Research Article

A customized 1D-CNN approach for sensor-based human activity recognition

Shilpa Ankalaki[1]* and Thippeswamy M. N.[2]

Research Scholar, Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India-560064[1]

Professor and Head, Department of Computer Science and Engineering, (AI&ML and Cyber Security), Ramaiah Institute of Technology, MSR Nagar, Bangalore, Karnataka, India-560054[2]

Abstract

Sensor-based human activity recognition (HAR) plays a major role in healthcare and security applications. The significance of this study is to understand the state-of-art of techniques for recognizing the activities of humans based on physiological signals acquired by a body-worn sensors. Accurate recognition of activities with just signals from the wearable sensor is a difficult task due to the inherent complexity of physical activities. Even though sensor-based HAR has been accomplished by using various algorithms exploiting machine and deep learning techniques, just a handful of researchers have made an extensive study on the contribution of various parameters on the accuracy of recognition of human activity. The main focus of this study is to perform a comparative evaluation of the state-of-the-art algorithms based on machine and deep learning techniques that have been proposed for HAR. Principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) methods are employed for data dimensional reduction and visualization. Machine learning algorithms like random forest (RF), kernel-based support vector machine (SVM), and deep learning algorithms such as convolutional neural networks (CNN) have been applied on the University of California, Irvine (UCI) HAR dataset. A comprehensive study has been conducted to understand the impact of changing parameters like pooling, activation functions, number of dense layers, and dropout percentage of CNN on the accuracy of recognition. The proposed work employed the swish activation function in the dense layer which was recently proposed by Google. The output layer of the last dense network employs the softmax function to classify human activities. The proposed CNN architecture (with max-pooling layer, the swish activation function in one dense layer, and softmax function at the output layer) achieved results of training and validation accuracy as 99.58% and 92.57% respectively for HAR.

1.Introduction

Human activity recognition (HAR) is an emerging area of research that is applicable in several fields such as health care, security, smart homes, fitness assessment, sports competition and assisted living. HAR is performed by acquiring data from wearable sensors and external sensors like cameras, pressure sensors, and video recorders. Vision-based HAR plays an active role in the detection of suspicious activity and events in video surveillance. The requirement of computationally expensive image and video analytics, expensive hardware, and lack of portability are the main limitations of vision-based HAR [1].

HAR system using wearable sensors overcomes these limitations. The contextual information required for HAR can be inferred using the raw data collected from various heterogeneous sensors which are placed at different positions on the human body and the sensors available in smartphones [2]. There are a couple of challenges associated with the processing of raw data obtained from different sensors. The first and foremost challenge is to devise a method for creating a discriminative feature space using the data gathered from the sensors. Due to its importance, many features engineering schemes have been proposed with the ultimate aim of improving the discriminatory function and consequently the accuracy of HAR [3]. The process of feature extraction involves the acquisition of signals from

*Author for correspondence

various sensors, processing them to extract data, segmenting the data, and determining the relevant features. The feature selection process is necessary in order to reduce the high dimensionality of data and provide feasible feature sets to enable activity classification easily. Capella et al. [3] have proposed Relief-F a fast correlation-based filter combined with a correlation-based feature selection method to address this issue. The engineered feature subsets were evaluated using naïve Bayes (NB), support vector machine (SVM), j48 decision tree classifiers. Another approach to reducing dimensionality is to employ principal component analysis (PCA) [4]. Unfortunately, PCA performs poorly, if input data is not normalized correctly, causing difficulty in accurate activity recognition. Researchers proposed an empirical cumulative distribution function (ECDF) to address the drawback of PCA. ECDF is an approach that has been proposed for feature representation that preserves structural information of features irrespective of the absolute feature range [3−5]. A major challenge is in evolving a method for synergistically combining the features from the multiple feature spaces in order to evolve one feature set which can consistently yield results of high accuracy. To address this challenge, researchers, such as in [6], have designed an approach to generate a unique set from the multiple set of features extracted from modality, position, and feature level data sets.

In the proposed work, the University of California, Irvine (UCI) HAR dataset is used. The dataset consists of a feature vector of length 561 attributes along with frequency and time domain variables with the respective class label of an activity. The t-Distributed stochastic neighbor embedding (t-SNE) method is used for feature representation and the results of t-SNE are compared with the results of PCA based feature learning method. Machine learning (ML) and deep learning (DL) algorithms are used for the classification of human activities.

Significant contributions and findings of the proposed work are as follows:
1. Analysis of PCA and t-SNE for feature reduction and visualization.
2. Evolution of a customized 1D-CNN for sensor-based HAR.
3. Performance analysis of ML algorithms like radial bias function (RBF) kernel-based SVM and random forest (RF) classifier for HAR on a UCI-HAR Dataset.
4. Performance analysis of different 1D-CNN architecture for HAR on a UCI-HAR dataset by

varying the parameters of CNN like its activation function, pooling function, and dropout percentage.

The rest of this article is arranged as follows: Section 2 Literature survey: It briefs about the state of art ML and DL techniques employed to accomplish HAR. Section 3 Materials and Methods: It describes dimensionality reduction methods, namely PCA and t-SNE; ML approaches and customized 1D-convolutional neural network models for HAR. Section 4 Experimental Results and Analysis: This section provides experimental results of the aforementioned ML and customized 1D-convolutional neural network models on a UCI-HAR benchmark datasets.

2.Literature survey
People have recognized the need for assessing their lifestyles and changing them as required in order to improve their quality of life. In this context, HAR plays a significant role in creating an accurate profile of the person based on which his lifestyle can be assessed and due adjustments can be done. HAR monitors the various activities of people using a wide range of hardware sensors, processing elements, and software components. These systems have evolved over a period of time incorporating more facilities and functionalities [7]. HAR plays a major role in assessing and improving the quality of life in both ambients assisted living (AAL) and activities of daily living (ADL) functions. The main objective of HAR with respect to the ADL is to recognize human activities by developing predictive models which assist in the classification of an individual's behavior as either normal or abnormal [8].

HAR can be accomplished by using a couple of approaches, namely manual feature extraction using handcrafted features combined with ML, and automatic feature extraction using DL techniques [9].

Ahmed et al. [10] have proposed an efficient and robust feature extraction scheme for processing sensor data to perform HAR. Researchers have employed enveloped power spectrum (EPS) for feature extraction. The feature size is reduced in size to a minimal set of discriminants using linear discriminant analysis (LDA). The reduced feature set based scheme was trained by using a multi-class support vector machine (MCSVM) on a UCI-HAR and DU-MD benchmark datasets. It was observed that an accuracy of 98.67% and 100% of UCI-HAR and DU-MD datasets respectively, was achieved.

Irvine et al. [11], proposed a neural network ensemble method for data-driven HAR, where human activities are identified as belonging to one of the four groups (models) namely morning, afternoon, evening, and miscellaneous (mixed model), and labelled accordingly. Each model is trained separately by a multi-layer perceptron (MLP) as a base classifier. The results of all MLP base classifiers are fused and subsequently, conflicts if any are resolved. The dataset for this research was created in UJAmI Smart Lab a set of binary sensors.

A comprehensive study has been carried out on various approaches for HAR, a few of them are summarized in *Table 1* [12].

Table 1 Different approaches for HAR

Approaches for HAR	Types of Sensors & Purpose	Advantages	Limitations
Computer Vision	Kinect sensors [13] – Identification of human activities based on the skeleton data extracted from the human body.	Provides good skeletal information for HAR.	Does not permit excessive mobility
RFID Based [14]	RFID – real-time object localization and then determining their movements and interactions	Identification of actions like making a sandwich, preparing coffee or tea, fetching a bowl of cereal	Capable of recognizing only a limited number of actions.
Motion-Based approach [12]	Accelerometer - senses the acceleration of the body in 3D (X, Y, and Z rotate). Used for estimating the rate of increase in speed to detect a free-fall condition.	Better in terms of mobility and simplicity.	Various kinds of noise in the data collected by wearable sensors complicate the task of activity classification [15].

HAR using ML and DL techniques has been the focus of a number of research groups in the recent past. Significant results of a selected few of them are summarized in this paper.

Anguitia et al. [16] focuses on the identification of movement of a person using sensors embedded in smartphones.

Le and Van [17] used naïve Bayesian (NB) and Decision tree classifiers to train time and frequency domain features for HAR. Motion-based activities like walking, jumping, and running has been detected and classified using NB and k-nearest neighbours (k-NN) [18].

Fazli et al. [19] proposed a hierarchical classification system to classify human activities based on a different level of abstraction using neural networks. In this work, activities are mainly categorized into flat and hierarchical activities. Hierarchical activities are initially classified into stationary and non-stationary in the first phase and deeper analysis in the second phase provides the final label of the activity.

