HAR dataset were used for experimentation and evaluation. Experimental findings show that the suggested method successfully reduced the dataset size for the benchmark datasets by about 40 % while preserving a recognition accuracy of over 94 %. The provided statement offers a concise representation of the process of eliminating outliers from a given set of instances.

The Convolution with Self-Attention Network (CSNet) and the Temporal-Channel Convolution with Self-Attention Network (TCCSNet) are two unique frameworks that Essa and Abdelmaksoud (2023) introduced. These frameworks were created to effectively categorize collections of data on human activities that were obtained from various sensor devices. Convolution and self-attention methods are combined by CSNet to effectively capture local and global dependencies present in the input data. On the other hand, TCCSNet employs two separate branches of convolutions and self-attentions to exploit inter-channel and temporal dependencies, enabling the extraction of time-wise and channel-wise information. The evaluation of the suggested approaches encompasses seven distinct datasets for HAR that rely on sensor data. These datasets include WISDM, USC-HAD, WHARF, UTD1, UTD2, PAMAP2, and MHEALTH. The evaluation is conducted using the leave-one-subject-out cross-validation protocol. The experimental results demonstrate that the proposed models exhibit superior performance compared to other contemporary methodologies, including transformers and models based on LSTM.

By allowing a model to continually learn on temporal input using a unique method based on attentive recurrent neural networks called Temporal Teacher Distillation (TTD), Yin et al. (2023) addressed the issue of temporal-based continual learning. TTD addresses the catastrophic forgetting issue by addressing the shortcomings of the current approaches to temporal-based continuous learning. TTD considerably beats state-of-the-art techniques on public datasets like a synthetic

dataset called Split-QuickDraw-100 and Wireless Sensor Data Mining (WISDM) by up to 45.1 % and 14.6 %, respectively, for metrics like forgetting and accuracy. Gupta (2021) investigated DL-based human activity recognition using CNN-GRU, a hybrid deep neural network model that combines convolutional and gated recurrent units for human activity detection. This model's accuracy is suggestively greater than that of other state-of-the-art deep neural network models like Inception Time and DeepConvLSTM created with AutoML, and it was successfully verified on the WISDM dataset.

In order to identify human activities using smartphone sensor data, Kumar et al. (2023) introduced clustering-based DeepTransHAR model and the Cross-Domain Activities Analysis (CDAA). To recognize the target activities, the suggested model used GRUs that automatically derived the useful properties from source sensory activity data. To evaluate the effectiveness of this method, analysis is conducted on the WISDM and KU-HAR benchmark datasets, two freely available test sets. The DeepTransHAR model outperformed the Bi-LSTM, LSTM, and the base RNN models in terms of average accuracy, F1 score, precision, recall, and elapsed average total training time with an average of 86.89 %, 18.30 s, and 55.33 % saved training time. To minimise the overfitting of a single network, Qu et al. (2023) studied a semi-supervised mutual learning technique. Using supervised data from one another, the main and auxiliary networks in this system are collaboratively trained. Second, it is recommended to use the distribution-preserving loss to close the gap between the class distribution of predictions and the labelled data in order to prevent the distribution from deviating. Finally, a context-aware aggregation module adopts the contextual data from the neighbour sequences. This module is able to retrieve more detailed information from a wider variety of sequences. mHealth, PAMAP2, WISDM, and UCI were the four datasets used for the validation. The experimental finding demonstrates that the suggested method outperforms four conventional semi-supervised HAR methods.

The Gated Recurrent Unit-Inception (GRU-INC) model, an Inception-Attention-based strategy that successfully utilizes the spatial and temporal information of the time-series data, was employed by Mim et al. (2023). On publicly accessible datasets, including UCI-HAR, OPPORTUNITY, PAMAP2, WISDM, and Daphnet, the proposed model received an F1-score of 96.27 %, 90.05 %, 90.30 %, 99.12 %, and 95.99 %, respectively. For the model's temporal component, GRU and the Attention Mechanism (AM) were used, and for its spatial component, Inception and the Convolutional Block Attention Module (CBAM) were used. For identifying human activities, Diykh et al. (2023) presented a novel hybrid technique combining adaptive boosting and hierarchical dispersion entropy (HDE) with convolutional neural networks (AdaB_CNN). A sliding window approach is used to segment HAR data into intervals, and the segmented data is then divided into several frequency bands. The dispersion entropy of several frequency bands is calculated to build a feature vector set. Using Joint Approximate Diagonalization of Eigenmatrices (JADE), the obtained features are lowered to further screen out extraneous information. The completed feature vector collection is then used to categorize human activities using the AdaB_CNN technique. Three publicly available datasets, PAMAP2, UCI-HAR 2012, and WISDM, are used to test the suggested methodology. The experimental findings show that the suggested model outperforms the majority of current techniques in HAR.

An attention-based multi-head model for HAR was put forth by Khan and Ahmad (2021). Each of the three compact convolutional heads in this framework was created using a one-dimensional CNN to extract features from sensory data. In order to improve CNN's capacity for representation, a lightweight multi-head model is introduced, enabling automatic selection of salient elements and suppression of unimportant ones. In order to assess the model's performance, ablation studies and experiments were carried out on the WISDM and UCI-HAR benchmark datasets. The experimental result shows how the suggested framework performs well in activity recognition and improves accuracy while maintaining the computing economy. To recognize human activities,

Sekaran et al. (2023) suggested the Lightweight Multiheaded TCN (Light-MHTCN) model of lightweight deep learning. Light-MHTCN uses parallelly organized Convolutional Heads to extract the multiscale features from the inertial sensor signals in order to collect more detailed data. Additionally, preserving longer-term dependency through the integration of dilated causal convolutions and residual connections can improve the performance of the model as a whole. Three well-known smartphone-based HAR databases are used to evaluate the performance of Light-MHTCN: UCI-HAR, WISDM V1, and UniMiB SHAR. On these databases, our lightweight model achieves state-of-the-art performance with recognition accuracies of 96.47 %, 99.98 %, and 98.63 % using just 0.21 million parameters.

DNN-based HAR designs were offered by Suwannarat and Kurdthongmee (2021) as the benchmarks and starting points for developing the candidate architectures. The Real World 2016, UCI-HAR, and the WISDM datasets were used to evaluate the experimental results. The recommended classifiers with optimized settings are advantageous because they require less CPU time and power to process acceleration data when it is received from the sensor. They also lower the memory needs for parameter saving and can be included into wearable technology. Climent-Pérez et al. (2022) suggested that it is still difficult to accurately and automatically assess daily living activities (ADLs) using ML algorithms, in part because there aren't many realistic datasets available to develop and test such algorithms on. 52 participants data with an equal number of males and females are included in the generation of the dataset. The data were gathered over the course of two periods, beginning with 33 people and ending with 19 more. The participants performed 24 distinct ADLs up to five times. First, a description of the dataset that was gathered while wearing the Empatica E4 wrist-worn measuring device. Second, the data collection process and the actual environment in which participants carried out the chosen activities. Finally, a few current and pertinent target applications, including lifelogging, behavioral analysis, and measurement equipment evaluation, where the gathered dataset can be employed.

3. Proposed methodology

In this study, a custom CNN framework called WISNet is proposed to recognize human activity in six different categories, including Jogging, Sitting, Standing, Walking, and Walking upstairs and downstairs. The proposed WISNet HAR system's overall flow is depicted in Fig. 1. The purpose of this work is to provide a computationally straightforward and precise learning model that combines the benefits of skip learning and the attention mechanism to precisely classify the HAR. An analysis of a deep neural network-based prediction system like SimpleRNN, GRU, and LSTM is tested to demonstrate the improved performance of the

proposed WISNet in HAR.

3.1. WISNet architecture

The architecture of the proposed WISNet is shown in Fig. 2. The accuracy of the WISNet model was increased by incorporating residual and skip connections that were fine-tuned by varying the number of stacked filters and filter size. The three phases of this model are Convolved Normalized Pooled (CNP_M) Block: extracts important features from the top layers; Identity and Basic (IDB_N) Block: extracts basic image-level features like edges and progresses to complex sequential data differences, and Channel and Spatial attention (CAS_b) Block: selects essential features with a high weight relative to other features. The learning is enhanced due to optimization between the optimal mapping and dilated block, which resulted O_{CNP_M} ss shown as Eq. (1).

$$O_{CNP_M} = f_{CNP_M}(O_{1DCN}) \quad (1)$$

The input 1D signal processed for classification f_{CNP_M} is the function representing the CNP_M. The input is fed to the 1D-convolutional layer which is a passive layer that receives the raw 1D signal followed by batch normalization, max pooling, and ReLU optimization function.

The IDB_{N1} retrieves the global features and local optimized residual progressive features, where I_N refers to the global and O_{IDB} local features. The implementation is represented as given in Eq. (2).

$$O_{IDB_{N1}} = f_{IDB_{N1}}(O_{CNP_M}) = (I_N, O_{IDB}) \quad (2)$$

Where O_{IDB} and I_N represents the output of residual identity and Basic block and $O_{IDB_{N1}}$ represents the output of IDB_{N1}, respectively. The progressive residual feature is fed to channel and spatial attention block f_{CAS_b} for feature refinement resulting O_{CAS_b} and is given in Eq. (3).

$$O_{CAS_b} = f_{CAS_b}(O_{IDB_{N1}}) \quad (3)$$

This resulting output is summed with the output from the convolution of O_{CAS_b} using element-wise summation resulting in attention-enhanced features $O_{CL_{IDB}}$ as shown in Eq. (4).

$$O_{CL_{IDB}} = O_{CNP_M} + O_{IDB_{N1}} \quad (4)$$

This normalized feature $O_{CL_{IDB}}$ is fed to the IDB_{N2} for generating elite progressive features shown as Eq. (5).

$$O_{IDB_{N2}} = f_{IDB_{N2}}(O_{CL_{IDB}}) \quad (5)$$

This resulting feature is normalized using global average pooling which is fed to fully connected layer incorporating Softmax activation layer resulting in enhanced human activity recognition.

