A Comparative Analysis on Sensor-Based Human Activity Recognition Using Various Deep Learning Techniques

V. Indumathi and S. Prabakeran

Abstract To forecast conditions of action or actions during physical activity, the issue of classifying body gestures and reactions is referred to as human activity recognition (HAR). As the main technique to determine the range of motion, speed, velocity, and magnetic field orientation during these physical exercises, inertial measurement units (IMUs) prevail. Inertial sensors on the body can be used to produce signals tracking body motion and vital signs that can develop models efficiently and identify physical activity correctly. Extreme gradient boosting, multi-layer perceptron, convolutional neural network, and long short-term memory network methods are contrasted in this paper to distinguish human behaviors on the HEALTH datasets. The efficiency of machine learning models is often compared to studies that better fit the multisensory fusion analysis paradigm. The experimental findings of this article on the MHEALTH dataset are strongly promising and reliably outperform current baseline models, comprising of 12 physical activities obtained from four separate inertial sensors. The best efficiency metrics were obtained by MLP and XGBoost with accuracy (92.85%, 90.97%), precision (94.66%, 92.09%), recall (91.59%, 89.99%), and F1-score (92.7%, 90.78%), respectively.

Keywords Human activity recognition · Sensors · IoT · LSTM · CNN · MLP · XGBoost

1 Introduction

Continuous growth in the health sector has led to astronomical advancements in the field of medicine. Due to this continuous growth, the quality of life has greatly increased when compared to one hundred years ago. Everything from life expectancy,

V. Indumathi
Department of Computer Science and Engineering, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India

S. Prabakeran (✉)
Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

physical health, education, safety, and freedom has vastly improved. This rise in health sector growth has led to an increase in healthcare costs. Steadily increasing medical expenses have led to dramatic cost-cutting steps employed by healthcare providers worldwide in tracking patients with chronic disease, monitoring the aged, along with many other cases. New technologies in healthcare will certainly lead to lowering the cost of health care by ensuring that physicians, surgeons, and other medical workers work and perform their everyday tasks in the vicinity of the hospital more effectively.

Choosing the right sensor to match the working environment or application is critical when faced with a human activity recognition challenge. With advanced sensors created each year, the number of different types of sensors to choose from is infinite. Continuous research in the area of behavior identification has driven more firms to try to reap the advantages of forecasting events to boost coordination and efficiency. The numerous types of tasks undertaken for behavior identification are discussed in this section. Body-worn inertial sensors monitor physical exercise activities like riding, jogging, standing still, and powerwalking in related jobs. When these tasks are completed, they generate a specific type of range of body motion, with calculated accelerations that are relatively identical when done by people with different characteristics.

Recognition of human behavior used with powerful technology will theoretically benefit from remote patient control, elderly people, chronic condition patients, and living with environmental assistance. Simple activities such as cycling, running, and jogging have been successfully recognized and classified to date. Complex activities are proving increasingly difficult to monitor, with continuous active research conducted in this area of HAR. The main goal of HAR is to predict common activities in real-life surroundings. Researchers are exploring pattern recognition and human–computer relationships due to their applicability in the real world, such as a human activity recognition healthcare framework. Successfully classifying human activities through wearable sensors generates endless individual information, which provides insight into the individuals' functional ability, lifestyle, and health. In this research, the MHEALTH dataset is analyzed using a variety of deep learning models. These models aim to classify activities performed by volunteers based on data gathered from on-body inertial sensors. The exploratory analysis distinguishes the differences and similarities between these deep learning models throughout this research.

The overall aim is to identify which algorithm best suits the data while discovering which algorithm best classifies each body movement of each person based on vital signs recordings. Two data clustering algorithms analysis identifies relationships between feature attributes and pleasantly visualize the data. Using wearable sensors, human activity recognition (HAR) involves recognizing the physical movements of a subject by analyzing data produced from on-body wearable sensors. Accelerometers, gyroscopes, and magnetometers are these inertial sensors, while the movements are known as everyday living activities (ADL). As described in [1], ADL involves self and body and emphasizes mobility specifically. Due to the applicability of sensor fusion, sensor-based HAR dominates the current study, which requires the incorporation of sensor data from multiple sensors, which drives analytical results in terms of

reliability, accuracy, and completeness. Deep learning methods continue to continuously advance and strengthen the HAR area in this regard. With its in-depth expertise and analytical capacity, XGBoost leads the way in taking data-oriented classification tasks and selecting and processing invaluable features from the data effectively. There are four deep learning (DL) models, which are applied to the HAR problem in this article. Using on-body sensor signal data created from four different sensors, several models were developed, trained, and analyzed for the results to identify which model best fits the data in terms of precision, accuracy, recall, F-score, and the total number of misclassified instances.

This article demonstrates its ability to perform parallel optimization and tree pruning while restricting overfitting and constantly learning sparse features, XGBoost is the highest performing model. Section 2 provides a summary of the work on the recognition of human behavior. The remainder of this paper is structured the following. Section 3 offers a description of the MHEALTH dataset, the design, and the research methodology. Section 4 addresses experiment performance. Finally, in Sect. 5, the conclusion and the future scope of the research work were presented.

