ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-15
Dr T. BENARJI *et. al.,* / International Journal of Engineering & Science Research

Clustering-classification method for human activity recognition using smart phone dataset

Dr T. BENARJI, 2 P. RENUKA

[1]Professor, Department of CSE, Indur Institute of Engineering and Technology, Siddipet, Telangana, Hyderabad. Email: tharinibenarji@gmail.com

[2]Associate Professor, Department of ECE, Indur Institute of Engineering and Technology, Siddipet, Telangana, Hyderabad. Email: renoostar@gmail.com

Abstract: The study of human activities is known as Human Activity Recognition (HAR). A few examples of the many real-world uses of HAR include healthcare systems, rehabilitation, and monitoring patients on a regular basis to determine how they are doing in terms of lowering the risk associated with activities of daily life. Traditional HAR algorithms abound, but they have a long way to go before they meet modern standards in areas like privacy and accuracy. Two main types of HAR exist: those that rely on vision, such surveillance footage and images, and those that rely on sensors, like smartphones, smart watches, and wearable gadgets. Concerns about privacy, data storage capacity, expense, infrastructure, and accuracy arise when using external devices for data collection in vision-based HAR. Data is collected by the sensor-based HAR using on-body smart devices worn in the bell, ankle, wrist, chest, and abdomen. The sensors built into these gadgets collect data that is both useful and sensitive to user privacy. This study proposes GSCV-RF and GSCV-SVM methods based on Random Forest (RF) and Support Vector Machine Classifier (SVMC), respectively, with the goal of addressing all facets of sensor-based human activity recognition, including datasets, pre-processing techniques, optimization techniques, classification models, and prediction accuracy. If we compare these approaches to more traditional models, we see that they provide far more accurate predictions. The following terms are used interchangeably: sensors, classification, machine learning, optimization, prediction, and human activity recognition.

I. INTRODUCTION

One way to anticipate complex movements is with the help of Human Activity Recognition, which uses video and sensor data gathered from security cameras, wearable devices, and the accelerometer, magnetometer, and gyroscope sensors in smartphones, smartwatches, and smart bands. Care for the elderly, assisted living, rehabilitation, sports activity tracking, and smart health care are just a few of the many HAR uses. Machine learning models use sensor data to forecast a person's physical actions, which aids in patient rehabilitation and enhances athletic performance. While the gyroscope detects rotation, the accelerometer detects x, y, and z-axis orientation (Figure 1.a and 1.(b) respectively).

Figure 1.(A) Accelerometer Sensing X,Y,Z Axis Orientation 1.(B) Gyroscope Sensing Angular Orientation.

This study proposes a new optimization model that uses machine learning to forecast people's physical activity levels. A literature overview on machine learning methods for human activity identification was provided in section 2. Human activity recognition prediction model is detailed in Section 3. We covered the experimental findings that were acquired using Python program in section 4. Section 5 concludes with the results and planned efforts.

II. LITERATURE REVIEW

Information on data sources, data kinds, pre-processing, optimization methods, classification and prediction models, and metrics for fitting the model are all covered in this area of the health care literature, with a focus on the identification of human physical activities. Collecting data about people's physical activity is possible with the use of wearable and body sensors. Machine learning optimization methods like Adaptive Learning rate, Stochastic Gradient descent, Gradient descent, conjugate gradient method, derivative free optimization, zeroth order optimization, meta learning, and other population algorithms like Genetic Algorithm and Particle Swarm Optimization (PSO) are commonly employed for parameter optimization in classification models.

2.1 Pre-processing techniques B. A.

Several classifiers, including Naïve Bayes, Random Forest (RF), Deep learning, and k-Nearest Neighbor, use the quick feature dimensionality reduction approach that Mohammed Hashim and R. Amutha[1] suggested for identifying human physical activities. Results from experiments with the UCI machine learning repository's human activity dataset demonstrate a random forest classifier's high classification accuracy of 98.72%.