De et al. [20] evaluated the accuracy provided by five HAR classification techniques. For this evaluation, input signals were acquired using the MTx miniature magnetic and inertial measurement unit (MIMU) sensor manufactured by Xsens Technologies. A 5-s sliding window with no overlap was used to segment the input signal. Twenty-one-time domain, three frequency domain, and fourteen time-frequency domain features were extracted for each window. The evaluation showed that the result of the k-NN method had an accuracy of 97%. This outperformed feedforward neural network (FNN) (95.8%), SVM (96.6%), NB (96.5%), and decision tree (91%) for classification of eight activities. Such ML algorithms depend on handcrafted features. As there might be considerable overlap and hence a lower degree of discrimination in handcrafted features the ML algorithms may fail to classify various activities accurately. The design of highly discriminatory handcrafted features relies on the expertise of a high order and the typical techniques used to aid in handcrafting are: assessing the statistics of raw data, transform coding, and symbolic representation [21].

The critical issue of requiring a high degree of expertise for designing handcrafted feature sets can be overcome by using a CNN for feature extraction. Such features are referred to as black box features [6].

Ignatov [22] proposed an online HAR with a user-independent deep learning approach. The authors used both statistical features and local features

extracted by CNN. The experiment was conducted on the wireless sensor data mining (WISDM) and UCI datasets.

Cho and Yoon [23], proposed a test data sharpening method along with divide and conquer based 1D-CNN for HAR. The exercise was conducted using the OPPORTUNITY and UCI HAR datasets and as expected the accuracy of the 1D CNN with test sharpening was found to be better than 1D CNN without test sharpening.

Avilés-Cruz et al. [24] proposed a novel framework for performing HAR using feature-fusion-based multi-CNN. The framework consists of three parallel CNNs which consider the filtered image using the fine-CNN, medium-CNN, and coarse CNN to extract local features independently. Subsequently, the three sets of features are fused to form one holistic picture for the classification of perceived Human activities.

HAR is a challenging task due to the issues like significant variations in each of the actions, the resemblance between different actions, the presence of a large percentage of the Null class, and long computation time. To alleviate some of these problems recent research work has focused on deriving systematic feature set representations and efficient classification techniques.

Ordóñez and Roggen [25] have proposed a hybrid approach combining CNN and long short-term memory (LSTM) to address complications involved in sequential HAR and have achieved a high degree of precision. However, the efficiency of this hybrid approach is very poor and it rarely satisfies the real-time requirements. This low efficiency has been improved to some extent by integrating of extreme learning machine (ELM) into the CNN-LSTM [21].

The CNN-GRU is another hybrid deep learning scheme proposed by [26] for classifying both simple and complex human activities contained in the WISDM dataset and it has achieved an accuracy of 96.54%.

Uddin and Soylu [27] have applied HAR on healthcare datasets. They proposed a LSTM based deep neural structured learning (NSL) scheme for activity modeling and obtained an accuracy of 99% on the MHEALTH dataset.

Ye et al. [28] have proposed a hybrid approach combining deep learning with micro-electro-

mechanical system (MEMS) based sensors for HAR. In this study, researchers designed LSTM and CNN for smartphone posture recognition, pedestrian navigation, and real-time pedestrian activity recognition and obtained comparatively better accuracy than that obtained with traditional ML algorithms.

Gao et al. [29] proposed a multi-branch CNN, which utilizes a selective kernel convolutional mechanism for HAR. The selective kernel convolutional has been evaluated on benchmark datasets like UCI-HAR, PAMAP2, UNIMIB SHAR, WISDM, and OPPORTUNITY. Accuracy improvements of the order of 0.23%, 0.80%, 0.71%, and 0.31%, while processing UCI-HAR, UNIMIB SHAR, WISDOM, and OPPORTUNITY dataset respectively.

Usually, deep neural networks and CNN are trained with a global loss. Backward locking precludes memory reuse, which is a significant constraint for wearable activity recognition. To overcome this issue, the researcher of [30] proposed the Layer-wise CNN architecture with local loss and the architecture has been evaluated on WISDM, OPPORTUNITY, PAMAP2, UCI-HAR, and UniMib-SHAR datasets and obtained comparatively better results than the existing methods.

The performance of the CNN has been improved by an ensemble approach.

Mukherjee et al. [31] proposed an ensemble classifier called EnsemConvNet by combining three CNN-Net, Encoded-Net, and CNN-LSTM.

Upon the comprehensive study on the literature of HAR, it has been observed that ML algorithms greatly depend on handcrafted features which require a high degree of expertise in the domain to successfully build the HAR system. This limitation has been overcome by adopting deep learning techniques. The research on the deep learning CNN approach has more scope of the application of HAR. Many researchers have proposed CNN architecture which is complex in nature and ensemble. In this study, we are proposing the simple and yet effective customized 1-D CNN architecture and finetuning it by tuning the CNN parameters to accomplish the HAR task on the UCI-HAR benchmark dataset.

3.Materials and methods
On the comprehensive study of the literature research papers, it has been observed that RF [1, 13] and SVM

[10, 20, 32] algorithms comparatively outperform other machine learning algorithms for HAR on most of the benchmark datasets, hence these algorithms are employed in this work to evaluate HAR on the UCI-HAR dataset.

The methodology followed in this paper is outlined in *Figure 1*. This work is focused on the evaluation of traditional classification algorithms namely, RBF kernel-based SVM and RF, and a DL approach namely, 1-D CNN for HAR on the UCI-HAR dataset. *Figure 2* depicts the comprehensive workflow of the proposed system.
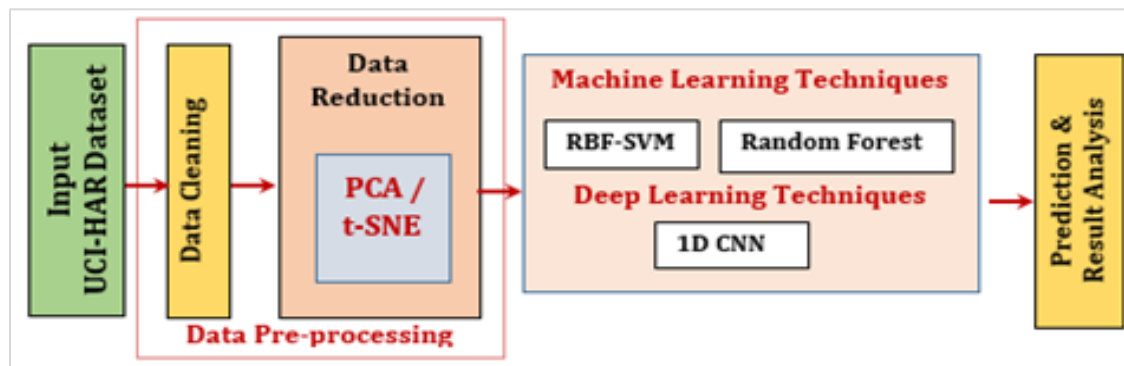
Figure 1 Workflow of proposed system

3.1 Input-UCI-HAR dataset

The UCI-HAR dataset was created by experimenting on 30 individuals who were in the age range of 19-48 years [16]. Each individual was asked to wear a smartphone and was asked to perform activities like walking, climbing down the stairs, climbing up the stairs, standing, sitting, and lying down. The data is periodically gathered from the accelerometer and gyroscope sensors embedded in the smartphone. Noise filters were applied to preprocess the sensor signals. Sensor signals were sampled by using fixed-width sliding windows of 2.56 sec with a 50% overlap. Each record in the dataset contains the following information:

"total acceleration, estimated body acceleration, triaxial angular velocity from the gyroscope, a 561-feature vector with time stamp, frequency domain variables, its activity label, and an identifier of the subject who was the participant in the experimented. The source link of the dataset is https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones#".

3.2 Data cleaning

The data cleaning has been conducted on the dataset by checking for duplicate values, null values, and data balance of data corresponding to each activity.

