IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 12, 2020, accepted November 24, 2020, date of publication December 2, 2020, date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041822

Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model

FURQAN RUSTAM[1], **AIJAZ AHMAD RESHI**[2], **(Member, IEEE), IMRAN ASHRAF**[3], **ARIF MEHMOOD**[4], **SALEEM ULLAH**[1], **DOST MUHAMMAD KHAN**[4], **AND GYU SANG CHOI**[3]

[1]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan 64200, Pakistan
[2]Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia
[3]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[4]Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

Corresponding authors: Gyu Sang Choi (castchoi@ynu.ac.kr) and Imran Ashraf (ashrafimran@live.com)

This work was supported in part by the Industrial Technology Innovation Program under Grant 10063130, and in part by the National Research Foundation of Korea, by Ministry of Trade, Industry and Energy Korean Government under Grant NRF-2019R1A2C1006159.

ABSTRACT The recent development of machines exhibiting intelligent characteristics involves numerous techniques including computer hardware and software architecture development. Many different hardware devices, wearable sensors, machine learning, and deep learning model implementations are being applied in human activity recognition (HAR) applications in recent times. However, to develop high accuracy classification systems for activity recognition using ilow-cost hardware technology is of significant importance. To achieve this goal this study uses sensor data from two low-cost sensors, gyroscope and accelerometer along with the implementation of an Artificial Neural Network (ANN) based deep learning model for HAR. In particular, Deep Stacked Multilayered Perceptron (DS-MLP) has been proposed. In the implementation of DS-MLP, an ANN model has been used as a meta-learner while five MLP models have been used as base-learners. In this study, these base-learners and meta-learner have been combined using a stack ensemble technique. The performance evaluations have been done first on the applicability of individual base-models followed by the application of DS-MLP, the results prove the high accuracy of 97.3% and 99.4% for heterogeneous datasets used for testing. The performance of the proposed DS-MLP models has been compared to some existing machine learning classifiers and several state-of-the-art activity recognition systems. The comparative result analysis also proves that the proposed system performed better than these classification approaches in terms of important performance metrics such as accuracy, precision, recall, F-score, Cohen's Kappa, and Mathew correlation coefficient.

INDEX TERMS Human activity recognition, artificial intelligence, neural networks, sensors data, multilayered perceptron, stacked learning, supervised machine learning.

I. INTRODUCTION

Artificial Intelligence (AI) is a vibrantly vast field whose technologies are used in a variety of fields ranging from expert systems to computer vision. Neural computing is a closely associated area of AI that attempts to mimic the human brain [1]. In the last couple of decades, AI has progressed significantly towards the computerization of human reasoning. The attempt to build the computer architectures and the way of information processing as an imitation of human brain functionality results in neural computing systems, or artificial neural networks (ANNs). The ANNs are based on knowledge representations, massive information processing, fast information retrieval, and capability of pattern recognition (PR) based on experience. ANNs can be thus referred to as emulation models of the biological neural system in which artificial neurons are interconnected as a network. The ANNs are analogous to biological neural networks as these neurons similarly receive inputs and send the output to the other level of the network. Thus the ANNs are composed of processing elements (PEs) and a Network. Each PE also referred to as a neuron takes input data, then processes it to generate a single output. These inputs can be raw data or output from other PEs. On the other hand,

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

the output can be an input to other PE or the final output. The network is composed of a set of these PEs grouped in layers. The layers are categorized as a single input layer, a set of intermediate layers or hidden layers, and a single output layer. A deep neural network (DNN) is a special type of ANNs that comprises a comparatively large number of intermediate or hidden layers. In recent years HAR has surfaced as a vital area of research in a variety of application areas such as medical, fitness, and rehabilitation applications. HAR is a typical PR problem which can be a suitable candidate to be solved using ANNs. In many PR applications, ANNs have performed efficiently and accurately as compared to conventional machine learning models like support vector machine (SVM), Naive Bayes, and hidden Markov models (HMM) [2].

The human body consists of a skeleton system that consists of numerous bones connected by joints. This complex system has various functions in the human body of which the most apparent function is that it facilitates body movements. To exploit the movement function of skeleton joints experiments in [3] prove that HAR can be performed by attaching light spots attached to the major joints of the human body.

In this study, we attempt to integrate and investigate the state of art methods such as human body wearable sensor technology and ANNs to develop an innovative system for HAR. The proposed system can be used in a variety of HAR applications like patient activity monitoring in healthcare, personal activity recognition in fitness applications, and person activity recognition in rehabilitation applications, etc. The set of wearable sensors read the real-time data from the joint movements of the human body. This real-time data can be used as input to ANNs based systems for the prediction of the present body position of the person. Specifically, the features related to six daily activities such as standing, sitting, lying, walking, moving upstairs, and moving downstairs have been focused on in this study. The feature set used for ANN classification to detect the current body activity is being collected by the gyroscope, and accelerometer sensors. To design an accurate HAR application stacked ensemble learning method has been used in this study. This approach uses several DNN models, trains them individually, and then ensemble them through the stacked method for achieving more accurate results compared to an individual model. To evaluate the learning model performance parameters such as accuracy, F_1 score, recall, precision, Cohen's kappa, and Matthews correlation coefficient have been used. The proposed HAR process pipeline is shown in Figure 1. In brief, the main contributions of this study are as follows:

- A stack ensemble approach has been proposed for human activity classification using DNNs. The given output of the five base-learners has been used to train the meta-learner in this regard. Six activities are classified using DS-MLP such as walking, running, sitting, standing, moving downstairs, and moving upstairs.
- The proposed approach is tested with two different datasets that contain data from sensors like gyroscope and accelerometer. The performance is analyzed both with a larger feature vector, as well as, smaller feature set.
- Performance analysis has been done for machine learning algorithms such as LR, SVM, KNN, RF as well as gradient boosting machine and random forest. Besides, the performance of the proposed DS-MLP has been compared with several of the state-of-the-art HAR systems.

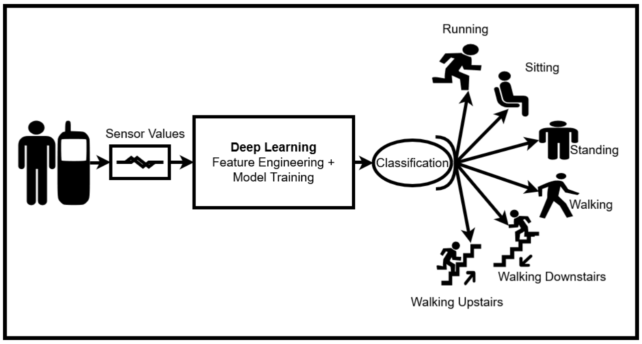

FIGURE 1. HAR process pipeline.

The paper is further presented in the form of many sections, Section II presents the discussions of research works related to the current study. Materials and methods used in the study have been given in Section III. Section IV contains Results and discussion of the study and lastly, Section V presents the conclusion.

II. RELATED WORK

Wearable sensors, ANNs, and HAR technologies have a significant impact on the quality of daily life in the current time. Numerous researchers and organizations are focused on these technologies to make human life better in all aspects. Numerous research studies have been carried out and many more are in progress in these application domains. The contributions include the inventions of smart environments and devices in varied fields like health care recovery, human well being, safety, surveillance, as well as, military operations [4]. There are certain research studies published related to HAR applications including home behavior analysis [5], video surveillance [3], [6] gait analysis [7], and gesture recognition [4]. Video-based HAR and Sensor-based HAR are two types of HAR on which many studies have been done. Video-based HAR can be performed with video or images that contain human motion while sensor-based HAR is based on the motion data extracted from smart sensors [8]. Due to exponential development and better privacy provided by sensor-based HAR, studies have been performed using many efficient supervised machine learning algorithms. Multiple sensors are used in sensor-based HAR such as gyroscope, accelerometer, sound sensors, and Bluetooth, etc. The authors in [7] provide comparative analysis results of conventional supervised and unsupervised machine learning models like KNN, RF, HMM, and Gaussian mixture models. The study reports that the supervised models provide higher accuracy than the unsupervised models. However, unsupervised models are computationally efficient than those of supervised models. There is tremendous work in HAR by using supervised machine learning approaches [9], especially deep learning is more efficient and accurate in pattern recognition. Deep learning relieves the effort of designing features and can learn high-level and meaningful features by training an end-to-end neural network which makes deep learning a preferred choice in HAR applications. Such as the study [10] uses a recurrent neural network (RNN) to recognize the daily human activities by using body-worn sensors and the study [11] uses smartphone sensor data for deep learning-based Long Short Term Memory (LSTM) network to detect human activities.