2 Related Work

Recognition of human behavior using wearable or mobile sensors supports a variety of applications such as health care, exercise, smart home, etc. For example, medical teams may track elderly people's health conditions based on information about their activity. Daily energy expenditure was estimated and provided with good advice according to the level of operation of the users. [2]. The growth of smartphones and the increase of access to technologies like the availability of high-speed internet and network infrastructure have significantly changed the lives of people. Today, several smartphones contain a range of powerful sensors, including orientation, location, network, and direction sensors. Specifically, motion or inertial sensors (e.g., accelerometers) were commonly used to detect the physical movements of the users. [2]. The sensor is a device that senses and obtains the changes that occurred in the environment and redirects the collected information to the operating system. Smartphone sensors are categorized into three major groups called motion, environmental, and position sensors. Motion sensors use axis-based sensing approach to finding the measurement. Environmental parameters are measured by environmental sensors, for example, temperature, humidity, and light. Position sensors are used for measuring the distance of the reference position. Widely used smartphone sensors are,[3]

1. Accelerometer: It detects variations in smartphone orientation concerning the x, y, and z-axis.
2. Ambient light sensor: It senses the light density of the environment. Auto-brightness adjustment in the mobile phone is the best example of this type of sensor.

3. Barometer sensor: Atmospheric pressure is sensed by the barometer, this sensor assists the GPS to track the location in an efficient manner.
4. Gyroscope sensor: It finds the axis-based motion along with angular rotation so the clean data can be obtained.

Human activity recognition can be achieved through a video-based or sensor-based approach. Video-based HAR investigates videos or photographs that include human movements, while sensor-based HAR focuses on movement data from smart sensors such as an accelerometer gyroscope, Bluetooth, sound sensors, etc. HAR approaches are generalized into a certain type of body-worn sensor, object sensor, ambient sensor, and hybrid sensor. Physical activities that are performed by users can be detected by using body-worn or ambient sensors embedded in smartphones. The physical activities of the users are directly related to the movement and resting of the human body. The body-worn sensors, such as the accelerometer, magnetometer, and gyroscope, are the sensors that users may wear. The environmental changes can be identified by ambient sensors. There are few such ambient temperature sensors, radars, motion sensors, and sound sensors [4]. Human activities are categorized into seven groups such as ambulation, transportation, daily activities, exercise, kitchen activities, transitional activities, and self-care activities. Based on the category, activities are shown below [5]

(a) Ambulation: Sitting, standing, running, lying, falling
(b) Transportation: Driving a car, riding a bicycle.
(c) Daily activities: Watching TV, drinking, eating, using a phone, using a computer, reading the book, listening to music, sleeping.
(d) Transitional activities: Walking upstairs and downstairs, lying down and getting up, sitting down and getting up.
(e) Self-care activities: Combing hair, shaving, brushing teeth, washing hands, washing face, washing clothes, drying hair.
(f) Kitchen activities: Adding tea-bag, add sugar, add milk, removing tea-bag, pour milk, making coffee and tea, cooking pasta, cooking rice, feed fish.

Wesllen et al. stated the essential steps required to recognize the human activity: Data collection, segmentation, feature extraction, and activity classification.

1. Data collection: Extraction of raw data from different sensors embedded in a smartphone. The data must be adequate to produce good models for classification activities. To ensure the correctness of the activity model, some parameters must be considered like sampling frequency, the position of the smartphone, and orientation from the user, and data collection time [6]. But Foerster et al. [7] stated that the technique pursued to collect the raw data from the user is very difficult for any human activities. The accuracy level of ambulation activities for controlled data collection is 95.6% but the accuracy level is dropped to 66% for the natural environment. The proper analysis would recognize a significant number of users with different characteristics. This will give a better result for the new user without obtaining extra training data.

2. Segmentation: The raw data or inertial sensor signals are not recommended to take a decision in the classification process. The raw data, therefore, requires other transformations, such as breaking the continuous raw sensor data into the windows over a certain period. Noise removal from the signal is another important role of segmentation. Inertial sensor signal might have a noise which leads to misclassification; obviously, it affects the model accuracy. Using signal processing techniques like low-pass, high-pass, and Kalman filters, the noisy data can be removed easily [6].

3. Features: Settings defined in the segmentation stage played a vital role to extract the sensor features. The extraction algorithm takes the input from time windows. The selection of good features is a very important factor to classify the labels correctly. For any classification model, the accuracy can be directly affected because of low-quality features. Wesllen et al. described a new principal domain. It is divided into three groups such as time, frequency, and discrete domain. Time-domain uses mathematical approach statistical data are extracted from the signals. Repetitive patterns are captured through frequency domain. The discrete domain makes the signal pattern by converting sensor signals into symbols.

Thus, although the main domain permits for the chaining of features, it is very important to remember that all of them are orientation-dependent features when the time features are handled alone, and when these features are chained with the magnitude or vertical–horizontal components, they all become orientation-independent features. Some of the time domain features are min, max, amplitude, amplitude peak, sum, absolute sum, Euclidian norm, mean, absolute mean, mean square, mean, cross-validation, auto-correlation, skewness, kurtosis. Example of frequency domain: Energy, energy normalized, power, centroid, entropy, domain component [6].

Bashar discussed the deep neural networks, and the model accuracy is expelled even human performance. This survey on the deep learning neural network architectures utilized in various applications for having an accurate classification with an automated feature extraction especially in CNN [8]. Prabhakaran et al. describe the various clustering framework for predicting kidney disease [9]. Indumathi et al. explain the utilization of various machine learning models [10]. This project details a deep learning comparative study of the:

(a)   Convolutional neural network model.
(b)   Long short-term memory (recurrent neural network) model.
(c)   Extreme gradient boosting (XGBoost) model.
(d)   Multilayer perceptron.