2.2 Classification models

Decision trees, neural networks, Bayesian classification, and lazy learners are some of the classification models used in machine learning. With the use of accelerometer sensors included in smartwatches, Min-Cheol Kwon and Sunwoong Choi[3] demonstrated a 95% success rate in classifying 11 different human physical activities. In order to enhance the classification performance for smartphone-based human activity identification, Mekruksavanich S. and Jitpattanakul A.[4] suggested a CNN-LSTM hybrid model. When fine-tuning hyperparameters, the Bayesian optimization method is used. Classification accuracy in the CNN-LSTM hybrid model may reach 2.24%. For their presentation of the Ensemble Learning Algorithm with Smartphone Sensor Data using fully connected DNN, Tan T-H, Wu J-Y, Liu S-H, and Gochoo M. [5] demonstrated a prediction accuracy of 96.7% as determined by the f1 measure. To identify qualitative human activity, Ysenllari, E., Ottenbacher, J., and McLennan[6] presented a 2D-convolutional neural network (2D-CNN) and achieved an accuracy ranging from 96.57% to 99.28%. From 0.96 to 0.99, they were able to derive Cohen's k-value. The CNN-GRU model, suggested by Dua, N. et.al. [7], obtained 96.20 percent accuracy on the UCI-HAR dataset, 97.21% on the WISDM dataset, and 95.27% on the PAMAP2 dataset, respectively. For the UCI-HAR dataset, Mutegeki R. and Han DS.[8] obtained an accuracy of 92.13 using a CNNLSTM method for human activity prediction.In order to identify human activities, Prasad et al.[9] suggested a convolutional neural network (CNN) method. They used the WISDM dataset, which includes data from the accelerometers in mobile phones, and achieved an accuracy rate of 89.67% in their predictions. An optimization strategy for crop suitability prediction was proposed by Vasanthanageswari, Prabhu P. [11]. Using publicly available datasets, the model is shown to be inferior to more traditional models. One well-designed k-nearest neighbour (KNN) method for human activity identification was given by Mohsen, S et.al. [11]. A higher level of classification accuracy is achieved by fine-tuning the parameters of this method. Experimental analysis is performed on this method using real-world datasets. Classifiers were found to attain a classification accuracy of 90.37 percent when tested using conventional performance criteria. Using a cardiac dataset, Sivakami.M. and Prabhu P.[12] used machine learning methods to enhance the precision of testing and classification. When tested against more traditional models, the model proves to be superior. Using a frequent item set and fuzzy logic, Prabhu P et.al.[13–14] suggested business intelligence models for an e-commerce recommender system. If you own an online store, this model can help you choose what products to sell. When it comes to prediction accuracy, sparsity, and scalability, traditional classifiers fall short. The increased reliance on sensors and other devices in these models raises concerns about patient privacy and potential adverse consequences. As a result, better health care approaches and parameter-optimized accurate prediction should be made available via mobile devices.

There are approaches that can solve the HAR problem, such as those based on decision trees, Bayesian statistics, and neural networks. To make reliable predictions about physical activities, this paper proposes to employ optimization approaches before adjusting categorization. According to this research, prediction models still have room for improvement in order to reach high accuracy and go beyond problems with sensitivity and scalability. The health care applications field is particularly in need of improved categorization.

III. METHODOLOGY

Optimal classification model for human activity recognition is described in this section. Figure 2 depicts the design of the created model.

