International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparative Analysis of Various Hybrid Neural Network Models to Determine Human Activities using Inertial Measurement Units

***S. Sowmiya[1], D. Menaka[2]**

Submitted: 25/11/2023 **Revised**: 28/12/2023 **Accepted**: 10/01/2024

Abstract: Human Activity Recognition (HAR) holds a pivotal role in a diverse range of applications that impact various aspects of human life. Advancements in sensor technology and the integration of IoT have expanded the scope of research in HAR through the utilization of deep learning algorithms. End-to-end learning is provided by the advanced deep learning paradigm from complex and amorphous data. Smartphones and IoT wearables are now widely employed in Ambience Assisted Living, e-health monitoring, fitness tracking, biometrics, smart cities, IIoT and other applications. Wearables and Smartphones employ Inertial measurement units (IMU) for the detection of human activities. This research proposes different hybrid neural network model built using GRU, bidirectional GRU, LSTM and bidirectional LSTM with CNN. WISDM, USCHAD, and MHEALTH activity recognition datasets are used to test the method. The hybrid model outperforms the other activity recognition algorithms in terms of accuracy.

Keywords: *Human Activity Recognition , IMU ,Hybrid Deep Neural network,Wearables, Phones, CNN, BiLSTM ,BiGRU*

1. Introduction

Physical activity can provide important details about a person's daily schedule, habits, and mental state. Monitoring how and when people engage in physical activity can yield information that can be used to tailor workout regimens, spot potential health problems, and promote wellbeing. Human activity recognition (HAR) is used in numerous different sectors, including Healthcare, AAL, Psychiatry, Sports and fitness, Entertainment such as online game console, Home automation, Transportation, Fall detection, Industrial applications, Groupware, Biometrics.

Human activity recognition is a technique for recognizing human actions from raw data of activities that has been gathered over a given period. The sequential data can be represented as images, videos, or discrete readings from accelerometers, gyroscopes, magnetometers, and other sensors used by wearables and other portable electronic devices. Smartphones, wearable devices, ambient sensors, device-free systems, object tagged sensors, and video-based systems all have the potential to be employed in the recognition of various human activities. The HAR system frequently makes use of a number of additional sensors, such as sound sensors, motion sensors, proximity sensors and GPS tracking for location tracking. According to a recent review [1], majority of HAR focuses on sensor-based data as opposed to RFID, vision based, and Wi-Fi data. Sensor-based human activity recognition is widely used because it offers a balance between accuracy, non-intrusiveness, real-time monitoring, adaptability, privacy preservation, cost–effective, low power consumption, scalabilty, robustness and the availability of sensors in Smartphones.

Fig 1. Sensors in Smartphone

[1,2]*Noorul Islam Centre for Higher Education, Tamil Nadu, India.,*
[1]*ORCID ID : 0000-0001-6502-4095, Email: nicherde@niuniv.com*
[2]*ORCID ID : 0000-0001-6539-0247,Email:menaka@niuniv.com*

The common sensors used for motion detection ,accelerometer and gyroscope is widely present in recent smartphones. Gyroscope which is used to detect orientation is also present. The various environment sensors like Barometer Sensor,Ambient Light Sensor, Ambient Temperature Sensor, Air Humidity Sensor ,Harmful Radiation Sensor, Proximity Sensor etc are equipped within. The ambient light sensor is responsible for detecting the surrounding light conditions in your environment. It plays a crucial role in the functioning of your screen's automatic brightness adjustment feature. This sensor is designed to capture the ambient light and transmit this data to the operating system. Subsequently, the operating system utilizes this information to dynamically regulate the screen's brightness. The various Application oriented Sensors included are Finger Print Sensor,NFC Sensor,Pedometer Sensor,Compass sensor etc.

Deep neural network technology has also advanced, making it simple to classify data from sensors efficiently, which will likely provide the HAR operations a tremendous boost. There has been a substantial boom in the field of HAR because of the advancements in the field of IoT, which is based upon multiple sensors. HAR has numerous applications in context-aware computing, Human Computer Interaction, surveillance, security and industrial manufacturing. Behavioural analysis using HAR is useful in areas like health monitoring, shopping and security. Personal biometrics obtained from HAR can be utilized for forensic, access control, security and healthcare. Monitoring patient status to support medical diagnoses, promoting quicker healing, and assisting old and chronic patients are a few important applications for HAR in contemporary medical practises. Applications for HAR include video game consoles, sports, and fitness training. The majority of past research in this field has utilised different Machine Learning (ML) techniques and has attained an accuracy rate exceeding 80 %. Machine learning and signal processing techniques have powerful non-linear feature extraction techniques and a shallow architecture that are not adaptive to variations. In controlled situations with few labelled data or little domain expertise needed (such as disease named entity recognition), these simple strategies produce pleasing results. As machine learning systems for HAR tasks traditionally depend on manually designed features derived from domain-specific expertise, their performance is limited by the extent of human knowledge in that domain [2].These methods [3] can only learn superficial aspects from some statistical numbers, which undermines performance and only allows them to identify simple activity. Deep learning (DL) is a subcategory of ML techniques that uses several layers of neural networks in a hierarchical manner to improve pattern recognition, enhance feature learning, and improve accuracy. Improvements in cloud computing and GPU processing power, the automatic extraction of features from enormous datasets, and latest developments in hybrid DL models are three key factors in the popularity of DL. The generative nature of the model is complemented by the addition of a discriminative top layer, allowing it to excel in tasks that require classification. It facilitates learning in an unsupervised manner from a massive amount of unlabeled training data, which is used to extract features at a profound level. Further benefit of DL is that the models simultaneously perform the feature extraction and model building processes. The deep network enables autonomous feature learning without any operator input. Deep learning networks, which can learn a hierarchy of increasingly abstract features and retrieve high-level properties in multiple hidden layer, can be used to recognise complex activities.

Hybrid architectures encompass a wide spectrum of designs, incorporating both parallel and series configurations, and sometimes a combination of the two. Within these architectures, the series learning approach plays a pivotal role in delivering solutions to prognostic problems. . In a recent research endeavor, a novel hybrid model was introduced, integrating Gated Recurrent Units (GRU)[4] in conjunction with Recurrent Neural Networks (RNN) featuring a gating mechanism, aimed at enhancing activity recognition. The GRU block is incorporated to model dynamic shifts in the temporal data so that possible characteristics in time sequence data can be better learned. The processing of spatiotemporal matrices and their mapping into the feature vector are done using the Convolutional Neural Networks (CNN) module. Transfer learning is a subfield within deep neural networks where researchers leverage pre-trained models to reduce training time, enhance learning rates, and improve accuracy by utilizing a strong initial model. This study aims to explore the feasibility of utilizing a both CNN and GRU for the effective recognition of similar pairs of human activities using accelerometer data.

