Real-time human activity recognition from smart phone using linear support vector machines

Kamel Maaloul[1,2], Lejdel Brahim[2], Nedioui Med Abdelhamid[2]
[1]LIAP Laboratory, Department of Computer Science, Faculty of Exact Sciences, University of El-Oued, El-Oued, Algeria
[2]Department of Computer Science, Faculty of Exact Sciences, University of El-Oued, El-Oued, Algeria

Article Info

ABSTRACT

The recognition of human activity (HAR) the use of cell devices embedded in its exten sively disbursed sensors affords guidance, instructions, and take care of citizens of smart cities. Consequently, it became essential to analyze human every day sports. To examine statistical models of human conduct, synthetic intelligence strategies such as machine studying can be used. Many studies have not studied type overall performance in real-time due to statistics series. To remedy this trouble, this paper proposes a structure primarily based on open supply technology and platforms consisting of Apache Kafka, for messages to flow over the internet, method them and provide shape for existing facts in real-time and formulates the trouble of identifying human pastime by using a smartphone tool as a type hassle using statistics collection by telephone sensors. The proposed version is skilled by some machine learning algorithms. The algorithm that has proven superior and quality results helps a linear vector machines.

Corresponding Author:

Kamel Maaloul
LIAP Laboratory, Department of Computer Science, Faculty of Exact Sciences
University of El-Oued, El-Oued, Algeria
Email: maaloul-kamel@univ-eloued.dz

1. INTRODUCTION

The proliferation of portable non-public devices including smartphones and smartwatches, to the emergence of technology has made a huge data. This led to the urgent need for non-public customization of use from the popularity of human activity [1]. Therefore, the recognition of human pastime is an essential area since includes several applications such as healthcare, protection, monitoring, health, and extra. Human activity recognition (HAR) especially within the healthcare and army realms requires real-time class to identify customers' movements to provide real-time feedback that enables users in real-time. Automated HAR systems used cameras, accelerometers, gyroscopes, and acoustic sensors to detect user motion. Recent years have seen the introduction of a variety of biosensors to identify human activities in HAR structures, including electromyography (EMG), electrooculography (EOG), and electroencephalography (EEG).

The most common non-vision-based sensors such as accelerometers, motion sensors, biosensors, gyroscopes, and pressure sensors are wearable and can be attached to the users like a daily-use object. The most common approach of using multimodal sensors is by placing them in a person's living environment such as in the kitchen or the living room to record their daily routine activity. The sensors monitor activities by achieved sensor data, such as opening doors (switch sensor), sitting down on the couch [2].

The emergence of massive records has caused main changes in lots of regions, inclusive of human activity recognition systems with a number of packages in smart towns, to decorate the protection of citizens. Clever metropolis mega information amassed from numerous resources are characterised by using size, range, and pace. The volume of data is too massive, it is able to be measured in petabytes or terabytes.

Diversity refers to many sources and types of based and unstructured records records pace refers to the rate of data technology. It could be in batch, close to-time, or real-time [3].

Sensors being used to learn about human activities are becoming more feasible. Three perpendicular axes are measured by the acceleration sensor. Gravity affects everything on the planet. The effect of accelerating a device's motion, omitting the effect of Earth's gravity on the device, is measured using linear acceleration. The gyroscope uses the Earth's gravity to determine the orientation of the smartphone. The data collected from these sensors can be used to detect the state and change of the smartphone's actual movement in space. The acquired huge data contains a wealth of information on human physical activity [4].

Most inquiries regarding real-time recognition of human activity need to be a swered inside sure transition intervals. However, there may be some inquiries that do not impose any strict timing necessities; recall, for example, analytical inquiries associated with city making plans [5]. For this reason, this work ambition to deal with and answer the subsequent questions: what devices can procedure very huge amounts of saved, decided on, and processed huge facts? what model can examine these very huge amounts of facts collected from the net of factors that HAR? the recognition of HAR is more important and popular with several devices including the accelerometer built into the smartphone well in the literature. Due to the increasing demand of customers in the field of the Internet of Things, many researchers have developed many methods to meet customer requests based on consideration of resource efficiency and accuracy of results. We mention here some of the work that has been done in this regard.

Fong et al. [6] proposed a comprehensive approach to data flow mining relying on parallel flow and inference, called stream-based holistic analytics and reasoning in parallel (SHARP). It aims to apply improvised methods in data flow mining. So that mining experiments of two types of data flow with a state of recognition of human activity were carried out. After the results, it was found that improvised methods have dramatically improved data stream extraction performance.

Wei et al. [7] propose a genetic algorithm (GA)-based finite feature selection method for real-time human activity recognition in generalization applications, which is modified by integrated resolution and fast sequential forward selection algorithm (FSFS). The features selected are extracted from all signals of four IMU (miniature wireless inertial measurement unit) sensors fixed on the foot and thigh of subject. The final classifier can run on the experimental platform in real time, with high accuracy for new users. Observable activity may be recognized with greater than 98% accuracy.