A broader explanation of the advantages and disadvantages of the classification models in HAR based on static and shallow features using body-worn sensors is given in [12]. The deep learning models for HAR were provided earlier by the authors in [13]. Researchers in [13] use Convolutional Neural Networks (CNNs) for the HAR application. Raw data have been used from the set of motion and position sensors in the study. The study also investigated the performance of various CNN architectures based on different sets of sensors [14]. InnoHAR a deep learning model for HAR has been proposed by researchers in [15]. The model uses the multi-channel waveform end-to-end sensor data. The approach uses a fusion of RNN and inception neural network (INN) which extracts the multidimensional features using inception-like modules. The feature extraction is being done using kernel-based convolution layers. Their experiments are performed on three public HAR datasets and present a promised generalized performance.

Besides the above-mentioned research works, several studies focus on using LSTM and its improved variations for HAR tasks. For example, the authors propose a residual bidirectional LSTM to address HAR problems in [16]. Owing to the capability of the proposed approach for concatenating forward and backward states, it outperforms CNN, baseline LSTM, and bidirectional and residual LSTM. The accuracy is improved by 4.8% and reaches 90.5% for HAR tasks. Another study that uses stacking for activity recognition tasks is [17] where sound and accelerometer data are utilized. Multi-view stacking is used to fuse the data from the smartphone. Different models are trained for each sensor's views and the output is combined through stacking. Stacking performs better than that of aggregation and achieves an accuracy of 0.925 for HAR tasks. Similarly, a stacked LSTM network is presented in [18] for identifying six human behaviors using smartphone data. The network consists of five LSTM cells and an L_2 regularization layer is used for network generalization. Experiments on the UCI dataset indicates improved performance than those of previous approaches with an accuracy of 0.93%.

Along the same direction, an approach using stacking denoising autoencoder (SDAE) and light gradient boosting machine (LGBM) is presented in [19]. SDAE helps to reduce noise in sensor data and extract features for unsupervised learning. LGBM is used to fetch inherent feature dependencies to enhance accuracy. The performance analysis indicates the superior performance of the proposed approach over XGBoost, CNN+statistical features, and single SDAE. Achieved accuracy for HAR tasks is 0.95. Similarly, the authors propose a stacked ensemble model for human activity recognition in [20]. A neural network is used as a meta learner while latent Dirichlet allocation, decision tree, Gaussian naive Bayes, k nearest neighbor, SVM, and multi-layered perceptron are used as the base learners for the proposed approach. The stacked ensemble model can achieve an accuracy of 0.96 for various HAR tasks. HAR task prediction accuracy is further improved in [21] where Boruta, a wrapper-based all-relevant feature selection method is used for feature extraction, before model training. The stacking approach is composed of various machine learning algorithms like the random forest, multi-layer perceptron, logistic regression, and SVM with linear kernel and Boruta shows good performance with an accuracy of 0.97%.

The current study presents an approach similar to what is presented in [22]. Authors present a novel HAR approach that leverages an ensemble model of several extreme learning machines (ELM) to classify human tasks from sensor data. Initially, several ELMs are trained with bootstrap sampling and are later pre-pruned to increase accuracy. Glowworm swarm optimization is used to identify the optimal ensemble. The reported achieved accuracy is 96.7% which is better than traditional ensemble approaches like bagging and Adaboost, etc.

Nonetheless, many research works focusing on HAR tasks lack in accuracy, robustness, and latency and depend on the data sources containing noise like sound, which requires data cleaning or filtering. Additionally, several works detect the various number of HAR activities from three to six which also impacts the accuracy. In addition to that, the accuracy is compromised with the mechanism followed during the data collection. For example, the authors used sensors hung on the chest, waist, and arm of the participants to collect the data. There is a striking difference between the quality of hung sensors and smartphone sensors. The data from the smartphone sensors are noisy than those of other sensors. Besides, users' changing the orientation of the smartphone during the data collection further complicates the process of HAR classification. Anyhow, there is a scope to further improve the overall performance of the learning model which can make the HAR application more reliable. In this study innovative set of experiments on deep learning, approaches have been used to significantly increase the performance of the classification system. In particular stack ensemble learning method has been used in which multiple DNN models are trained individually to achieve higher accuracy.

III. MATERIAL AND METHODS

A. DATASET

In this study, the HAR application has been developed using supervised learning models. The most important part of the study related to supervised learning is the dataset. This study uses two datasets, i.e., "Human Activity Recognition Using Smartphones Dataset" [23] and "Heterogeneity Human Activity Recognition Dataset" obtained from UCI Machine Learning Repository [24]. Both datasets named as 'Dataset1', and 'Dataset2' has been used to train and test the learning models. Dataset1 contains features related to six activities of daily life. These features have been extracted by wearing a smartphone (Samsung Galaxy S II) on the subject's waist. Smartphone embedded sensors accelerometer and gyroscope had been used to capture 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. A group of 30 volunteers within the age of 19-48 years had participated in this data collection task, each person performed all six activities by wearing a smartphone on their waist. The noise filters had been applied to the signals gathered from the accelerometer and gyroscope for signal pre-processing. The signal sampling had then been done using a 2.56 sec with a 50% overlap sliding windows, leading to 128 readings per window. A Butterworth low-pass filter had been used to separate different signal components. The components of the signal included body acceleration and gravity. Due to the assumption of gravitational force having low-frequency components, a 0.3 Hz cutoff frequency filter had been used in this case. The time and frequency domains had been used for calculations to extract a vector of features from each window. The dataset used has been split into two subsets training set and testing set after data extraction with the ratio of 70:30. Thus the characteristics of the dataset are depicted in Table 1, as can be seen in Table 1, there are fluctuations in the label counts, the labels are quite equally distributed

TABLE 1. Description of characteristic of Dataset1.

Examples	Dataset	Training set	Testing Set
Total Examples	10297	7351	2946
Walking (1)	1722	1226	496
Walking_Upstairs (2)	1544	1073	471
Walking_Downstairs (3)	1406	986	420
Sitting (4)	1777	1286	491
Standing (5)	1904	1373	531
Laying (6)	1994	1407	537
Total Features	561	561	561

Dataset2 contains the data collected for daily human activity using smartphone sensors and smartwatch sensors. In this study, we use the smartphone sensor dataset. The two types of smartphone sensor gyroscope and accelerometer are used to extract the human activity data. The activities which are considered in this dataset are: 'Biking', 'Sitting', 'Standing', 'Walking', 'Stair Up' and 'Stair down'. The file of both gyroscope and accelerometer sensor consists of the following columns: 'Index', 'Arrival_Time', 'Creation_Time', 'x', 'y', 'z', 'User', 'Model', 'Device', 'gt'. We just extract x,y,z columns from both gyroscope and accelerometer files, and a 'gt' column. In this way, we have 6 columns for features and 1 for the target class. The dataset has 1,048,575 records but in this experiment, we use 10,000 records, i.e., 2,000 from each class. The description of dataset characteristics is given in Table 2.

TABLE 2. Description of characteristic of Dataset2.