While RNN structures are compatible for capturing temporal dependencies in sequential data, they can face challenges, particularly when it comes to performance and their reliance on available data from restricted situations. As a solution, we propose the development of a hybrid bidirectional model that integrates CNN with Bidirectional Long Short-Term Memory (BiLSTM) and GRU networks for the recognition of Activities of Daily Life (ADL).

The paper's primary contributions and key concepts can be succinctly summarized as follows:

In this research work, a novel Hybrid Deep Neural Network is introduced for recognizing activities using

Smartphones and IoT wearable sensors. This approach showcases the capability to improve the overall performance of research focused on human activity recognition based on inertial measurement units. The fundamental model used in this research is the hybrid model using CNN-BiGRU and CNN-BiLSTM. The CNN within the CNN-BiLSTM architecture is responsible for breaking down multidimensional temporal data into unidimensional version, facilitating the extraction of valuable features from the input data. These extracted features are then passed on to the bidirectional modules of GRU or LSTM. The BiLSTM component within the hybrid model plays a crucial role in learning long term dependencies in either directions from the output of the convolutional layer. By leveraging the strengths of both CNN and BiLSTM, this adopted basic model showcases the potential to significantly enhance the accuracy of activity recognition.The hybrid model including CNN and GRU is also proposed. The advantage of CNN and GRU is utilised .Bidirectional GRU is experimented to prove the power of bidirectional neural networks.

The proposed approach makes use of four models with overlapping windows: CNN-GRU, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU. The use of the sliding window approach improves classifier accuracy overall. Each basic model's hyper-parameters are set adaptively based on the length of the time frame. This will increase the basic model's diversity and make it more relevant to IoT-based real-time applications.

The structure of the rest of this paper is organized as outlined below. Section 2 provides an overview of previous research within this domain. Section 3 delves into the fundamental Human Activity Recognition (HAR) process. In Section 4, we introduce our proposed approach, which is succeeded by an assessment in Section 5 and concluding statements in Section 6.

2. Related Work

Shallow neural networks constitute the foundation for the majority of the previous models for HAR. Numerous shallow learning methods have been applied over time, including Decision Trees (DT), Naive Bayes, Hidden Markov Models (HMM), Random Forest(RF), Support Vector Machines (SVM) and K-Nearest Neighbours (KNN).Majority of the works for HAR using DL utilized CNN[5-7]. Alsheikh et al.'s research, as described in [8], demonstrated the potential application of Deep Belief Networks (DBNs) for activity recognition. RNN and LSTM are also a focus of this field of study. In their work, Jiho Park et al.[9], introduced a Residual Recurrent Neural Network (RNN) model designed for predicting human actions. Because it contains sequence data throughout time, the study made use of the properties of the (RNN) structure. Numerous research based on the GRU based models [10] and CNN-BiGRU, CNN-GRU [4] hybrid models were put out to demonstrate the efficacy of the hybrid model. Additionally, previous research efforts have explored the performance of LSTM and its various adaptations [11],[12],[13], showcasing their capability for robust classification when compared to baseline models. A recent study [14] has introduced a CNN-LSTM-Based Late Sensor Fusion technique, which notably enhances the accuracy in recognizing similar activities. In another work [15], an attention-based BiLSTM was employed to design a Wi-Fi system dedicated to activity recognition. In a separate study, a model grounded in transfer learning, utilizing Gated Recurrent Units (GRUs), was proposed for the identification of collateral, complex, interleaved, and diverse activities of human nature [16] Moreover, a HAR technique by Chen et al. [17] was introduced, centered on the fusion of data from diverse wearable sensors, offering portability and improved accuracy, making it suitable for real-time applications..In a recent work by Abbaspour et al. [18] they four different hybrid models that integrate CNNs with four RNNs. They evaluate model using PAMAP2 dataset and the hybrid models attain an exceptional level of performance.

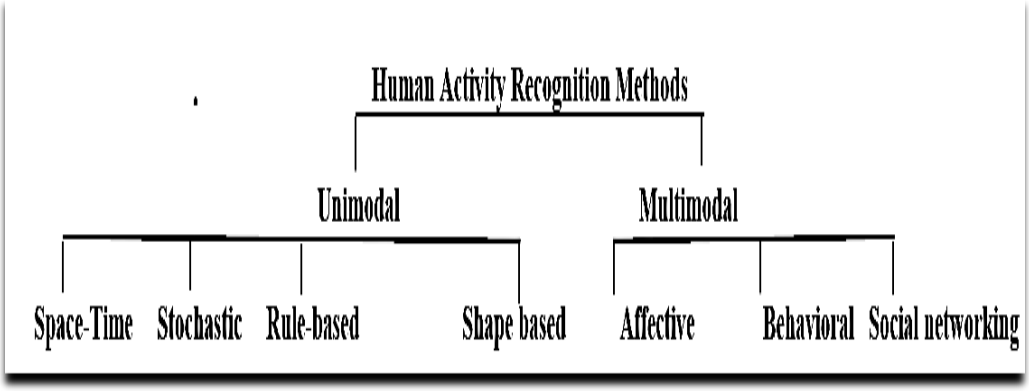

Fig 2. HAR methods

According to the type of data they use, HAR approaches can be divided into two primary divisions based on a hierarchical classification: namely multimodal and unimodal activity recognition systems. The classification of human activity recognition techniques is organized hierarchically, with further subcategories within each of the two primary categories, depending on how they represent human activities. Figure 2 illustrates this hierarchical structure. Unimodal approaches, which are further divided into the following classifications: (a) rule-based (b)space time (c) stochastic (d) shape based methods. Unimodal techniques depict human actions from data of a single modality, like images. Rule based approach involve the use of predefined rules and conditions to analyse the data and infer the activities being performed. Activity recognition techniques that describe human actions in terms of spatiotemporal features or trajectories fall under the category of space-time methods. Stochastic approaches use statistical models to represent human behaviour, such as HMM to identify activities. In HAR, they are effective for recognizing activities that have well-defined temporal patterns and sequential dependencies. Shape based approaches efficiently replicate advanced reasoning processes by simulating the mobility of different human body parts. Multimodal approaches are harnessed across both research and practical domains to attain a comprehensive comprehension of human behavior, emotions, and interactions.

3. HAR Process

The following steps constitute the algorithm for human activity recognition. Acquire the input signals from various sensors in Step 1. Step 2: Some data pre-processing is required as the data is obtained from raw sensor data. Noise removal using filters and segmentation of the sensor data readings is done.Step 3: The whole dataset is divided into train, test and validation data in the appropriate ratio. Step 4: In order to train the models, features are manually or automatically taken from the network. Step 5: Draw conclusions about activity from actual HAR tasks. What actions are meant to be acknowledged? The choice of sensor, where the sensor is placed, and how the data gathering environment is configured will all affect the response to this question. The ability to distinguish more complicated activities increases with the number of sensors included. Depending on the type of activity that has to be detected, the sensors need to be placed optimally. The process of HAR is depicted in Figure 3.