2.1   *Convolutional Neural Networks in HAR:*

Jiang and Yin [11] compare multiple deep convolutional neural network (DCNN) architectures using accelerometer and gyroscope data in classifying activities. Jiang et al. [11] perform analysis on the UCI MHEALTH dataset (UCI) [12], USC-SIPI

human activity dataset (USC) [13], and a dataset compiled by the fusion of smartphone motion sensors (SHO) [14]. Jiang et al. [11] compared performances to identify the architecture, which achieved the highest accuracy, recall, and precision along with the low computational cost. Jiang et al. [11] found that SHO achieved the highest accuracy, followed by USC, while UCI performing slightly lower. Hammerla, Halloran, and Plotz [15] compare multiple convolutional and recurrent approaches when using wearable sensors in classifying activities. Hammerla et al. [15] perform analysis on three datasets: The opportunity dataset, Pamap2 dataset, and Daphneit Gair (DG) dataset. Hammerla et al. [15] conduct thousands of experiments to identify the substantial effect of altering hyperparameters. Performance evaluation indicated that the approaches achieved the highest accuracy on DG, the lowest root mean squared error, and the highest F1-score, followed by Opportunity, while performance scores on Pamap2 were slightly lower. Kim and Moon [16] compare the use of deep convolutional neural networks (DCNNs) for activity recognition in classifying activities. Kim et al. [16] also compare DCNNs for human detection. A Doppler radar gathers data, which produces velocity data when placed on a human or near a human. Kim et al. [16] found that the DCNN achieved accuracy as high as 97.6% for human detection. They also found that human activity classification accuracy reached heights of up to 90.9%.

2.2 Long Short-Term Memory (LSTM) in HAR

LSTMs are intended on tackling the vanishing gradient problem. The main difference between LSTMs and RNNs is LSTM's use of memory cells. Memory cells allow for the sufficient storage and sequential processing of data. Time is not restricted and the data does not disappear back into the network. It enables the development of relationships in the data, leading to insightful knowledge regarding the output to be analyzed. Gating is at the core of LSTMs. Gating regarding LSTMs involves component-wise multiplication of the input as seen in related work [17]. This leads to consistent updates in each data cell, due to the gating calculation that applies to each cell. The data must encounter the write, read, and reset gates to process data correctly. The write gate is the input gate. The read gate is the output gate while the reset gate is the forget gate. LSTMs contain information in a gated cell, which is the key idea of these networks. LSTMs can add or delete information to the cell through the gates. These gates are composed of a sigmoid neural network layer and a point-wise multiplicative operator.

2.3 Extreme Gradient Boosting in HAR

Ayumi investigates if extreme gradient boosting is superior in classifying activities in the HAR domain over classical techniques such as support vector machine (SVM) and

Naïve Bayes (NB). The UT Kinect-Action3D Dataset, the Badminton Sports Action Dataset, and the Bali Dance Motion Dataset conduct analysis. XGBoost takes more computational time to run as opposed to the other two methods but prevails as the best method with higher accuracy, precision, recall, and F1-score. Zhang et al. [18] propose an XGBoost method to recognize activities on their own dataset, which they created themselves. The dataset consists of 40 volunteers performing multiple activities contained in an indoor facility. XGBoost outperforms other ensemble classifiers with a higher recognition rate in accuracy, F1-score, and precision. F-score reached heights of up to 84.41% while accuracy surpassed previous studies achieving a rate of 84.19%. Nguyen, Fernandez, Nguyen, and Bagheri [19] explained, the XGBoost model uses wrist-worn accelerometer data to identify events, RGB-camera data, and environmental sensor data. In contrast to previous research, this unique method produced an elevated performance of 38% precision. A Brier Score of 0.1346 was also obtained, which indicates that it predicts the right behavior 90% of the time.

2.4 Multilayer Perceptron in HAR

Mo et al. [20] stated that the classification of activities based on the CAD-60 Dataset compares convolutionary neural networks and multilayer perceptron efficiency. The CAD-60 Dataset [21] provides RGB-D video sequences of events undertaken by volunteers. Sensor signals are recorded by the Microsoft Kinect sensor. In order to produce highly precise performance results, this research focuses on data pre-processing along with feature extraction. By using CNN for feature extraction and using MLP for the classification of the operation, the model presented incorporates CNN and MLP. It proved highly successful with the model achieving 81.8% accuracy across twelve different types of activities. Catal, Tufekci, Pirmit, and Kocabag [22] compare the performance of a model integrating aspects of decision tree, multilayer perceptron, and logistic regression. Accelerometer data is analyzed to classify activities. Related work [23] performs analysis on the Wireless Sensor Data Mining (WISDM Dataset) which contains information from 36 volunteers performing activities as seen in related work [23]. The proposed model achieved state-of-the-art results while achieving a superior performance when compared to a multilayer perceptron approach in related work. Results prove that integrating an ensemble of a classifier yields outstanding results in the activity recognition domain.

Talukdar and Mehta [8] built a multilayer perceptron network to classify physical human activities through the automated analysis of video data. The volunteer performed six activities 25 different times wearing a variation of different clothes each time. The activities performed were; walking, jogging, running, boxing, hand waving, and hand clapping. Talukdar et al. [8] present an MLP network that trains the data through a recurrent neural network that led to a vast reduction in learning time for the features and labels. The model achieved an overall accuracy of 92%. A comparison of the classification efficiency of XGBoost, MLP, CNN, and LSTM is one feature that is absent from the previously mentioned linked work on the machine

and deep learning models. To determine which network best fits the MHEALTH and WISDM dataset, our goal of this article is to conduct an investigation and compare these six different machine and deep learning algorithms with each other. This project revolves around the topic of using deep learning to benefit the healthcare industry.

Remote patient management (RPM) is one feature that could benefit from deep learning. Sufficient real-time surveillance of the actions of distant individuals will provide tremendous benefits in medical settings. Through reviewing, data sent to them by RPM technology, physicians, nurses, and clinicians may establish good relationships with and strengthen the experience of their patients. The data sent to them via RPM, as seen in [2], will establish a customized care plan and participate in shared decision making to encourage better results. Wearable devices can feed data to a clinician in real time by producing this data, leading to a substantial reduction in continuous patient surveillance. This device may be useful for the elderly, those who are vulnerable to heart problems (or severe medical conditions) and those who suffer from chronic illness. According to [4], the most frequent cause of readmission for patients in the USA is chronic heart failure (CHF). It is calculated that up to 84% of readmissions were considered preventable over a seven-day duration, while 76% of readmissions over a 30-day period were still considered preventable [4].