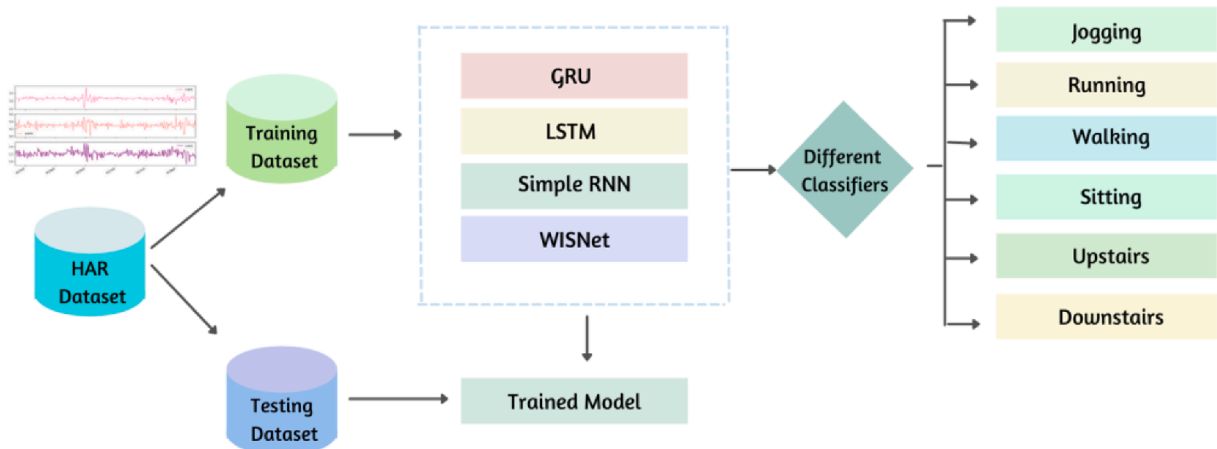


Fig. 1. Overall workflow of the proposed system.

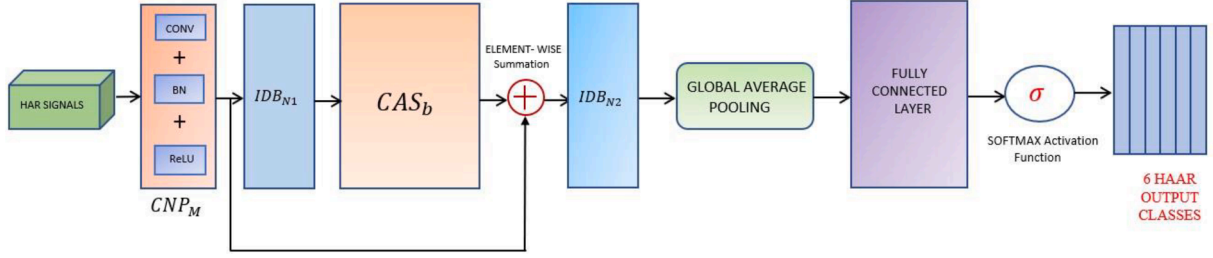


Fig. 2. Architecture of the proposed WISNet.

3.1.1. Convolved normalized pooled (CNP_M) block

In the proposed work the input 1D signal ‘l’ is being fed into the Convolved Normalized Pooled (CNP_M) block adapted with ReLU and batch normalization layer overcomes the vanishing gradient during the initial phase, with a filter size 5. A series of convolutions are performed by the CNN layer, producing features that are then transferred to the activation function, and the subsampling procedure. The 1D filter kernels are chosen with a size of 5 and a sub-sampling factor of 2. The input signal a^l is processed using a kernel w^{l+1} resulting in convolved features $C_n(l)$ as shown in Eq. (6).

$$C_n(l) = (a^l * w^{l+1})(l) \tag{6}$$

The generated features are normalized in-order to speed up the training, resulting in decreasing the initial weight importance w^l and regularizes the model. Each layer of a neural network identifies a unique feature from the previous layer after each gradient update on a batch of data. The data distribution of this input feature map changes dramatically throughout training because the parameters of the preceding layers are adjusted. This significantly affects the training rate and demands the employment of various strategies to select parameter initialization. A typical method for dealing with this shift in internal covariate is batch normalization B_n . Batch normalization is accomplished by implementing a normalizing process that adjusts the means β_n and variances γ_n^2 of the inputs for each layer ‘n’.

The features \widehat{C}_j are normalized to improve the computational speed as shown in Eq. (7) where ϵ an arbitrarily small constant is included for numerical stability. By adjusting the parameters to a range between -3 and 3, batch normalization solves the vanishing/exploding

gradient issue by fitting a maximum likelihood estimate for a normal distribution to the line of channel activations over a batch.

$$\widehat{C}_j = \frac{C_j - \beta_n}{\sqrt{\gamma_n^2 + \epsilon}} \tag{7}$$

This block consists of a skip connection with convolution and a batch normalizing layer supplied with information from a max pooling layer with a 1x7 filter size. The values are scaled and shifted, where σ is learned in the optimization process. The normalized output CO_j is subjected to the IDB_N block referred to in Eq. (8).

$$CO_j = \widehat{C}_j + \sigma \equiv \in B_n(C_n) \tag{8}$$

3.1.2. Identity and Basic (IDB_N) Block

The IDB_N block includes two identity and basic blocks (IDB_{N1}, IDB_{N2}) as shown in Fig. 3. The CNP_M layer CO_j characteristics are fed into the Identity and Basic blocks. The basic block $F_b(CO_j)$ includes 1D convolution layer, max pooling, batch normalization and ReLU layer. One variable filter for each channel is included in the convolution layer, which clearly convolves across the face of the feature map that has been padded to the same size. The ‘Conv’ layer’s filter sizes are sensitive to a narrower region of interest along the block’s line (3 for the first block, 5 for the second, and 7 for the third). $F_b(CO_j)$ performs a set of fed forward progressive feature optimization operations to the next blocks. $F_s(CO_j)$ represents skip connection that includes a set of convolution, max pooling and batch normalization operations, which is summed with $F_b(CO_j)$ features resulting in elite features of O_{BB1} .

The Basic block mainly addresses problems with the loss derivate

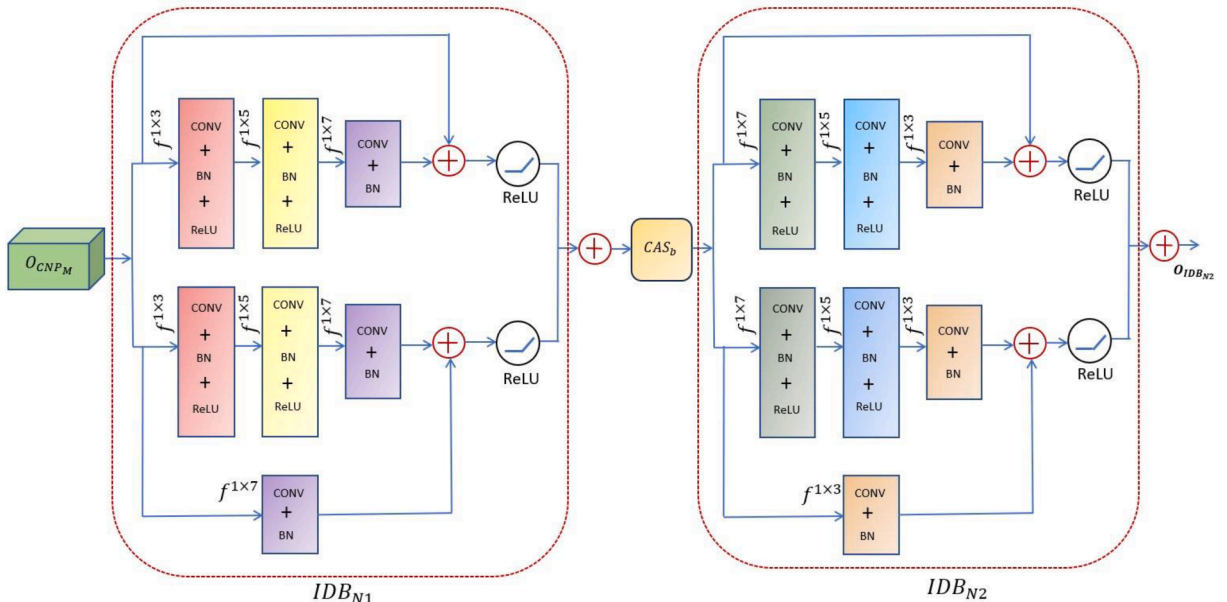


Fig. 3. The representation of Identity and Basic (IDB_N) Block.

approaching towards zero. Similarly, the Identity block $G_{id}(CO_j)$ is similar to a basic block where the skip connections are not included, represented as $G_{id}(CO_j) + CO_j$ resulting in O_{ID1} features. Identity block enhances the spatial and semantic feature maps. The resultant identity and basic block features O_{BB1} and O_{ID1} are separately subjected through Rectified Linear Unit (ReLU) activation function to rectify the convolved features, thereby eliminating any negative values. The ReLU activation function was employed due to its ability to maintain a consistent gradient, even when dealing with greater activation values. This characteristic contributes to the stabilization of the learning process as represented in Eqs. (9) to (11).

$$O_{BB1} = (\mathcal{L}_{BB}(w_n(r^3(\mathcal{L}_{BB}(w_n(r^5(\mathcal{L}_{BB}(w_n(O_j))))))))) \odot O_j r^7 \quad (9)$$

$$O_{ID1} = (\mathcal{L}_{ID}(w_n(r^3(\mathcal{L}_{ID}(w_n(r^5(\mathcal{L}_{ID}(w_n(r^7(O_j)))))))))) \odot (\mathcal{L}_{ID}(w_n(O_j))) r^7 \quad (10)$$

$$O_{IDB_{N1}} = O_{BB1} + O_{ID1} = \{F(O_j, w_n) + O_j\} + \{(O_j, w_n) + w_s O_j\} \quad (11)$$

In contrast, an increasing number of filters are stacked backward in the Identity and Basic block IDB_{N2} to avoid loss derivative squashing. The input to this block is the attention-enhanced features from CAS_b block. Similarly, the convolution kernel sizes of 64, 128, and 256 were incorporated to extract the rich semantic features resulting $O_{IDB_{N1}}$ through exploiting the spatial locality thus encompass all of its distinct components inside a single frame.