Examples	Dataset	Training set	Testing Set
Total Examples	12000	8400	3600
Bike	2000	1392	608
Sit	2000	1383	617
Stairsdown	2000	1404	596
Stairsup	2000	1425	575
Stand	2000	1384	616
walk	2000	1412	588

1) DATA VISUALIZATION

Data visualization plays an important role to understand the importance, as well as, the relationship of various features. As the dataset is geared towards classifying the activity of the daily life performed by 30 participants, dataset characteristics have been plotted for more investigation by using T-distributed Stochastic Neighbor Embedding (TSNE) a machine learning algorithm for visualization which shows the separability of the classes. To visualize the high dimensional data in a 2D space, the TSNE method has been used. Since the TSNE is a tool to which a high dimensional feature set has been passed and it returns the features set with reduced dimensions with only two dimensions. These two features have been plotted on *s*-axes and *y*-axes of the given graph and are labeled as 'TSNE: Dimension 1', and 'TSNE: Dimension 2', respectively. Figure 2 (top) indicates that all activities are almost separable while Figure 2 (bottom) plots the distribution of personal information of all participants. For example, various persons have a unique/separable walking style as shown in the top plot. Therefore, such features can be used to detect various activities of a person, provided the data of such activities.

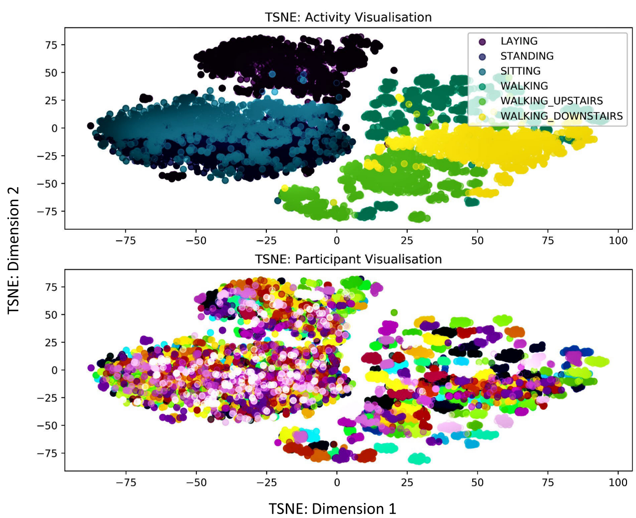

FIGURE 2. Participant activity visualization for Dataset1.

Despite the separability of various activities, each activity may have a different set of features that are important to determine that activity. Analyzing the importance of a feature(s) helps to elevate the performance of HAR tasks. Features importance is evaluated by dividing the data according to sensor type, i.e., accelerometer and gyroscope, because each sensor may possess various features and associated importance. Figure 3 shows that the features extracted from Dataset1. It shows that features from the accelerometer sensor are more important for activity classification because the accelerometer supplying slightly more information. However, both sensors are contributing to classification, and each should be used to enhance the performance of HAR tasks.

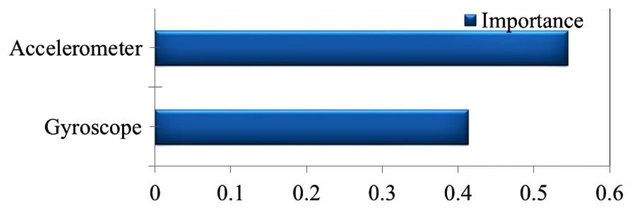

FIGURE 3. **Sensor importance for classifying participants by walking style (feature importance sum) for Dataset1.**

Similarly, the characteristics of Dataset2 are plotted in Figure 4. Task separability is less evident for Dataset2 than that of Dataset1, so, it would be challenging to achieve high performance with Dataset2. Few activities have distinct features like 'stairsup', and 'walk', while others are less separable like 'bike' and 'sit', etc.

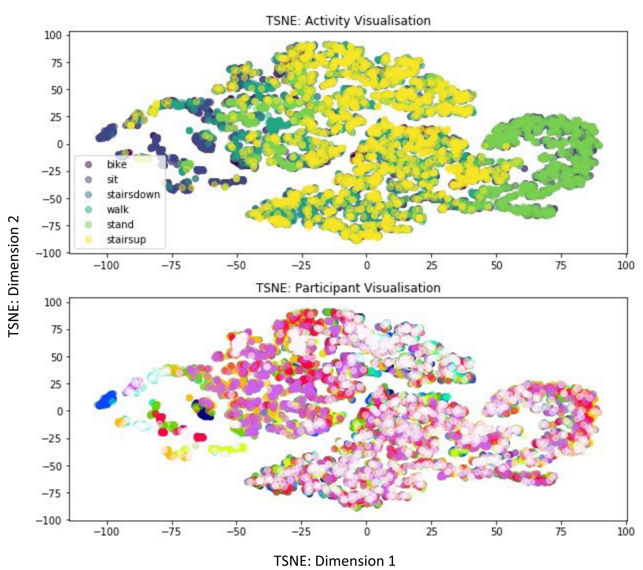

FIGURE 4. **Participant activity visualization for Dataset2.**

Feature weights (importance) are portrayed in Figure 5 that indicates the importance of both accelerometer and gyroscope data to determine user various activities. The importance of data features from accelerometer and gyroscope sensors weighs less for Dataset2 than that of Dataset1. It is due to less separability of features for various tasks as shown in Figure 4.

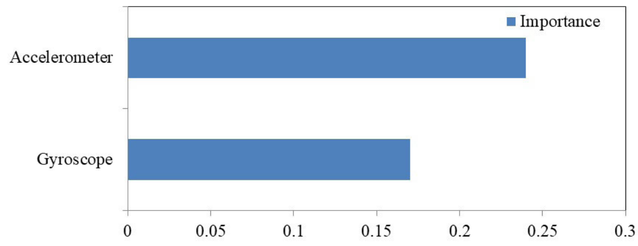

FIGURE 5. **Sensor importance for classifying participants by walking style (feature importance sum) for Dataset2.**

B. PROPOSED MODEL

We propose the DS-MLP model by assembling multiple MLP models as base-models and a neural networks Model as the meta-model using stack ensemble techniques. In this section, we will describe the method and the architecture of the proposed model.

1) STACKING

Staking sometimes known as a stacked generalization is an ensemble learning technique used to combine multiple models via a meta-model. Stacking combines the multiple models that lead to the idea of meta learner [25]. Although stacking is a very attractive approach, however, its usage is still very limited in comparison to techniques like bagging and boosting. It is very effective in various application scenarios as is the case in our study for activity detection. The stacking ensemble technique combines different types of models that makes it different from bagging and boosting ensemble learning. In staked generalization, there are two kinds of models: the level-0 model called base-models and the Level-1 model called a meta-model. meta-model learns from the outputs of base models [26]. Consequently, stacked learning outperforms the best base-model in prediction results accuracy. Proposed stacked generalization architecture is illustrated in Figure 6. It shows the general architecture of the proposed stacked model wherein the names of based and meta-model are not given. Detailed architecture with base and meta-models' name is shown in Figure 11.

Following is the step-wise procedure for stacked ensemble learning:

1) Divide the training set into two sub-sets.
2) Use the first sub-set to train base models.
3) Use the second subset to test the base models.
4) Use the outputs (predictions) of base models as the training input of the meta-model.

As can be seen in the above set of steps, steps 1 through 3 remain the same as in the cross-validation method. The difference thus lies in the fact that instead of a winner-take-all method, it combines the base models non-linearly. This difference leads to the increased accuracy provided by stacked algorithms.

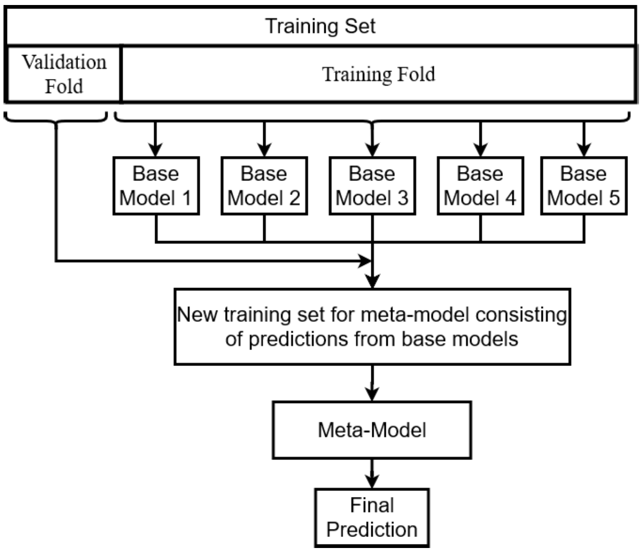

FIGURE 6. Stacked ensemble learning approach.