3.3 Data reduction

Data reduction has been performed by using popular data reduction techniques such as PCA [4] and t-SNE. PCA is a linear data reduction algorithm that

can be used to diminish the dimensions of the original features to extract optimal features. PCA is an unsupervised method (data without labeling), which employs correlation between features to recognize the patterns from the data. The purpose of using PCA is to produce a new subspace of features with a lesser dimension than that of the original feature set by preserving the significant features in the dataset [33]. The features representing the original data set are translated to the principal components that are linear combinations of the existing features. The first principal component is the feature that causes the greatest variance and the feature with the second-highest variance is the second principal component and so on. Steps involved in the PCA data reduction method are summarized in [33]. However, in case a non-linear high dimension feature dataset is needed to model the data, then modeling it using parameters determined by conducting PCA on the data set will generate a very bad model and will result in less accurate recognition results. This limitation can be overcome by performing data reduction using the t-SNE data reduction methodology.

t-SNE is presented by Van et al. [34]. It is a non-linear method for dimensionality reduction and visualization of data. In contrast to PCA, it is based on probability and it retains only local similarities whereas PCA aims to preserve large pair distances to maximize variance. On performing t-SNE on the UCI-HAR dataset, the data is clustered in such a way that each cluster is formed by grouping similar data

from disparate data, the resultant dataset will be simpler to analyze than the original dataset. The working of the t-SNE algorithm depends on a couple of parameters namely: (i) perplexity: denotes the number of neighbors of any data-point in high dimensional space, (ii) the number of iterations. The steps involved in the t-SNE algorithm are as follows [34]:

Algorithm: t-SNE algorithm
Input: Dataset: D= {d1, d2,…dn}
Parameters: Perplexity, No. of iterations, momentum $\alpha(t)$, and learning rate (η)
Output: Data representation in low dimensional space S(0) ={s1,s2,...,sn}
Start
Step 1: Calculate pairwise similarity measures in high dimensional space using Equation 1 and 2.

$$P_{ij} = \frac{P_{j|i} + P_{i|j}}{2n} \tag{1}$$

Where,

$$P_{j|i} = \frac{\exp(-\|d_i - d_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|d_i - d_k\|^2 / 2\sigma_i^2)} \tag{2}$$

Where σ_i is the variance of the Gaussian that is centered on data point x_i.

Step 2: Sample the preliminary solution S(0) ={s1,s2,...,sn} from N (0,10−4I)
Step 3: Repeat until till reaches the maximum number of iterations.
 a. Calculate low-dimensional similarity measure q_{ij} using Equation 3.

$$q_{ij} = \frac{(1 + \|s_i - s_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|s_k - s_l\|^2)^{-1}} \tag{3}$$

 b. Calculate gradient using Equation 4 and 5.

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(s_i - s_j)(1 + \|s_i - s_j\|^2)^{-1} \tag{4}$$

Set

$$s^{(t)} = s^{(t-1)} + \eta \frac{\delta C}{\delta y} + \alpha(t)(s^{(t-1)} - s^{(t-2)}) \tag{5}$$

 end
end

In the proposed work, t-SNE has been applied to the dataset, and results are analyzed by varying its parameters like the number of iterations and perplexity factor.
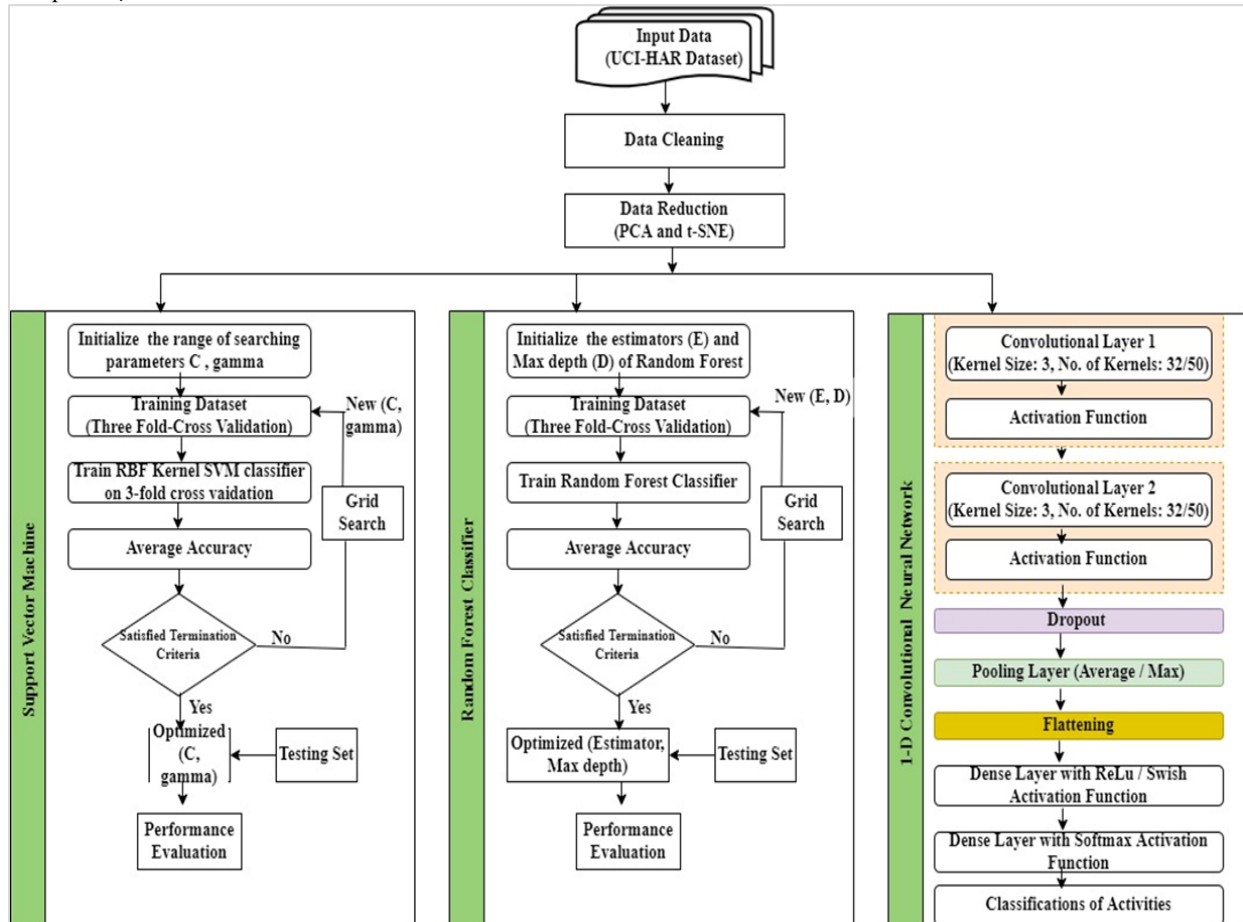

Figure 2 Comprehensive workflow of proposed system

3.4 Machine learning algorithms

Plenty of ML algorithms have been proposed as solutions for the classification of human activities. Important among them are SVM [1, 4, 5, 15, 35 and 36], decision tree and its variations [1, 4, 17 and 36], k-NN [1−5] [15, 17, 18 and 36], MLP [1, 5 and 12], NB [1, 4, 21, 14, 18 and 36], RF [1, 15]. RBF kernel-based SVM grid search and RF algorithms are used in the proposed work for activity recognition. RBF kernel-based SVM is a classifier for non-linearly separable data, which uses the RBF as the kernel function. The optimum hyperplane can be determined by the vector W in a nonlinear space, which reduces the function given in Equation 6. [37].

$$\varphi(W, \xi) = \frac{1}{2}\|W\|^2 + C \sum_{i=1}^{t} \xi_i \qquad (6)$$

Where ξ represents the slack variable and C represents the cost of misclassification. The representation of the kernel function is given by Equation 7.

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2) \qquad (7)$$

The selection of optimal parameters is a challenging task in SVM, which can be achieved by using the grid search SVM. In the case of RBF kernel-based

SVM, two parameters (C and γ) can be optimized using grid search.

Another ML algorithm that is used in the proposed work is RF Algorithm. It is an ensemble tree-based learning algorithm. In this, a cluster of decision trees from an arbitrarily selected subset of the training set is created. This is used to classify and label the object by taking aggregates of votes from various decision trees.

The flow of the SVM and RF algorithm is summarized in *Figure 2*.

3.5 Customized 1D-convolutional neural network

Customized 1D CNN is a DL algorithm that has overcome the requirement of handcrafted feature extraction. Proposed work employed 1-D CNN to accomplish the task of HAR using the UCI-HAR dataset. 1D CNN is used for the recognition of both static activities (sitting, lying, and standing) and dynamic activities (climbing stairs, walking, and descending stairs). *Figure 3* depicts the architecture of the CNN model employed to train both static and dynamic activities.
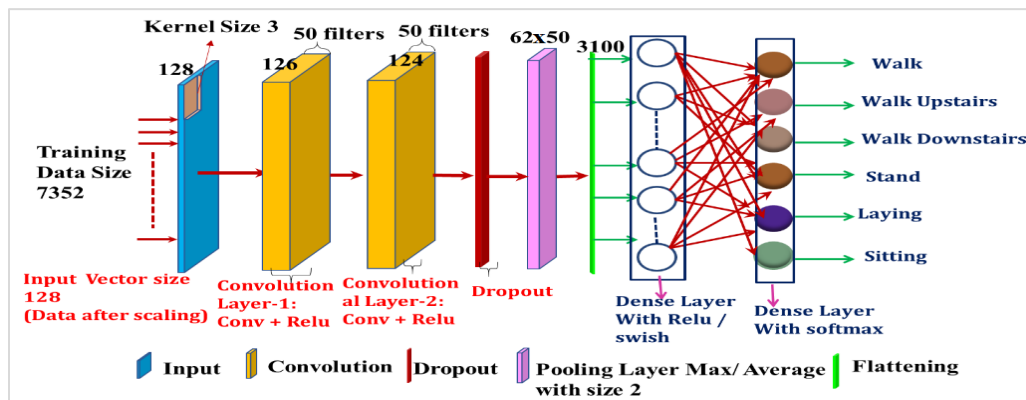

Figure 3 Architecture of customized 1D- convolutional neural network

As shown in *Figure 3*, 1-dimensional sensor data is passed through the two convolutional layers followed by the activation function. The convolutional layer is solely responsible for the feature extraction. Features are computed by performing the convolutional operations on input data using 1-dimensional kernels (filters).