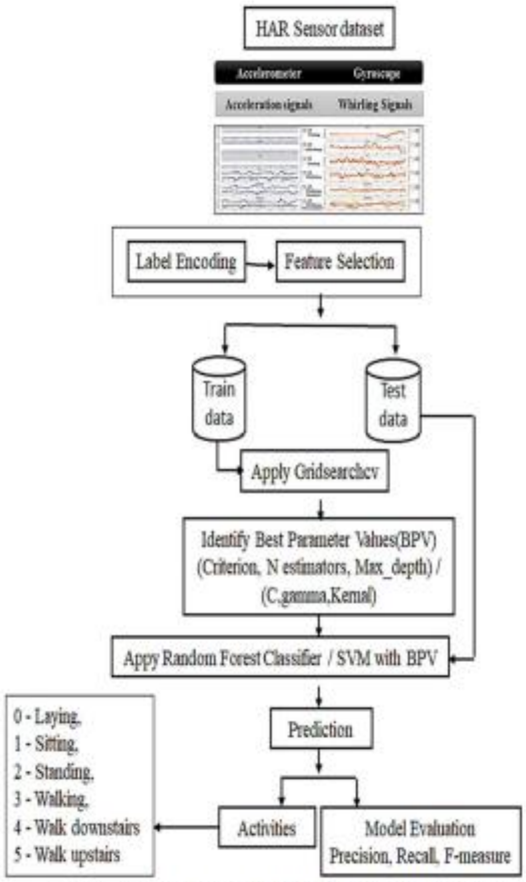


Figure 2. Optimization Based Prediction Model Design.

Preprocessing, modeling, and assessment are the three main parts of this model. As an initial step, the suggested model loads the accelerometer and gyroscope data used for human activity identification. Label encoding and dimension reduction are the next steps in getting this dataset ready for modeling. Train data and test data are subsets of the input dataset. Applying an optimization strategy to determine the optimal values for the prediction model's parameters is the next stage in the suggested model. The rest of this section goes into more depth on the suggested prediction model. 3.1 preparatory Inconsistencies, missing values, and irrelevant data are checked for during the pre-processing of the dataset. When it comes to lowering the amount of characteristics, the dimensionality reduction approach is crucial for eliminating the least significant ones. The input for the model is a list of features extracted from the dataset using the highly accurate and interpretable Random Forest feature selection approach. Separate "train" and "test" datasets are used for model testing. Part 3.2: Model Training. Training the model is the next stage of this planned project. A classifier's prediction accuracy may be greatly improved by hyperparameter tuning, which involves identifying the optimal values for these parameters. To fine-tune the hyperparameters, you may use either the GridSearchCV or RandomizedSearchCV methods. While RandomizedSearchCV defines combinations at random, GridSearchCV allows the user to specify all potential values. Because it uses a random selection process, the RandomizedSearchCV model reduces the amount of time required to find a value when there are many possible combinations of parameters.

3.2.1 GridSearchCV optimization

To determine the optimal classifier parameters from the provided list, this study used the GridsearchCV method, a cross validation tool. It tests several values for each hyperparameter and their combinations to see which one works best, then picks the best one. Hyperparameters for random forest classifiers include forest size, maximum tree levels, feature count, and impurity criterion (Gini and entropy) for decision tree generation. The optimal parameters have been determined to be criterion='entropy,' max_depth=11, n_estimators=200, and random_state=42. Regularization parameter, gamma, and kernel are the hyperparameters that support VGG19 classifiers. We found that the SVM approach works best with the following parameters: C=100, gamma=0.001, kernel='rbf'.

3.3 Activity Prediction and Evaluation

Predicting the appropriate class for the new input data (i.e., test data) is the final step in this model. After receiving the test data, the model compares it to the train data and produces a result. Applying the optimal parameters to the classifier model allows it to make predictions on test data once they have been determined via an optimization approach.