2) MULTILAYERED PERCEPTRON (MLP)

In recent times the deep learning approaches are used in several different fields including natural language processing (NLP), medical image processing, automatic speech recognition, computer vision, etc. [2]. In this study, a deep learning model MLP which is a feed-forward class of ANNs [27] has been used. The MLP contains at least three layers of nodes: an input layer, a hidden layer, and an output layer. In an MLP, the data flow is unidirectional, that is from the input layer to the output layer. All other nodes except input layer nodes are neurons that use a nonlinear activation function to learn from the data. Characteristics like multiple layers and non-linear activation function of the MLP differentiate it from a linear perceptron. It can distinguish data that is not linearly separable.

As mentioned earlier, the MLP is a type of ANN which is a feed-forward layered network. In this type of ANN, the nodes in the input layer contain the input features of the network while the hidden layers gather the weighted inputs from the input layer and forward their output data to the successive layer. Finally, the output layer contains the classification results for input data [28]. There are multiple algorithms for learning steps of MLP and a common supervised learning technique is called back-propagation [29]. Back-propagation comprises of four distinct stages: 1) initializing weights, 2) feed-forward, 3) back-propagation of errors 4) weight update.

3) DEEP STACKED MULTILAYERED PERCEPTRON (DS-MLP)

The DS-MLP model is the novelty adopted in this study for the HAR application. This model is a combination of stacked generalization and a deep learning model MLP. In this study, five MLP models have been used as base learners and a neural networks model "Model" as a meta learner. The architecture of DS-MLP is illustrated in Figure 6. As shown

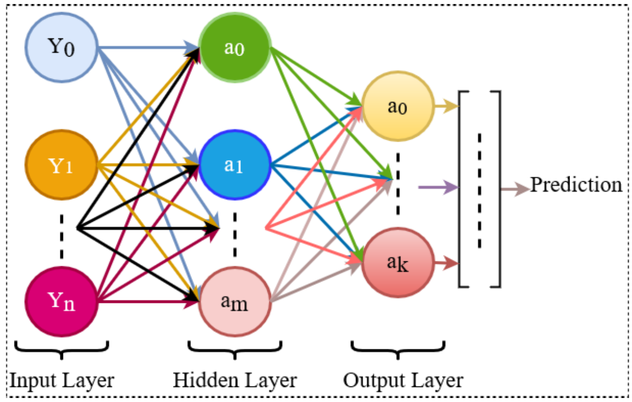

FIGURE 7. The architecture of MLP.

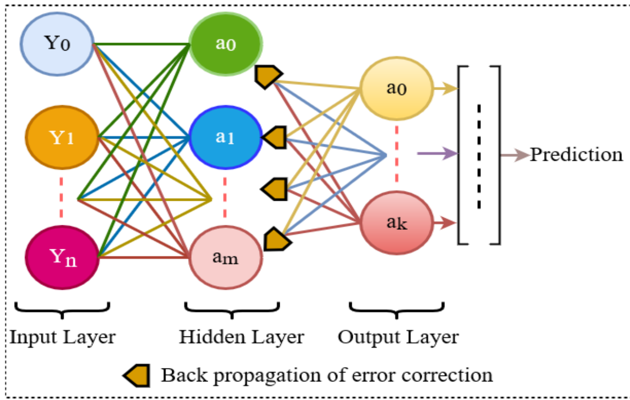

FIGURE 8. The MLP back-propagation.

in the figure DS-MLP training set has been split into two folds/sets, training fold and validation fold. The training fold is used to train five base-models (MLP) and the validation fold is passed to each base-model to make predictions on validation fold data. Then a matrix is constructed by using outputs of these base-models on validation folds. After this, a meta-model is trained using the outputs produced by the base-models to make a strong ensemble learning model (DS-MLP).

Each base-model consists of five layers: 3 dense layers, 2 dropout layers. The first dense layer in each base-model which is also an input layer consists of 64 neurons which will take 561 input features to train the base-model. The first dense layer also contains the activation function ReLU (Rectified Linear Unit). ReLU activation functions are used to introducing non-linearity into the ANN model so that the ANN can progressively learn more effective feature representations. ReLU is the most commonly used activation in all convolutional neural networks or deep learning models [30]. ReLU is a non-linear function that can backpropagate the errors easily. One main advantage of the ReLU activation function is that it does not activate all neurons at the same time so activation of few neurons at a time makes sparse neural networks that make it efficient and easy for computation. The second layer of each base-model is a dropout layer with a 0.2 dropout rate. The dropout layer avoids the over-fitting of models by ignoring randomly selected neurons during training [31]. The third layer is again a dense layer with 128 neurons and the ReLU activation function. The fourth layer in each base-model is again a dropout layer with a 0.2 dropout rate. The fifth or last layer is the output layer with 6 neurons and a softmax activation function. The softmax has been used in the output layer of each base-model and meta-model. The softmax function that has been used at the output layer is mostly used to build a model for multi-class classification [32] because it gives the probability for each target class. Class with a high probability score will be the final class. Figure 9 illustrates the layers parameters of each base-learner.

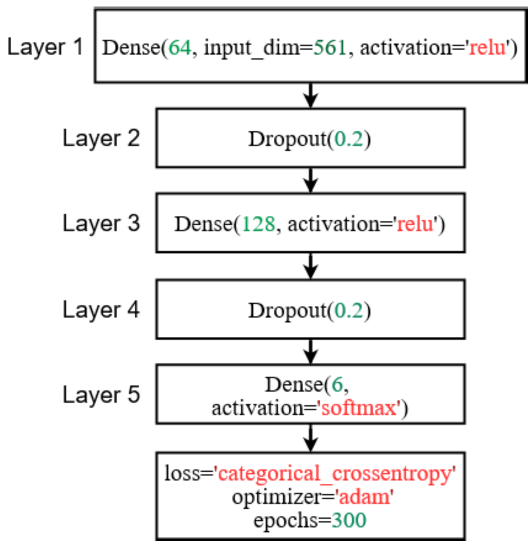

FIGURE 9. **The layer architecture diagram for each base-learner (MLP).**

The meta-model in DS-MLP consists of three-layer: 2 dense layers (input layer and output layer) and 1 dropout layer. The input layer consists of 16 neurons and the ReLU activation function, the second layer is a dropout layer in the meta-model. The last and the output layer of the meta-model is comprised of 6 neurons and a softmax activation function. Figure 10 shows layers parameters of meta-learner.

In all the base-models and the meta-model, we used the 'Adam' optimizer (an adaptive learning rate optimization algorithms) because 'Adam' is designed specifically for the training of ANN models [33]. Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. Adam is relatively easy to configure where the default configuration parameters do well on most problems. The loss function 'categorical_crossentropy' has been used in all base-models and the meta-model. The loss function estimates the loss of the model so the loss on the next evaluation can be reduced by updating weights. Additionally, 'categorical_crossentropy'' has been used because it increases the effectiveness of the stacking model [34].