The study of Hassan et al. [8] outlines a method for using smartphone sensors to identify human activity. The raw data was used to extract significant features. Core component analysis (KPCA) and linear differentiation analysis (LDA) are used to strengthen these features. The deep belief network (DBN) was then trained to detect the successful actions using the characteristics. The new method's superiority was demonstrated by a comparison with the conventional multi-layered support vector machines (SVM) method.

The architecture of an unattended online learning algorithm was demonstrated by Qi et al. [9] using a system based on adaptive recognition and real-time monitoring of human activities (ada-HAR) away from the direction of the smartphone. The authorized hierarchical classification and grouping methods can also be used to categorize other activities individually. The quickest method for modeling evolution turns out to be a workbook based on decision trees.

2. PROPOSED MULTI-LAYER ARCHITECTURE

Within the smart domestic, human pastime is continuously monitored by means of sending the data generated via the sensors of the cellular phones to find out the information in real-time. Metropolis smart city device requires a strong gadget with parallel processing of records evaluation and real-time decision making. Thus, the Hadoop environment is used, which contains the grasp nodes, and numerous facts nodes below the grasp node. The essence of this architecture is that it makes use of Kafka as an intermediary between the numerous information sources from which characteristic statistics is accumulated, the version building surroundings where the model is fit, and the production utility that serves predictions. Eventually, the decision is made based totally on outcomes from the Hadoop environment [10]. The decision-making method makes use of gadget getting to know, sample popularity, tender computing, and choice models, has been illustrated in Figure 1.

Smart city layer: the fast development of the cellular network, which is the cornerstone of huge records, has brought about explosive growth in the number of cellphone customers and the era of huge quantities of information. The facts generated by way of the cellular smartphone sensors is accrued and gathered and then sent to Kafka for processing. He receives orders that are lower back thru Kafka [11].

Big data layer: the big data layer is designing the data pipeline with the various requirements of a batch processing or data storage system. Includes the general data management infrastructure, usually cloud-based, and the big data analytics part that will require high-performance computing clusters. The collected data is separated and loaded based on the metadata and prepared for the transformation which is done with different components. This structure consists of two layers, which ensure the safe flow of data:

− Data streaming and storage: the mobile phone (app) sends (JSON) messages to Kafka. It is processed by the predictor and then publishes the message with the prediction result back to Kafka, which will eventually be received by the application. After several messages, the predictor will publish a "retrain topic" message [12]. The trainer will receive the "retrain topic" message and start retraining the algorithm. In the meantime, the predictor will not stop serving predictions. When the algorithm is trained, the trainer publishes a message to the predictor that the retraining is complete, and the predictor will download the new model. The Hadoop layer stores the data then perform fast data processing and provides efficient management of storage, availability, performance, and scalability (via MapReduce) [13].

− Machine learning: the many apps and databases that host feature data are fetched into Kafka. Models are created using this information. Depending on the required abilities and selected tool set, the environment will change. A data warehouse or a massive data environment like Hadoop [14] could be used to develop the model. The model can be made public so that any production software that receives the same model parameters can use it to process incoming examples (perhaps using Kafka streams to help index the feature data for easy usage on demand). The production app could be a Kafka streams application or simply a pipeline that receives data from Kafka. When it comes to feeding, creating, using, and monitoring analytical models, Kafka acts as the machine learning (ML) architecture's central nervous system [15].
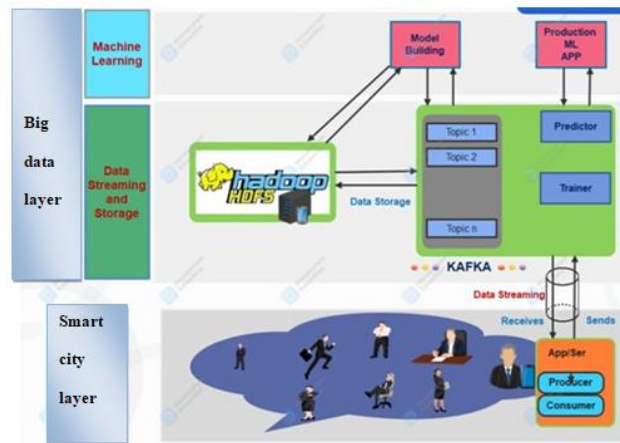