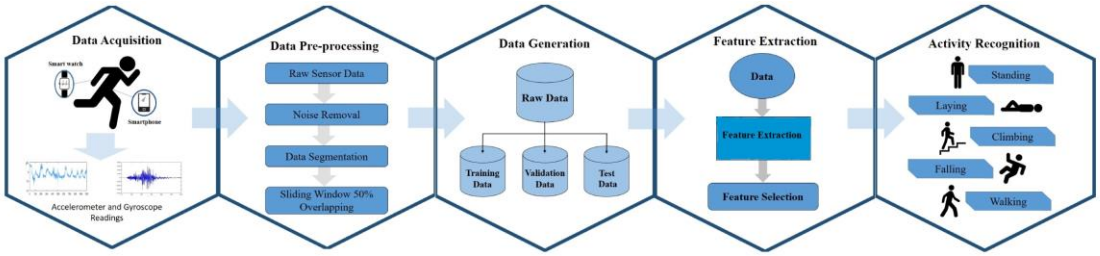

Fig 3. HAR process

In their research, Zhang and Sawchuk (2012) [19] highlighted that the accelerometer stands out as the most effective motion sensor for recognizing a wide range of daily activities. This includes activities such as sitting ,walking, using elevators (both ascending and descending), climbing stairs, and even tasks like brushing teeth. They found that the accelerometer is particularly well-suited for these purposes.

Furthermore, their study indicated that gyroscope measurements, which capture rotation angles, offer enhanced results when it comes to fall detection. The reliability of the recognition process is substantially improved by combining data from both the gyroscope and accelerometer, in contrast to relying on other sensors.

Shoaib et al.'s (2013)[13] research underscores the merits and limitations of various sensors in the context of activity recognition. Their study underscores the importance of selecting the appropriate sensor based on the distinct demands of the activity being identified. Accelerometers, for instance, prove adept at recognizing static activities such as standing, whereas gyroscopes demonstrate a superior capacity for capturing the dynamic nature of actions like stair climbing. In contrast, the magnetometer's reliance on magnetic fields and orientations renders it less suitable for these applications, potentially resulting in complications during classifier training and undermining the overall effectiveness of activity recognition systems..It should be mentioned that the number of sensors used and their distribution over the body will have an impact on classification accuracy. With good classification accuracy, sensors placed on the wrist, knee, waist, and thigh can identify the majority of daily activities. Additionally, it is shown that for the research cited, most of the activities can be identified using a combination of arm and leg positioned readings. According to research investigations by (Bao and Intille, 2004)[21], sensors must be dispersed across the body in order to detect complicated activities, with at least one sensor on the lower body and one on the upper body. The

combination of a smartphone and an IMU sensor, placed at the right ankle along with the deep neural network model, led to a substantial improvement of 23% as opposed to only using smartphone[22].This work utilised CNN and achieved a good F1-score of 96.89 %.

3.1.DataAcquisition

Ambient sensors/Environmental sensors like temperature, humidity sensors or embedded sensors, like smartphones and wearables, are the most often utilised types of sensors for activity identification. Wearable sensors are among them and are effective at sensing activities. The most frequently used sensors include magnetometers, barometers, proximity sensors, accelerometers, and gyroscopes. Smart watches, activity trackers, and smart clothing are examples of wearable technology. The ease of use, portability, continuous monitoring capabilities, and even the ability to recognise complicated activities make wearables appealing for HAR. Several biosensors, in addition to inertial measurement sensors like an accelerometer, gyroscope, or magnetometer, are utilised for activity recognition. Since muscle activity is a crucial component of the majority of human activities and the electromyogram detects the bioelectrical impulses released by muscles, it has previously been widely employed in HAR research. Barometers and other microphone kinds, like piezoelectric (PZT) or airborne, are also included in the devices for HAR.

Ambient sensors are immobile or permanent sensors positioned within spaces where human activities occur. Ambient sensors use a variety of technologies, such as microphones, cameras, IR motion detectors, and pressure mats. They are designed to collect information related to movements, audio patterns, or interactions with objects in the given environment. Importantly, these sensors enable the system to detect and identify activities without requiring individuals to wear specialized devices, typically in confined or specific coverage areas.

Deviceless HAR, often referred to as device-free HAR, tries to identify human activities without forcing people to wear any specific sensors or carry any electronic devices. Rather it makes advantage of already-existing infrastructure to track changes in wireless signals brought on by human movement, such as Wi-Fi, RFID, or radio frequency signals. Algorithms can be used to analyse the fluctuations in these signals to identify routine activities or even more complex ones like dancing or working out. Device-free recognition systems can be noninvasive, but the surroundings may have an impact on its accuracy. CSI is used in device-free HAR to track changes in the wireless network. In a wireless communication system, channel state information (CSI) is knowledge of the wireless channel characteristics between the transmitter and receiver [3]. CSI is used in device-free HAR to track changes in wireless signals brought on by the presence and motion of human bodies in the environment.

The accelerometer is discovered to be more effective than the other sensors out of a variety. The majority of studies have focused on this area since smartphones contain the majority of the sensors including IMU used for activity recognition and because smartphones have become an integral aspect of our daily existence.

3.2. Data Pre-processing

The standardisation and transformation preparation procedures used in HAR are both common. The stages of standardisation are as follows: Relabeling is the process of labelling unknown activities by looking at nearby data. Trimming is used to achieve balance in training. Interpolation is the process of replacing lost data with nearby contiguous observations. Denoising is the process of removing unnecessary components. The process of transformation includes various methods a) Normalisation to a uni-dimensional vector magnitude b) Augmentation of data c) Separation to divide the signal into gravitational and linear components d)Resampling and e)Dimensionality reduction.Activity recognition performance is influenced by the choice of window length, type of window and window overlap percentage. In situations where subject independent cross validation is employed, overlapping sliding windows outperform non-overlapping sliding windows, according to research by (Dehghani et al., 2019) [23]. The accuracy is significantly influenced by the window length selection. A wider window size is required for the recognition of complicated operations, whereas a shorter window enables faster recognition with less resource use. However, each window must contain least ways one occurrence of a recurrent action of the activities in order to distinguish one activity from the others. e.g., stepping forward to run or jog. But employing a large window size does not ensure that performance will improve. Janidarmian et al., (2017)[24]. The study conducted by (Banos et al., 2014)[25] revealed that the maximum level of performance in the recognition of activities is attained when employing small window size, lasting less than two seconds. Moreover, their research indicated that the most precise activity recognition results are achieved using very short windows of 0.25 seconds. Transforming raw sensor data into different domains, such as multi-channel plots and spectrograms allows for a richer representation of the data. This enhanced representation can lead to more accurate and robust recognition of human activities in specific applications. The multi-channel approach has the highest accuracy and requires minimal training. Longer segment lengths result in a slower rate of accuracy improvement. According to the study's findings, accelerometer readings from the shin obtained the peak

accuracy of 90.51%. By the studies of X. Zheng et al., (2018), the forearm and shin data were combined, and a respectable accuracy of 93 % was obtained. [26]