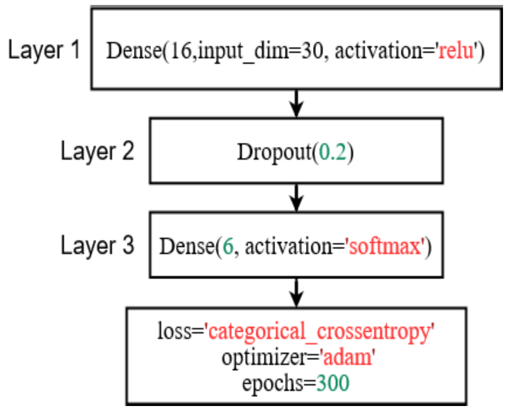

FIGURE 10. **The layer architecture diagram for meta-learner.**

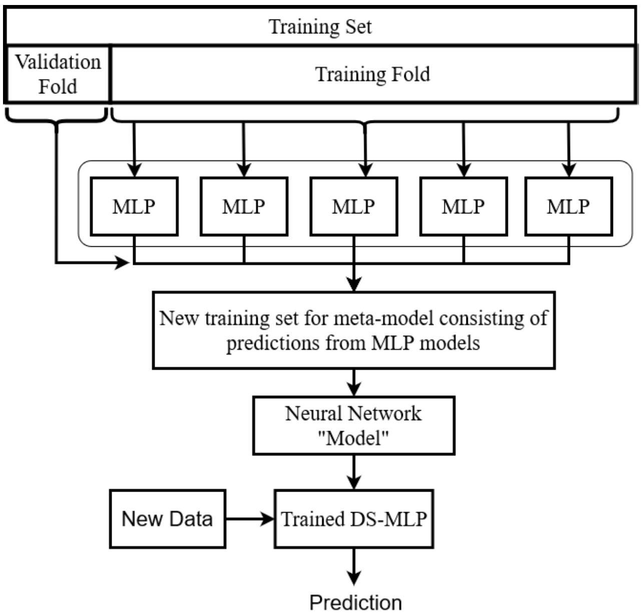

FIGURE 11. **The DS-MLP model architecture diagram.**

The inputs of the base model are the features of the two datasets. In the case of dataset1, we have 561 features and 6 target classes. The input of each model was 561 features and each such model produced 6 predictions against each example in output. In the proposed methodology we have five models and each model makes six predictions per example, which becomes 30 (6×5) outputs for each example provided to the sub-models. Then these 30 outputs became input features to the meta-model. Thus the [7351, 5, 6] shaped predictions from the sub-models have been transformed into a [7351, 30] shaped array to be used to train a meta-learner and flattening the final two dimensions.

Similarly, in the case of dataset2 total of 6 features have been used with 6 target classes. The input of each model was 5 features and each such model produced 6 predictions against each example in output. In the proposed methodology we have five models and each model makes six predictions per example, which becomes 30 (6 × 5) outputs for each example provided to the sub-models. Then these 30 outputs became input features to the meta-model. Thus the [8400, 5, 6] shaped predictions from the sub-models have been transformed into a [8400, 30] shaped array to be used to train a meta-learner and flattening the final two dimensions.

Five base models are given the input features from dataset1 and dataset2. The input of the deep neural network or meta-model in DS-MLP is the output of these five base models. As mentioned earlier, each of the five base models gives 6 predictions on each example which becomes a total of 30 predictions per example. So these 30 predictions are then given as input to the meta-model of DS-MLP which then produces a prediction for 6 target classes.

C. EVALUATION PARAMETERS

This study performs experiments for HAR by using a supervised learning approach. We compare the performance of the proposed learning model DS-MLP with previous approaches in terms of accuracy. In this study, the four most important evaluation metrics have been used to measure the performance of DS-MLP such as accuracy, precision, recall, and f_1 score [35]. These performance measures are given in equations 2 through 5:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F_1\ Score = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{4}$$

where TP denotes True Positive, TN denotes True Negative, FP denotes False Positive and FN denotes False Negative.

The most important measure of performance is the accuracy that is used to conclude the overall success of any model. Accuracy represents the ratio of correct classification instances (TP+TN) to the total sample instances (TP + TN + FP + FN).

This study solves the multi-class classification problem so we also use Cohen's kappa (CK) and Matthews correlation coefficient (MCC). Here, CK can be interpreted to indicate the performance of the classifier as given in Table 3. On the other hand, the MCC score can be between −1 to 1, where −1 represents the poor performance of the classifier while 1 indicates the perfect performance of a classifier.

TABLE 3. CK values and interpretations.

Value	Interpretation
≤ 0	poor performance
0.01-0.20	none to slight
0.21-0.40	fair
0.41- 0.60	moderate
0.61-0.80	substantial
0.81–1.00	perfect

D. EXPERIMENT FLOW

This study is related to predicting human activity using a supervised machine learning approach. The workflow of the proposed approach is illustrated in Figure 12. In our experiment, we obtained the HAR dataset from the UCI data repository (see Section III-A). First, the dataset has been divided into two subsets: a training set and testing sets with a ratio of 70% and 30%. The ratio of the records after splitting the dataset is shown in Table 1. After splitting the dataset into training and testing sets, the training set contains a total of 7351 records and the testing set contains 2946 records. We train our proposed model DS-MLP using the training set and after the training DS-MLP, we evaluate the performance of the trained model. To evaluate the performance of the trained model we pass test data and measure the performance of the DS-MLP. We measure the performance of the DS-MLP in terms of accuracy, precision, recall, F_1 score, CK, and MCC (see Section III-C).

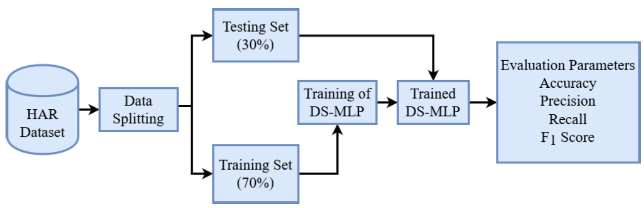

FIGURE 12. Our proposed methodology diagram.

IV. RESULTS AND DISCUSSION

This study implements an ensemble learning model (DS-MLP) for HAR applications. DS-MLP has been used with five base-models that have been trained on the dataset, then a meta-model has been trained on the outputs of these five base-models. The proposed ensemble model proves to achieve high accuracy than the individual models and secure 97.3% accuracy.

Table 4 shows the accuracy of all base-models individually and also shows the accuracy of DS-MLP. DS-MLP achieved high accuracy because the meta-model has been trained on the outputs of base-models which boosts the overall performance as shown in Figure 13.

TABLE 4. The DS-MLP model results.

Model	Model Type	Accuracy
MLP 1	base-model	0.962
MLP 2	base-model	0.959
MLP 3	base-model	0.957
MLP 4	base-model	0.957
MLP 5	base-model	0.958
DS-MLP	stacked model	0.973

Figure 13 illustrates the training and testing accuracy after each epoch and it can be seen that after each epoch accuracy is boosted till it reaches 250 epochs after which there is no further improvement in terms of accuracy. So, a total

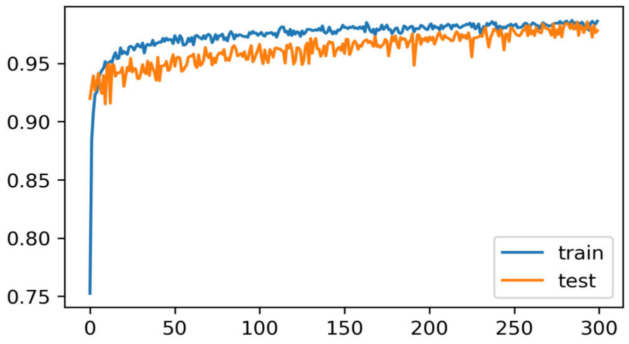

FIGURE 13. Training and testing accuracies of DS-MLP on each epoch.

of 300 epochs are used to train the models for the current study.

The confusion matrix in Figure 14 shows the classification results with DS-MLP. It accurately predicts 2867 motion activities out of a total of 2946 predictions on the test data. DS-MLP also shows high performance for each activity prediction as shown in Table 5.

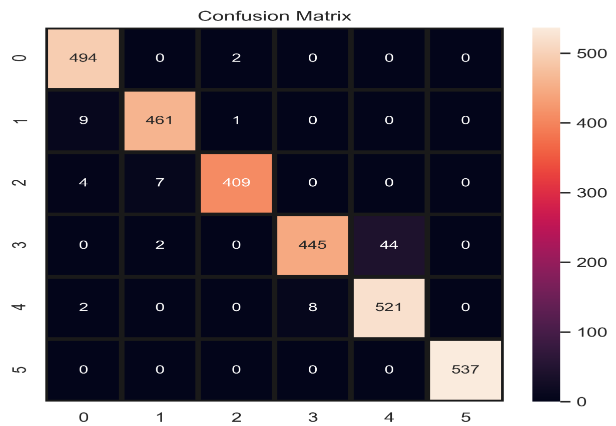

FIGURE 14. Confusion matrices of all DS-MLP, here 0 represents the Walking class, 1 represents Walking Upstairs class, 2 represents Walking Downstairs class, 3 represents Sitting class, 4 represents Standing class, 5 represents Laying class.