Figure 1. The architecture of the implemented system

3. METHOD

3.1. Describe of dataset

A group of 30 participants, ages 19 to 48, participated in the experiments. Every participant utilized a Samsung Galaxy S II smartphone for six various exercises (walking, walking upstairs, walking downstairs, sitting, standing, and laying). The accelerometer and gyroscope that included with the device were used to capture 3-axial angular velocity and 3-axial linear acceleration at a constant rate of 50 Hz. In order to manually classify the data, the exams have been filmed [16]. The resulting dataset was divided into two sets at random, with 30% of participants chosen to provide test data and 70% of volunteers chosen to produce training data. After applying noise filters as a pre-processing step, the data from the gyroscope and accelerometer sensors were sampled using movable windows with fixed widths of 2.56 seconds and 50% overlap (128 readings/window). Separated from each other were the body motion and gravity components of the sensor's acceleration data. A butter low-pass filter was used to adjust for the body's acceleration and gravity. Since it is believed that the gravitational force only consists of low-frequency components, a filter with a 0.3 Hz cutoff frequency was employed. Each window's time and frequency variables were calculated, and the results were converted into a vector of features [17]. The dataset includes the following information for each record:

− Triaxial acceleration is calculated from the body acceleration and the overall acceleration measured by the accelerometer.
− The gyroscope's triaxial angular velocity.
− A vector of five hundred and sixty-one (561) features that include time and frequency domain variables.
− The label of its activity.
− A unique identifier for the experimental subject.

3.2. Feature selection

Because feature selection is so crucial to machine learning, it is routinely performed as part of the pipeline. They're the automated or guide selection of a set of features to optimize the model and forecast output. It also influences the overall performance of the model in phrases of production time as well as accuracy. Feature profiling techniques eliminate the traits without affecting the relaxation. We mentioned a method for function choice that's L1-based function choice. L1-based totally function selection makes use of the coefficients of regression fashions for the choice and interpretation of functions [18]. Linear models sanctioned via the L1 criterion offer numerous separate solutions: lots of their envisioned coefficients are 0. If the aim is to reduce the scale of the information to be used with another workbook, it is able to be used with select from model to outline non-null parameters. Specifically, sporadic estimators useful for this reason are logistic regression and linear SVC for category.

3.3. Machine learning algorithms

Programs that can learn from data and get better with practice are known as machine learning algorithms. IoT can benefit from machine learning algorithms by saving money, getting things done faster, and performing better. Which machine learning algorithm type is most effective depends on the business problem you're trying to solve, the sort of dataset you're using, and the resources you have at your disposal. Here is an overview of the many machine learning techniques used.

3.3.1. Logistic regression (LR)

A basic sort of approach is logistic regression. It is specifically similar to polynomial and linear regression and is a member of the linear classifier organization. The tool of choice for learning binary-type algorithms is logistic regression. With a binomial reaction variable, it really goes much further than linear regression [19]. The ability to use explanatory variables continuously and the simplicity of addressing many explanatory variables at once are also advantages. LR is the best regression method to utilize if the reliant variable is dichotomous (binary). LR is a predictive analysis, just like many other types of regression. In order to explain statistics and the relationship between a single binary dependent variable and one or more independent variables, one or more nominal, ordinal, C programming language, or ratio-level independent variables, there may be one or more LR models used [20]. The LR model gives us a result based on distinct features.

3.3.2. Linear support vector classifier (linear SVC)

Linear SVC is the fit of the given records, returning the "most appropriate" superscript that divides or classifies your records. After you have the hyper stage, we can feed a few functions to the unique classifier to find out what the "expected" class is. On a dataset, a category called linear SVC can perform binary and multi-magnitude categories. If we compare it to the SVC version, the Linear SVC has extra parameters, such as loss feature and penalty normalization that applies "L1" or "L2" [21]. The kernel technique cannot be changed in linear SVC, due to the fact it's far based on the kernel linear technique. The intention of linear SVC is to maximize the margin width between training, has been illustrated in Figure 2.

3.3.3. Decision tree classifier (DTC)

Decisions tree a tree-like shape to symbolize selections, and it's a famous version for machine gaining knowledge of they may be built into a top-down structure using metrics which includes gene impurities and facts. selection timber are modeled for both, class and regression troubles. The selection tree is easy, but it over-splits into tendencies and trains seriously with the schooling records (see Figure 3). To keep away from this, they're typically pruned to save you them from developing further. A choice tree has nodes, such as the choice node and leaf node. Choice nodes are used to make any choice and have many branches, whereas leaf nodes represent the outcomes of those choices and no longer contain any similarly branched choice nodes [22].
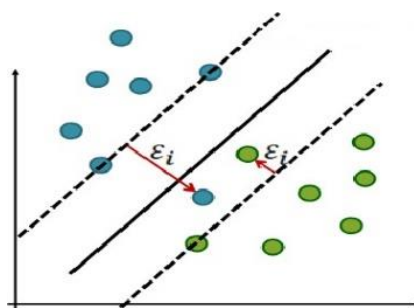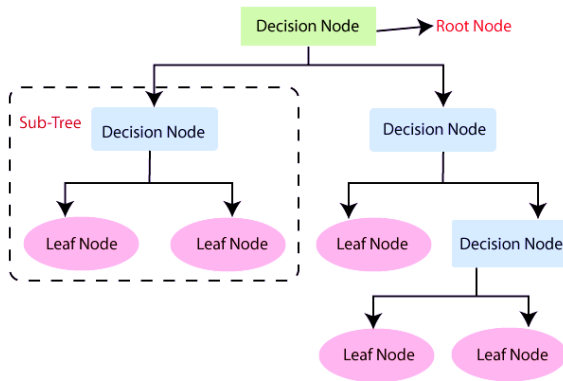