3.3 Feature Extraction

Feature extraction is a crucial stage in the process of converting raw sensor signals into a set of relevant and distinguishing features that are instrumental for classifying activities. For unprocessed sensor data, there are various feature representations available, including features in the frequency and time domains, multichannel spectrums obtained from multi-mode sensor signal channels, spectrums with shallow features, and deep features. The best accuracy is achieved by the multi-channel transformations since the image size is drastically lowered and training time is reduced compared to the raw plot technique. This approach can offer a more comprehensive set of features and offer valuable insights into the interconnections among various sensor measurements.

The study conducted by Ravi et al. [27] highlighted the necessity of utilizing a spectrogram representation for feature extraction. This approach is crucial to capture characteristics that enable the interpretation of intensity variations among adjacent inertial data points. By representing sampling time and frequency invariance within a spectrogram, the researchers achieved accurate and robust data classification, even in the presence of challenges like temporal shifts, changes in sampling rates, and variation in signal loudness. Furthermore, the use of the spectrogram domain proved to be highly successful in filtering out noise from the data.

3.4. Feature Classification

A variety of machine learning approaches were used for the majority of the classification up until recently. Deep neural classifiers are currently in use due to deep learning's advancements. The two most popular deep learning techniques are ensemble methods and hybrid techniques, which include AutoEncoder, CNN, and RNN. Data from sensors has been used in a variety of research studies.

In the field of transfer learning, there are primarily four areas into which the studies fall. Features representation and instance transfer methods are included in the first division. Shallow learning and active learning are examples of transfer learning approaches, which fall under the second group. The third classification is based on the types of learning: transductive ,inductive and unsupervised learning, with the unsupervised group receiving the majority of research. A popular approach for data analysis is to examine the point of transfer in domains across user, environment , sampling rate and task. In circumstances where there is a dearth of labelled data, active learning is utilised to tackle the problem.

The combining of data from a variety of various types of sensors is covered by multi-sensor fusion approaches. The information may consist of attributes that have been retrieved or judgements made based on these characteristics by various categorization methods [28]. This method aims to increase accuracy and inference quality through the usage of many sensors. The benefits of sensor fusion are numerous, including increased resolution, dependability, and robustness against interference. Fusion of multi-sensor is divided into 3 primary categories namely data level or observation level, decision level and feature level fusion, based on the abstraction level at which the data are fused. [29]. The selection of the fusion level depends on factors such as the nature of the sensor data, the particular application, and the inherent characteristics of the sensors being used.

Because the arrangement order of neural networks in an architectural design can profoundly impact the performance of hybrid approaches, substantial research is currently underway in this domain. One major concern with bidirectional networks such as BiLSTM and BiGRU is their computational demands.. Although this effective strategy has another side that involves computing complexity, it yields superior results. GRU represents an enhanced iteration of LSTM, and both models share a common architecture.. While LSTM employs both input and forget gates to manage information flow, GRU simplifies this process by using a reset gate and an update gate. This design difference contributes to GRU's computational efficiency and makes it a competitive alternative to LSTM for various sequential data tasks. The update gate is responsible for determining the amount of data to be transferred from the previous state to the next state, as well as what should be discarded. The reset gate, on the other hand, dictates the extent to which past knowledge should be forgotten. GRU is quicker to train since its architecture is simpler than that of LSTM. On each step, GRU completely uncovers its memory content and uses leaky integration with an adaptive time constant managed by the update gate to preserve a balance between the previous and subsequent memory contents. GRU trains more quickly than LSTM because it employs less training parameters. A type of generative deep learning called BiGRU creates a classifier in the forward direction and a generator in the other. The output layer of this model can simultaneously receive data from past state ,in backward time direction and future state towards forward time direction ,which is another benefit.

4. Approach

Bidirectional Recurrent Neural Networks (BRNN) belong to a class of deep learning architectures that connect 2

deep layers operating in opposing directions to obtain a single output. Architecture for Bidirectional RNN is given in Figure 4. This setup enables them to capture information from both preceding and subsequent states. BRNNs find valuable applications in sequence-to-sequence tasks, including Natural Language Processing(NLP), Speech Recognition and time series prediction.

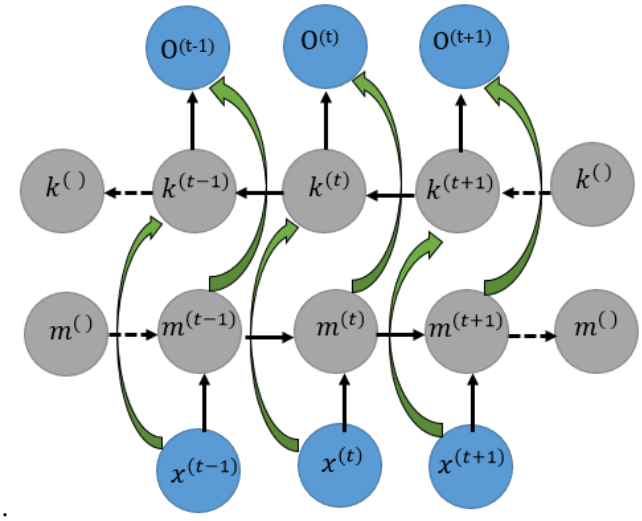

Fig 4. Birectional RNN

While BRNNs can be used for semi supervised or unsupervised learning, they are more commonly used in supervised learning approaches. This is because calculating a reliable probabilistic model for BRNNs can be demanding, particularly when dealing with complex and variable-length input sequences. Bidirectional Recurrent Neural Networks (BRNNs) are engineered to make predictions in both the positive and negative temporal directions simultaneously. In contrast to conventional recurrent neural networks, BRNNs segregate their neurons into two directions: one set for the forward states, representing the positive time direction, and another set for the backward states, denoting the negative time direction. Notably, the output states of each direction are not interconnected with the inputs of the opposite direction. This dual-directionality empowers BRNNs to incorporate input data from both the past and future relative to the current time frame for concurrent output calculation. In contrast, standard recurrent networks need an additional layer to incorporate future information. Therefore, BRNNs are unique because they can incorporate both past and future information without needing an extra layer.