3.4 Proposed GSV-RF and GSV-SVM Algorithm

```
Input:
  HAR dataset D (feature vectors:561; classes c= 6)
Output:
  Predicted c physical activities.
  Precision(P),Recall(R),F measure(F) values.

Begin
  Load HAR sensor dataset D
  Initialize number of classes (c = 6);
  Do Label encoding
  Apply dimensionality reduction technique
  Apply feature selection
  Spilt the D into train (TR) and test dataset (TS)
  Apply GridSearchCV_optimization(TR) in RFC
   Identify Best Parameter Values (BPV) for RFC
//RFC(criterion, max_depth, N_estimators)
// Criterion – (gini, entropy),
//max_depth of the tree,
//N-estimators (Number of trees to build)
Apply GridSearchCV_optimization(TR) in SVMC
// Support Vector Machine (SVM)
//   SVM(C, gamma, kernel)
// C is the penalty parameter of the error term
// Kernal(Linear, Polynomial, RBF, and Sigmoid)
 Perform RFC(BPV) on TR
 Perform SVM(BPV) on TR
 Test the GSV-RFC using test dataset (TS)
 Test the GSV-SVMC using test dataset (TS)
 Evaluate model using precision, recall and f-measure.
End
```

3.3.1 Evaluation Measures

We put the prediction model through its paces in a Python environment using real-world datasets and mathematical metrics to ensure its quality. A variety of metrics are used to assess the performance of the suggested prediction model. These include recall (sensitivity), precision, F-measure, macro average, and weighted average derived from the total number of TP, FP, and FN that were detected. A measure of accuracy is the proportion of results that are really positive out of all possible outcomes. The F-measure finds the harmonic mean of the recall and accuracy values, therefore balancing them.

$$Precision = \frac{TP}{(TP+F\ \ )} \quad (1)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

$$F\ measure = \frac{(2\ X\ Precision\ X\ Recall)}{Precision + Recall} \quad (3)$$

*Corresponding Author

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-15

IV. RESULTS AND DISCUSSION

Here we describe the experimental setup and the outcomes of the prediction models that were constructed for human activity recognition.

4.1 Experimental setup

4.1.1 Datasets description

The HAR datasets may be sourced from publicly accessible online databases as well as real-world sources. collection of machine learning models housed at UCI, Nottingham Trent University, WISDM[15,16], Groupware, Wrist Sensor Dataset, Chest Sensor Dataset, and Human Activity Recognition Using Smartphones Data Set. The following datasets are often made accessible to the public: HAR Single Chest-Mounted Accelerometer, SisFall, Cornell Activity Dataset, MobiSense, MobiAct, and PAMAP2. The purpose of this study is to evaluate the model's efficacy in human activity identification using a dataset extracted from smartphones and stored in the machine learning repository at the University of California, Irvine [14]. Thirty participants participated in six different physical activities, including 0-laying, 1-sitting, 2-standing, 3-walking, 4-walk downstairs, and 5-walk upstairs, and their raw data is included in this dataset. It contains data from the temporal and frequency domains and is comprised of 102,299 samples with 561 properties. Subject identifiers, target classes, time domain variables, triaxial angular velocities, and accelerations are all part of each sample. Using a dataset consisting of activities recognized by cellphones, figure 3(a) displays the distribution of these activities from the UCI Machine Learning library.

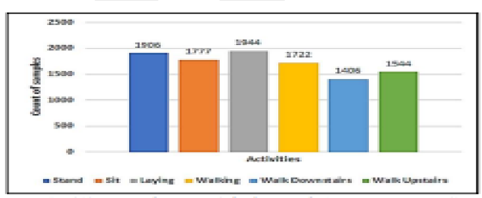


Figure 3. (A) Distribution Of Physical Activities In UCI – HAR Dataset.

Figure 3.(b) shows the sample of Accelerometer and gyroscope signals of six activities.