Figure 2. Linear SVC
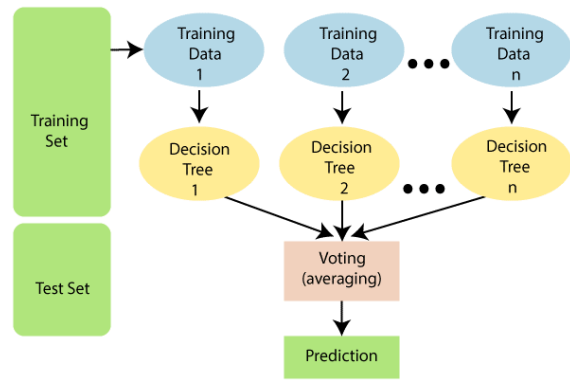
Figure 3. Classification decision tree structure   Figure 4. Random forest structure for classification

3.3.4. Random forrest classifier (RFC)

Random forest classifier is constructing a forest and is a set of selection trees. A random subset of the training records is used to generate a hard and fast of selection bushes, and the decisions from all of the bushes are combined to determine the outcome. The fact that some trees refuse to influence the final decision makes this strategy effective. Averaging all forecasts also eliminates prejudice [23]. Because it does not search for the best feature while dividing a node, random forest differs from selection trees. The collection of rules models both type and regression problems without any issues (see Figure 4).

3.4. Frameworks
3.4.1. NumPy

One of the most popular Python packages for scientific computing is called Numpy. NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, Fourier transform, and matrices. It offers multidimensional array objects and variants like masks that can be applied to a variety of mathematical operations. Numerous other well-liked Python packages, like pandas and matplotlib, are compatible with and depend on Numpy.

3.4.2. Pandas

NumPy is the foundation for the open-source library known as Pandas. Pandas enable quick analysis of data cleansing and preparation. Pandas perform and produce at a high level. Additionally, it offers internal visualization tools. It can handle handling data from a variety of sources.

3.4.3. Scikit-learn or sklearn

The most practical Python library for machine learning is definitely scikit-learn. Numerous effective methods for machine learning and statistical modelings, such as classification, regression, clustering, and dimensionality reduction, are available in the sklearn library. Machine learning models are constructed via Scikit-learn. It shouldn't be used to read, manipulate, and summarize the data.

4.   RESULTS AND DISCUSSION

Human activities are one of the major challenges faced in a smart city. That is due to the lack of resources and the rapid growth of the world's population. In this section, the experiments are presented, where the focus will be driven on the problem of addressing the minimizing of human activities in a city. The experimentation of the proposed architecture has been proposed in this part. The major goal of this research is to offer machine learning approaches for recognizing human activities using smartphone sensors.

4.1. Data information

The experiments had 30 participants, ranging in age from 19 to 48. Each participant completed six tasks while carrying a Samsung Galaxy S II smartphone around their waist (walking, walking upstairs, walking down-stairs, sitting, standing, laying). With the use of the device's internal accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant frequency of 50 Hz [24]. To enable manual labeling of the data, the tests were recorded. The acquired dataset was divided into two sets at random, with 30% of the participants chosen to produce test data and 70% of the volunteers chosen to provide training data.

Prior to sampling, noise filters were applied to the sensor data (accelerometer and gyroscope), which were then sampled in 2.56 second fixed-width sliding windows with 50% overlap (128 readings/window). A Butterworth low-pass filter was used to isolate the body acceleration from the gravitational component of the sensor acceleration data. The gravitational force is thought to only have low-frequency components, hence a filter with a cutoff frequency of 0.3 Hz was used. A vector of properties from each frame was produced by calculating variables in the time and frequency domain. We received 563 characteristics (columns).

− These time-domain signals (prefix 't' to represent time) were collected. To reduce noise, they were filtered with a median filter and a 3rd order low pass Butterworth filter with a 20 Hz corner frequency.

− Using a low pass Butterworth filter with a corner frequency of 0.3 Hz, the acceleration signal was separated into body and gravity acceleration signals ($tBodyAcc - XYZ$ and $tGravityAcc - XYZ$).

− The body linear acceleration and angular velocity were then calculated in time to provide jerk signals ($tBodyAccJerk - XYZ$ and $tBodyGyroJerk - XYZ$). In addition, the Euclidean norm was used to calculate the magnitude of these three-dimensional signals ($tBodyAccMag$, $tGravityAccMag$, $tBodyAccJerkMag$, $tBodyGyroMag$, $tBodyGyroJerk - Mag$).