3   Proposed System

See Fig. 1.

3.1   *Convolutional Neural Network*

In a neural network, neurons learn from each other as they are fully connected. Neurons in convolutional neural networks connect to a fraction of the neurons that are in the previous layer. This layer is the receptive field as seen in related work [24]. Neurons in convolutional neural networks have three dimensions. These dimensions are width, height, and depth.

3.1.1   Architecture

CNNs have a unique architecture. It contains many sequential layers such as the convolutional layer, pooling layer, rectified linear unit layer, normalization layer, and fully connected layer.
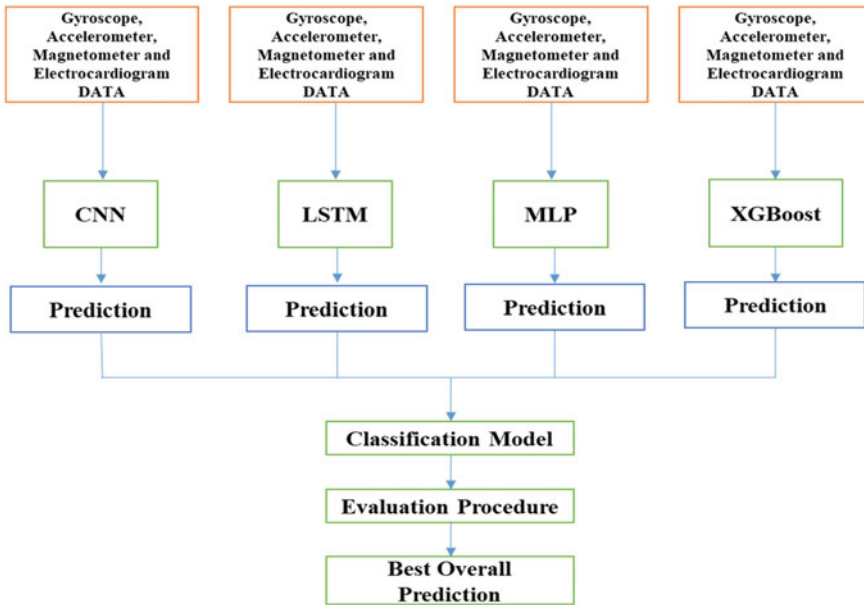
Fig. 1 Data flow diagram for the proposed model

3.1.2 Convolutional Layer

The convolutional layer is the focal point of a CNN. The convolutional layer's main objective is to extract high-level features about the data.

3.1.3 MaxPooling Layer

Pooling layers allow for the reduction in a number of parameters in the neural network. It reduces the number of descriptive parameters used to explain the structure of the neural network. It essentially avoids overfitting as it reduces the spatial size of the network. Training a neural network takes a great amount of time. Pooling ensures the number of computations needed to train the network is minimized. It ensures the classification task runs smoothly.

3.1.4 ReLU Layer

Relations in data are often nonlinear. The ReLU layer ensures there is an increase in nonlinearity. It applies the following element-wise non-saturating activation function to ensure that a neural network can build the nonlinear relation between data points. If there were no ReLU layer, a neural network would not be able to classify nonlinear

data points. When compared to tanh and sigmoid, the ReLU layer prevails in terms of speed, accuracy, and precision. The width, height, and depth of the neural network, also known as the spatial size, are left unchanged.

3.1.5 Dropout

Overfitting is a common problem neural networks face when training data. The dropout regularization technique successfully prevents overfitting. Fully connected layers in a neural network have many variations. Dropout identifies the nodes in a specific layer and removes them. A definitive probability, p, is then applied to the layer. The training process removes nodes linked with the removed layer. As seen in [25], after training, these nodes are placed back into the neural network and assigned their original value (weight). This, in turn, boosts the performance of the neural network. The validation training set benefits the most from this during the deployment of the model.

3.1.6 Optimizer: Adam

Adam is a gradient-based optimizer. It is straightforward, simple to implement, and is computationally inexpensive. It is suited to solving classification problems related to human activity recognition. The data involved in HAR is normally relatively large, leading to Adam to be a perfect fit. The hyperparameters require little or no tuning, which is why Adam is the most common optimizer in convolutional neural networks.

3.1.7 Softmax Activation Function

The softmax activation function is usually set in the output layer and loss layer. This is usually the final layer in the neural network before the output layer presents the result. The following equation is a detailed representation of the softmax activation function. The layers described above make up the full architecture of a convolutional neural network. The equation below incorporates all the layers and functions mentioned above to represent the typical architecture of a CNN (Fig. 2).

Feature extraction, encoding the labels to one-hot type along with separating the training and testing data, is done before the model is initialized.