3.1.3. Channel and spatial attention (CAS_b) block

The inter-channel and inter-spatial features are learned by the channel and spatial attention block (Fig. 4). After the channel attention map has been generated, the spatial attention is initially computed from the intermediate feature map. Each feature is multiplied using element-wise computation. Global averaging and max-pooling are employed to achieve more excellent feature representation. The feature map (Q_c) $C \times H \times W$ is fed to the global max and average pooling generating inter-channel features. These are then fed to fully connected layers and are concatenated and fed to the sigmoid function, which normalizes the features generated by the channel attention model. The inter-spatial features (Q_s) process the $C \times H \times W$ and generate $1 \times H \times W$ dimension. This is fed to a convolution block of 5 kernel and is normalized through the sigmoid function.

Let $O_{IDB_{N1}} \in R^{C \times H \times W}$ be the input to the channel and spatial attention module used in the CAS_b block, where the feature map's height referred as H, width as W and number of channels as C, respectively. The weights of the channel attention module W_{CA} are expressed in Eq. (12):

$$W_{CA} = \alpha (W_{CAB}^{i+1} (\text{ReLU}[W_{CAB}^i G_{ap}(O_{IDB_{N1}})])) + (W_{CAB}^{i+1} (\text{ReLU}[W_{CAB}^i G_{mp}(O_{IDB_{N1}})])) \quad (12)$$

Where W_{CAB} represents the channel attention weights, α refers to sigmoid activation function, ReLU refers to activation function and W_{CAB}^{i+1} , W_{CAB}^i refers to the weight matrices whose size is defined as $C \times C / x$

and $C/x \times C$ respectively. G_{ap} and G_{mp} are the global average pooling and global max pooling of the channel respectively.

Similarly, spatial attention W_{CS} weight is expressed in the Eq. (13).

$$W_{CS} = \alpha (\text{Con}^{1 \times 7} [G_{ap}(O_{IDB_{N1}}); G_{mp}(O_{IDB_{N1}})]) \quad (13)$$

Here $\text{Con}^{1 \times 7}$ refers to convolution operation with filter size of 1×7 , ‘;’ refers the concatenation of G_{ap} and G_{mp} which is global average pooling and global max pooling. To obtain essential spatial information, the channel spatial and attention layer CAS_b compresses channel ‘C’ into a single channel and spatial dimension $H \times W$ into a single pixel, ignoring cross-dimensional integration information between spatial and channel dimensions.

3.1.4. Output layer

The 1D CNN forward propagation is denoted as Eq. (14):

$$C_m^n = N_m^n + \sum_{j=1}^{z_j-1} \text{conv1D}(K_{we_{jm}^{n-1}}, A_j^{n-1}) \quad (14)$$

Here C_m^n refers to the input, N_m^n is the bias of the m^{th} neuron, A_j^{n-1} output of the j^{th} neuron of layer ‘n’, $K_{we_{jm}^{n-1}}$ is the kernel. conv1D refers convolution without zero-padding. Output arrays A_j^{n-1} dimension is more than the dimension of the input array C_m^n . The intermediate output I_{er} , can be expressed by passing the input C_m^n through the activation function F_{act} .

The back-propagation (BP) algorithm calculates the error from the output layer of the MLP. The error ϵ_v is computed using mean squared error (MSE) for the output layer. The learning factor ρ can be used to update biases N_m^n and weights K_{we} by computing the weight and bias sensitivities as given in 15 and 16:

$$K_{we_{jm}^{n-1}}(ta+1) = K_{we_{jm}^{n-1}}(ta) - \rho \frac{\delta \epsilon_v}{\delta K_{we_{jm}^{n-1}}} \quad (15)$$

$$N_m^n(ta+1) = N_m^n(ta) - \rho \frac{\delta \epsilon_v}{\delta N_m^n} \quad (16)$$

3.1.5. HAR classification loss function

The categorical cross-entropy loss function is applicable for the classification of six classes in the WISDM dataset, namely Climbing Upstairs and Downstairs, Walking, Standing, Sitting, and Jogging. One hot encoding is used to represent the true labels and expected probability associated with each signal. The categorical cross-entropy loss function C_{loss} for 6 human activity classification is shown in Eq. (17).

$$C_{loss}(P_i, \hat{P}_i) = - \sum_{\forall m_i} \sum (P_i, \log(\hat{P}_i)) \quad (17)$$

The difference between actual labels and expected probabilities is measured by the categorical cross-entropy loss, written as $C_{loss}(P_i, \hat{P}_i)$. P_i denotes the one-hot encoded true label for the i^{th} landmark in this

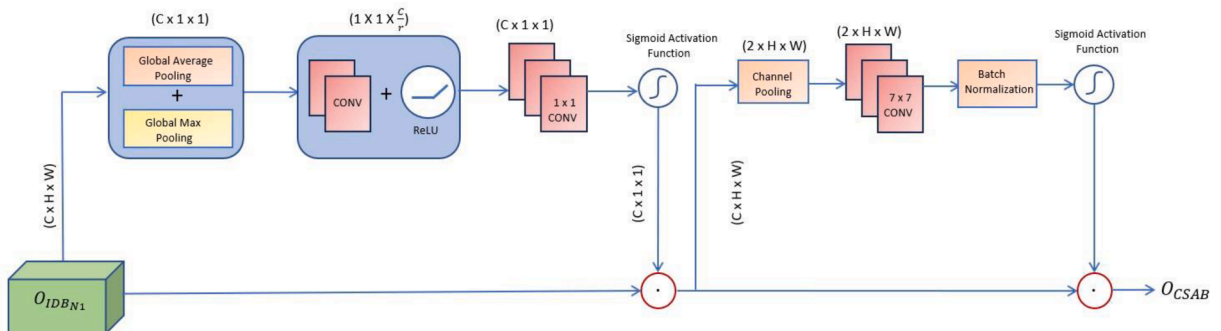


Fig. 4. Channel and Spatial attention (CAS_b) Block.

context, while \widehat{P}_i stands for the predicted probability for the same landmark. Two crucial dimensions are included in this formulation. The losses from all human actions are aggregated in the outer summation, which ranges from $i = 1$ to 6. The inner summation computes the loss for each class within this outer summation while considering the two possible class labels indicating the presence or absence of an activity. Comprehensive loss evaluation is ensured by this nested summing approach.

The loss for each landmark is calculated by multiplying the accurate label P_i by the anticipated probability's logarithm \widehat{P}_i , and using the negative sign to achieve a positive loss value. The objective of the training procedure is to modify the model's parameters to reduce this categorical cross entropy loss. In turn, this improves the model's ability to categorize the six activities in the WISDM dataset accurately by allowing it to assign a larger probability to the precise Human Activity Recognition classes.

4. Result and discussion

The implementation work for the human activity recognition model WISNet is described in this section. The following sections describe the dataset and data description, the division of data for training and testing data, the proposed custom WISNet architecture, transfer learning models like GRU, LSTM, and SimpleRNN results and discussion based on the evaluation metrics. This section also includes a comparison of the performance of the proposed method with the already existing works.

4.1. Experimental setup

Anaconda Navigator, a GUI program, was used to implement the designed WISNet. TensorFlow, an open-source Python framework, is used to train and test the WISNet model on a machine with an NVIDIA GeForce RTX 3050 with 4 GB of GDDR6 Dedicated Graphics and a maximum TGP 95 W VRAM, a 4.6 GHz Intel Core i7 – 11800H CPU, and 32 GB of memory. To recognizing human activity, the Adam optimizer initialized and modified the WISNet settings with 0.001 initial learning rate and 0.0001 wt decay across 30 epochs.

4.2. Experimental signal WISDM 1D dataset description

The WISDM dataset (Weiss, 2019) is used in this research work, which includes six classes: Walking, Upstairs, Standing, Sitting, Jogging and Downstairs. The dataset includes tri-axial accelerometer data samples from 36 volunteers who participated in a predetermined set of activities, leading to a total of 109,816 instances. Table 1 provides a detailed description of the dataset.

During each of the six activities, each volunteer was expected to carry a smartphone in order to collect data for this study. They were also instructed to walk, jog, sit, climb downstairs and upstairs, and do other activities at specific intervals. Furthermore, the supervisor was able to regulate the type of data from various sensors, which included gyroscope, accelerometer, as well as the frequency of data. The sample signal for each class is given in Fig. 5.

Table 1
WISDM dataset description.