The study of human activity recognition is still underway. Since machine learning models can't create features directly from unprocessed time series data in this experiment. Machine learning models are trained using the features that specialists extract from the raw data. Machine learning algorithms are used to recognize the accuracy of various human actions. For additional clarification and analysis of the findings. The number of actions for the training data set are presented in Figure 5.
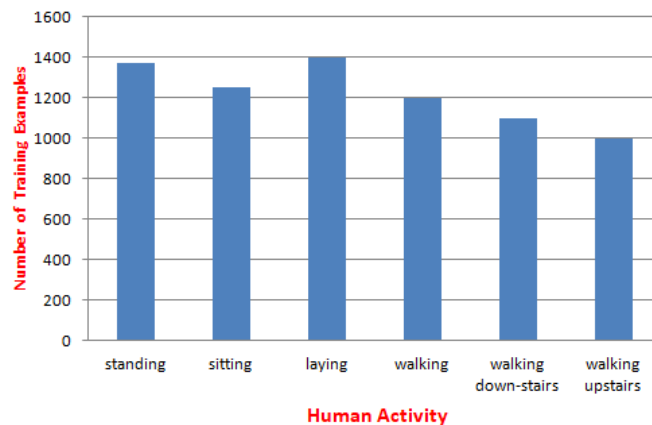

Figure 5. Frequency of activities in training dataset

4.2. Machine learning model evaluation

After selecting the feature, we used a comparison analysis on four classifiers to refine the model. Logistic regression classifier, linear SVC classifier, decision tree classifier, and random forrest classifier were the four classification algorithms tested. To reduce the dimensions of the feature dataset, we employed L1-based feature selection methods, and we compared the outcomes against a scale of the time it took to create and train the model, as well as the model's accuracy. The L1-based feature selection algorithm picked 563 features from the original dataset's 7352 features. "n estimators" was evaluated with values of 100, 50, 200, 250, and 350 to develop the random forrest classifier model. After the figure of 200, it was found that the precision did not rise. (see the result in Table 1).

Table 1 shows the outcomes of the various ML classifiers. Individual classifiers findings outcomes are represented by their ranks in Table 1. The value of the recognition accuracy is used to define the classifier's rank. The classifier with the highest recognition accuracy receives rank 1, followed by the classifier with the second-highest recognition accuracy and the classifier with the lowest recognition accuracy, which receives rank 2. According to Table 1, the classifiers SVM, LR, RFC, and DTC are ranked from 1 to 4 based on their respective recognition accuracy values of 99.06%, 95.83%, 92.09%, and 87.21%. It is clearly demonstrated that SVC performs better among the classifiers chosenfor this study.

Two of the machine learning models better classified all four basic human activity labels. All in all linear support vector classifier performed exceptionally well on expert-generated features with 99.06% accuracy. While the logistic regression model performed comparatively less with an accuracy of 95.83%. Figure 6 model accuracy scores provide further insight and a deeper study of the findings.

We have contrasted our categorization model's results with those from related studies in order to put them into perspective. Table 2 compares the accurate accuracy of categorization rates for various classification techniques with earlier study findings. We have compared the correct accuracy classification rate among various previous studies. We also showed the method of classification adopted in the treatment.

Table 1. Using performance analysis to rank the classifiers

Model name	Accuracy (%)	Precision (%)	Recall (%)	Rank
Logistic regression	95.83	90.22	89.76	2
Linear SVC	99.06	96.45	95.42	1
Decision tree	87.21	82.74	80.94	4
Random forest	92.09	87.38	86.71	3