CNN activity recognition overview:

(a)   One input layer containing 23 features.
(b)   Two separable convolution 1D layers with max pooling.
(c)   Three hidden layers that were big enough to train the data well.
(d)   Two dropout layers that yielded positive results.
(e)   Quite good performance, moderately slow.
(f)   One output layer of 12 results (labels).

$$Input \rightarrow [[Conv \rightarrow ReLU] * N \rightarrow Pool?] * M \rightarrow [FC \rightarrow ReLU] * K \rightarrow FC$$
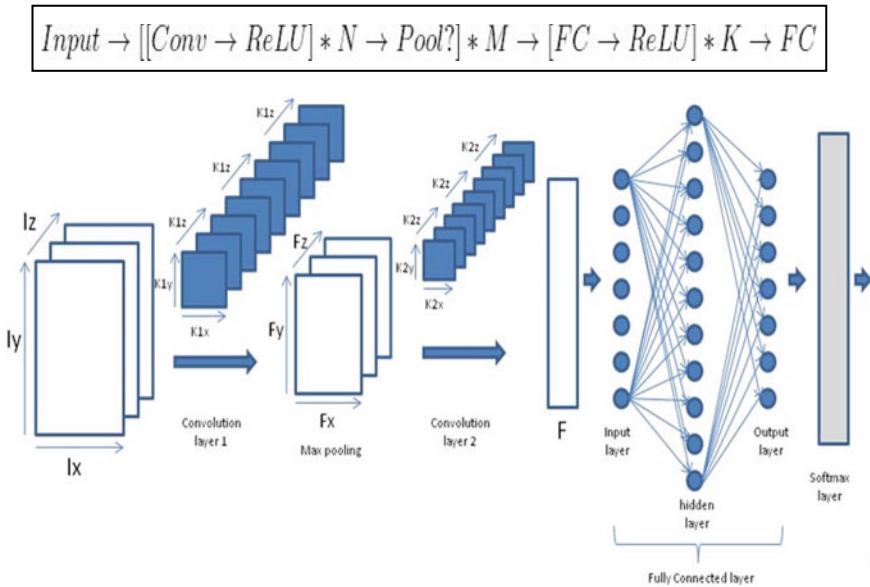

Fig. 2 Typical architecture of a CNN [26]

To classify (predict) the class variable, which is the motion that each subject executes, the neural network found in the CNN model utilizes the data values given for each of the 23 signals reported. This section gives an overview of training the model and hyperparameter setting.

(a) In the training process, the 'fit()' function is used to train the CNN model.
(b) 'X_train' represents the training data.
(c) 'y_train' refers to the target data.
(d) 'X_test, y_test' represent the validation data.
(e) The model is trained on a total of 245,584 parameters.
(f) While training the model:
(g) The Learning rate is set as 0.0005.
(h) Batch size is set as 32.
(i) The training process is run for 20 epochs.

3.1.8 Algorithm Speed

CNN performed excellently on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance is achieved, the time the model took to train the data was 242 min 18 s.

3.2   Long Short-Term Memory

LSTM activity recognition overview:

(a)   One input layer containing 23 features.
(b)   Two LSTM layers.
(c)   Two dropout layers.
(d)   Three hidden layers and one output layer.
(e)   One output layer of 12 results (labels).

This section gives an overview of training the model and hyperparameter setting.

(a)   In the training process, the 'fit()' function is used to train the CNN model.
(b)   'X_train' represents the training data.
(c)   'y_train' refers to the target data.
(d)   'X_test, y_test' represent the validation data.
(e)   The model is trained on 175,373 parameters. While training the model:
(f)   The Learning rate is set as 0.0005.
(g)   Batch size is set as 32.
(h)   Training process is run for 20 epochs.

3.2.1   Algorithm Speed

LSTM performed very poorly on the classification problem. It processed the data very slowly. Upon evaluating all six algorithms, LSTM is the poorest performing algorithm and processes data ten times slower than the other algorithms. The total amount of time the model took to process the data was 10 h. At first, the LSTM model was set to run for 20 epochs. The hard drive used to conduct each experiment was not strong enough to process data for such a significant amount of time. Considering the level of performance and speed of processing, CNN was the poorest algorithm applied to the MHEALTH dataset.

3.3   Extreme Gradient Boosting

Gradient boosting machines are associated with a distinctive type of machine learning branch called ensemble learning. The objective of ensemble learning is to train and predict a variety of models at the same time, while each model aims to outperform each others with respect to their output. Consider for example, the route from Paris to Berlin. There are many alternative travel options. As you proceed to take each route, you begin to learn which route is faster and more efficient, leading to the 'superior' route. Taking time to learn, each model has led to the conclusion that X is the superior route. Ensemble learning simply implies this strategy.

XGBoost implements the boosting. Boosting aims to convert weak learners to strong learners. During boosting, iterations lead to the weights of weak learners to

adjust accordingly. Bias reduces allowing for an increase in performance. Accuracy, precision, and recall benefit from the implementation of boosting greatly, as well as a range of evaluation techniques. Extreme gradient boosting (XGBoost) is the best performing boosting algorithm. XGBoost is a decision-tree-based algorithm that utilizes the use of gradient boosting and ensemble learning. XGBoost performs so well on data due to its ability to transform weak learners into strong learners. It utilizes boosting within the gradient descent architecture. It allows the framework to develop dramatically with its quick and easy to learn optimization techniques and parameter enhancements.

Feature extraction and splitting of the training and testing data are performed before the model is initialized. XGBoost activity recognition overview:

(a)   Multiclass classification using the softmax activation function.
(b)   Evaluation: multiclass classification error rate.
(c)   Learning rate is set to 0.05.
(d)   Trained on 161,959 parameters.

The parameters regarding the XGBoost model are discussed.

(a)   The model is trained using the parameter list and the training data.
(b)   The model is trained for 10 rounds.
(c)   Early stopping.
(d)   The learning rate of the model is 0.05 while the number of estimators is set to 1000.
(e)   Early stopping is used in the validation set to identify the appropriate amount of boosting rounds. This is usually the optimal number of boosting rounds needed. It is set to stop within 5 rounds. It will train until 'validation_0-merror' has not improved in 5 rounds.
(f)   The XGBoost model will train until the 'validation_0-merror' has not improved anymore. The 'validation_0-merror' must continue to decrease in order for the 'early_stopping_rounds' to continue to train the XGBoost model.