Actions	Category	Instances
Walking	0	42,433
Upstairs	1	12,274
Standing	2	4839
Sitting	3	5989
Jogging	4	34,225
Downstairs	5	10,056

4.3. Data splitting

The entire dataset was split into two sets: training and testing. When splitting the data, it was made sure that the data in the testing did not alter the data in the training set, which is accomplished by allocating 80 % of data to training and 20 % to testing. This method is also used to evaluate the model's overall performance throughout training and testing. Before splitting, the shape of the training set was (109816, 50, 3). After splitting into train and test, the shape of the training set was (87852, 50, 3), and the testing set was (21964, 50, 3).

4.4. Hyperparameter tuning

The tuning approach employed the following five hyperparameters: learning rate, epochs, dropout rate, batch size, and optimization units used in the gradient. The dropout factor for the hyperparameter was 30 %, and the number of epochs was 30. The used search space's batch size was 1024, and the chosen gradient optimizers was Adam. In most iterations of the model, a dropout probability of 30 % was found to work well. The number of epochs to be trained was chosen as 30 with a batch size of 1024, resulting in improved results. Adams was the gradient optimizer with the best performance for WISNet. To reduce overfitting and improve the impact of generalization, a 30 % dropout for regularization was used. As stated in our proposal, the WISNet model underwent training with hyperparameters. These included $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 0.001. A channel and a spatial attention block were also incorporated into the model architecture to allow for adaptation to the problem's various levels of complexity. To dynamically enhance the channel-wise characteristics throughout the network, these blocks were selectively used in conjunction with the convolutional and max pooling layers.

4.5. Ablation study

An ablation study was done to evaluate the impact of different modules within the proposed WISNet model, which is intended for the classification of six different human activity classes. Table 2 displays the performance of these network elements: CNP_M , IDB_N , and CAS_b . The WISDM dataset is used to conduct these experiments. The quantitative assessment findings of these various components when applied to the WISDM dataset, are shown in Table 2. The objective of identifying the six different activity classes was to assess the performance of the proposed WISNet classification architecture utilizing important metrics such as Confusion Matrix, ROC Curve, F1-Score, Recall, Precision, and Accuracy.

The core module of our suggested model, CNP_M , had an 87.52 % classification success rate. There are only two convolutional layers in it, which are followed by a max pooling layer. These two convolutional layers can extract complex and abstract features from the input data because they use higher filter sizes and fewer feature maps. As a result, on the testing data, this configuration resulted in a classification rate that resulted in 83.26 %. We incorporated both CNP_M and IDB_N modules to further improve the network's depth and ability to recognize intricate patterns and data linkages. The deeper network architecture produced by this integration increased classification accuracy to 89.49 %. This advantage is justified by the fact that each layer of a neural network is skilled at extracting unique landmark features from the input data, which boosts overall performance. Subsequent layers can then integrate and recombine the features learned in earlier layers to create more abstract representations of the data. Without significantly losing context, IDB_N aids in improving progressive features while also reducing the possibility of overfitting. Regardless of the input size, this effectively equates to sliding a classifier over the input signal and making predictions at each window.

The proposed model's convolutional layers incorporate the CAS_b block, which allows the network to concentrate on the most informative

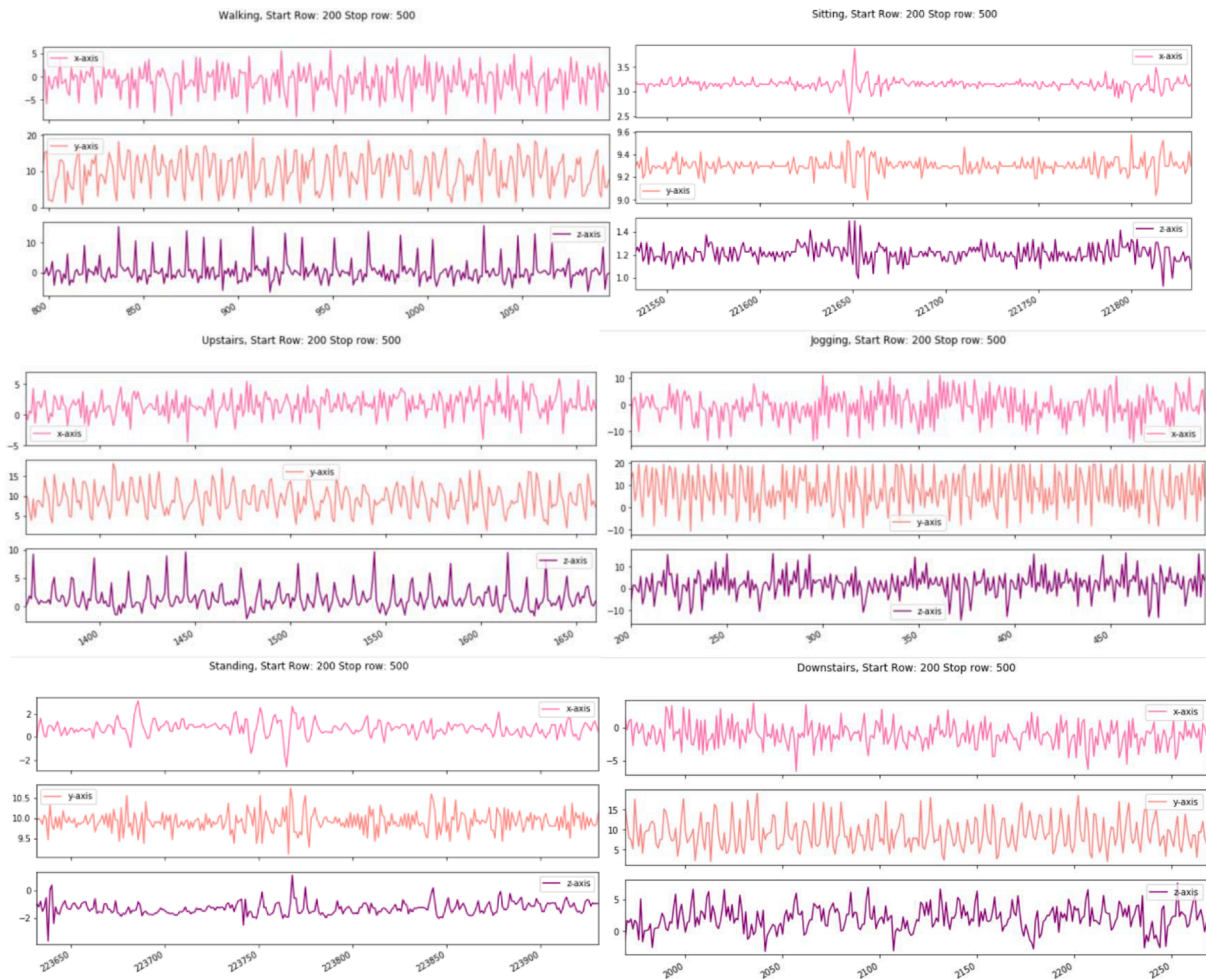


Fig. 5. Sample signal for six classes: Walking, Sitting, Upstairs, Jogging, Standing and Downstairs.

Table 2

The classification accuracy of 6 human activity classes in train and test set for proposed modules.

Proposed Modules	Accuracy: HAR Classification	
	Train Set	Test Set
CNP _M	87.52 %	83.26 %
CNP _M + IDB _N	89.49 %	85.83 %
CNP _M + CAS _b	92.37 %	87.61 %
CNP _M + IDB _N + CAS _b	95.62 %	91.31 %
Proposed	96.41 %	94.52 %

channels and suppress the less informative ones. This improves discriminative power and raises the accuracy and prediction rates of human activities. The classification accuracy improved with the addition of the CAS_b block to CNP_M, reaching 92.37 % and 87.61 %, respectively, for training and testing HAR data. When the CAS_b block was added along with CNP_M and IDB_N the classification accuracy increased significantly to 95.62 % and 91.31 %. To effectively understand the characteristics of the infestation, context-relevant features are effectively recorded by learning attention coefficients for each pixel in the feature map.

The maximum accuracy in HAR categorization was attained by the proposed WISNet, scoring 96.41 %. According to this finding, the performance of HAR classification and the success rate can be greatly enhanced by combining several modules, such as CNP_M, IDB_N, and CAS_b. Several inferences can be made from the results in Table 2. First off, the inclusion of these elements greatly enhances CNN's classification

performance in terms of classification accuracy. Second, it can be inferred from the comparison of the IDB_N and CAS_b modules that the calibration of spatial dimension features by spatial attention is advantageous to more reliable feature selection. Third, the addition of a deep supervision mechanism can help the network be further guided to have better-tuned features for classification.

4.6. Analysis of custom WISNet and transfer learning architecture.

The WISNet model was trained and tested on 109,816 instances of six classes to see how effective it was at categorization. For both the training and validation sets, Table 3 illustrates the accuracy and loss curves for each model. While training accuracy increased with time, it did so at first at a quicker pace. Validation accuracy improved over time; however, it fluctuated during the training. The model obtained a mean validation accuracy of 94.52 % across all the testing processes.