Hybrid neural network models are used in this experiment. Four different hybrid models are proposed i) CNN-GRU ii) CNN-LSTM iii) CNN-BiGRU iv) CNN-BiLSTM.

The full structure of the hybrid model is as shown in Figure 5.Two stages of convolutional layer with following Maxpooling layer is implemented.After the flatten layer it is forwarded to two consecutive Bidirectional GRU or Bidirectional LSTM layer. Final output is then obtained after passing through the dense layer with activation as Softmax. Output comprises of the number of classes of different activity. Adam is the optimiser used with a batch size of 64.Hyperparameters are tuned using Optuna.

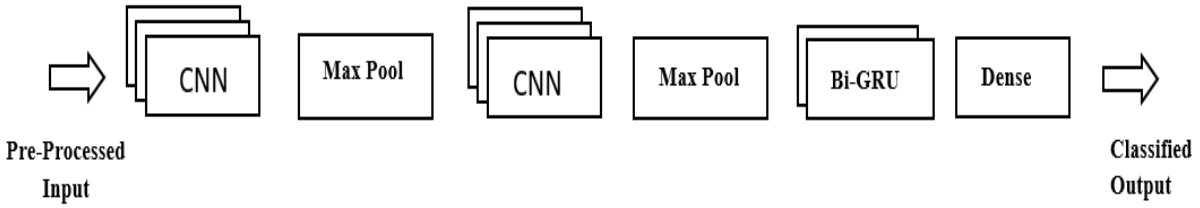

Fig 5 Proposed Hybrid architecture using CNN-BiGRU

4. Experimental Setup

Three datasets that are publicly available consisting of daily life activities recorded using IMU are utilised for this study; USC-HAD [30], WISDM [31] and MHEALTH[32].The MHEALTH is a data collection that includes information from three sensor devices that track physical activity and vital indicators. The dataset makes use of magnetometer, ECG, gyroscope, accelerometer,

and measurements from each. Ten individuals with various backgrounds were given 12 distinct physical tasks to complete. These exercises included standing, sitting, jogging, running, walking, crouching, lying down, cycling, jumping, escalating stairs, bending forward at the waist, and raising the arms in front of one's body. Each individual had an accelerometer attached to their lower right arm, left ankle, and chest, while their right lower arm and left ankle were used to collect magnetometer and gyroscope measurements. Additionally, it offers 2-lead ECG readings for observing the heart from locations where the sensors are placed on the chest. The resulting signals are sampled at the HAR standard sampling rate of 50 Hz. A smartwatch and a smartphone were used to capture everyday activities in the WISDM dataset. Each of the 51 participants in the study carried a Google Nexus or Samsung Galaxy S5 smartphone in their right pocket and wore an LGG smartwatch with a software ,Android Wear version 1.5 on their left wrist. The phone's screen was facing away from the body and was positioned right-side up. Each participant completed 18 different tasks using their smartwatch and smartphone for three consecutive minutes each.A sampling rate of 20 discrete samples in a second using a sliding window approach is used. The sensor data from each device comprises values obtained from both its gyroscope and accelerometer. The target or dependable variable is the code for different activity and the independent or predictor variables are the gyroscope and accelerometer sensor readings in the x, y, and z directions. The 18 actions in WISDM are divided into three groups for better analysis. 1) Ambulation-related actions, such as those conducted without using hands 2) Hand-oriented activities, which include hand-only, non-eating activities 3) Hand-oriented activities, which are made up of eating-related activities. Using 14 participants, 12 different activities from the USC-HAD dataset are chosen and classed as long term, short term, low level, and high level activities.

In order to recognise human actions from data collected using sensors from a smart watch and a smartphone, we combined CNN with the benefits of the Gated Recurrent Unit (GRU) , and its modifications.

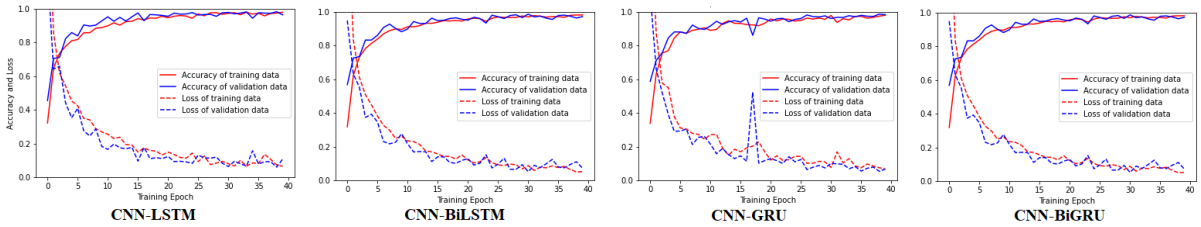

Fig 6. Accuracy-Loss plot for Mhealth using different hybrid models

Proposed algorithm Method

Step 1:- Apply the overlapping sliding window method to preprocess the raw data.

Step 2:-Divide the data set into the following sections namely; train, test and validation sets.

Step 3:- The different deep neural network models are considered and integrate two models to form the hybrid architecture with proper design.

Step 4: -Train and validate the model using the designed neural network.

Step 5:-Determine performance matrices using testing.

Table 1. Evaluation for Mhealth dataset

Model	Parameters	MHealth						
		Stand	Climb	Cycling	Jogg	Run	Knee Bend	Walk
CNN-GRU	Precision	0.953	0.941	1.000	1.000	0.942	0.939	0.981
	Recall	1.000	0.914	0.993	0.938	1.000	0.838	0.982
	F1-score	0.935	0.927	0.996	0.968	0.970	0.886	0.979

CNN-BiGRU	Precision	0.993	0.984	1.000	0.979	0.973	0.992	1.000
	Recall	1.000	0.935	1.000	0.972	0.979	0.946	0.994
	F1-score	0.976	0.959	1.000	0.976	0.976	0.969	0.974
CNN- LSTM	Precision	0.941	0.978	1.000	0.960	0.993	0.971	0.988
	Recall	1.000	0.957	1.000	0.993	0.959	0.926	0.994
	F1-score	0.969	0.967	1.000	0.976	0.975	0.948	0.991
CNN-BiLSTM	Precision	0.953	0.978	1.000	0.986	0.979	0.985	0.977
	Recall	1.000	0.957	1.000	0.979	0.986	0.939	0.994
	F1-score	0.976	0.967	1.000	0.982	0.983	0.962	0.985