The nonlinear activation functions like Rectified Linear Unit (ReLu) or Leaky Relu functions are applied to the resultant features obtained by the convolutional layer to make features highly nonlinear. To reduce the dimensionality, down

sampling has been performed on the feature matrix by using average pooling and max pooling. The resultant features are flattened into a vector and it is considered as trainable learning which will be given as input to the fully connected network. The commonly used approach for the fully connected network is MLP.

4. Experimental results and analysis
4.1 Data cleaning
UCI-HAR dataset does not contain duplicates and null values. Data distribution for all six classes is

almost balanced. *Figure 4* depicts the data distribution for all six classes.
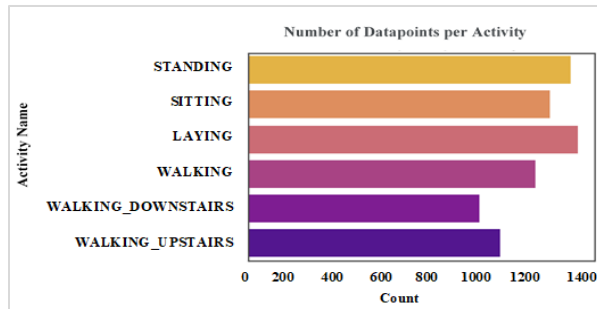

Figure 4 Data points distribution in each class

4.2Data reduction

PCA and t-SNE methods are used for dimension reduction of data. PCA has been applied to the dataset with 2 principal components (PC), the outcome of PCA is depicted in *Figure 5* which provides the visualization of data after reducing it to 2 principal components. It can be observed that it has failed to cluster similar class data in some cases. The x and y-axis of the graph represent PC1 and PC2 respectively.
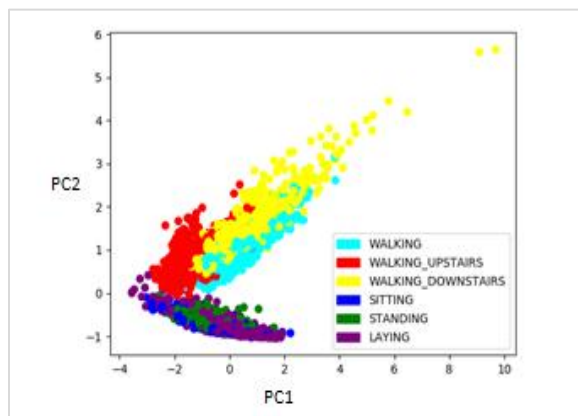

Figure 5 Data reduction using PCA with 2 components PC1 and PC2

To overcome the limitation of PCA, t-SNE has been used for data reduction and visualization. *Figure 6* depicts the visualization results of t-SNE with perplexities values of 10, 20, 30, 40, and 50 over 1000 and 2000 iterations respectively. P denotes the perplexity value and N represents the number of iterations. The results show that the t-SNE algorithm generates robust results in a perplexity value of 50 over 1000 iterations. *Figure 6 (i)* depicts the clear separation of static and dynamic activities by forming three clusters such that cluster-1 represents the

standing and sitting postures, while cluster-2 represents the activities for laying where both clusters 1 and 2 represents static activities. Cluster-3 represents all activities of walking such as walking, walking_upstairs, and walking_downstairs.

4.3RBF-Kernel-based support vector machine grid search approach

The proposed work employed the RBF kernel-based SVM using the grid search approach to optimize the parameter values of SVM. In this experiment, the RBF kernel is used with parameter C set iteratively to 2,6,10 and γ set to 0.0065123, 0.125, 2. *Figure 7* depicts the normalized confusion matrix of kernel-based SVM. The classifier correctly classifies all instances of the class "LAYING". The class "SITTING" has a total of 491instances, out of which 439 instances are classified correctly, 50 instances are misclassified as "STANDING" and 2 instances are misclassified as "WALKING_UPSTAIRS". Out of 532 instances of the class label "STANDING", 98% of (522 instances) total instances are classified correctly, remaining 2% of instances (10 instances) are misclassified as the class label "SITTING". The class label "WALKING" has a total of 496 instances, out of which 2 and 5 instances are misclassified as class labels "WALKING_DOWNSTAIRS" and "WALKING_UPSTAIRS" respectively. Out of 420 instances of the class "WALKING DOWNSTAIRS", 19 and 4 instances are misclassified as "WALKING_UPSTAIRS" and "WALKING" respectively. The last class "WALKING_UPSTAIRS" consists of 471 instances, out of which 17 instances are misclassified as label "WALKING" and the remaining 2 instances are misclassified as the class label "WALKING DOWNSTAIRS". The observation of the confusion matrix shown in *Figure 7*, clearly depicts that the model is biased more on the "SITTING" class as 50 instances are misclassified as "STANDING". *Figure 8* depicts the classwise precision, recall, and F1-score. Parameters C=10 and gamma=0.0065123 are the best estimators.

4.4Random forest

RF classifier has been applied on the dataset with following parameter values, 'n_estimators'= (10,201,20), max_depth=3,15,2. RF gives an accuracy of 90.83% with the best parameter estimators those are 'max_depth' =13 and 'n_estimators'=70. *Figure 9* depicts the normalized confusion matrix of the RF classifier. From *Figure 9* we can infer that the classifier correctly classifies all instances of the class "LAYING". The class

"SITTING" has a total of 491instances, out of which 443 instances are classified correctly, 48 instances are misclassified as "STANDING". Out of 532 instances of the class label "STANDING", 491 instances are classified correctly, remaining 41 instances are misclassified as the class label "SITTING". The class label "WALKING" has a total of 496 instances, out of which 10 and 6 instances are misclassified as the class label "WALKING_DOWNSTAIRS" and "WALKING_UPSTAIRS" respectively. Out of 420 instances of the class "WALKING DOWNSTAIRS", 22 and 43 instances are misclassified as "WALKING" and "WALIKNG_UPSTARIS" respectively. The last class "WALKING_UPSTAIRS" consists of 471 instances, out of which 44 instances are misclassified as label "WALKING" and the remaining 7 instances are misclassified as the class label "WALKING DOWNSTAIRS". The observation of the confusion matrix shown in *Figure 9* clearly depicts that the model is biased more on the "WALKING_DOWNSTAIRS" class as 65 instances are misclassified. Performance evaluation methods such as precision, recall, and f1-score are computed to analyze the performance of RF classifiers for each class of the dataset. *Figure 10* depicts the classwise precision, recall, and F1-score. The average precision, recall, and f1-score of the RF model are 0.93, 0.92, and 0.92 respectively. Data reduction and visualization results of t-SNE are comparatively better than PCA as t-SNE handles non-linear data and are capable of preserving the local and global structure of data.