*Corresponding Author

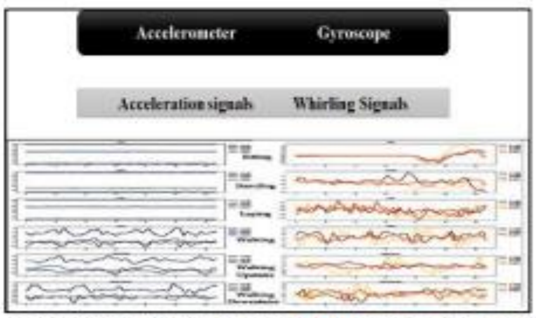


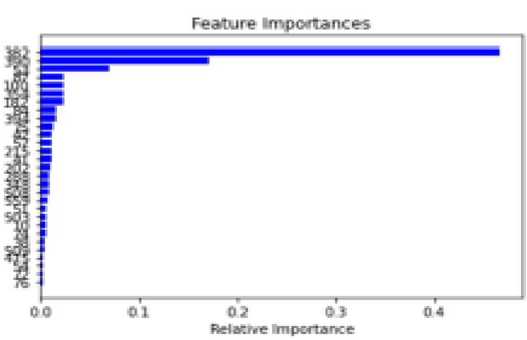
Figure 3.(B) Accelerometer And Gyroscope Signals Of Six Activities

The Public UCI datasets [15] collected from website can also be used to test the performance of this prediction model.

4.1.2 Feature selection Results

4.1.2.1 Random forest feature selection

The selected features using Random Forest feature selection method are [76 72 54 475 509 38 74 10 503 51 559 508 348 288 202 41 215 57 42 75 394 84 182 354 100 87 53 390 382]. Figure 3.(c) shows the features selected using random forest feature selection method.



3.(C) Feature Selection Using Random Forest.

4.1.2.2 Best fit feature selection

Selected attributes Using Best fit method: [11,24,38,39,42,43,45,51,52,53,54,55,58,67,69,71,73,76,79,80,81,105,106,109,122,159,160,161,181,204,232,298,304,305,309,312,368,370,371,383,388,413,450,451,452,453,492,506,520,539,558,559,560, 561 : 54].

4.1.2.3 Greedy stepwise feature selection

Selected attributes using greedy stepwise method: [11,24,38,39,42,43,45,51,52,53,54,55,58,67,69,71,7 3,76,79,80,81,105,106,109,122,159,160,161,181,20 4,232,298,304,305,309,312,368,370,371,383,388,4 13,450,451,452,453,492,506,520,539,558,559,560, 561 : 54]

4.1.2 Software implementation description

Python 3.10 for Windows, an Intel(R) Core (TM) i7-8565U CPU running at 1.80 GHz with a turbo boost to 1.99 GHz, and 8 GB of RAM were used to execute this study. Created with the help of Scilab, Numpy, Pandas, Tensorflow, sklearn, matplotlib, and seaborn, the optimization-based prediction model for Human Activity Recognition is well named.

4.2 Results and discussion

4.2.1 Performance comparison

To test how well a prediction model works, one often looks at its accuracy, recall, F measure, sensitivity, specificity, ROC curve, MAE, and confusion matrix. The methods for implementing a HAR system to identify potential health hazards using two kinds of sensor data were laid forth by Abdulhamit Subasi et al., [2]. The experimental findings of analyzing the performance of several machine learning approaches in recognizing critical conditions are shown and analyzed.

Table 1 shows the accuracy comparison of various conventional classifiers calculated using HAR dataset.

Table.1 Classification Accuracy Comparison Of Conventional Classifiers

Classifier	Measure		
	Precision	Recall	F-Measure
k_NN	96.32	96.27	96.26
Decision Tree	93.56	93.48	93.51
Naïve Bayes	93.00	93.00	93.00
AdaBoost	95.35	95.36	95.35
Random Forest	97.66	97.67	97.67
SVM	97.65	97.68	97.67
Multilayer Perceptron	98.00	98,00	98,00
Logistic Regression	98.12	98.12	98.12

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-15
Dr T. BENARJI *et. al.*, / International Journal of Engineering & Science Research

Table 2 shows the accuracy comparison of various improved conventional classifiers calculated using HAR dataset. The Logistic Regression (LR) Classifier gives high f-measure accuracy of 98.12 when compared with other conventional models.