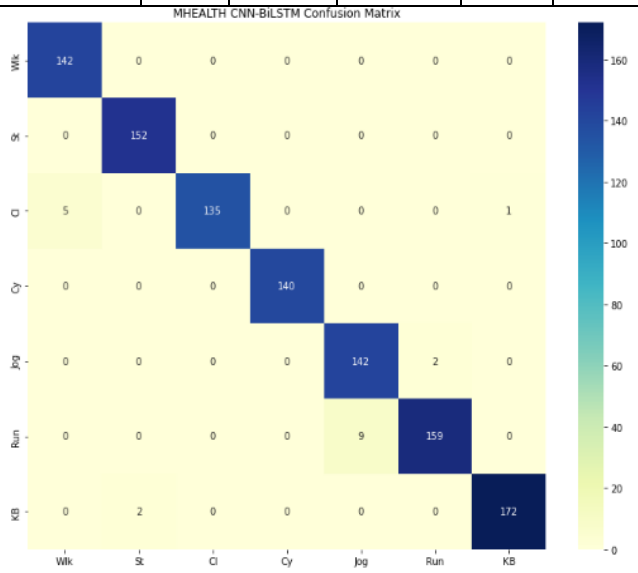

Fig 7. MHEALTH CNN-BiGRU Confusion Matrix

Table 2. Evaluation for WISDM dataset

Model	Parameters	WISDM Smartphone			WISDM SmartWatch		
		AO	HOE	HOG	AO	HOE	HOG
CNN-GRU	Precision	0.916	0.899	0.893	0.967	0.945	0.978
	Recall	0.944	0.859	0.900	0.964	0.971	0.961
	F1-score	0.930	0.879	0.896	0.965	0.958	0.969

CNN-LSTM	Precision	0.912	0.870	0.889	0.971	0.953	0.976
	Recall	0.934	0.852	0.885	0.959	0.972	0.973
	F1-score	0.923	0.861	0.887	0.965	0.962	0.974
CNN-BiGRU	Precision	0.935	0.917	0.917	0.935	0.917	0.917
	Recall	0.957	0.890	0.919	0.957	0.890	0.919
	F1-score	0.946	0.903	0.918	0.946	0.903	0.918
CNN-BiLSTM	Precision	0.934	0.916	0.871	0.975	0.954	0.975
	Recall	0.928	0.854	0.926	0.967	0.961	0.976
	F1-score	0.931	0.884	0.898	0.971	0.958	0.976

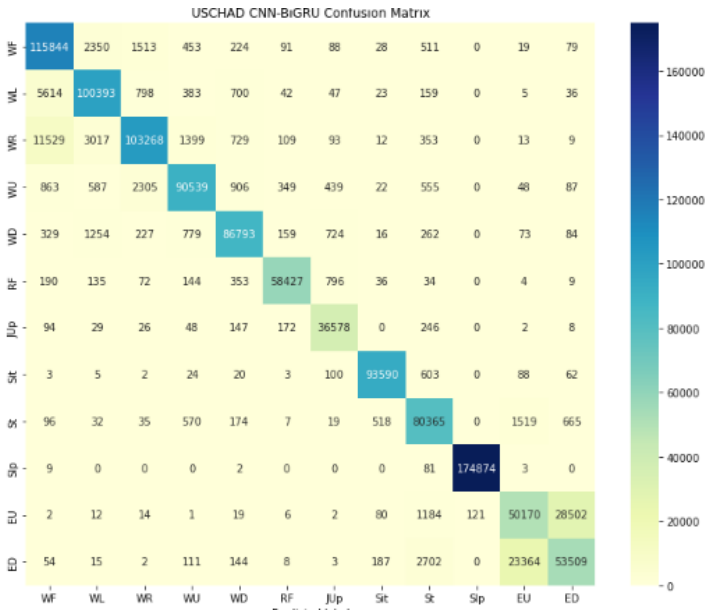

Fig 8. USCHAD CNN-BiGRU Confusion Matrix

In this study, we introduce a novel hybrid model that combines both the 1D Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (BiGRU) models. In this structural design, the CNN layer precedes the GRU layers, with the output from the CNN being further handled by the custom-designed GRU layers.What distinguishes our proposed model is the utilization of bidirectional GRU layers, each comprising two hidden

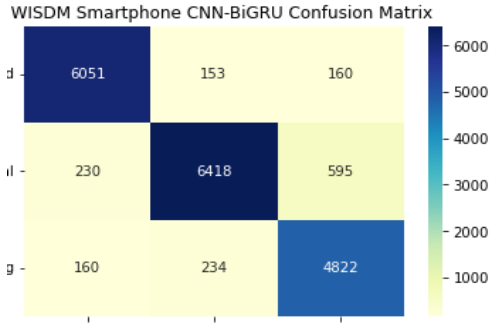

Fig 9. CNN-BiGRU Confusion Matrix WISDM Smartphone

layers operating in opposing directions with respect to the output. The initial segment manages the input in a positive orientation, whereas the following component governs the input in the negative direction. The dataset is split into three sections for model creation and evaluation: 60% training data, 20% validation data, and 20% test data. The

validation data and the training data are mixed together. Additionally, the data is shuffled to make sure that all of the likely occurrences defining the problem are included in the testing, training, and validation sets.

Convolutional layers make up its input. After two stages of CNN, a maxpooling layer is simulated, and then the bidirectional layer follows. ReLu and Softmax layers are applied in the end. The sensor readings are segmented into a window size of 5 seconds with an overlap of 50 % employing the sliding window approach. The smallest window size that still gives the classifier decent performance is 5 seconds, thus that is the value that is used. The dataset is split into three groups—train, validation, and test sets—with a ratio of 6:2:2 to guarantee that all classes are represented in each segment. The experiment uses appropriate batch size for each dataset and a weight decay parameter of 0.0001. During the training process in this study, the optimizer employed is Adam, featuring a learning rate set at 0.001.