From the analysis, we can conclude that precision, recall, and f1-score values of RBF-SVM are higher than the RF algorithm. Hence RBF-SVM algorithm with an accuracy of 96.23% outperform the RF algorithm which has an accuracy score of 92.5%. Compared to the RBF-SVM, the RF classification model has more incorrect predictions in dynamic activities such as "WALKING_DOWNSTAIRS and WALKING_UPSTAIRS". Hence, the RF classification model performs poorly when compared to the kernel-based RBF SVM model.
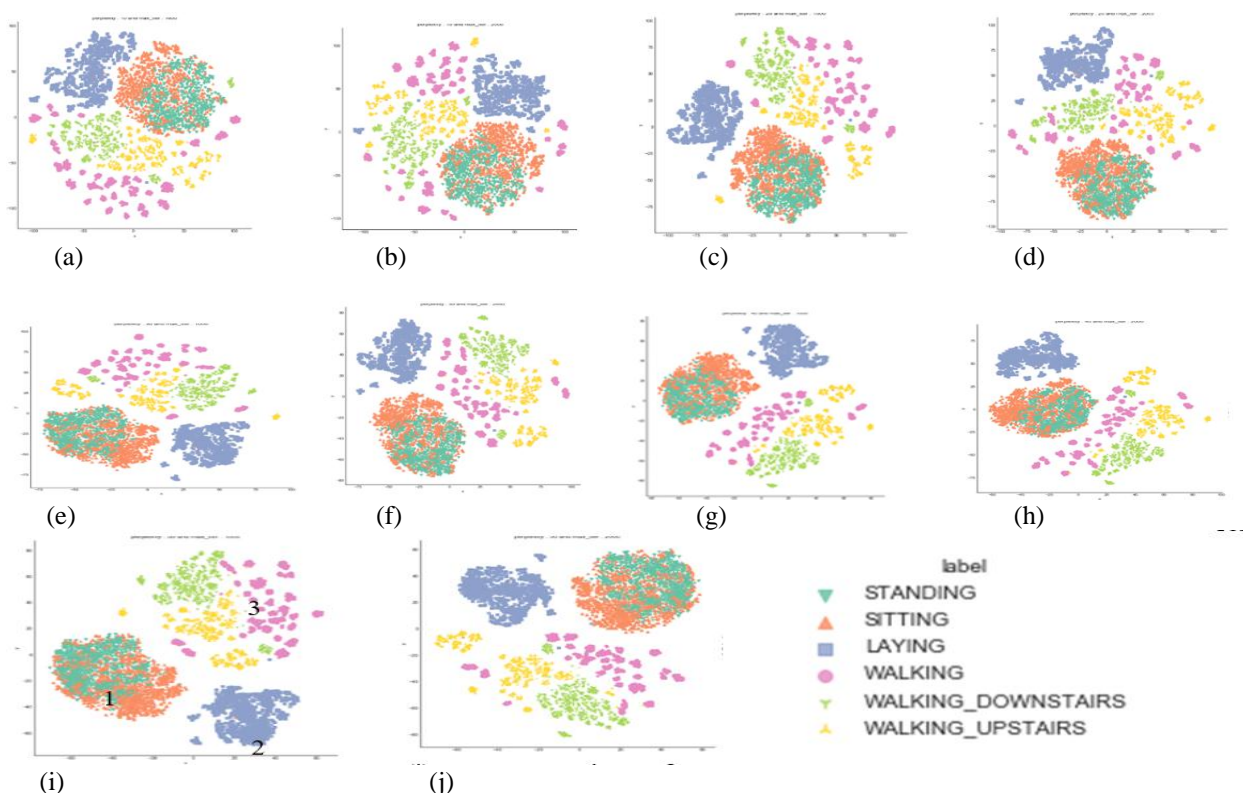

Figure 6(a-j) Visualization results of t-SNE [a: Demonstration of t-SNE result with P =10 N =1000 b: Demonstration of t-SNE results with P =10 N =2000 c: Demonstration of t-SNE result with P =20 N=1000 d: Demonstration of t-SNE result with P =20 N=2000 e: Demonstration of t-SNE result with P =30 N =1000 f: Demonstration of t-SNE result with P=30 N =2000 g:

Demonstration of t-SNE result with P=40   N =1000 h: Demonstration of t-SNE result with P=40   N=2000 i: Demonstration of t-SNE result with P =50   N=1000 j: Demonstration of t-SNE result with P=50   N=2000]
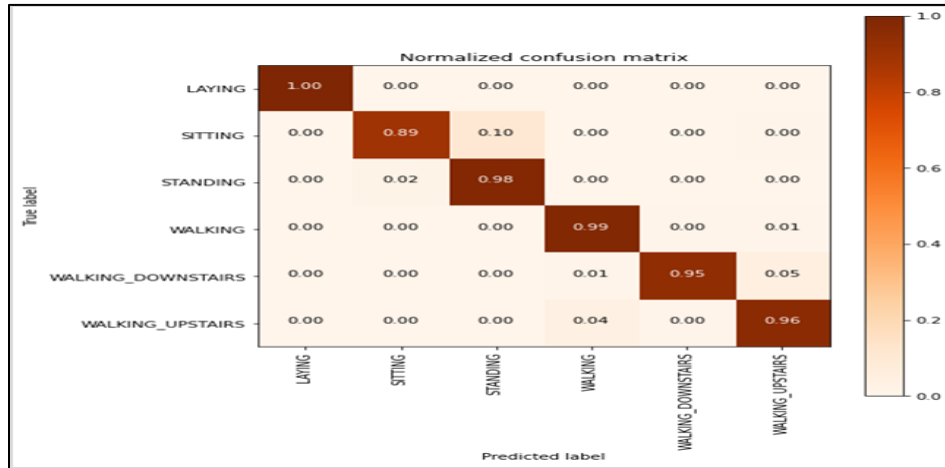

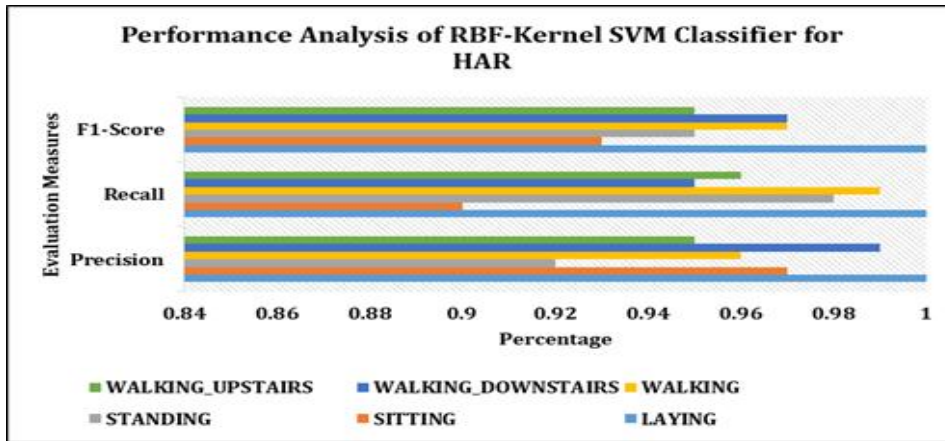**Figure 7** Normalized confusion matrix of RBF- SVM classifier

Figure 8 Performance analysis of RBF-Kernel SVM classifier
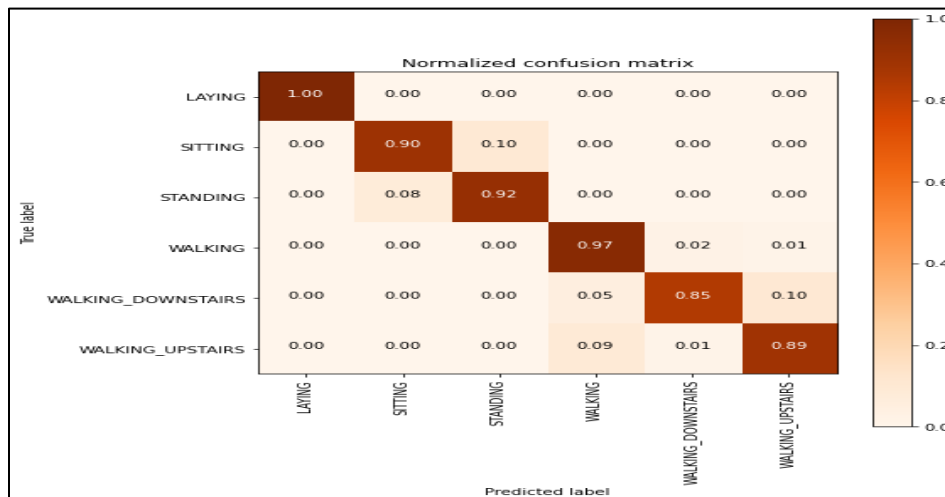

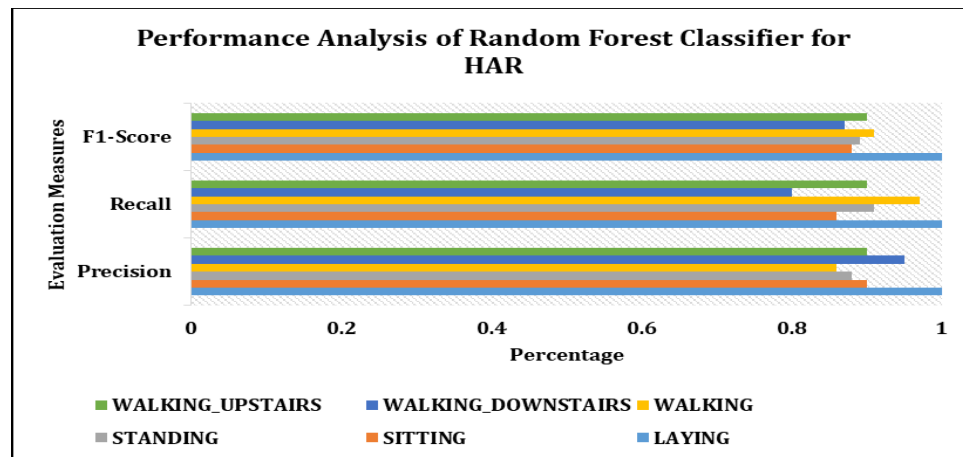**Figure 9** Normalized confusion matrix of RF classifier