3.3.1   Algorithm Speed

XGBoost performed very well on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance it achieved, XGBoost was the second-best performing algorithm.

3.4   Multilayer Perceptron

Feature extraction, encoding the labels to one-hot form along with splitting the training and testing data, is performed before the model is initialized. MLP activity recognition overview:

(a)   One input layer containing 23 features.
(b)   Four hidden layers that were big enough to train the data well.
(c)   Two dropout layers that benefited the results greatly. The top-class algorithm is based on the concept of gradient descent.
(d)   One output layer of 12 results (labels).

This section gives an overview of training the model and hyperparameter setting.

(a)   In the training process, the 'fit()' function is used to train the CNN model.
(b)   'X_train' represents the training data.
(c)   'y_train' refers to the target data.
(d)   'X_test, y_test' represent the validation data.
(e)   The model is trained on 706,317 parameters. While training the model:
(f)   The learning rate is set as 0.0005.
(g)   Batch size is set as 32.
(h)   Training process is run for 20 epochs.

3.4.1   Algorithm Speed

MLP performed excellently on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance it achieved, the time the model took to train the data was 86 min 40 s.

3.5   MHEALTH Dataset

The MHEALTH dataset consists of body motion and vital signs recordings. Ten volunteers conducted the experiment, each with different characteristics. The subjects' task is to perform 12 different types of activities. The accelerometer, gyroscope, and magnetometer placed on the subjects' body measure acceleration, rate of turn, and magnetic field orientation. These sensors measure the range of motion experienced by each body part. The electrocardiogram sensor positioned on the chest also provides 2-lead ECG measurements. ECG can assist in the basic heart monitoring, checking for various arrhythmias, or looking at the effects of exercise on the ECG.

4   Result and Performance Analysis

Accuracy, precision, recall, F1-score of each DL architecture. Figure 1 outlines the accuracy of the MLP, XGBoost, and CNN machine learning models on the

MHEALTH dataset. The figures are extracted from the model's related confusion matrix. MLP attains the highest values, followed by MLP, then CNN (Table 1).

The four classification models presented in this research perform well when compared to existing state-of-the-art baselines. MLP and XGBoost achieve excellent performance measures, challenging many research papers with improved accuracy, precision, recall, and F1-score. The XGBoost model is the best performing model in terms of overall performance and is highly suited to mobile health data.

Figure 3 outlines the accuracy of the MLP, XGBoost, and CNN machine learning models on the MHEALTH dataset. The figures are extracted from the model's related confusion matrix. MLP attains the highest values, followed by MLP, then CNN (Table 2).

Figure 4 examines the performance of each deep learning model on the classification task, the figures are taken from each approaches confusion matrix output. MLP performs the best in comparison to the other models. The CNN and LSTM models achieved an average performance of 66.2% and 48.9%, respectively. The hybrid and

Table 1 Performance metrics evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score
MLP	90.55	91.66	90.55	90.7
CNN	83.91	83.47	83.91	82.98
LSTM	78.09	74.86	78.09	75.6
XGB	89.97	90.09	89.97	89.78
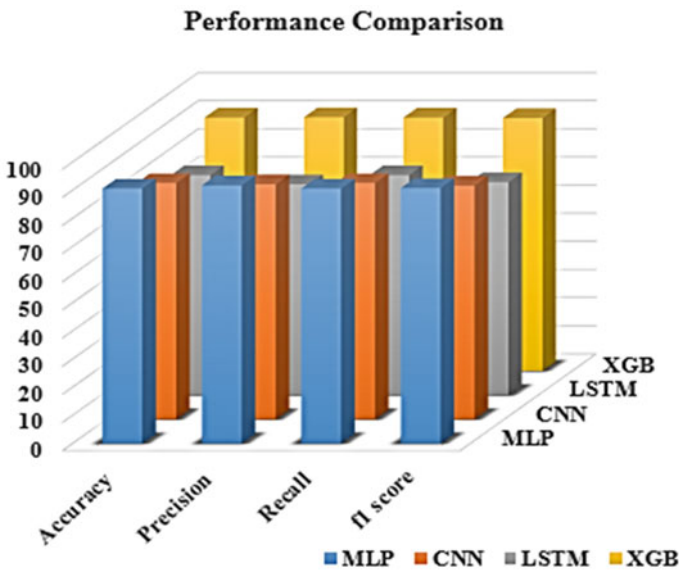
Fig. 3 Performance comparison

Table 2 Performance metrics evaluation

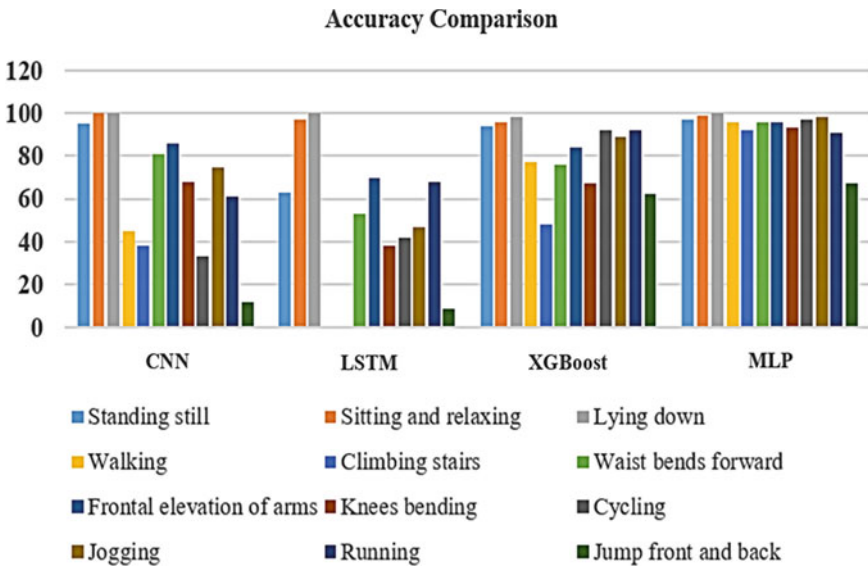| Activity	CNN	LSTM	XGBoost	MLP
Standing still	95	63	94	97
Sitting and relaxing	100	97	96	99
Lying down	100	100	98	100
Walking	45	0	77	96
Climbing stairs	38	0	48	92
Waist bends forward	81	53	76	96
Frontal elevation of arms	86	70	84	96
Knees bending	68	38	67	93
Cycling	33	42	92	97
Jogging	75	47	89	98
Running	61	68	92	91
Jump front and back	12	9	62	67
Average	66.2	48.9	81.25	93.5