Table 2. Accuracy Of Various Improved Conventional Classifiers

Classifier	Type of classifier	Accuracy %
k-NN[10]	Nearest Neighbour	90.37
k-NN-FFDRT[1]	Nearest Neighbour	98.45
Random Forest-FFDRT[1]	Decision Tree	98.72
Naïve-Bayes-FFDT[1]	Bayes Theorem	93.00
ANN[3]	Neural Network	95.00
2D-CNN[6]	Neural Network	96.57 to 99.28
Deep learning-FFDRTT[1]	Neural Network	98.68
CNN-LSTM	Neural Network	92.13
CNN[9]	Neural Network	89.67

The 2D-CNN[6] based method provides high accuracy of classification up to 99.28%. The figure 4 shows the ROC curve shows false positive rate vs. True positive rate value of 0.998 of sitting activity classified using Random Forest Method.

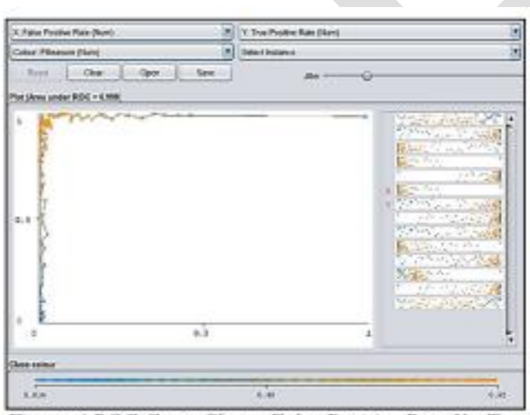


Figure 4 ROC Curve Shows False Positive Rate Vs. True Positive Rate Value Of Sitting Activity Classified Using Random Forest Method.

The confusion matrix is a performance evaluation metrics using for visualizing the performance of classifiers.

Figure 5 shows the confusion matrix of proposed GridSearchCV + SVM classifier using UCI machine learning HAR smartphone dataset.

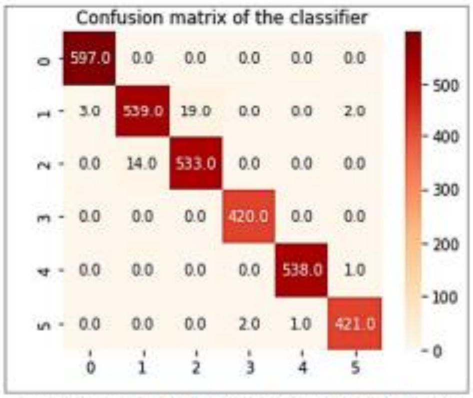


Figure 5. Confusion Matrix Of Proposed GSCV+SVM Classifier.

Table 3 shows the activity wise precision, recall and f1 score accuracy of SVM classifier calculated using HAR smartphone dataset. The Walk-downstairs activity gets high f1 score of 0.9950. The overall accuracy obtained for all activities is 0.9754.

Table 3 Activity Wise Accuracy Of SVM Classifier Using HAR Dataset.

Activity	Precision	Recall	F-Score
Laying	0.9917	0.9983	0.9950
Sitting	0.9395	0.9378	0.9387
Standing	0.9482	0.9378	0.9430
Walking	0.9929	0.9952	0.9941
Walk-downstairs	0.9963	0.9963	0.9963
Walk-upstairs	0.9906	0.9953	0.9929
Accuracy			0.9754
Macro average	0.9765	0.9768	0.9767
Weighted Average	0.9753	0.9754	0.9753

Table.4 shows the activity wise precision, recall and f1 measure accuracy of proposed GridSearchCV+ SVM classifier calculated using HAR dataset. The Walk-downstairs activity gets high f1 score of 0.9981. The overall accuracy obtained is 0.9764.

Table 4 Activity Wise Accuracy Of Gridsearchcv+Svm Classifier Using Har Dataset.