Figure 10 Performance analysis of RF classifier

4.5 Activity recognition using customized 1D-CNN

The CNN method has been employed for human activity detection on the UCI-HAR dataset using the architecture given in *Figure 2* and *Figure 3*. An extensive experiment has been conducted by varying the parameters of CNN Architecture which is shown in *Figure 3*. The CNN architecture consists of two convolution layers with ReLu activation followed by drop out, pooling layer with average or max function, flattening, dense layer with ReLu or swish activation function, and last dense layer with softmax activation function for activity classification. The following are the training parameters used for training all architectures: Adam optimizer, number of epochs 30, and categorical cross-entropy (CCE) loss function. During the experimentation, there was an obvious increase in the validation loss of the model when the number of epochs was increased beyond the 30 epochs which caused the model overfitting. The softmax activation function [38, 39] is employed in the last layer of the network to compute the probability distribution of the given data over 'n' different classes. In the proposed work, we have considered six different activities, hence 6 classes are represented as 6 nodes in the output layer of the network. The softmax activation function is computed Equation 8.

$$Softmax(\vec{X}_i) = \frac{e^{X_i}}{\sum_{c=1}^{n} e^{X_c}} \; for \; j = 1, \dots, n \quad (8)$$

Where,

$\vec{X}$ is the input vector, e^{X_i} represents an exponential function of the input vector, n represents the number of classes and e^{X_c} represents the exponential function of the output vector.

In HAR using CNN, the error of the current state must be estimated repeatedly. In order to accomplish this,

the proposed system employs CCE to compute the loss /error of the model so that the weights can be updated accordingly to diminish the loss on subsequent evaluation. CCE measures the loss between the probability distribution of the target class and the predicted distribution of the current data of the model [40, 41]. CCE also depends on the activation function which is used in the output layer of the network as shown in *Figure 11* and CCE is computed using the relationship given in Equation 9. In this work, as softmax is used in the output layer the loss computed is also called softmax-loss.

$$CCE = -\sum_{i=1}^{n} y_i \cdot \log(f(X_i)) \quad (9)$$
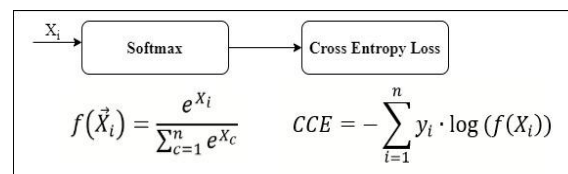

Figure 11 Categorical cross-entropy

The experiment has been carried out by varying the number of kernel functions in convolution layers, type of function for pooling, and type of activation function in the first dense layer of CNN. The detailed architectures of CNN used for the experiment are shown in *Table 2*.

The input vector of size 561 was subjected to scaling and reduced the dimensionality to 128. The input vector comprising of 128 attributes is fed into the CNN for all architectures used in the proposed work.

The architectures shown in *Table 2* have been applied to the benchmark dataset UCI-HAR dataset. *Figures 12* to *16* depict the results of performance analysis of the above-shown architectures in terms of training and

validation accuracy; and training and validation loss. So, in this experimentation, the number of epochs is restricted to 30. The CNN architecture 1-4 produced very good training and validation accuracy, but the validation loss of those architectures was huge compared to the training loss. Using architecture 1, we obtained the training accuracy of 99%, validation accuracy of 91.99%, 2.62% of training loss, and 75.18% of validation loss. The validation loss is comparatively more than the training loss. In the case of architecture 2, the training accuracy is 98.44%, validation accuracy is 90.26%, training loss is 4.3% and validation accuracy is 39.47%. Compared to architecture 1, architecture 2 shows slightly less training and validation accuracy, but shows a drastic reduction in validation loss.

From architecture 2, the number of kernels in Conv1D_1 is reduced from 50 to 32 and trained the model. The results obtained after reducing the number of kernels is depicted in *Figure 14*. As observed in *Figure 14,* the training and validation accuracy is 99.05% and 90.09% respectively. The training and validation loss is 3.5% and 90.94%. Even though the training and validation accuracy is good, the validation loss of the network is very huge which leads to overfitting. Similarly, architecture 4 obtained 99.6%, 92.43%, 1.21%, and 54.5% of training accuracy, validation accuracy, training loss, and validation loss respectively. Architecture 5 obtained a very high validation loss as shown in *Figure 16.*

From the above observation, we can conclude that both accuracy and loss of the architectures are increasing, the network is starting to overfit. A good prediction model should have less loss and high accuracy. In order to reduce the overfitting and validation loss of the network, architecture 5 is fine-tuned by using regularization techniques. The regularization techniques help to reduce the overfitting by penalizing coefficients of the high value of the model. In the proposed work, we have applied the L2 regularization to the convolution layer 1 and 2 of the network with weights of 0.1 and 0.06 respectively. *Figure 17* depicts the performance of the regularized CNN. The regularized CNN obtained the training and validation accuracy of 99.58% and 92.5% respectively. The training and validation loss is 7.2% and 28.3%. *Figure 18* depicts the confusion matrix of the CNN model.
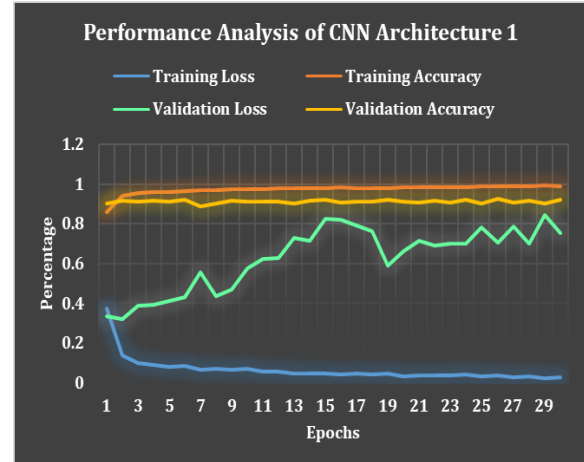

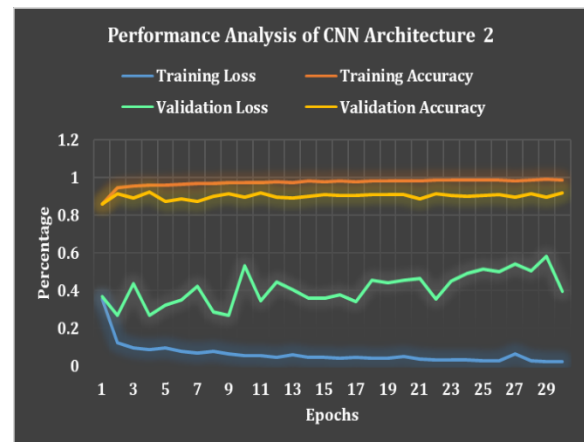**Figure 12** Training accuracy, loss and validation accuracy and validation loss of CNN Architecture 1

Figure 13 Training accuracy, loss and validation accuracy and validation loss of CNN Architecture 2

Figure 14 Training accuracy, loss and validation accuracy and validation loss of CNN Architecture 3