4.6.1. Performance analysis

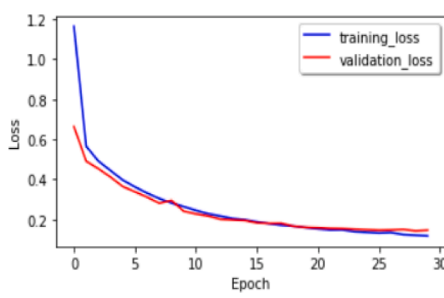
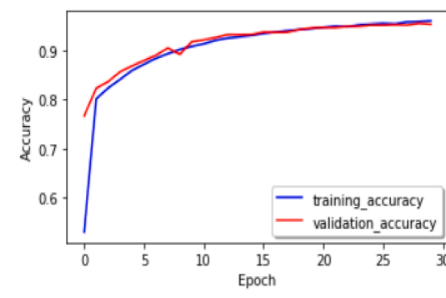
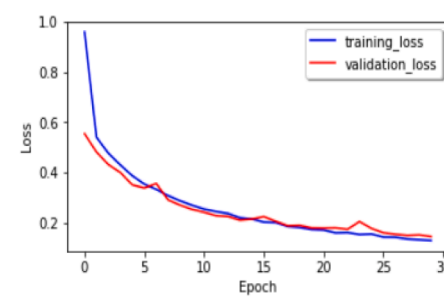
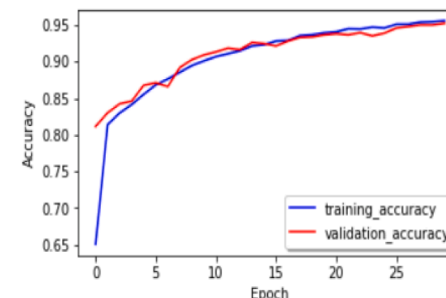
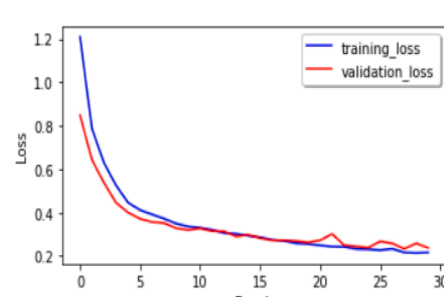
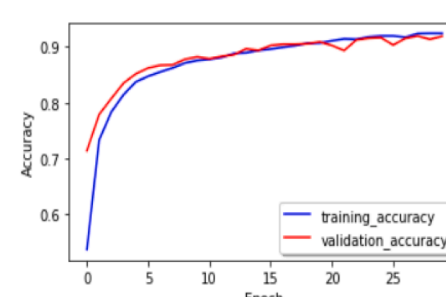
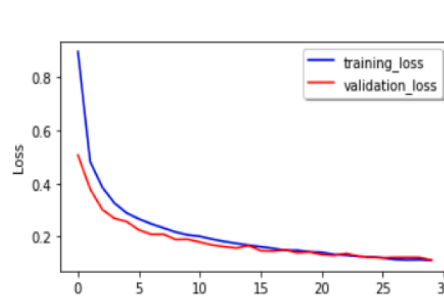
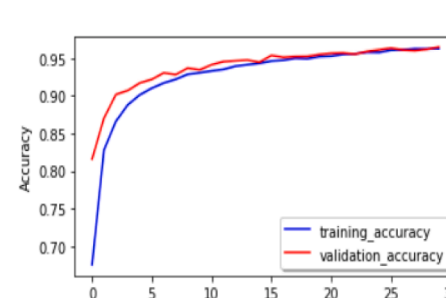
On the training set, custom WISNet had a 96.41 % accuracy, whereas GRU, LSTM, Simple RNN had 95.27 %, 95.15 % and 91.18 % accuracy, respectively. Table 4 presents confusion matrix for the proposed WISNet Architecture. GRU, LSTM, Simple RNN experienced few deviations during validation, where the WISNet performed well in both training and validation.

4.6.2. Effectiveness of the WISNet architecture

Based on the results obtained on the human activities prediction of the six class data of GRU, LSTM, SimpleRNN, and WISNet, the

Table 3

The learning process in terms of the epochs, models loss, and model accuracy curve of GRU, LSTM, Sample RNN and WISNet model.

MODEL	LOSS	ACCURACY
GRU		
LSTM		
SIMPLE RNN		
Proposed WISNet		

evaluation matrix is shown in Table 5. The sensitivity and specificity obtained for WISNet were 0.99 and 0.99, respectively. The training parameters of GRU were 41,906, whereas LSTM resulted with 52,306, SimpleRNN used 21,106 parameters, and the proposed WISNet generated 49,666 parameters. Table 3 describes the learning rate of the DL models and the proposed model used for this study in terms of accuracy and loss plots. Table 4 presents the confusion matrix and ROC curve of all the trained models on the six class signals. A graph that displays how well a classification model works at every level of classification is called a receiver operating characteristic curve (ROC curve). The rate of True Positives and False Positives are two metrics that are plotted on this graph. Attaining higher true positive and true negative rates than other model shows that the proposed WISNet model performs better than other models. The area under the ROC Curve (AUC) of the WISNet model

was found to be 0.999, which is the best compared to other models. AUC is a performance indicator that incorporates all feasible classification limits.

4.6.3. WISNet validation with similar open source dataset

The evaluation and testing of the proposed WISNet encompass a range of HAR tasks. Table 6 delineates the analysis of two distinct HAR datasets utilized in the study for various human recognition tasks. The inclusion of the open-source UCI-HAR and KU-HAR datasets aims to validate WISNet capacity for generalized 1D HAR classification.

The UCI-HAR dataset (Anguita et al., 2013) involved 30 volunteers aged between 19 and 48, engaging in six activities (Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, Laying) while wearing a smartphone (Samsung Galaxy S II) positioned at the waist. This dataset

Table 4
Confusion matrix recorded by the GRU, LSTM, Sample RNN and WISNet model.

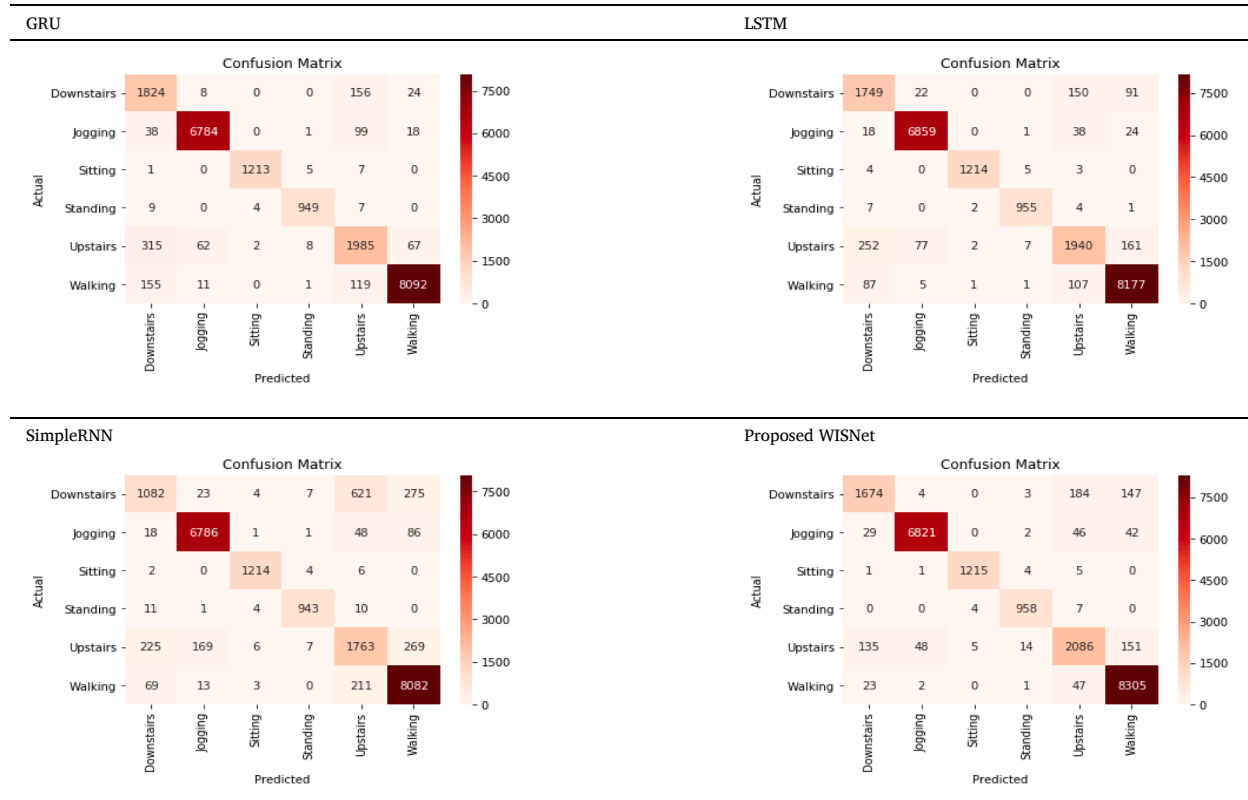


Table 5
Comparison of evaluation matrix of all the models.

Activity	GRU			LSTM			SimpleRNN			WISNet		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Downstairs	0.87	0.82	0.84	0.81	0.88	0.84	0.72	0.64	0.68	0.88	0.88	0.88
Jogging	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.98	0.98	0.99	0.99	0.99
Sitting	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.98	0.99	1.00	0.99	0.99
Standing	0.99	0.98	0.99	0.98	0.99	0.98	0.96	0.98	0.97	0.98	0.99	0.98
Upstairs	0.85	0.83	0.84	0.85	0.82	0.84	0.74	0.78	0.76	0.89	0.87	0.88
walking	0.96	0.99	0.97	0.98	0.97	0.97	0.95	0.96	0.96	0.98	0.99	0.98
Sensitivity	0.995			0.987			0.977			0.993		
Specificity	0.997			0.994			0.993			0.995		
AUC	0.997			0.997			0.993			0.998		
Accuracy (%)	95.27			95.15			91.8			96.41		

Table 6
Parameters of the WISNet model on UCI-HAR and KU-HAR dataset.