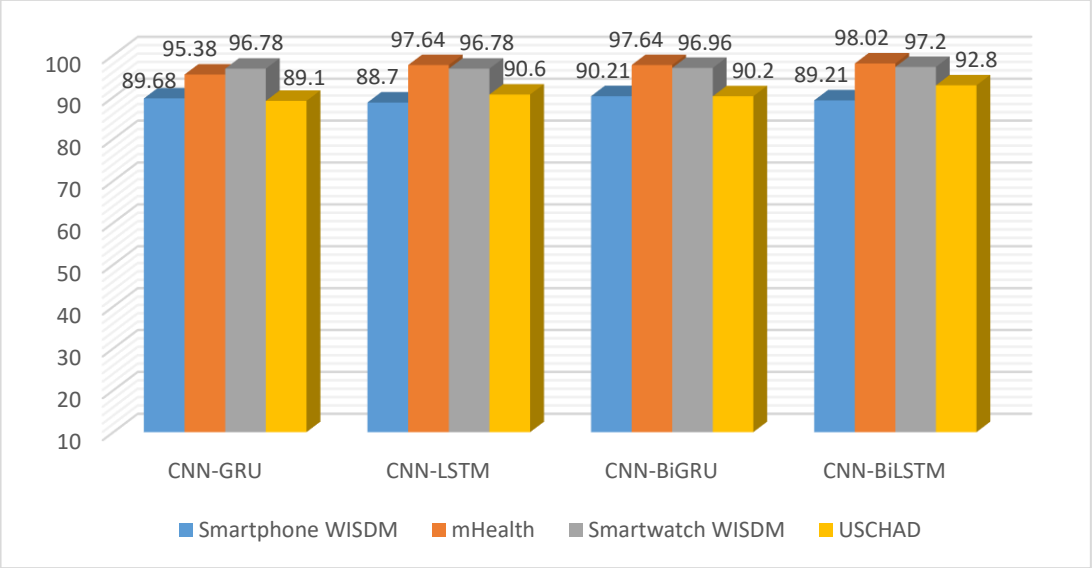

Fig 10. Accuracy graph using various models

The learning rate is initially configured at 0.0001 and subsequently reduced through the utilization of Keras' ReduceLROnPlateau library. The model undergoes training for a total of 40 epochs.. The experiments were conducted within the Google Colab environment, utilizing Keras, and were executed on a GPU equipped with approximately 12 GB of RAM.. The advantage of Colab's GPU processor is utilised, resulting in a quicker completion of the training phase. Various classifiers employing a hybrid DL approach are utilized for data categorization. For classification purposes, four distinct classifiers; CNN-GRU, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU models have been developed.

5. Evaluation

The evaluation of a model is a pivotal step in gauging its effectiveness, reliability and recognising potential issues or areas for improvement. Accuracy is a frequently utilized metric in classification tasks, assessing the general correctness of predictions. However, to gain a more nuanced understanding of a model's performance, it is essential to consider additional metrics such as sensitivity (recall), precision, and the F1-score. These metrics hold particular value in multiclass classification scenarios, and their computation relies on the information furnished by the confusion matrix.

The confusion matrix, a fundamental tool for assessing classification models, plays a central role in the calculation of various performance metrics. These metrics encompass the different parameters; recall (also referred to as sensitivity), specificity, accuracy , precision and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). In addition, confusion matrix helps to identify errors, handling imbalance datasets and hyper tuning. Among the multitude of evaluation metrics, the F1-score is commonly utilized to achieve equilibrium between precision and recall. This equilibrium proves particularly advantageous in scenarios involving imbalanced class distributions.

In this research study, the confusion matrix is constructed and analysed. The confusion matrix is a vital instrument for evaluating and enhancing the performance of classification models, enabling practitioners to make informed decisions regarding model adjustments and refinements. This matrix serves as the foundation for computing essential evaluation parameters. These metrics, when considered collectively, offer a

comprehensive evaluation of the model's classification performance. The mathematical expressions for the metrics mentioned above are in this manner:

$$Accuracy = TP+TN/TP+FN+TN+FP$$

$$Recall = TP/TP+FN$$

$$Precision = TP/TP+FP$$

$$F\text{-}measure = 2*Precision*Recall/Precision+Recall$$

For the three separate data sets, we have included the confusion matrices for the different hybrid models. For the MHEALTH dataset, the parameters F1 score, Precision and Recall are calculated for various hybrid models . The values are shown in Table 1. Table 2 presents the performance details for the parameters F1 score, Recall and Precision for WISDM . In Figure 6, the loss and accuracy plot for the four distinct models is presented for the MHEALTH dataset. The confusion matrix furnishes a comprehensive breakdown of the model's predictions and their correspondence with the actual class labels. The confusion matrices are illustrated for the various hybrid deep neural models in Figures 7, 8, and 11.Figure 10 shows the comparison graph for accuracy for different models, with the help of three distinct data sets; USCHAD, WISDM, and MHEALTH. The activity recognition problem responds best to the bidirectional learning approach. The accuracy of various classification models must be examined and the accuracy is visualised in Figure 9.

6. Conclusion and Future Scope

The aim of this study is to present various hybrid DL models for identifying complicated human behaviours, including CNN-LSTM, CNN-GRU, CNN-BiLSTM, and CNN-BiGRU. This study makes use of the MHEALTH and WISDM datasets. The original WISDM dataset was divided into different datasets for smart watches and smart phones. The algorithm is also verified using the MHEALTH dataset. The preprocessing of the data involved data transformation using the sliding window method. This paper also mentions the benefits of deep neural learning, which has autonomous feature extraction. The train, test, and validation results are used to validate the outcomes. Bidirectional models have a sophisticated architecture and outperform unidirectional models in terms of performance. More complicated models will be investigated for our upcoming studies in addition to the hybrid models utilising CNN, BiGRU, and BiLSTM. Future research endeavors have the potential to delve into the utilization of transformers for temporal data classification. Transformers, being neural networks based on self attention mechanisms, exhibit the ability to comprehend relationships in sequential input and sensor data with heightened precision and swiftness.

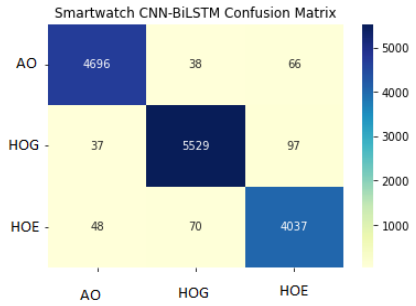

Fig 11. WISDM Smartwatch CNN-BiLSTM Confusion Matrix

7. Declarations

Solid State Ionics' Ethical Statement

I, S. SOWMIYA hereby affirm that the aforementioned conditions are met with regard to the manuscript with the title indicated:

1) The authors' original work, which has never been published before, is the subject matter of this article.

2) The paper accurately and thoroughly reflects the authors' own research and analysis.

3) The meaningful contributions of co-authors and fellow researchers are duly recognized within the work.

4) All sources are correctly cited and fully reported. If a text is copied verbatim, it is acknowledged as such by using quotation marks and providing the appropriate citation.

5) Throughout my effort, I did not knowingly engage in or take part in any type of deliberate injury towards an anyone or an animal.

Conflicts of interest

The authors state that they do not have any recognized financial or interpersonal conflicts that could have been perceived as influencing the research findings presented in this study.

Contributions from authors

The content study was done by SS. All the data were gathered by SS for analysis. The approach suggested by MD was accepted by SS. Based on predetermined stages, SS and MD finished the analysis. Discussions and writing of the conclusions follow the results. Both authors have reviewed and consented to the final manuscript.

Funding

Not applicable

Availability of data and materials

Three publicly accessible datasets were employed for the purpose of this study, namely, MHEALTH, USCHAD and WISDM.