Activity	Precision	Recall	F-Score
Laying	0.9950	1.0000	0.9975
Sitting	0.9747	0.9574	0.9659
Standing	0.9656	0.9744	0.9700
Walking	0.9953	1.0000	0.9976
Walk-downstairs	0.9981	0.9981	0.9981
Walk-upstairs	0.9929	0.9929	0.9929
Accuracy			0.9864
Macro average	0.9869	0.9871	0.9870
Weighted Average	0.9864	0.9864	0.9864

Figure 6 shows the accuracy of activities calculated using GridSearchCV+SVM classifier. The Walk-downstairs activity gets high f1 score of 0.9981 in the bar chart. The overall accuracy obtained is 0.9764.

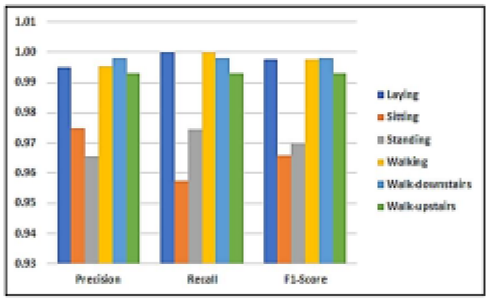


Figure 6 Accuracy Of Activities Calculated Using Gridsearchcv+SVM Classifier

Table.5 shows the accuracy comparison of various conventional classifiers with proposed GridSearchCV classifiers calculated using HAR dataset.

Table 5. Classification Accuracy Of Conventional Vs. Proposed Method

Classifiers	Accuracy %
k-NN	96.26
Decision Tree	93.51
Naïve Bayes	93.00
Random Forest (RF)	97.67
GSCV+ RF	97.72
Support Vector Machine (SVM)	97.67
GSCV+SVM	98.70

Figure.7 shows the accuracy comparison of various conventional classifiers with proposed GridSearchCV classifiers calculated using HAR dataset.

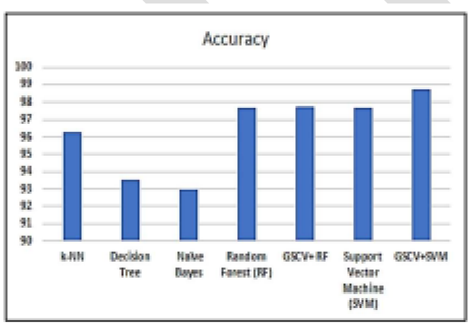


Figure 7 Accuracy Of Conventional Classifiers With Proposed Gridsearchcv Classifiers.

Using Random Forest and SVM classifier, we assess the performance of the proposed enhanced grid search optimization model. With grid search optimization, the accuracy of the Random Forest model went up from 97.67% to 97.72%, and that of the SVM model went up from 97.67% to

98.70%. GridSearchCV (GSCV) improves the accuracy of SVM models. It is possible to evaluate this model's efficacy by testing it with real-world data acquired from hospitals, although there are still some challenges and unanswered research questions. This work is in the subject of health care systems, thus there may be a need for greater improvement in prediction accuracy. Finding the best settings for several classifiers is the goal of this proposed study, which employs three feature selection philosophies and optimization techniques including grid search and random forest.

CONCLUSION:

Care for the elderly and other medical conditions presents significant challenges to the research of human activity recognition. Predicting human physical activity is the focus of this research, which proposes a machine learning approach based on grid search optimization. We also identify and examine the accuracy of different classifiers. The paper's meat and potatoes are the recommended model for predicting people's physical activity from data gathered by cellphones' sensors. When it comes to predicting human actions, the experimental findings reveal that the suggested model works better. For the purpose of making predictions, the enhanced classifier employs optimum parameters. This approach offers superior accuracy when compared with other traditional optimization methods. Additional optimization methods may be used to test this prediction model. It is possible to suggest models using deep neural networks, which are based on neural networks. The performance benchmark may be tested with datasets retrieved from various public repositories or actual data from the real world.