Parameters	UCI-HAR				KU-HAR			
	GRU	LSTM	SimpleRNN	WISNet	GRU	LSTM	SimpleRNN	WISNet
Acc (%)	91.82	88.87	90.06	95.66	93.17	90.17	90.78	94.01
Pre (%)	0.92	0.90	0.90	0.96	0.92	0.91	0.88	0.94
Rec (%)	0.92	0.89	0.90	0.96	0.92	0.91	0.88	0.94
Sen (%)	0.94	0.95	0.93	0.99	0.96	0.92	0.89	0.98
Spe (%)	0.95	0.94	0.92	1.0	0.93	0.92	0.90	0.96
F1-Score	0.92	0.89	0.90	0.96	0.93	0.92	0.90	0.94
AUC	0.98	0.93	0.94	0.996	0.95	0.94	0.93	0.99

comprises 10,299 instances, divided into training and testing sets. The training set has the shape of (7352, 564), while the testing set has the shape of (2947, 564).

The KU-HAR dataset (Sikder & Nahid, 2021) was carried over from 90 participants (75 male and 15 female) and encompasses data on 18

distinct activities, recorded through smartphone sensors (Accelerometer and Gyroscope). It comprises 1,945 raw activity samples directly collected from the participants, along with 9,185 subsamples derived from them. The activities include Stand, Sit, Talk-sit, Talk-stand, Stand-sit, Lay, Lay-stand, Pick, Jump, Push-up, Sit-up, Walk, Walk-backward,

Walk-circle, Run, Stair-up, Stair-down, and Table-tennis. This dataset contains 20,750 instances, split into training and testing sets, with the training set having a shape of (14,525, 1,800) and the testing set having a shape of (6,225, 1,800).

In this study, the proposed WISNet, along with established models like GRU, LSTM, and SimpleRNN, underwent training and testing using both the UCI-HAR and KU-HAR datasets. Fig. 6 illustrates the confusion matrix depicting the classification performance across various activities. Table 5 provides a comparison of precision, recall, F1-score, sensitivity, specificity, AUC, and accuracy for the classification outcomes on the UCI-HAR and KU-HAR datasets. WISNet exhibited outstanding performance across various metrics for both the UCI-HAR and KU-HAR datasets. WISNet demonstrated the highest accuracy among the models compared, achieving accuracy rates of 95.66 % for UCI-HAR and 94.01 % for KU-HAR. The GRU model achieved the second-highest accuracy of 91.82 % and 93.17 % for the UCI-HAR and KU-HAR datasets, respectively, trailing behind WISNet and outperforming the LSTM and SimpleRNN models. The obtained precision and recall values of 0.96 and 0.94 for both datasets underscore WISNet’s ability to accurately identify positive cases and minimize false positives. Notably, WISNet showcased remarkable sensitivity and specificity, particularly evident in the UCI-HAR dataset, where it attained a sensitivity of 0.99 and a specificity of 1. Moreover, WISNet outperformed other models in terms of F1-Score, indicating a harmonious balance between precision and recall. Consistently achieving the highest Area Under the Curve (AUC) values, WISNet demonstrated excellent overall performance in classifying activities. These results affirm WISNet’s robustness and effectiveness in human activity recognition tasks, positioning it as a promising model for such applications.

4.6.4. Comparison with state-of-the-art architectures

The proposed WISNet model performance is compared with other state-of-the-art architectures for HAR, as shown in Table 7. Wan et al. (Wan et al., 2020) proposed a CNN model for automatically classifying human activities using UCI-HAR and Pamap2 datasets and attained an accuracy of 92.71 %. Inoue et al. (2018) explored different parameters and architectures of the Deep Recurrent Neural Network (DRNN), utilizing a HASC dataset consisting of 432 trials across six activity classes and achieved recognition rates of 95.42 % and 83.43 %, respectively. Mekruksavanich and Jitpattanakul (2021) analyzed the performance of CNN and LSTM deep learning models in classifying 12 activities. The CNN model attained an accuracy of 91.77 %, while the LSTM model achieved 92.43 % accuracy on both the UCI-HAR and USC_HAD (Zhang

Table 7 Performance comparison with other state of art methods.

Method	Data	Accuracy (%)	Algorithm
Andrey et al. (Ignatov, 2018)	WISDM	93.32 %	CNN
	UCI-HAR	94.35 %	
Wan et al. (Wan et al., 2020)	UCI-HAR, Pamap2	92.71 %	CNN
Kun et al. (Xia et al., 2020)	UCI	95.78 %	LSTM
	WISDM	95.85 %	
	OPPORTUNITY	92.63 %	
Inoue et al. (Inoue et al., 2018)	HASC (Kawaguchi et al., 2011)	95.4	DRNN
Mekruksavanich (Mekruksavanich & Jitpattanakul, 2021)	UCI-HAR	91.78 %	CNN
	USC_HAD	92.43 %	LSTM
Sikder et al. (Sikder, 2019)	UCI-HAR	95.25 %	Multi-channel CNN
Agarwal et al. (Agarwal & Alam, 2020)	Six activities performed by 29 subjects accounted to 1,098,207 samples(Private Data)	95.78 %	Lightweight RNN-LSTM
Akter et al. (Akter, 2023)	UCI-HAR	93.48 %	Attention-Mechanism-based Deep Learning
	WISDM	93.89 %	
	KU-HAR	96.86 %	
Proposed	WISDM	96.41 %	WISNet
	UCI-HAR	95.66 %	
	KU-HAR	94.01 %	

et al., 2012) datasets. Xia et al. (2020) introduced an LSTM model designed to automatically extract activity features and efficiently classify them across UCI, WISDM, and OPPORTUNITY datasets. Their model exhibited remarkable robustness and activity detection capability, achieving high accuracies of 95.78 %, 95.85 %, and 92.63 %, respectively. Ignatov (2018) presented a CNN-based deep learning approach on WISDM and UCI-HAR datasets, yielding accuracies of 93.32 % and 94.35 %, respectively. The proposed WISNet model achieved an accuracy of 96.41 %, 95.66 % and 94.01 %, respectively, for WISDM, UCI-HAR and KU-HAR datasets. Sikder (2019) introduced a two-channel CNN strategy for extracting frequency and power information from raw time-domain accelerometer signals, which achieved 95.25 % accuracy on the UCI-HAR dataset. Agarwal and Alam (2020) created a lightweight RNN-LSTM model that accounted for six activities carried out by 29 individuals with an accuracy of 95.78 %. Akter (2023) used an attention-mechanism-based deep learning model together with feature

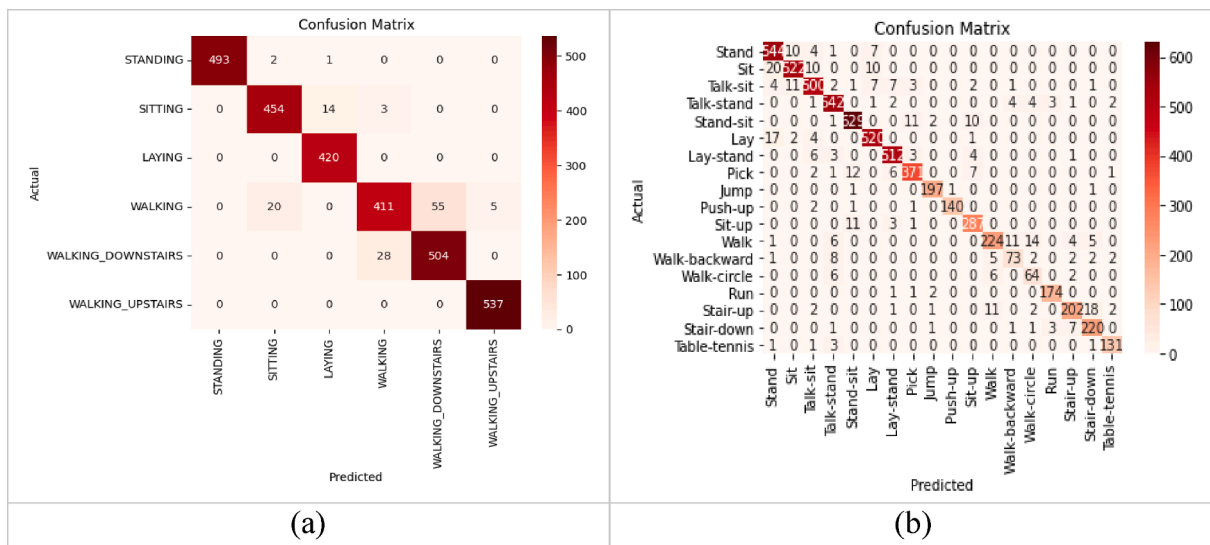


Fig. 6. Confusion Matrix of the WISNet on (a) UCI-HAR and (b) KU-HAR dataset.

combination to achieve accuracies of 93.48 % on UCI-HAR, 93.89 % on WISDM, and 96.86 % on KU-HAR datasets.

4.6.5. WISNet validation with different open source datasets

The proposed WISNet undergoes evaluation and testing across various dissimilar open-source datasets such as the Fall detection dataset, Sleep state detection and ECG Heartbeat dataset. The inclusion of the open-source Fall, Sleep state and ECG Heartbeat datasets serves to validate WISNet's generalization capability in the realm of 1D data classification.

Fall detection datasets (Grimaldi, 2024) are of significant concern in public health, especially among the elderly population, where approximately 30 % of individuals aged over 65 living independently experience a fall annually. The dataset comprises two primary signals: accelerometer and gyroscope, each with three axes (x, y, z), resulting in multidimensional time-series records. The features include fall, lfall (lateral severe fall), rfall (reverse (back) severe fall), light (light fall), sit, step (stairs walk) and walk. This dataset includes 96,800 instances, including train and test sets where the shape of the training set was (7352, 564) and the testing set was (2947, 564).