TABLE 5. The DS-MLP model results for each activity in Dataset1.

Activity	Total Prediction	Correct Prediction	Wrong Prediction
Walking	496	494	2
Walking Upstairs	471	461	10
Walking Downstairs	420	409	11
Sitting	491	445	46
Standing	531	521	10
Laying	537	537	0
All activities	2946	2867	79

A. DS-MLP PERFORMANCE ON HETEROGENEITY HUMAN ACTIVITY RECOGNITION DATASET

The performance of the proposed approach is tested using Dataset2 as well. For this purpose, we extracted x,y,z columns from both gyroscope and accelerometer files, and a 'gt' column. So, training and testing are based on six features, three each from accelerometer and gyroscope data. We train DS-MLP on the training set and then we evaluate the performance of these models by using test data. Table 6 shows the performance of all base-models and DS-MLP on Dataset2.

TABLE 6. The DS-MLP model results on Heterogeneity Human Activity Recognition Dataset.

Model	Model Type	Accuracy
MLP 1	base-model	0.981
MLP 2	base-model	0.980
MLP 3	base-model	0.978
MLP 4	base-model	0.982
MLP 5	base-model	0.984
DS-MLP	stacked model	0.994

Results shown in Table 6 indicate that the performance of each base model is high for Dataset2 than that of Dataset1. Since base models perform better, the accuracy of the proposed DS-MLP is also elevated for the 2nd dataset, i.e., 0.994. Classification results for each activity in Dataset2 are given in Table 7. Results are indicative of enhanced accuracy, with walking activity having the highest accuracy. Even though a balanced dataset i.e., an approximately equal number of samples are used for training and testing, various activities have disparate accuracy score. Accuracy is affected due to the features less distinctive and many activities like sit and bike may have similar data features. Despite that, the proposed approach performs well and achieves very good average accuracy.

TABLE 7. The DS-MLP model results for each activity in Dataset2.

Activity	Total Prediction	Correct Prediction	Wrong Prediction
Total Examples	3600	3579	21
Bike	608	608	0
Sit	617	610	7
Stairsdown	596	591	5
Stairsup	575	568	7
Stand	616	616	0
walk	588	586	2

B. COMPARISON WITH OTHER MACHINE LEARNING MODELS

We also used machine learning models to perform comparison between the results if machine learning models and the proposed DS-MLP. Five machine learning models such as logistic regression (LR) [35], [36], random forest (RF) [37], [38], decision tree (DT) [39], support vector machine (SVM) [40], K nearest neighbour (KNN) [41] and Gaussian Naive Bayes (GNB) [42] algorithms are used for this purpose.

We use RF with 300 n_estimators which means that 300 decision trees are constructed by RF on each example to give the predictions and then perform majority voting between 300 predictions to make the final prediction. SVM with the 'linear' kernel is better when data is linearly separable and is preferred to be used when the dataset has large features as this study has 561 features. LR is used with the 'liblinear' solver because it is considered fast for computation and 'multinomial' parameter which is considered better for multi-class classification problems. DTC is used with a max_depth parameter that reduces the complexity of the tree. We use this parameter with a 300 value which means that the decision tree will grow maximum to 300 level depth. We use KNN with n_neighbors parameters and set its value equal to 6. We show all the hyper-parameters of machine learning models in Table 8.

TABLE 8. Hyper-parameters for machine learning algorithms.

Models	Hyper-parameters
LR	solver='liblinear', multi_class='multinomial'
SVM	kernel='linear', C=2.0
RF	n_estimators=300, random_state=5, max_depth=300
GNB	Used with default setting
DTC	max_depth=300
KNN	n_neighbors=6, weights='uniform'

Table 9 shows the comparison between the machine learning models and DS-MLP. Results indicate that the DS-MLP outperforms all other machine learning algorithms in terms of accuracy, precision, recall, F1 score, CK, and MCC. The performance of the statistical machine learning model LR is close to that of the DS-MLP's. On the other hand, DT is the worst performer in this study with a 0.852 accuracy score.

TABLE 9. Comparison between DS-MLP and other learning algorithms.

Model	Accuracy	Precision	Recall	F_1 score	CK	MCC
LR	0.960	0.964	0.956	0.959	0.953	0.953
SVM	0.950	0.949	0.947	0.943	0.940	0.941
RF	0.924	0.931	0.922	0.926	0.911	0.912
GNB	0.879	0.876	0.876	0.874	0.851	0.853
DT	0.852	0.855	0.851	0.852	0.848	0.849
KNN	0.889	0.890	0.886	0.881	0.868	0.864
DS-MLP	0.973	0.977	0.966	0.971	0.970	0.969

TABLE 10. Models performance on Heterogeneity Human Activity Recognition Dataset.

Model	Accuracy	Precision	Recall	F_1 score	CK	MCC
LR	0.681	0.6824	0.679	0.679	0.599	0.597
SVM	0.822	0.823	0.821	0.821	0.787	0.781
RF	0.981	0.983	0.981	0.981	0.978	0.979
GNB	0.802	0.797	0.81	0.797	0.746	0.750
DT	0.950	0.949	0.950	0.949	0.936	0.936
KNN	0.967	0.966	0.965	0.995	0.956	0.957
DS-MLP	0.994	0.994	0.993	0.993	0.992	0.992

Table 10 shows comparison results for 2nd dataset used in this study. It reveals that the performance of the DS-MLP is better than those of machine learning algorithms. While tree-based models perform well on this dataset but the stacked learning technique DS-MLP achieves an accuracy of 0.994 which is higher than those of other models.

Figure 15 shows the performance of machine learning modes on both datasets. Here dataset 1 is the 'Human Activity Recognition Using Smartphones Data Set' and dataset 2 is the 'Heterogeneity Human Activity Recognition Dataset'. Figure 15 illustrates the significant performance of DS-MLP as compared to other models. DS-MLP achieves high accuracy on both datasets which shows that this model can perform well on these types of datasets and also proves that the ensemble learning can perform better than those of individual models.

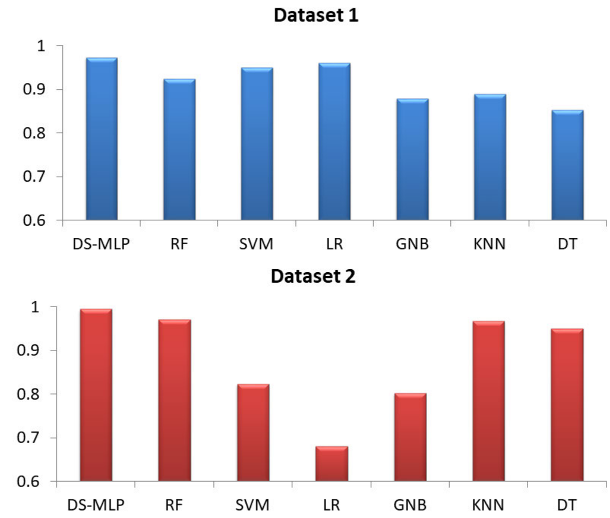

FIGURE 15. Models performance comparison on both datasets.

C. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART STUDIES

We carried out a performance comparison of the proposed DS-MLP with other state-of-the-art approaches proposed for HAR classification. For this purpose, studies including [43]–[46] that use deep convolutional neural networks (CNN) and the study [47] that uses deep convolutional long short term memory (ConvLSTM) are selected. The selection of these studies is based on the similarity of the task, i.e. HAR classification, and the used technique, i.e. deep neural networks. Besides, six other approaches have been added to the performance comparison. Four of these approaches are based on a stacked ensemble, similar to ours, while two others follow an ensemble of deep neural networks with various machine learning algorithms like SVM, RF, etc. Table 11 contains the results and the models of the above-mentioned studies. Results demonstrate that the proposed DS-MLP outperforms the state-of-the-art HAR classification techniques.