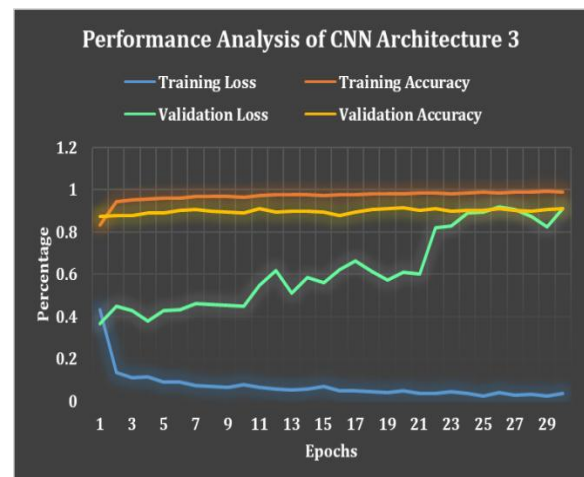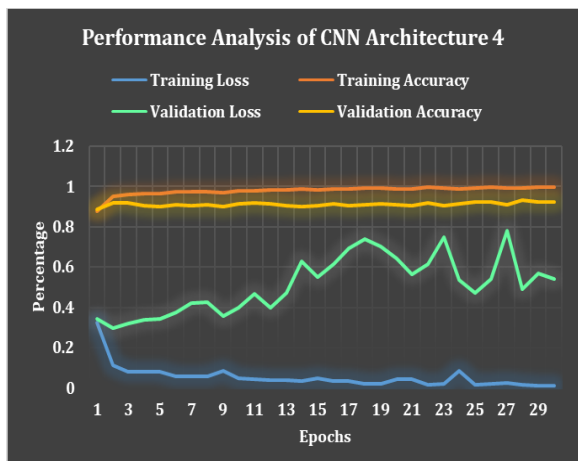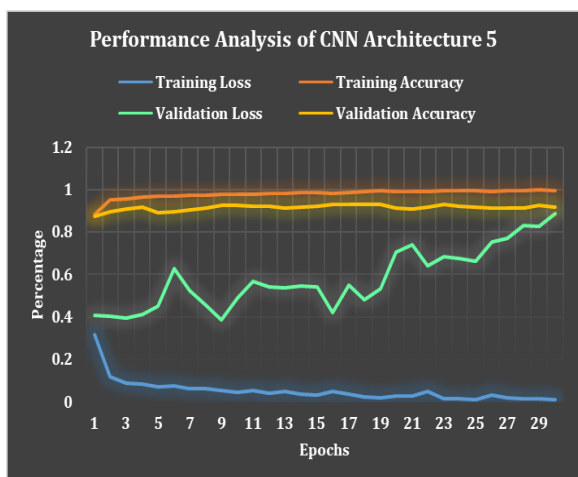**Figure 15** Training accuracy, loss and validation accuracy and validation loss of CNN architecture 4

Figure 16 training accuracy, loss and validation accuracy and validation loss of CNN architecture 5
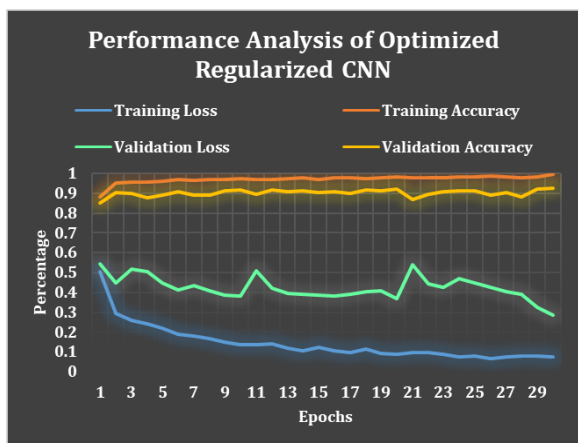

Figure 17 training accuracy, loss and validation accuracy and validation loss of regularized CNN architecture
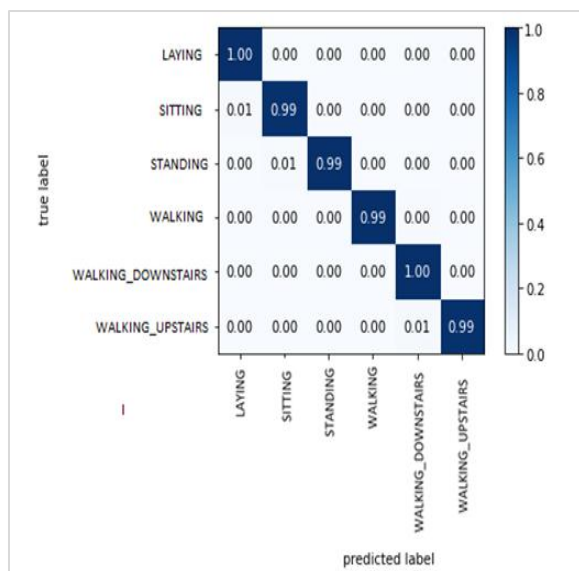

Figure 18 Confusion matrix of regularized CNN architecture

The above architectures employed a swish activation function. Swish is an activation function that was proposed by the authors of [42]. The activation functions of swish and Relu are shown in Equations (10) and (11) respectively.

$$f(x) = x \, \text{X}(1 + e)^{-x} \qquad (10)$$

$$y = \max(0, x) \qquad (11)$$

The swish activation function outperforms Relu because of its properties such as being unbounded above, bounded below, non-monotonic, and smooth function.
Analysis depicted in *Figure 17* shows that the accuracy of CNN Architecture with Swish Activation Function and regularization is comparatively better than other architectures mentioned in *Figure 12* to *Figure 16*. Analysis of these experimental results shows that the Swish Activation Function gives better results than Relu.

Table 3 summarizes the comparison of performance evaluation of state-of-art-techniques for HAR. The observation shown in *Table 4*, proves that the accuracy obtained from the proposed Regularized CNN outperformed the state-of-art-techniques discussed in the literature.

A complete list of abbreviations is shown in *Appendix I.*

Table 4 Comparison of performance evaluation

Reference	Method	Accuracy
[10]	Multi-class Support Vector Machine	98.67%
[22]	CNN + stat. features, intervals of size 50	94.35%
	CNN + stat. features, intervals of size 128	97.63%
[23]	1D-CNN	97.62%
[30]	CNN-Local Loss	96.90%
[43]	DRNN	96.50%
[44]	PCA + SVM	91.82%
[45]	Convolution Neural Network	94.79 %
[46]	Handcrafted features + RF	77.81%
[47]	Recurrent Neural Network	95.03%
Proposed Method	CNN (Swish Activation) Regularized	99.58%

5.Conclusions

The proposed work presented a performance analysis of RF and Kernel-based SVM machine learning algorithms and various models of CNN for HAR using PCA and t-SNE. t-SNE performed betters visualization results compared to PCA. RF achieved 92.5% accuracy for HAR on the UCI-HAR dataset and RBF–kernel-based SVM achieved higher accuracy than RF that is 96.23%. A detailed analysis has been carried out for the performance analysis of CNN with various parameters. The 1D CNN with 2 convolution layers with 50 filters followed by ReLu, Max pooling, dense layer with Swish Activation function, and dense layer with softmax for classification has achieved very high accuracy of the order of the highest accuracy of 99.58%. The limitation of this work is that the 1D-CNN is evaluated only on the UCI-HAR dataset. This work can be further extended by evaluating the 1D-CNN on other HAR benchmark datasets.

Conflicts of interest

The authors have no conflicts of interest to declare.

Authors contribution statement

Shilpa Ankalaki: Conceptualization, Investigation, Data curation, Writing – original draft, Writing – review and editing. **Thippeswamy M. N.**: Study conception, Design, Data collection, Supervision, Investigation on challenges and Draft manuscript preparation.

References

[1] Yazdansepas D, Niazi AH, Gay JL, Maier FW, Ramaswamy L, Rasheed K, et al. A multi-featured approach for wearable sensor-based human activity recognition. In international conference on healthcare informatics 2016 (pp. 423-31). IEEE.

[2] Bhattacharya S, Lane ND. From smart to deep: robust activity recognition on smartwatches using deep learning. In international conference on pervasive computing and communication workshops 2016 (pp. 1-6). IEEE.

[3] Capela NA, Lemaire ED, Baddour N. Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. PloS one. 2015; 10(4):1-18.

[4] Plötz T, Hammerla NY, Olivier PL. Feature learning for activity recognition in ubiquitous computing. In twenty-second international joint conference on artificial intelligence 2011(pp-1729-34).

[5] Zebin T, Scully PJ, Ozanyan KB. Inertial sensor based modelling of human activity classes: feature extraction and multi-sensor data fusion using machine learning algorithms. Ine health 360° 2017 (pp. 306-14). Springer, Cham.

[6] Zhang X, Wong Y, Kankanhalli MS, Geng W. Hierarchical multi-view aggregation network for sensor-based human activity recognition. PloS one. 2019; 14(9):1-20.

[7] De-la-hoz-franco E, Ariza-colpas P, Quero JM, Espinilla M. Sensor-based datasets for human activity recognition–a systematic review of literature. IEEE Access. 2018; 6:59192-210.

[8] Al MF, Ranasinghe S, Plattner J, Jnoub N. Human activity recognition based on real life scenarios. In international conference on pervasive computing and communications workshops 2018 (pp. 3-8). IEEE.

[9] Nafea O, Abdul W, Muhammad G, Alsulaiman M. Sensor-based human activity recognition with spatio-temporal deep learning. Sensors. 2021; 21(6):1-20.

[10] Ahmed BR, Ahmed N, Amiruzzaman M, Islam MR. A robust feature extraction model for human activity characterization using 3-axis accelerometer and gyroscope data. Sensors. 2020; 20(23):1-17.