The Sleep state detection dataset (Esper et al., 2023) was sourced from the Child Mind Institute (CMI) as part of a competition aimed at detecting sleep onset and wake phases. This initiative aims to facilitate more efficient analysis of wrist-worn accelerometer data for sleep monitoring, enabling sleep experts to conduct large-scale studies more easily. By enhancing the understanding of the importance and function of sleep, this dataset contributes to improving overall research in this field. This dataset includes 551,154 instances, which included train and test sets where the shape of the training set was (413,365) and the testing set was (137,789).

ECG Heartbeat Categorization Dataset (Fazeli, 2024; Kachuee et al., 2018) consists of two collections of heartbeat signals obtained from two well-known datasets in heartbeat classification, namely the MIT-BIH Arrhythmia Dataset and The PTB Diagnostic ECG Database. The Arrhythmia dataset included five categories and 109,446 instances, and PTB Diagnostic ECG Database included 14,552 instances of 2 categories.

Table 8 provides a comparison of precision, recall, F1 score, sensitivity, specificity, AUC, and accuracy coefficients for classification results on open-source Fall detection, Sleep state detection and ECG Heartbeat datasets, assessing the performance of WISNet across various tasks. For sleep state detection, the WISNet model achieved an accuracy of 93.96 %, with precision and recall scores of 0.94. The sensitivity and specificity values reported were 0.95 with an F1-score of 0.92 and an AUC of 0.986. In contrast, the WISNet architecture demonstrated higher performance, achieving an accuracy of 97.52 % along with precision and recall scores of 0.98 for the Fall detection dataset. The proposed model exhibited sensitivity and specificity values of 0.95 and 0.88 with an enhanced F1-score of 0.96 and an AUC of 0.99. For the ECG Heartbeat Categorization Dataset, the WISNet model achieved an accuracy of 98.47 %, with precision and recall scores of 0.95 and 0.88. The sensitivity and specificity values reported were 0.95 with F1-score of 0.92 and an AUC of 0.986. These findings highlight the effectiveness of the WISNet architecture across diverse datasets, notably excelling in its classification performance. Fig. 7 illustrates the confusion matrix

Table 8
Comparative study of WISNet with sleep state detection, fall detection.

Parameters	Sleep State Detection	Fall Detection	ECG Heartbeat Categorization Dataset
Acc (%)	93.96	97.52	98.47
Pre (%)	0.94	0.98	0.95
Rec (%)	0.94	0.98	0.88
Sen (%)	0.95	1.00	0.99
Spe (%)	0.95	1.00	0.88
F1-Score	0.92	0.98	0.96
AUC	0.99	0.99	0.99

depicting the classification performance across various activities.

Unlike traditional Deep Convolutional Neural Networks, which typically consist of convolutional layers followed by pooling layers and fully connected layers, this framework integrates specialized blocks tailored for HAR, namely the CNP_M Block, IDB_N Block, and CAS_b offers unique functionalities for feature extraction, information flow facilitation, and dynamic feature focus, respectively, enhancing the model's ability to accurately recognize human activities from sensor data. LSTM and GRU a type of recurrent neural network (RNN), focuses on capturing temporal dependencies and long-range dependencies in sequential data, whereas the proposed framework leverages convolutional operations, batch normalization, and attention mechanisms to extract features and improve discriminative representation in HAR tasks.

5. Conclusion and future work

Smartphones and smartwatches are widely used for activity recognition in different areas of daily life, such as workplace monitoring, emergency identification, and healthcare applications. This study introduces a highly effective methodology for identifying various human activities, such as Jogging, Walking Downstairs, Sitting, Standing, Walking, and Climbing Upstairs. The architecture comprises three essential blocks: the Convolved Normalized Pooled (CNP_M) Block, responsible for extracting noteworthy features from the early layers; the Identity and Basic (IDB_N) Block, designed to extract progressive residual features and adept at capturing intricate sequential data dependencies; and finally, the Channel and Spatial Attention (CAS_b) Block, which prioritizes significant features by assigning them higher weights compared to other features. The incorporation of an attention mechanism into the Identity and Basic block, together with the inclusion of skip connections, enhanced the process of feature learning. The receptive field at a specific layer is expanded to include feature maps from multiple layers of the processing hierarchy allowing the current layer to enrich its input processing with additional contextual information driven by the occurrence of backpropagation of tensors along the skip connections. The proposed network has approximately 700,000 parameters for activity classification, which is significantly fewer than the number of parameters used in previous studies employing LSTM, GRU, and SimpleRNN learning methods. The experimental findings demonstrate that the WISNet architecture obtained an accuracy rate of 96.43 % in accurately distinguishing different human activities. The architecture and optimization strategies of the WISNet model have been meticulously crafted to reduce computational overhead, ensuring efficient multi-class human activity recognition. This research study indicates promising scalability for WISNet architecture through optimized algorithms and processing techniques to manage significant workloads without sacrificing performance. When compared to similar open-source datasets, WISNet exhibited the highest level of accuracy, achieving 95.66 % for UCI-HAR and 94.01 % for KU-HAR. WISNet model achieved an accuracy of 93.96 %, 97.52 % and 98/47 % for Sleep state detection, Fall detection and ECG Heartbeat Categorization dissimilar open source datasets. Thus, it is demonstrated that WISNet could also be utilized for other tasks.

Resilient classification models in human activity recognition can be developed by combining deep learning and ensemble learning techniques. Moreover, the integration of explainable artificial intelligence techniques can provide users with valuable insights, thereby improving their comprehension of the decision-making process in these models and promoting confidence and adoption. Implementing resilient categorization systems customized for distinct activity subcategories in real-world scenarios enables individualized and streamlined approaches, ultimately bolstering the efficacy of activity identification, and leading to improved results for individuals.

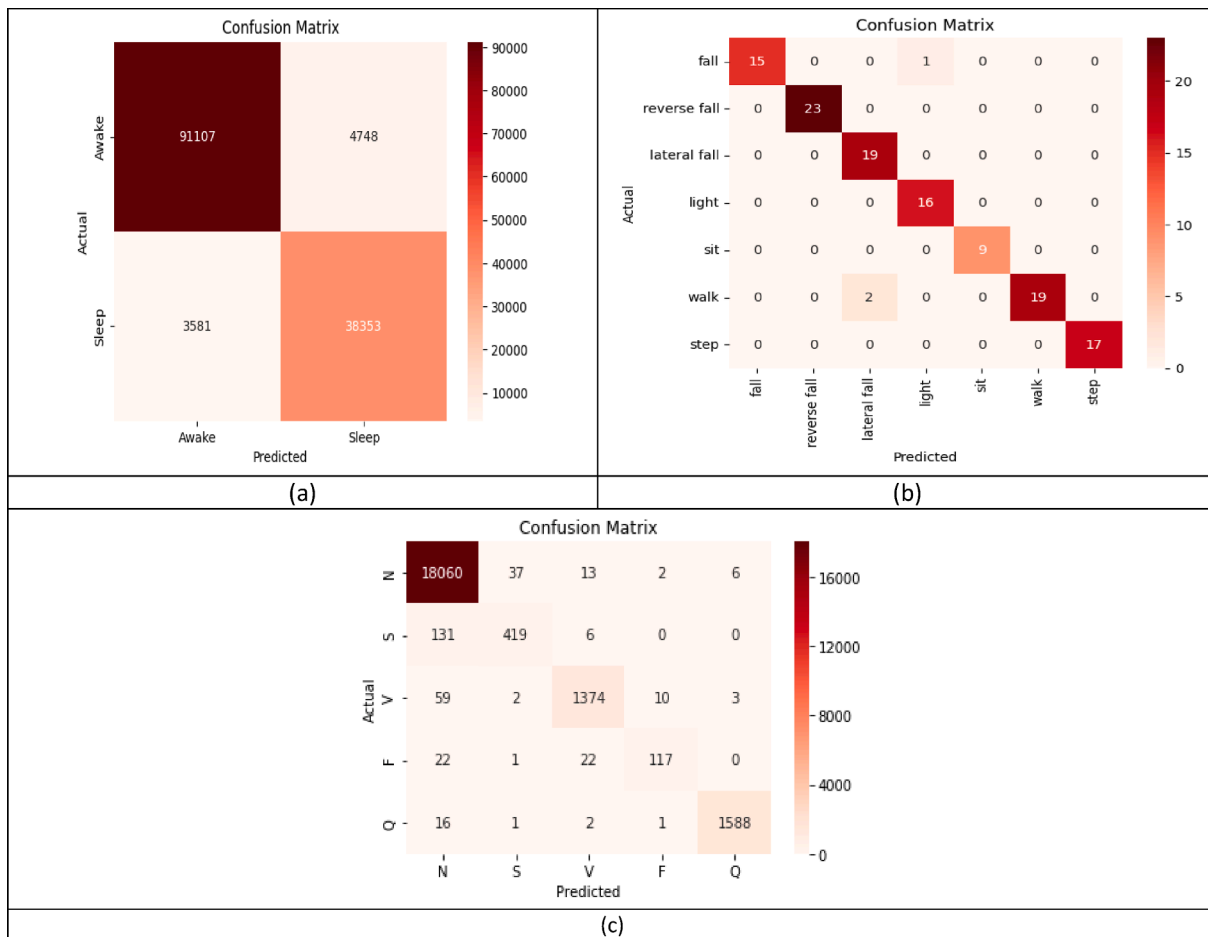


Fig. 7. Confusion Matrix of the WISNet on (a) Sleep State detection (b)Fall detection dataset and (c) ECG Heartbeat Categorization.

6. Compliance with ethical standards

Human participants and/or animals: None.

