

## Research Article

# An Analysis of Audio Features to Develop a Human Activity Recognition Model Using Genetic Algorithms, Random Forests, and Neural Networks

**Carlos E. Galván-Tejada,<sup>1</sup> Jorge I. Galván-Tejada,<sup>1</sup>  
José M. Celaya-Padilla,<sup>1</sup> J. Rubén Delgado-Contreras,<sup>2</sup>  
Rafael Magallanes-Quintanar,<sup>1</sup> Margarita L. Martínez-Fierro,<sup>3</sup> Idalia Garza-Veloz,<sup>3</sup>  
Yamilé López-Hernández,<sup>3</sup> and Hamurabi Gamboa-Rosales<sup>1</sup>**

<sup>1</sup>Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas,  
Jardín Juárez 147 Centro, 98000 Zacatecas, ZAC, Mexico

<sup>2</sup>Instituto Tecnológico Superior Zacatecas Sur, Av. Tecnológico 100, Las Moritas, 99700 Tlaltenango, ZAC, Mexico

<sup>3</sup>Unidad Académica de Medicina Humana y Ciencias de la Salud, Universidad Autónoma de Zacatecas,  
Jardín Juárez 147 Centro, 98000 Zacatecas, ZAC, Mexico

Correspondence should be addressed to Carlos E. Galván-Tejada; [ericgalvan@uaz.edu.mx](mailto:ericgalvan@uaz.edu.mx)

Received 1 June 2016; Revised 31 August 2016; Accepted 29 September 2016

Academic Editor: Daniele Riboni

Copyright © 2016 Carlos E. Galván-Tejada et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work presents a human activity recognition (HAR) model based on audio features. The use of sound as an information source for HAR models represents a challenge because sound wave analyses generate very large amounts of data. However, feature selection techniques may reduce the amount of data required to represent an audio signal sample. Some of the audio features that were analyzed include Mel-frequency cepstral coefficients (MFCC). Although MFCC are commonly used in voice and instrument recognition, their utility within HAR models is yet to be confirmed, and this work validates their usefulness. Additionally, statistical features were extracted from the audio samples to generate the proposed HAR model. The size of the information is necessary to conform a HAR model impact directly on the accuracy of the model. This problem also was tackled in the present work; our results indicate that we are capable of recognizing a human activity with an accuracy of 85% using the HAR model proposed. This means that minimum computational costs are needed, thus allowing portable devices to identify human activities using audio as an information source.

## 1. Introduction

The capacity to recognize the activity currently being performed by either oneself or someone else is an inherent behavior of an intelligent system, reason why the human activity recognition (HAR) is currently a relevant research topic [1–5]. There is a wide range of areas in which HAR can be applied, such as automatic vigilance, elderly care, entertainment, and residential activities support [6–11].

HAR is the principal component that allows people to recognize a high level human behavior and therefore to identify routines and social interactions. In consequence,

proposals using different techniques and information sources have been previously published. Additionally, efforts have also aimed at fusing different well-known techniques and information sources in order to increase the accuracy of the systems and the coverage ratio. Some of these analyses are presented next.

Moore et al. [12] presented a framework titled *ObjectSpaces* that uses familiar object-oriented constructs like classes and inheritance to manage object context and allows for the classification of activities. In said work the authors demonstrated that both familiar and previously unseen objects could be recognized using action and context

information. However, there were some constraints in this proposal. For example, the association of actions with objects can reduce the activities being recognized, even with a well-done domain of action, since sometimes people perform activities using unusual objects. Additionally, video is a complex signal and needs cameras deployed in the environment, therefore two information sources are required.

An increasing number of portable devices contain multiple sensors, all with capability of recording information from different sources. These devices have been used in several analyses to recognize human activities. The one presented by Lester et al. [13] proposes the gathering of data from different sources using one device and the use of a modified version of the Adaptive Boosting algorithm (AdaBoost) for feature selection. A similar work proposed the use of Hidden Markov Models (HMM) to classify certain activities [14].

Other approaches are based on the methodology and theories of social psychology, collecting audio data that can be tagged with Essential Social Interaction Predicates (ESIP). A relevant example of this technique is proposed by Lester et al. [13], who built a model named Discriminative Conditional Restricted Boltzmann Machine (DCRBM), this model combines a discriminating approach with the capabilities of the Conditional Restricted Boltzmann Machine (CRBM). The model allows the discovery of actionable components from ESIP to train the DCRBM model and use it to generate low-level data corresponding to the ESIP with a high degree of accuracy.