Fig. 4 Accuracy comparison

XGBoost models achieved an average performance of 70.6% and 81.25%, respectively. For each subject, the MLP model attained the best performance obtaining an average performance of 93.5%. The figure identifies the superiority of the MLP model over others by showing a significantly higher performance.

4.1 Accuracy and Loss Results

Fine-tuning each model is to classify the data accurately, which is significant to maximizing accuracy while minimizing loss. MLP, XGBoost, CNN, and LSTM are fine-tuned with hyperparameters, respectively. The loss and accuracy of each model are visualized throughout this section. After fine-tuning each model and using a total of 20 epochs, MLP prevailed as the network with the highest accuracy (91%) and minimal loss. The average accuracy for each model is relatively high. MLP and XGBoost performed better and more consistently than CNN, LSTM. MLP and XGBoost achieved accuracies of 91% and 89%, respectively. MLP and XGBoost were able to converge far more easily as outlined in Fig. 3 which depicts the multilayer perceptron model achieving 91% accuracy.

The following two plots depict the convolutional neural network model achieving 84% accuracy. CNN performed moderately well in terms of accuracy (84%) and loss. MLP and XGBoost increased their accuracy along with diminishing their loss as the number of training iterations increased. Overall, MLP outperformed the other networks with the highest accuracy and lowest loss. The LSTM model misclassified too many instances, which lead it to become the poorest performing model with the least accuracy (78%) and highest loss.

LSTM networks performed poorly with 20 training iterations as depicted in Figs. 5. They both achieved 78% and 84% accuracy, respectively. The following two plots depict the LSTM model achieving 78% accuracy (Figs. 6 and 7).
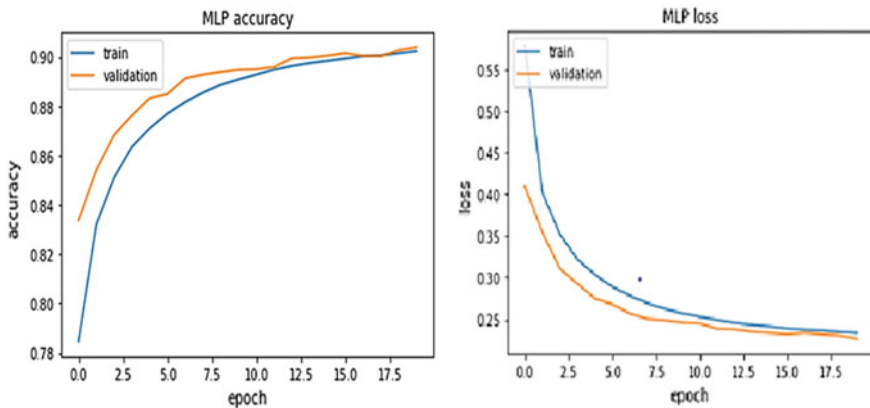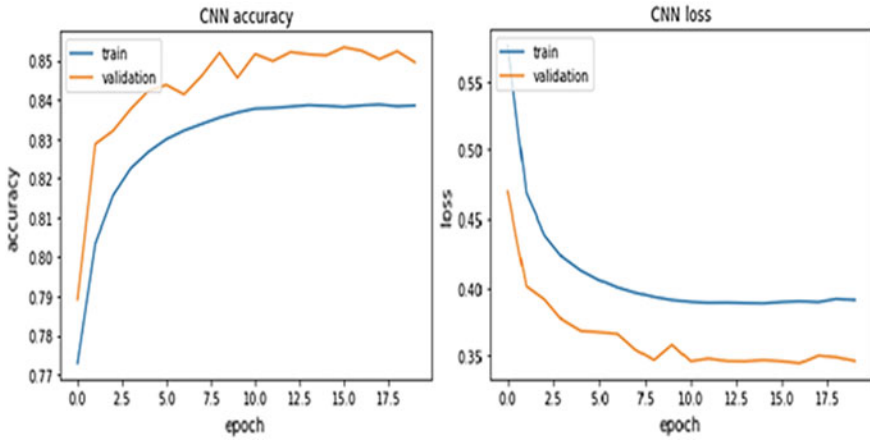

Fig. 5 MLP accuracy and loss
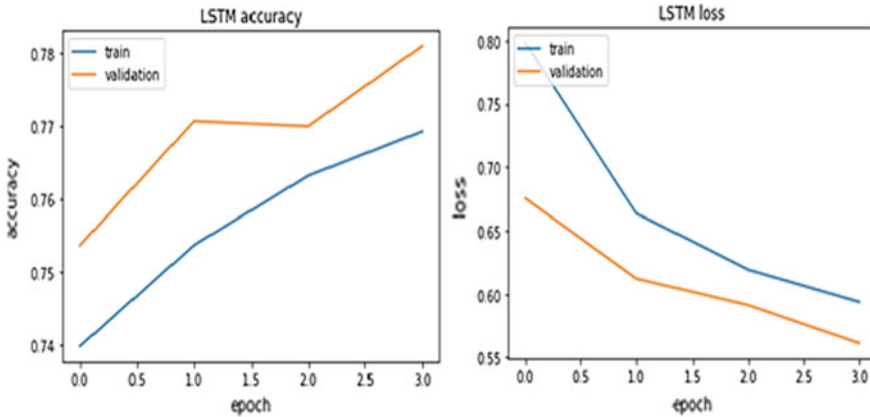
Fig. 6 CNN accuracy and loss