CRediT authorship contribution statement

H. Sharen: Conceptualization, Data curation, Methodology, Writing – original draft. L. Jani Anbarasi: Data curation, Investigation, Writing – original draft, Supervision. P. Rukmani: Data curation, Writing – original draft. Amir H. Gandomi: Visualization, Writing – review & editing, Supervision. R. Neeraja: Investigation, Methodology, Writing – original draft, Visualization, Validation. Modigari Narendra: Investigation, Methodology, Writing – original draft, Visualization, Validation.

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.

Data availability

Data are borrowed from a reference as cited in the paper.

References

Agarwal, P., & Alam, M. (2020). A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science*, 167, 2364–2373.
 Akter, M., et al. (2023). Human activity recognition using attention-mechanism-based deep learning feature combination. *Sensors*, 23(12), 5715.

Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J. L., et al. (2013). A public domain dataset for human activity recognition using smartphones. *Esann*, 3, 3.
 Arif, A., & Jalal, A. (2021). Automated body parts estimation and detection using salient maps and gaussian matrix model. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCASCAT)* (pp. 667–672). IEEE.
 Athota, R. K., & Sumathi, D. (2022). Human activity recognition based on hybrid learning algorithm for wearable sensor data. *Measurement: Sensors*, 24, Article 100512.
 Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., & Kechadi, T. (2019). Human activity recognition with convolutional neural networks. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part III 18* (pp. 541–552). Springer.
 Climent-Pérez, P., Muñoz-Antón, Á. M., Poli, A., Spinsante, S., & Florez-Reuelta, F. (2022). Dataset of acceleration signals recorded while performing activities of daily living. *Data in Brief*, 41, Article 107896.
 Dahou, A., Al-qaness, M. A., Abd Elaziz, M., & Helmi, A. (2022). Human activity recognition in ioh applications using arithmetic optimization algorithm and deep learning. *Measurement*, 199, Article 111445.
 Diykh, M., Abdulla, S., Deo, R. C., Siuly, S., & Ali, M. (2023). Developing a novel hybrid method based on dispersion entropy and adaptive boosting algorithm for human activity recognition. *Computer Methods and Programs in Biomedicine*, 229, Article 107305.
 Esper, N., Demkin, M., Hoolbrok, R., Kotani, Y., Hunt, L., Leroux, A., van Hees, V., Zipunnikov, V., Merikangas, K., Milham, M., Franco, A., & Kiar, G. (2023). Child mind institute - detect sleep states. <https://www.kaggle.com/competitions/child-mind-institute-detect-sleep-states>, 2023. Accessed Date : 15-05-2024.
 Essa, E., & Abdelmaksoud, I. R. (2023). Temporal-channel convolution with self-attention network for human activity recognition using wearable sensors. *Knowledge-Based Systems*, 278, Article 110867.
 Fazeli, S. (2024). Heartbeat dataset. <https://www.kaggle.com/datasets/shayanfazeli/heartbeat>, Year. Accessed: Date.May 2024.
 Gao, W., Zhang, L., Teng, Q., He, J., & Wu, H. (2021). DanHAR: Dual attention network for multimodal human activity recognition using wearable sensors. *Applied Soft Computing*, 111, Article 107728.
 Grimaldi, E. (2024). Falls vs normal activities dataset. <https://www.kaggle.com/datasets/enricogrimaldi/falls-vs-normal-activities>, Year. Accessed: Date. May 2024.

- Gupta, S. (2021). Deep learning based human activity recognition (har) using wearable sensor data. *International Journal of Information Management Data Insights*, 1(2), Article 100046.
- Hu, D. H., & Yang, Q. (2008). Cigar: Concurrent and interleaving goal and activity recognition. *AAAI*, 8, 1363–1368.
- Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, 62, 915–922.
- Inoue, M., Inoue, S., & Nishida, T. (2018). Deep recurrent neural network for mobile human activity recognition with high throughput. *Artificial Life and Robotics*, 23, 173–185.
- Jalal, A., Kim, J. T., & Kim, T.-S. (2012). Human activity recognition using the labeled depth body parts information of depth silhouettes. In *Proceedings of the 6th international symposium on Sustainable Healthy Buildings* (pp. 1–8).
- Jia, Y., Guo, Y., Wang, G., Song, R., Cui, G., & Zhong, X. (2021). Multifrequency and multi-domain human activity recognition based on sfcw radar using deep learning. *Neurocomputing*, 444, 274–287.
- Kachuee, M., Fazeli, S., & Sarrafzadeh, M. (2018). ECG heartbeat classification: A deep transferable representation. In *2018 IEEE international conference on healthcare informatics (ICHI)* (pp. 443–444). IEEE.
- Kautz, R. L. (1987). Activation energy for thermally induced escape from a basin of attraction. *Physics Letters A*, 125(6–7), 315–319.
- Kawaguchi, N., Ogawa, N., Iwasaki, Y., Kaji, K., Terada, T., Murao, K., Inoue, S., Kawahara, Y., Sumi, Y., & Nishio, N. (2011). Hasc challenge: Gathering large scale human activity corpus for the real-world activity understandings. In *Proceedings of the 2nd augmented human international conference* (pp. 1–5).
- Khan, A. M. (2013). Recognizing physical activities using wii remote. *International Journal of Information and Education Technology*, 3(1), 60.
- Khan, Z. N., & Ahmad, J. (2021). Attention induced multi-head convolutional neural network for human activity recognition. *Applied Soft Computing*, 110, Article 107671.
- Kim, E., Helal, S., & Cook, D. (2009). Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, 9(1), 48–53.
- Kumar, P., Suresh, S., & S. (2023). DeepTransHAR: A novel clustering-based transfer learning approach for recognizing the cross-domain human activities using GRUs (Gated Recurrent Units) Networks'. *Internet of Things*, 21, Article 100681.
- Mekruksavanich, S., & Jitpattanakul, A. (2021). Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models. *Electronics*, 10(3), 308.
- Mim, T. R., Amatullah, M., Afreen, S., Yousuf, M. A., Uddin, S., Alyami, S. A., Hasan, K. F., & Moni, M. A. (2023). GRU-INC: An inception-attention based approach using GRU for human activity recognition. *Expert Systems with Applications*, 216, Article 119419.
- Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 115.
- Panja, A. K., Rayala, A., Agarwala, A., Neogy, S., & Chowdhury, C. (2023). A hybrid tuple selection pipeline for smartphone based human activity recognition. *Expert Systems with Applications*, 217, Article 119536.
- Ponce, H., Martínez-Villasenor, M. D. L., & Miralles-Pechuán, L. (2016). A novel wearable sensor-based human activity recognition approach using artificial hydrocarbon networks. *Sensors*, 16(7), 1033.
- Qu, Y., Tang, Y., Yang, X., Wen, Y., & Zhang, W. (2023). Context-aware mutual learning for semi-supervised human activity recognition using wearable sensors. *Expert Systems with Applications*, 219, Article 119679.
- Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications*, 59, 235–244.
- Sekaran, S. R., Han, P. Y., & Yin, O. S. (2023). Smartphone-based human activity recognition using lightweight multiheaded temporal convolutional network. *Expert Systems with Applications*, 227, Article 120132.
- Sharma, A., Lee, Y.-D., & Chung, W.-Y. (2008). High accuracy human activity monitoring using neural network. In *2008 third international conference on convergence and hybrid information technology* (pp. 430–435). IEEE.
- Sikder, N., et al. (2019). Human activity recognition using multichannel convolutional neural network. *2019 5th International conference on advances in electrical engineering (ICAEE)*. IEEE.
- Sikder, N., & Nahid, A.-A. (2021). KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognition Letters*, 146, 46–54.
- Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., & Holzinger, A. (2017). Human activity recognition using recurrent neural networks. In *Machine learning and knowledge extraction: First IFIP TC 5, WG 8.4, 8.9, 12.9 International Cross-Domain Conference, CD-MAKE 2017, Reggio, Italy, August 29–September 1, 2017, Proceedings 1* (pp. 267–274). Springer.
- Suwannarat, K., & Kurdthongmee, W. (2021). Optimization of deep neural network-based human activity recognition for a wearable device. *Heliyon*, 7(8).
- Tapia, E. M., Intille, S. S., & Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. In *International conference on pervasive computing* (pp. 158–175). Springer.
- Thapa, K., Abdullah Al, Z. M., Lamichhane, B., & Yang, S.-H. (2020). A deep machine learning method for concurrent and interleaved human activity recognition. *Sensors*, 20(20), 5770.
- Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2020). Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25(2), 743–755.
- Wang, J., Hu, F., & Li, L. (2017). Deep bi-directional long short-term memory model for short-term traffic flow prediction. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part V 24* (pp. 306–316). Springer.
- Weiss, G. M. (2019). Wisdm smartphone and smartwatch activity and biometrics dataset. *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set*, 7, 133190–133202.
- Wu, W., Dasgupta, S., Ramirez, E. E., Peterson, C., Norman, G. J., et al. (2012). Classification accuracies of physical activities using smartphone motion sensors. *Journal of Medical Internet Research*, 14(5), e2208.
- Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN architecture for human activity recognition. *IEEE Access*, 8, 56855–56866.
- Xiao, Z., Xu, X., Xing, H., Song, F., Wang, X., & Zhao, B. (2021). A federated learning system with enhanced feature extraction for human activity recognition. *Knowledge-Based Systems*, 229, Article 107338.
- Yin, S.-Y., Huang, Y., Chang, T.-Y., Chang, S.-F., & Tseng, V. S. (2023). Continual learning with attentive recurrent neural networks for temporal data classification. *Neural Networks*, 158, 171–187.
- Zhang, M., & Sawchuk, A. A. (2012). USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 1036–1043).