Performance analysis suggests that the approaches based on deep CNN have an accuracy between 0.90 to 0.95. Ensemble approaches, on the other hand, tend to have higher accuracy than that of simple deep learning approaches. Moreover, the stacked ensemble proves to show elevated accuracy in ensembles. The accuracy of the proposed DS-MLP is 0.973, which is higher than that of state-of-the-art approaches for HAR tasks. This accuracy is for Dataset1, while DS-MLB can achieve an accuracy of 0.994 with 2nd dataset used

TABLE 11. Performance comparison of the DS-MLP model and state-of-art approaches.

Model	Reference	Accuracy
[43]	Deep CNN	0.946
[44]	Deep CNN	0.951
[45]	Deep CNN	0.947
[46]	Deep CNN	0.900
[47]	Deep ConvLSTM	0.958
[16]	Residual Bi LSTM	0.905
[17]	Multiview stacking	0.925
[18]	Stacked LSTM	0.930
[19]	SDAE+GBM	0.959
[20]	Stacked ensemble	0.960
[21]	Stacked ensemble	0.968
DS-MLP	Deep Stacked Ensemble	0.973

for training and testing. Performance is elevated due to the stacked ensemble-based structure. The meta-model model in DS-MLP is trained on the results of base-models which makes it more efficient and accurate in comparison to other learning models.

D. PERFORMANCE COMPARISON WITH ENSEMBLE MODEL

In addition to traditional machine learning approaches and state-of-the-art approaches, this study analyzes the performance of the proposed approach against [22] which performs HAR using an ensemble of ELMs. The current research follows a similar approach to that given in [22] with few striking differences.

- The TRIGNOTM wireless system is used to capture data for various activities. Its practical application in the real life is difficult as the person needs to carry the device. The present study, on the other hand, utilizes the smartphone sensors data which is easy, convenient, and simple to implement in real-life scenarios.
- The data used for training and testing is collected from five participants only with ages between 20 to 36 years. Consequently, the dataset is not diversified. The data used in the current study contains data from 30 participants between the ages of 19 to 48 years.
- Research [22] uses a triaxial accelerometer sensor to collect the activity data while the current study used data from accelerometer and gyroscope sensors.
- Research [22] performs the training and testing on a single dataset and requires further analysis to generalize the results. At the same time, the current study analyzes the performance of the proposed approach on two different datasets which makes the results generalized.
- The research under discussion used 100 to 150 based classifiers (ELMs) initially which were later pruned. The use of such a large number of base classifiers makes the model complex, computationally expensive, and pruning time-consuming. Our model can achieve higher accuracy with only six base learners on datasets that are diversified and complex than the dataset used in [22].

TABLE 12. Performance comparison of ELM model [22] and the proposed DS-MLP model.

Model	Accuracy	
	Dataset 1	Dataset 2
ELM [22]	0.72	0.61
DS-MLP	0.992	0.973

The model given in [22] which is based on ELM is implemented in the current study to analyze its efficacy. ELM is similar to a neural network with few differences like hidden units, and weight calculation mechanism, etc. The number of hidden units is higher in neural networks than that of ELM. Similarly, the weights from input to hidden layer are generated randomly and the output neurons are linear in ELM [48], [49]. The ELM model is built with 1000 hidden units and 561 input shapes. After that, the dot product between inputs and inputs-to-hidden layers weights are calculated, and ReLU is used as the activation function. We compute output weights and minimize the least square error between the predicted labels and the training labels. The model is tested with both the datasets that are used in the current study. Results are given in Table 12.

Results demonstrate that our model performs much better than that of [22]. The primary reason for performance degradation of [22] is the dataset used for training and testing. Research [22] used the dataset that was collected with hung sensors that generate smooth data with clear distinction in features for various activities. On the other hand, the data that are collected from the smartphone sensors are noisy and features are less distinctive. Consequently, the performance of the ensemble model presented in [22] is affected adversely.

V. CONCLUSION

The study proposes a Deep Stacked Multilayered Perceptron (DS-MLP) model for human activity recognition tasks. The implementation uses five MLP models as base-learners and a neural network as a meta-learner. Evaluation is done on two different datasets that contain the data for various human activities like ‘walking’, ‘sitting’, ‘biking’, and ‘stairsup’, etc. The gathered data is from the accelerometer and gyroscope sensors of the user’s smartphone. Performance is evaluated using the measures of accuracy, precision, recall, F1 score, as well as, Cohen’s Kappa and Mathew Correlation Coefficient.

Experimental results demonstrate that the proposed model can achieve an accuracy of 0.973 and 0.994 for two datasets, respectively. Performance comparison is made with five most commonly used machine learning algorithms including logistic regression, decision trees, support vector machine, k nearest neighbor, and Gaussian naive Bayes. Comparison analysis indicates that the DS-MLP performs better than that of selected machine learning algorithms. Although statistical machine learning models like LR and SVM also perform well, yet, the accuracy of the DS-MLP is high.

Moreover, performance analysis with eleven state-of-the-art HAR approaches shows that the DS-MLP yields high accuracy. Stacking improves the efficiency of HAR tasks than those of individual models. Another finding is that the DS-MLP performs better on both large features set, as well as, small features set. It is safe to conclude that stacked ensemble learning can be adapted to achieve higher performance in multi-class classification problems for HAR and pattern recognition tasks.

ACKNOWLEDGMENT

(Furqan Rustam, Aijaz Ahmad Reshi, and Imran Ashraf contributed equally to this work.)

REFERENCES

[1] E. Turban and L. E. Frenzel, *Expert Systems and Applied Artificial Intelligence*. Upper Saddle River, NJ, USA: Prentice-Hall Professional Technical Reference, 1992.

[2] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, ''Deep learning for sensor-based activity recognition: A survey,'' *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.

[3] J. K. Aggarwal and L. Xia, ''Human activity recognition from 3D data: A review,'' *Pattern Recognit. Lett.*, vol. 48, pp. 70–80, Oct. 2014.

[4] C. Jobanputra, J. Bavishi, and N. Doshi, ''Human activity recognition: A survey,'' *Procedia Comput. Sci.*, vol. 155, pp. 698–703, Aug. 2019.

[5] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, ''A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities,'' in *Proc. IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2015, pp. 1–6.

[6] J. Qin, L. Liu, Z. Zhang, Y. Wang, and L. Shao, ''Compressive sequential learning for action similarity labeling,'' *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 756–769, Feb. 2016.

[7] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, ''Physical human activity recognition using wearable sensors,'' *Sensors*, vol. 15, no. 12, pp. 31314–31338, Dec. 2015.

[8] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu, and X. Li, ''Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori,'' *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 3992–4001, Jul. 2020.

[9] O. D. Lara and M. A. Labrador, ''A survey on human activity recognition using wearable sensors,'' *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[10] M. Inoue, S. Inoue, and T. Nishida, ''Deep recurrent neural network for mobile human activity recognition with high throughput,'' *Artif. Life Robot.*, vol. 23, no. 2, pp. 173–185, Jun. 2018.

[11] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, ''LSTM networks for mobile human activity recognition,'' in *Proc. Int. Conf. Artif. Intell., Technol. Appl.* Paris, France: Atlantis Press, 2016, pp. 1–4.

[12] A. Bulling, U. Blanke, and B. Schiele, ''A tutorial on human activity recognition using body-worn inertial sensors,'' *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, Jan. 2014.

[13] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, ''Deep activity recognition models with triaxial accelerometers,'' in *Proc. Workshops 30th AAAI Conf. Artif. Intell.*, 2016, pp. 8–13.

[14] A. Bevilacqua, K. MacDonald, A. Rangarej, V. Widjaya, B. Caulfield, and T. Kechadi, ''Human activity recognition with convolutional neural networks,'' in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2018, pp. 541–552.