Fig. 7 LSTM accuracy and loss

5   Conclusion and Future Enhancement

Each model performs human activity recognition from wearable sensors such as gyroscopes, accelerometers, magnetometers, and electrocardiograms. To the author's knowledge, for the MHEALTH dataset, XGBoost has not been performed to classify the activities in question. MLP prevailed as the best performing model achieving accuracy, precision, recall, and F1-score of 90.53%, 91.71%, 90.53%, and 90.76%, respectively. XGBoost was the next best performing model that achieves accuracy, precision, recall, and F1-score of 89.98%, 90.14%, 89.98%, and 89.78%, respectively. While MLP outperformed XGBoost in terms of precision, accuracy, recall, and F1-score, 471 instances were misclassified by MLP, while XGBoost misclassified just

281.1341, 2533, and 2742 instances were misclassified, respectively, by CNN, and LSTM. XGBoost is the highest performing model in terms of total precision, consistency, recall, F1-score, and number of correctly categorized instances. This outlines the established area of appropriateness for the XGBoost system, which was never documented using wearable sensors on the MHEALTH dataset. This section presents an account of future work in human activity recognition using deep learning. Some of the ways in which human activity recognition models using deep learning will benefit healthcare in remote patient monitoring: clinician decision support, ambient assisted living / aiding the elderly, drug discovery, developing regions whose healthcare services are limited, app with patient data, diagnostic abilities, reduce need for electronic health records, creating more precise analytics for diagnosis, clinical decision making, risk scoring, and early alerting. Common challenges on human activity recognition presented by this research are unmeasurable uncertainty factors, activity similarity, and the null class problem.

References

1. Heart.org (2019) [Online]. Available: https://www.heart.org/-/media/files/about-us/policy-res earch/policy-positions/clinical-care/remote-patient-monitoring-guidance-2019.pdf
2. Khusainov R, Azzi D, Achumba IE, Bersch SD (2013) Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations. Sensors
3. Roobini S, Fenila Naomi J (2019) Smartphone sensor based human activity recognition using deep learning models. Int J Recent Technol Eng (IJRTE) 8(1)
4. Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: a survey. Pattern Recog Lett 119:3–11
5. Slim SO, Atia A, Elfatta MM, Mostafa MSM (2019) A survey on human activity recognition based on acceleration data. Int J Adv Comput Sci Appl 10:84–98
6. Sousa W, Souto E, Rodrigres J, Sadarc P, Jalali R, El-Khatib K (2017) A comparative analysis of the impact of features on human activity recognition with smartphone sensors. In: Proceedings of the 23rd Brazillian symposium on multimedia and the Web, Gramado, Brazil, 17–20 Oct 2017; pp 397–404
7. Foerster F, Smeja M, Fahrenberg J (1999) Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comput Hum Behav 15(5):571–583
8. Bashar A (2019) Survey on evolving deep learning neural network architectures. J Artif Intell 1(2):73–82
9. Prabakeran S (2018) In-depth survey to perceiving the effect of kidney dialysis parameters using clustering framework. J Comput Theor Nanosci 15(6–7):2233–2237
10. Indumathi V (2018) Utilizing data mining classification technique to predict kidney diseases. J Comput Theor Nanosci 15(6–7):2193–2196
11. Jiang W, Yin Z (2015) Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM international conference, pp 1307–1310
12. Banos O, Garcia R, Saez A (2019) UCI machine learning repository: MHEALTH Dataset Data Set, Archive.ics.uci.edu
13. Zhang M, Sawchuk AA (2012) Human activities dataset. Sipi.usc.edu
14. Shoaib M, Bosch S, Durmaz Incel O, Scholten H (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors
15. Hammerla NY, Halloran S, Plotz T (2016) Deep, convolutional, and recurrent models for human activity recognition using wearables. IJCAI

16. Kim Y, Moon T (2015) Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks
17. Javier Ordóñez F (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors 16:115 [69]
18. Zhang W, Zhao X, Li Z (2019) A comprehensive study of smartphone-based indoor activity recognition via Xgboost. IEEE Access 7:80027–80042
19. Nguyen T, Fernandez D, Nguyen Q, Bagheri E (2017) Location-aware human activity recognition. Adv Data Min Appl 821–835
20. Mo L, Li F, Zhu Y, Huang A (2016) Human physical activity recognition based on computer vision with deep learning model. In: 2016 IEEE international instrumentation and measurement technology conference proceedings
21. Cornell Activity Datasets: CAD-60 & CAD-120 | re3data.org, Re3data.org, 2019
22. Catal C, Tufekci S, Pirmit E, Kocabag G (2015) On the use of ensemble of classifiers for accelerometer-based activity recognition. Appl Soft Comput
23. Talukdar J, Mehta B (2019) Human action recognition system using good features and multilayer perceptron network
24. Luo W, Li Y, Urtasun R, Zemel R (2019) Understanding the effective receptive field in deep convolutional neural networks
25. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
26. Mody M, Mathew M, Jagannathan S, Redfern A, Jones J, Lorenzen T (2019) CNN inference: VLSI architecture for convolution layer for 1.2 TOPS. Ieeexplore.ieee.org