REFERENCES

[1] Mohammed Hashim, B.A., Amutha, R. Human activity recognition based on smartphone using fast feature dimensionality reduction technique. J Ambient Intell Human Comput 12, 2365–2374 (2021). https://doi.org/10.1007/s12652-020- 02351-x.

[2] Abdulhamit Subasi, Kholoud Khateeb, Tayeb Brahimi and Akila Sarirete, Human activity recognition using machine learning methods in a smart healthcare environment, Innovation in Health Informatics (2020) Elsevier Inc. pp123- 144.

[3] Min-Cheol Kwon and Sunwoong Choi, Recognition of Daily Human Activity Using an Artificial Neural Network and Smartwatch, Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 2618045, https://doi.org/10.1155/2018/2618045.

[4] Mekruksavanich S, Jitpattanakul A., Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models. Electronics. 2021; 10(3):308. https://doi.org/10.3390/electronics10030308.

[5] Tan T-H, Wu J-Y, Liu S-H, Gochoo M. Human Activity Recognition Using an Ensemble Learning Algorithm with Smartphone Sensor Data. Electronics. 2022; 11(3):322. https://doi.org/10.3390/electronics1103032.

[6] ysenllari, E., Ottenbacher, J. & McLennan, D. Validation of human activity recognition using a convolutional neural network on accelerometer and gyroscope data. Ger J Exerc Sport Res 52, 248–252 (2022). https://doi.org/10.1007/s12662- 022-00817-y.

[7] Dua, N., Singh, S.N. & Semwal, V.B. Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing 103, 1461– 1478 (2021). https://doi.org/10.1007/s00607-021-00928-8.

[8] R. Mutegeki and D. S. Han, "A CNN-LSTM Approach to Human Activity Recognition," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 362-366, doi: 10.1109/ICAIIC48513.2020.9065078.

[9] Prasad, A.; Tyagi, A.K.;Althobaiti, M.M.; Almulihi, A.;Mansour, R.F.; Mahmoud, A.M. Human Activity Recognition Using Cell PhoneBased Accelerometer and Convolutional Neural Network. Appl.Sci. 2021, 11, 12099. https://doi.org/10.3390/app112412099.

[10] Vasanthanageswari, Prabhu P, Improving Svm Classifier Model Using Tree Structured Parzen Estimator Optimization for Crop Prediction, Journal of Theoretical and Applied Information Technology, 2022, 100(22), pp. 6808–6818

[11] Mohsen, S., Elkaseer, A., Scholz, S.G. (2022). Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm. In: Scholz, S.G., Howlett, R.J., Setchi, R. (eds) Sustainable Design and Manufacturing. KESSDM 2021. Smart Innovation, Systems and Technologies, vol 262. Springer, Singapore. https://doi.org/10.1007/978-981-16-6128-0_29.

[12] M. Sivakami, P. Prabhu, Classification of Algorithms Supported Factual Knowledge Recovery from Cardiac Data Set, International Journal of Current Research and Review, vol.13. issue 6. March 2021 pp161-166.

[13] Prabhu P., Anbazhagan N. (2013) FI-FCM Algorithm for Business Intelligence. In: Prasath R., Kathirvalavakumar T. (eds) Mining Intelligence and Knowledge Exploration. Lecture Notes in Computer Science, vol 8284. Springer, Cham, pp 518-528.

[14] Paulraj P., Neelamegam A. (2014) Improving Business Intelligence Based on Frequent Itemsets Using k-Means Clustering Algorithm. In: Meghanathan N., Nagamalai D., Rajasekaran S. (eds) Networks and Communications (NetCom2013). Lecture Notes in Electrical Engineering, vol 284. Springer, Cham, DOI : 10.1007/978-3-319-03692-2_19, pp 243-254 .

[15] UCI machine learning repository: https://archive.ics.uci.edu/ml/index.php. [16] UCI machine learning repository WISDM dataset: https://archive.ics.uci.edu/ml/datasets/WISDM+ Smartphone+and+Smartwatch+Activity+and+Bi ometrics+Dataset