[15] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, ''InnoHAR: A deep neural network for complex human activity recognition,'' *IEEE Access*, vol. 7, pp. 9893–9902, 2019.

[16] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, ''Deep residual bidir-LSTM for human activity recognition using wearable sensors,'' *Math. Problems Eng.*, vol. 2018, pp. 1–13, Dec. 2018.

[17] E. Garcia-Ceja, C. E. Galván-Tejada, and R. Brena, ''Multi-view stacking for activity recognition with sound and accelerometer data,'' *Inf. Fusion*, vol. 40, pp. 45–56, Mar. 2018.

[18] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, ''Stacked lstm network for human activity recognition using smartphone data,'' in *Proc. 8th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Oct. 2019, pp. 175–180.

[19] X. Gao, H. Luo, Q. Wang, F. Zhao, L. Ye, and Y. Zhang, ''A human activity recognition algorithm based on stacking denoising autoencoder and lightGBM,'' *Sensors*, vol. 19, no. 4, p. 947, 2019.

[20] A. Nandy, J. Saha, and C. Chowdhury, ''Novel features for intensive human activity recognition based on wearable and smartphone sensors,'' *Microsyst. Technol.*, vol. 26, pp. 1889–1903, Jan. 2020.

[21] A. Bhavan and S. Aggarwal, ''Stacked generalization with wrapper-based feature selection for human activity recognition,'' in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 1064–1068.

[22] Y. Tian, J. Zhang, L. Chen, Y. Geng, and X. Wang, ''Selective ensemble based on extreme learning machine for sensor-based human activity recognition,'' *Sensors*, vol. 19, no. 16, p. 3468, 2019.

[23] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, ''A public domain dataset for human activity recognition using smartphones,'' in *Proc. Esann*, 2013, p. 3.

[24] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, ''Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition,'' in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, 2015, pp. 127–140.

[25] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA, USA: MIT Press, 2020.

[26] Z. Ma, P. Wang, Z. Gao, R. Wang, and K. Khalighi, ''Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose,'' *PLoS ONE*, vol. 13, no. 10, pp. 1–12, 2018.

[27] S. B. Driss, M. Soua, R. Kachouri, and M. Akil, ''A comparison study between MLP and convolutional neural network models for character recognition,'' *Proc. SPIE*, vol. 10223, May 2017, Art. no. 1022306.

[28] O. Tarkhaneh and H. Shen, ''Training of feedforward neural networks for data classification using hybrid particle swarm optimization, mantegna Lévy flight and neighborhood search,'' *Heliyon*, vol. 5, no. 4, Apr. 2019, Art. no. e01275.

[29] A. Manik, A. Adiwijaya, and D. Q. Utama, ''Classification of electrocardiogram signals using principal component analysis and Levenberg Marquardt backpropagation for detection ventricular tachyarrhythmia,'' *J. Data Sci. Appl.*, vol. 2, no. 1, pp. 29–37, 2019.

[30] C. Banerjee, T. Mukherjee, and E. Pasiliao, ''An empirical study on generalizations of the ReLU activation function,'' in *Proc. ACM Southeast Conf.*, 2019, pp. 164–167.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ''Dropout: A simple way to prevent neural networks from overfitting,'' *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. R. Rao, ''Deep learning based representation for face recognition,'' in *ICCCE 2019*. Singapore: Springer, 2020, pp. 419–424.

[33] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[34] J. He, I. Pedroza, and Q. Liu, ''Metanet: A boosting-inspired deep learning image classification ensemble technique,'' in *Proc. Int. Conf. Image Process., Comput. Vis., Pattern Recognit. (IPCV)*, 2019, pp. 51–54.

[35] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, ''Classification of shopify app user reviews using novel multi text features,'' *IEEE Access*, vol. 8, pp. 30234–30244, 2020.

[36] A. P. López, N. C. Ramírez, E. M. Montes, and A. G. Hernández, ''Discriminative parameter learning of Bayesian networks using differential evolution: A preliminary analysis,'' *Res. Comput. Sci.*, vol. 149, no. 5, pp. 75–82, May 2020.

[37] L. Yang and A. Shami, ''On hyperparameter optimization of machine learning algorithms: Theory and practice,'' *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020.

[38] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, ''Tweets classification on the base of sentiments for US airline companies,'' *Entropy*, vol. 21, no. 11, p. 1078, Nov. 2019.

[39] H. Al-Hadeethi, S. Abdulla, M. Diykh, R. C. Deo, and J. H. Green, ''Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications,'' *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113676.

[40] W. S. Noble, ''What is a support vector machine?'' *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.

[41] A. Lin, Q. Wu, A. A. Heidari, Y. Xu, H. Chen, W. Geng, Y. Li, and C. Li, "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.
[42] I. Rish, "An empirical study of the naive Bayes classifier," in *Proc. IJCAI Workshop Empirical Methods Artif. Intell.*, 2001, vol. 3, no. 22, pp. 41–46.
[43] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.
[44] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015, pp. 1307–1310.
[45] C. A. Ronao and S.-B. Cho, "Deep convolutional neural networks for human activity recognition with smartphone sensors," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2015, pp. 46–53.
[46] C. A. Ronaoo and S.-B. Cho, "Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors," *Korean Inf. Sci. Soc.*, vols. 6–215, pp. 858–860, 2015.
[47] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
[48] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 985–990.
[49] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

FURQAN RUSTAM received the M.C.S. degree from the Department of Computer Science, The Islamia University of Bahawalpur, Pakistan, in October 2017. He is currently pursuing the master's degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is also serving as a Research Assistant for the Fareed Computing Research Center, KFUEIT. His research interests are related to data mining, machine learning, and artificial intelligence, mainly working on creative computing, and supervised machine learning.

AIJAZ AHMAD RESHI (Member, IEEE) is currently working as an Assistant Professor with the Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia. His research interests are related to machine learning and deep learning-based IoT, and text mining tasks.

IMRAN ASHRAF received the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in 2018, and the M.S. degree in computer science from the Blekinge Institute of Technology, Sweden, in 2010. He worked as a Postdoctoral Fellow with Yeungnam University, Gyeongsan, South Korea. He is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University. His research interests include indoor positioning and localization, indoor location-based services in wireless communication, smart sensors (LIDAR) for smart cars, and data mining.

ARIF MEHMOOD received the Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in November 2017. He is currently working as an Assistant Professor with the Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. His research interests are related to data mining, mainly working on AI and deep learning-based text mining, and data science management technologies.

SALEEM ULLAH was born in Ahmedpur East, Pakistan, in 1983. He received the B.Sc. degree in computer science from The Islamia University of Bahawalpur, Pakistan, in 2003, the M.I.T. degree in computer science from Bahauddin Zakariya University, Multan, in 2005, and the Ph.D. degree from Chongqing University, China, in 2012. From 2006 to 2009, he worked as a Network/IT Administrator in different companies. From August 2012 to February 2016, he worked as an Assistant Professor with The Islamia University of Bahawalpur. He has been working as an Associate Dean with the Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, since February 2016. He has almost 14 years of industry experience in the field of IT. He is also an Active Researcher in the field of adhoc networks, IoTs, congestion control, data science, and network security.

DOST MUHAMMAD KHAN received the M.Sc. degree (Hons.) in computer science from Bahauddin Zakariya University (BZU), Multan, in 1990, and the Ph.D. degree from the School of Innovative Technologies and Engineering (SITE), University of Technology, Mauritius (UTM), in 2013. He is currently working as an Assistant Professor with the Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Pakistan. His research interests include data mining and data mining techniques, multiagent systems (MASs), object-oriented database systems, and formal methods in software engineering.

GYU SANG CHOI received the Ph.D. degree from the Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA, in 2005. He was a Research Staff Member with the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics Company Ltd., from 2006 to 2009. Since 2009, he has been a Faculty Member with the Department of Information and Communication, Yeungnam University, South Korea. His research interests include non-volatile memory and storage systems.