Table 2. Comparison correct accuracy classification rate between different

| Reference	Classification method	Accuracy rate
Anguita *et al.* [25]	SVM-Gaussian kerne	96.50%
De Leonardis *et al.* [26]	K-nearest neighbor - feedforward neural network - SVM - decision tree - Naïve Bayes	97.00%
Nurhanim *et al.* [27]	SVM polynomial kernel - one versus all	98.57%
Agarwal and Alam [28]	SVMs- k-nearest neighbor- linear discriminant analysis	98.00%
Minarno *et al.* [29]	SVM+LR	98.00%
Jindal *et al.* [30]	SVM, KNN, and LR	92.78%
Patel and Shah [31]	Long short-term, LR	92.00%
Navita and Mittal [32]	SVM	98.03%
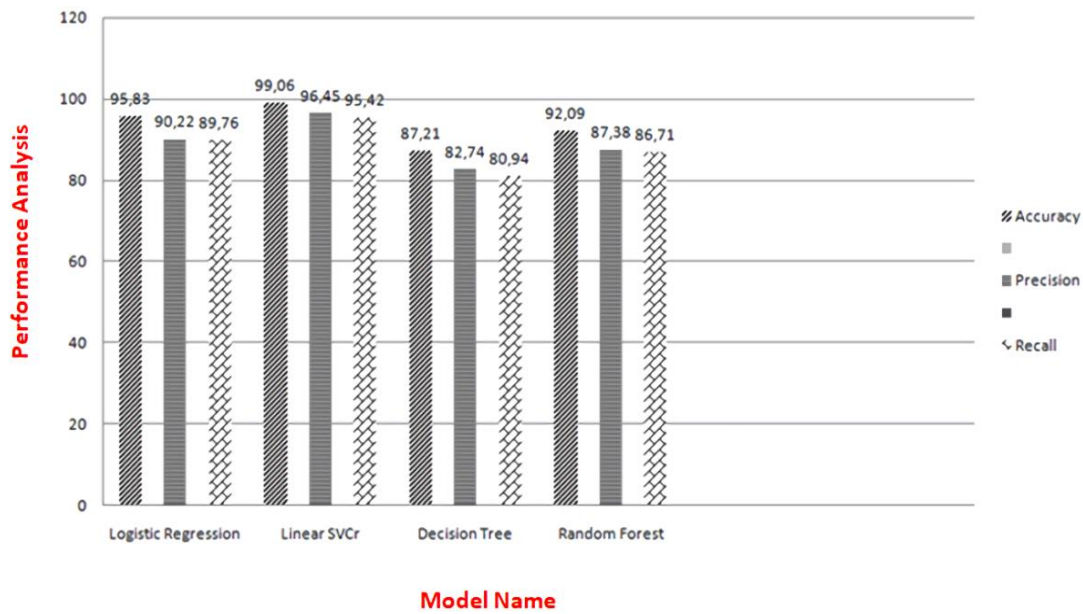
Figure 6. Model accuracy scores

4.3. Discussion

Linear SVC allowed for the classification of human activities to have an accuracy of up to 99%. As displayed in the preceding table. Because the linear SVC technique is based on finding the distance (usually Euclidean distance) between the features extracted from the new window to be categorised and the data in the training set, it is straightforward to apply. It permits an even lower reduction in the necessary characteristics and has low memory and processing requirements. In the majority of applications, we want HAR to be executed in real-time. This calls for the classifier to be developed on the microcontroller of a wearable device and to process the collected data as soon as possible. To construct a more universal physiological activity classifier that can be used on an unknown person without training on themselves first, advancements in feature extraction and classifier design are required. From the previous Table 1 and Table 2, it can be seen that.

−   The logistic regression approach works well when the dataset can be linearly separated and has a good accuracy for many simple data sets. Although it is less likely to do so, it can overfit in datasets with high dimensions. To prevent over-fitting in these cases, regularization (L1 and L2) approaches may be taken into consideration. Independent and dependent variables are, however, linearly connected in linear regression. However, for logistic regression, independent variables must have a linear relationship with the log chances ($log(p/(1-p))$). Because of this, we produced highly respectable and effective results in our work.

−   Decision trees take less work to prepare the data during pre-processing than other methods do. But because of its intricacy and length, decision tree training is relatively expensive. Due to these factors, our research produced average but respectable results.

−   For some datasets with noisy classification/regression tasks, random forests have been seen to overfit. Random forests are biased in favor of attributes with higher levels for data containing categorical variables with various numbers of levels. It provides really helpful feature importance output. Although not the finest, our outcomes are good and acceptable.

−   The linear SVC algorithm is an useful way to solve difficult issues since it is adaptable, provides for a better solution, and unifies several regions. Different regularizations (L1, L2) might be used in the formulation. SVC is a convex optimization problem for which there exist effective solutions. Finding the nature of the relationship between variables is frequently done using linear regression, which almost perfectly fits linearly separable datasets. The more samples there are, the quicker linear SVC tends to converge. This is because Liblinear is tuned for certain cases like the linear kernel, whereas Libsvm is not. The findings are therefore quite good and the best when compared to other algorithms for these aspects and attributes.

In our work, we discovered some limits. Each smart home must train a specific SVM classifier to distinguish between distinct environments' activity. Furthermore, because the SVM classifier must be trained several times from data samples, human tagging is an expensive operation.

5.   CONCLUSION

In this paper, we provided a comprehensive shape for building real-time human activity recognition structures with Kafka and we applied several exclusive algorithms for machine learning of popular and vital activity analysis of human activity. So that we conducted a comparative have a look at between the unique strategies implemented logistic regression classifier, linear SVC classifier, the choice tree classifier, and random forrest classifier on actual wearable sensors' statistics from University of California Irvine (UCI) database are carried out to confirm the effectiveness of the classification algorithms for body activity recognition experiments show that the linear SVC approach is acceptable and near-to-date results for difficult classification troubles whilst being tons quicker than any other algorithm. The set of rules can be carried out to huge information units, despite the fact that complexity increases with the scale of the statistics set. We can use those consequences to song user activity and notify them of their daily activity log, or for instance to screen elderly human beings. Future work would possibly deal with the recognition of greater complicated activities. Many human activities, like cooking, analyzing and looking TV, do not induce significantly unique acceleration traces. For these, ambient light and sound based methods can be explored through using the alternative sensors included in modern-day cell phones. Additional actions and the use of a real-time system on a smartphone might also be taken into consideration.