References

[1] Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS. Human activity recognition in artificial intelligence framework: a narrative review. *Artif Intell Rev*. 2022;55(6):4755-4808. doi:10.1007/s10462-021-10116-x

[2] Yang, J.B., Nguyen, M.N., San, P.P., Li, X.L., Krishnaswamy, S., 2015. Deep convolutional neural networks on multichannel time series for human activity recognition, in: IJCAI, Buenos Aires, Argentina, pp. 25–31

[3] Bengio, Y., 2013. Deep learning of representations: Looking forward, in: International Conference on Statistical Language and Speech Processing, Springer. pp. 1–37

[4] M. S. Siraj and M. A. R. Ahad, "A Hybrid Deep Learning Framework using CNN and GRU-based RNN for Recognition of Pairwise Similar Activities," 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2020, pp. 1-7, doi: 10.1109/ICIEVicIVPR48672.2020.9306630.

[5] Morales, F.J.O., Roggen, D., 2016. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations, in: Proceedings of the 2016 ACM International Symposium on Wearable Computers, ACM. pp. 92–99

[6] Ravi D., Wong, C., Lo, B., Yang, G.Z., 2016. Deep learning for human activity recognition: A resource efficient implementation on low-power devices, in: Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on, IEEE. pp. 71–76.

[7] Chen, Y., Xue, Y., 2015. A deep learning approach to human activity recognition based on single accelerometer, in: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE. pp. 1488–1492.

[8] Al-sheikh, M.A., Selim, A., Niyato, D., Doyle, L., Lin, S., Tan, H.P., 2016. Deep activity recognition models with triaxial accelerometers. AAAI workshop .

[9] Park, Jiho et al. "Deep neural networks for activity recognition with multi-sensor data in a smart home." *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (2018): 155-160.

[10] Chen, J.X., Jiang, D., & Zhang, Y.N. (2019). A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification. IEEE Access, 7, 118530-118540.

[11] Murad A, Pyun J-Y. Deep Recurrent Neural Networks for Human Activity Recognition. Sensors. 2017; 17(11):2556. https://doi.org/10.3390/s17112556

[12] L. Al-awneh, B. Mohsen, M. Al-Zinati, A. Shatnawi and M. Al-Ayyoub, "A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition," 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2020, pp. 1-6, doi: 10.1109/PerComWorkshops48775.2020.9156264.

[13] Q. Tao, F. Liu, Y. Li and D. Sidorov, "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU," in IEEE Access, vol. 7, pp. 76690-76698, 2019, doi: 10.1109/ACCESS.2019.2921578.

[14] Zartasha Baloch, Faisal Karim Shaikh, Mukhtiar Ali Unar, "CNN-LSTM-Based Late Sensor Fusion for Human Activity Recognition in Big Data Networks", Wireless Communications and Mobile Computing, vol. 2022, ArticleID 3434100, 16 page s, 2022

[15] Ding, Jianyang & Wang, Yong. (2019). WiFi CSI based Human Activity Recognition Using Deep Recurrent NeuralNetwork. IEEE Access. PP.1-1.10.1109/ACCESS.2019.2956952

[16] Thappa, Kshav & Zubaer, Md & Lamichhane, Barsha & Yang, Sung-Hyun. (2020). A Deep Machine Learning Method for Concurrent and Interleaved Human Activity Recognition. Sensors. 20. 5770. 10.3390/s20205770.

[17] Chen J, Sun Y, Sun S. Improving Human Activity

Recognition Performance by Data Fusion and Feature Engineering. *Sensors*. 2021; 21(3):692. https://doi.org/10.3390/s21030692

[18] Abbaspour S, Fotouhi F, Sedaghatbaf A, Fotouhi H, Vahabi M, Linden M. A Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition. Sensors. 2020; 20(19):5707. https://doi.org/10.3390/s20195707

[19] Mi Zhang and Alexander A. Sawchuk. 2012. USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12). Association for Computing Machinery, New York, NY, USA, 1036–1043. DOI:https://doi.org/10.1145/2370216.2370438

[20] M. Shoaib, H. Scholten and P. J. M. Havinga, "Towards Physical Activity Recognition Using Smartphone Sensors," 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, 2013, pp. 80-87, doi: 10.1109/UIC-ATC.2013.43.

[21] Bao L., Intille S.S. (2004) Activity Recognition from User- Annotated Acceleration Data. In: Ferscha A., Mattern F. (eds) Pervasive Computing. Pervasive 2004. Lecture Notes in Computer Science, vol 3001. Springer, Berlin,Heidelberg. https://doi.org/10.1007/978-3-540-24646-6_1.

[22] Rahn, V. X., Zhou, L., Klieme, E., & Arnrich, B. (2021). Optimal Sensor Placement for Human Activity Recognition with a Minimal Smartphone-IMU Setup. In SENSORNETS (pp. 37-48).

[23] Dehghani, A., Sarbishei, O., Glatard, T. and Shihab, E., 2019. A quantitative comparison of overlapping and non- overlapping sliding windows for human activity recognition using inertial sensors. Sensors, 19(22), p.5026.

[24] Janidarmian, Majid & Roshan Fekr, Atena & Radecka, Katarzyna & Zilic, Zeljko. (2017). A Comprehensive Analysis on Wearable Acceleration Sensors in Human Activity Recognition. Sensors. 17. 529 10.3390/s17030529.

[25] Banos, Oresti & Galvez, Juan & Damas, Miguel & Pomares, Hector & Rojas, Ignacio. (2014). Window Size Impact in Human Activity Recognition. Sensors (Basel, Switzerland). 14. 6474-99 10.3390/s140406474.

[26] Zheng, Xiaochen, Meiqing Wang, and Joaquín Ordieres-Meré. 2018. "Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0" Sensors 18, no. 7: 2146.

[27] https://doi.org/10.3390/s18072146

[28] D. Ravi, C. Wong, B. Lo and G. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2016, pp. 71-76, doi: 10.1109/BSN.2016.7516235

[29] Atrey, Pradeep & Hossain, M. & El Saddik, Abdulmotaleb & Kankanhalli, Mohan. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Syst 16. 345-379.10.1007/s00530-010-0182-0.

[30] Liggins, M.E.; Hall, D.L.; Llinas, J. Handbook of Multisensor Data Fusion:Theory and Practice; CRC Press: Boca Raton, FL, USA, 2009

[31] USC-HAD Dataset. http://sipi.usc.edu/HAD

[32] Weiss, Gary. WISDM Smartphone and Smartwatch Activity and Biometrics Dataset. UCI Machine Learning Repository,2019,https://doi.org/10.24432/C5HK59. http://archive.ics.uci.edu/ml/datasets/mhealth+dataset

[33] Banos,Oresti, Garcia,Rafael, and Saez,Alejandro. (2014). MHEALTH Dataset. UCI Machine Learning Repository. https://doi.org/10.24432/C5TW22.