[11] Irvine N, Nugent C, Zhang S, Wang H, Ng WW. Neural network ensembles for sensor-based human activity recognition within smart environments. Sensors. 2020; 20(1):1-26.

[12] Voicu RA, Dobre C, Bajenaru L, Ciobanu RI. Human physical activity recognition using smartphone sensors. Sensors. 2019; 19(3):1-18.

[13] Han J, Shao L, Xu D, Shotton J. Enhanced computer vision with microsoft kinect sensor: a review. IEEE Transactions on Cybernetics. 2013; 43(5):1318-34.

[14] Fortin-simard D, Bilodeau JS, Bouchard K, Gaboury S, Bouchard B, Bouzouane A. Exploiting passive RFID technology for activity recognition in smart homes. IEEE Intelligent Systems. 2015; 30(4):7-15.

[15] Ponce H, Martínez-villaseñor MD, Miralles-pechuán L. A novel wearable sensor-based human activity recognition approach using artificial hydrocarbon networks. Sensors. 2016; 16(7):1-28.

[16] Anguita D, Ghio A, Oneto L, Parra PX, Reyes OJL. A public domain dataset for human activity recognition using smartphones. In proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning 2013 (pp. 437-42).

[17] Le TD, Van NC. Human activity recognition by smartphone. In 2nd national foundation for science and technology development conference on information and computer science 2015 (pp. 219-24). IEEE.

[18] Gupta P, Dallas T. Feature selection and activity recognition system using a single triaxial accelerometer. IEEE Transactions on Biomedical Engineering. 2014; 61(6):1780-6.

[19] Fazli M, Kowsari K, Gharavi E, Barnes L, Doryab A. HHAR-net: hierarchical human activity recognition using neural networks. In international conference on intelligent human computer interaction 2020 (pp. 48-58). Springer, Cham.

[20] De LG, Rosati S, Balestra G, Agostini V, Panero E, Gastaldi L, et al. Human activity recognition by wearable sensors: comparison of different classifiers for real-time applications. In international symposium on medical measurements and applications 2018 (pp. 1-6). IEEE.

[21] Sun J, Fu Y, Li S, He J, Xu C, Tan L. Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. Journal of Sensors. 2018.

[22] Ignatov A. Real-time human activity recognition from accelerometer data using convolutional neural networks. Applied Soft Computing. 2018; 62:915-22.

[23] Cho H, Yoon SM. Divide and conquer-based 1D CNN human activity recognition using test data sharpening. Sensors. 2018; 18(4):1-24.

[24] Avilés-cruz C, Ferreyra-ramírez A, Zúñiga-lópez A, Villegas-cortéz J. Coarse-fine convolutional deep-learning strategy for human activity recognition. Sensors. 2019; 19(7):1-16.

[25] Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors. 2016; 16(1).

[26] Gupta S. Deep learning based human activity recognition (HAR) using wearable sensor data. International Journal of Information Management Data Insights. 2021; 1(2):1-18.

[27] Uddin MZ, Soylu A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. Scientific Reports. 2021; 11(1):1-15.

[28] Ye J, Li X, Zhang X, Zhang Q, Chen W. Deep learning-based human activity real-time recognition for pedestrian navigation. Sensors. 2020; 20(9):1-30.

[29] Gao W, Zhang L, Huang W, Min F, He J, Song A. Deep neural networks for sensor-based human activity recognition using selective kernel convolution. IEEE Transactions on Instrumentation and Measurement. 2021; 70:1-13.

[30] Teng Q, Wang K, Zhang L, He J. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition. IEEE Sensors Journal. 2020; 20(13):7265-74.

[31] Mukherjee D, Mondal R, Singh PK, Sarkar R, Bhattacharjee D. EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimedia Tools and Applications. 2020; 79(41):31663-90.

[32] Yang R, Wang B. PACP: a position-independent activity recognition method using smartphone sensors. Information. 2016; 7(4):1-18.

[33] Dharavath R, Madhukarrao G, Khurana H, Edla DR. t-SNE manifold learning based visualization: a human activity recognition approach. In advances in data science and management 2020 (pp. 33-43). Springer, Singapore.

[34] Van DML, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008; 9:2579-605.

[35] Anguita D, Ghio A, Oneto L, Parra X, Reyes-ortiz JL. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In international workshop on ambient assisted living 2012 (pp. 216-23). Springer, Berlin, Heidelberg.

[36] Ravi N, Dandekar N, Mysore P, Littman ML. Activity recognition from accelerometer data. In AAAI 2005 (pp. 1541-6).

[37] Amami R, Ayed DB, Ellouze N. Practical selection of SVM supervised parameters with different feature representations for vowel recognition. arXiv preprint arXiv:1507.06020. 2015.

[38] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.

[39] Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378. 2018.

[40] Zhou Y, Wang X, Zhang M, Zhu J, Zheng R, Wu Q. MPCE: a maximum probability based cross entropy

loss function for neural network classification. IEEE Access. 2019; 7:146331-41.

[41] Gordon-rodriguez E, Loaiza-ganem G, Pleiss G, Cunningham JP. Uses and abuses of the cross-entropy loss: case studies in modern deep learning. Proceedings of machine learning research 2020(pp. 1-10).

[42] Ramachandran P, Zoph B, Le QV. Swish: a self-gated activation function. 2017. Arxiv:1710.05941v1 [Cs.NE]: 1-12.

[43] Murad A, Pyun JY. Deep recurrent neural networks for human activity recognition. Sensors. 2017; 17(11):1-17.

[44] Li Y, Shi D, Ding B, Liu D. Unsupervised feature learning for human activity recognition using smartphone sensors. In mining intelligence and knowledge exploration 2014 (pp. 99-107). Springer, Cham.

[45] Ronao CA, Cho SB. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications. 2016; 59:235-44.

[46] Kolosnjaji B, Eckert C. Neural network-based user-independent physical activity recognition for mobile devices. In international conference on intelligent data engineering and automated learning 2015 (pp. 378-86). Springer, Cham.

[47] Inoue M, Inoue S, Nishida T. Deep recurrent neural network for mobile human activity recognition with high throughput. Artificial Life and Robotics. 2018; 23(2):173-85.

Shilpa Ankalaki received an engineering degree in Computer Science and Engineering from the SGBIT, VTU, Belagavi in 2012 and a Master of Technology in Computer Science and Engineering from Nitte Meenakshi Institute of Technology, Bangalore in 2014. Currently, she is pursuing a Ph.D. in Computer Science and Engineering. Her field of research is based on Machine Learning, Deep Learning, Data Mining, and Artificial Intelligence applications.
Email: shilpaa336@gmail.com

Thippeswamy M. N. currently working as Professor and Head, Department of Computer Science and Engineering,(AI&ML and Cyber Security), Ramaiah Institute of Technology Bangalore. He holds a BEng degree in Computer Science and Engineering from Kuvempu University, India, M.Tech degree in Computer Science and Engineering from VTU, India, and Ph.D. in Engineering from School of Engineering (Electrical, Electronic and Computer Engineering), Howard college campus, University of KwaZulu-Natal, Durban, South Africa.

He has spent more than 20 years in research and development, teaching, and administration at well-known universities and engineering institutions in India and Africa. His field of research is based on the Internet of Things, Big Data Analytics, Machine Learning, Deep Learning, Wireless Ad Hoc & Sensor Networks, and Cognitive Radio with emphasis on the design and analysis of MAC and Routing protocols.
Email: mntswamy@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	AAL	Ambients Assisted Living
2	ADL	Activities of Daily Living
3	CCE	Categorical Cross-Entropy
4	CNN	Convolutional Neural Networks
5	DL	Deep Learning
6	ECDF	Empirical Cumulative Distribution Function
7	ELM	Extreme Learning Machine
8	EPS	Enveloped Power Spectrum
9	FNN	Feedforward Neural Network
10	HAR	Human Activity Recognition
11	k-NN	k-Nearest Neighbours
12	LDA	Linear Discriminant Analysis
13	LSTM	Long Short-Term Memory
14	MC-SVM	Multi-Class Support Vector Machine
15	MEMS	Micro-Electro-Mechanical System
16	MIMU	Miniature Inertial Measurement Unit
17	ML	Machine Learning
18	MLP	Multi-Layer Perceptron
19	NB	Naïve Bayes
20	PC	Principal Component
21	PCA	Principal Component Analysis
22	RB	Radial Bias Function
23	ReLu	Rectified Linear Unit
24	RF	Random Forest
25	SVM	Support Vector Machine
26	t-SNE	t-Distributed Stochastic Neighbour Embedding
27	UCI-HAR	UC Irvine HAR
28	WISDM	Wireless Sensor Data Mining