REFERENCES

[1]   S. Bian, M. Liu, B. Zhou, and P. Lukowicz, "The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey," *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124596.

[2]   A. S. Albahri *et al.*, "IoT-based telemedicine for disease prevention and health promotion: State-of-the-Art," *Journal of Network and Computer Applications*, vol. 173, 2021, doi: 10.1016/j.jnca.2020.102873.

[3]   A. A. Musa, S. Dulawat, K. T. Saleh, and I. A. Alhassan, "Applicability of Utilizing Blockchain Technology in Smart Cities Development," in *Smart City Infrastructure: The Blockchain Perspective*, Scrivener Publishing LLC, 2022, pp. 317-339, doi: 10.1002/9781119785569.ch13.

[4]   B. Zhou, P. Wu, Z. Gu, Z. Wu, and C. Yang, "XDRNet: Deep Learning-based Pedestrian and Vehicle Dead Reckoning Using Smartphones," in *2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2022, pp. 1-8, doi: 10.1109/IPIN54987.2022.9918132.

[5]   C. -Y. Li, M. Renda, F. Yusuf, J. Geller, and S. A. Chun, "Public Health Policy Monitoring through Public Perceptions: A Case of COVID-19 Tweet Analysis," *Information*, vol. 13, no. 11, 2022, doi: 10.3390/info13110543.

[6]   S. Fong, K. Liu, K. Cho, R. Wong, S. Mohammed, and J. Fiaidhi, "Improvised methods for tackling big data stream mining challenges: case study of human activity recognition," *The Journal of Supercomputing*, vol. 72, pp. 3927-3959, 2016, doi: 10.1007/s11227-016-1639-5.

[7]   S. -J. Wei, B. Zhang, X. -W. Tan, X.-G. Zhao, and D. Ye, "A Real-time Human Activity Recognition Approach with Generalization Performance," in *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 6334-6339, doi: 10.23919/CCC50068.2020.9188860.

[8]   M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307-313, 2018, doi: 10.1016/j.future.2017.11.029.

[9]   W. Qi, H. Su, and A. Aliverti, "A Smartphone-Based Adaptive Recognition and Real-Time Monitoring System for Human Activities," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 5, pp. 414-423, 2020, doi: 10.1109/THMS.2020.2984181.

[10]   D. J. Cook, G. Duncan, G. Sprint, and R. L. Fritz, "Using Smart City Technology to Make Healthcare Smarter," in *Proc. of the IEEE*, vol. 106, no. 4, pp. 708-722, 2018, doi: 10.1109/JPROC.2017.2787688.

[11]   M. Kuzlu, H. Kalkavan, O. Gueler, N. Zohrabi, P. J. Martin, and S. Abdelwahed, "An End to End Data Collection Architecture for IoT Devices in Smart Cities," in *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2022, pp. 1-5, doi: 10.1109/ISGT50606.2022.9903049.

[12]   C. Martín, P. Langendoerfer, P. S. Zarrin, M. Díaz, and B. Rubio, "Kafka-ML: Connecting the data stream with ML/AI frameworks," *Future Generation Computer Systems*, vol. 126, pp. 15-33, 2022, doi: 10.1016/j.future.2021.07.037.

[13]   G. Demirdöğen, Z. Işık, and Y. Arayici, "BIM-based big data analytic system for healthcare facility management," *Journal of Building Engineering*, vol. 64, 2023, doi: 10.1016/j.jobe.2022.105713.

[14]   J. S.-Pereira, L. Gruenwald, and J. Bernardino, "Top data mining tools for the healthcare industry," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no, 8, pp. 4968-4982, 2022, doi: 10.1016/j.jksuci.2021.06.002.

[15]   P. Unal, Ö. Albayrak, M. Jomâa, and A. J. Berre, "Data-Driven Artificial Intelligence and Predictive Analytics for the Maintenance of Industrial Machinery with Hybrid and Cognitive Digital Twins," in *Technologies and Applications for Big Data Value*, Éd. Cham: Springer International Publishing, 2022, pp. 299-319, doi: 10.1007/978-3-030-78307-5_14.

[16]   C. Bedon and S. Mattei, "Facial Expression-Based Experimental Analysis of Human Reactions and Psychological Comfort on Glass Structures in Buildings," *Buildings*, vol. 11, no. 5, 2021, doi: 10.3390/buildings11050204.

[17]   D. Thakur and S. Biswas, "Guided regularized random forest feature selection for smartphone based human activity recognition," *Journal of Ambient Intelligence and Humanized Computing*, 2022, doi: 10.1007/s12652-022-03862-5.

[18]   I. Tsamardinos, "Just Add Data: automated predictive modeling for knowledge discovery and feature selection," *Npj Precision Oncology*, vol. 6, no. 38, 2022, doi: 10.1038/s41698-022-00274-8.

[19]   N. Naz *et al.*, "Ensemble learning-based IDS for sensors telemetry data in IoT networks", *Mathematical Biosciences and Engineering*, vol. 19, no. 10, pp. 10550-10580, 2022, doi: 10.3934/mbe.2022493.

[20]   D. Weisburd, D. B. Wilson, A. Wooditch, and C. Britt, "Logistic Regression," in *Advanced Statistics in Criminology and Criminal Justice*, Éd. Cham: Springer International Publishing, 2022, pp. 127-185, doi: 10.1007/978-3-030-67738-1_4.

[21]   D. Jiang, W. Lin, and N. Raghavan, "A Novel Framework for Semiconductor Manufacturing Final Test Yield Classification Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 197885-197895, 2020, doi: 10.1109/ACCESS.2020.3034680.

[22]   S. Dargad, P. Thakkar, and S. Giri, "Machine Learning for Classification of DOS Attack in Smart Healthcare Networks," in *International Conference on Computing Science, Communication and Security*, 2022, pp. 184-197, doi: 10.1007/978-3-031-10551-7_14.

[23]   M. Achite, S. Samadianfard, N. Elshaboury, and M. Sharafi, "Modeling and optimization of coagulant dosage in water treatment plants using hybridized random forest model with genetic algorithm optimization," *Environment, Development and Sustainability*, 2022, doi: 10.1007/s10668-022-02523-z.

[24]   V. Vijayan, J. P. Connolly, J. Condell, N. McKelvey, and P. Gardiner, "Review of Wearable Devices and Data Collection Considerations for Connected Health," *Sensors*, vol. 21, no. 16, 2021, doi: 10.3390/s21165589.

[25]   D. Anguita, A. Ghio, L. Oneto, X. P. Perez, and J. L. R. Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013, pp. 437-442. [Online]. Available: https://upcommons.upc.edu/handle/2117/20897

[26]   G. De Leonardis *et al.*, "Human Activity Recognition by Wearable Sensors: Comparison of different classifiers for real-time applications," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438750.

[27]   K. Nurhanim, I. Elamvazuthi, L. I. Izhar, and T. Ganesan, "Classification of human activity based on smartphone inertial sensor using support vector machine," in *2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA)*, 2017, pp. 1-5, doi: 10.1109/ROMA.2017.8231736.

[28]   P. Agarwal and M. Alam, "Quantum-Inspired Support Vector Machines for Human Activity Recognition in Industry 4.0," in *Proc. of Data Analytics and Management*, 2022, pp. 281-290, doi: 10.1007/978-981-16-6289-8_24.

[29]   A. E. Minarno, W. A. Kusuma, and H. Wibowo, "Performance Comparisson Activity Recognition using Logistic Regression and Support Vector Machine," in *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2020, pp. 19-24, doi: 10.1109/ICoIAS49312.2020.9081858.

[30]   S. Jindal, M. Sachdeva, and A. K. S. Kushwaha, "Performance evaluation of machine learning based voting classifier system for human activity recognition," *Kuwait Journal of Science*, pp. 1-12, 2022, doi: 10.48129/kjs.splml.19189.

[31]   A. D. Patel and J. H. Shah, "Performance Analysis of Supervised Machine Learning Algorithms to Recognize Human Activity in Ambient Assisted Living Environment," in *2019 IEEE 16th India Council International Conference (INDICON)*, 2019, pp. 1-4, doi: 10.1109/INDICON47234.2019.9030353.

[32]   Navita and P. Mittal, "Machine Learning (ML) based Human Activity Recognition Model using Smart Sensors in IoT Environment," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2022, pp. 330-334, doi: 10.1109/Confluence52989.2022.9734152.

BIOGRAPHIES OF AUTHORS

Kamel Maaloul obtained an engineer in computer media from Skikda University, Skikda, Algeria in 2004, and a master's in artificial intelligence from El-Oued University, El-Oued, Algeria in 2017. He is currently a PhD student from El-Oued University, El-Oued, Algeria. And a principal engineer in computer media at El-Oued University, El-Oued, Algeria. He can be contacted at email: maaloul-kamel@univ-eloued.dz.

Lejdel Brahim currently a Full Professor in the Faculty of Exact Sciences, University of EL-Oued (Algeria). He has held a Magister Degree in Computer Science from University of Ouargla since 2009. He held his PhD in computer science in 2015. He is the author of more than 120 published articles in refereed journals and International Conferences. He publishes four books in his research domain. He is the chair of many international conferences. He works as an invited editor, reviewer, and others in many journals and conferences. He is the supervisor of many students in computer science in License, master and doctorate. He can be contacted at email: lejdel.brahim@gmail.com.

Nedioui Med Abdelhamid is an associate Professor in the Faculty of Exact Sciences, University of EL-Oued (Algeria). He has held a Magister Degree in Computer Science from University of Biskra since 2015. He held his PhD in computer science in 2021. He can be contacted at email: nedioui3904@gmail.com.