Binary silhouettes have also been used to represent the different human activities. Uddin et al. proposed a system based on Generalized Discriminant Analysis on Enhanced Independent Component, obtaining features from binary silhouettes information to be used with Hidden Markov Models for training and recognition [15]. Likewise, principal component analysis (PCA) [16–18] and independent component analysis (ICA) [19] have also been used for this purpose.

The process of feature extraction is used to represent a signal using several derived values or features intended to be informative and nonredundant. This process could easily result in very large sets of features that describe a signal. Nevertheless, not all extracted features will be useful in discerning between different kinds of signals (i.e., signals representing different activities), reason why feature selection is also needed. In order to determine which set of features could accurately classify different activities, a feature selection analysis that included the use of a genetic algorithm, forward selection and backward elimination steps, and a random forest (RF) or neural network (NN) algorithm was implemented.

Briefly, we attempted to identify a small set of features that could accurately classify eight different activities using only features derived from audio recordings. Additionally, the accuracy of such a model was compared against models with no size restrictions and against models obtained using a different state-of-the-art methodology.

## 2. Materials and Methods

Three main tasks were performed to generate the HAR model: audio collection, feature extraction, and feature selection.



FIGURE 1: Typical distance between mobile phone and activity.

Feature extraction and selection were performed using R (<https://www.r-project.org/>).

*2.1. Data Description.* The dataset is comprised of seven activities and additionally a collection of no activity noises, commonly performed in a residential setup: Namely, brewing coffee, cooking, using the microwave oven, taking a shower, dish washing, hand washing, teeth brushing, and no activity noises. We generated individual sounds to collect information from activities. Table 1 shows the type and a short description. It is worth mentioning that four of these activities have in running water a similar background sound, adding to the complexity of the HAR problem. All recordings have been made available through the AmiDaMi research group page at <http://ingsoftware.reduaz.mx/amidami/>.

*2.1.1. Recording Devices.* The devices used to record all the audio clips were chosen given the different specifications from the microphones embedded in these. In Table 2 are shown system on chip (SoC) and operating system from the selected mobile phones, to know the hardware and software involved in internal audio recording and preprocessing process.

*2.1.2. Spatial Environments.* In order to cover a wide gamma of sounds, all were recorded in different houses meaning different spatial environments, audio reflections, and background sounds. Additionally different home facilities mean different cookware, home appliances, and running water reflections. Mobile phone laid near where the activity was being performed. An example of which, shown in Figure 1, was determined in order to record the sounds as clearly as possible.

*2.1.3. Metadata.* Audio clips with a sample rate between 8,000 Hz and 44,100 Hz and Mono and Stereo recordings were done depending on the device used to record the audio clip. The range of the sample rate assured that most mobile phones were able to be included, allowing for a future

TABLE 1: Activities general description.

Activity	Description
Brew coffee	Brewing coffee in stove using coffee pots and in coffee machines
Cooking	Cooking meat and scrambled eggs in stove
Use microwave oven	Using several microwave ovens to heat up water and meal
Take a shower	Taking a shower in different environments; in some cases water fall was interrupted on intervals
Dish washing	Dishes were washed by hand individually or in groups of different dishes; water noise is in background
Hand washing	Washing hands with bar soap
Teeth brushing	Audio clips start from opening the tap to closing it
No activity	No activity audio clips which are mostly noises added by the device used to record

TABLE 2: Selected mobile phones system on chip and operating system.

Smartphone	System on chip (SoC)	Operating system
Lanix Ilium s600	Qualcomm Snapdragon 210 MSM8909	Android 5.1
LG G Pro Lite	MediaTek MT6577	Android 4.1.2
iPhone 4	Apple A4 APL0398	iOS 4
iPhone 3GS	Samsung S5PC100	iOS 3
HTC One M7	Qualcomm Snapdragon 600 APQ8064T	Android 4.1.2

expansion of the database. In Table 3 is shown the summary of metadata for each activity performed in this dataset.

*2.1.4. Data Preparation.* In this work, all the audio samples have no other preprocessing than trimming the samples in 10-second clips, no other audio processing was performed in order to simplify the implementation in any device.

*2.2. Feature Extraction.* In order to obtain information that could potentially distinguish which activity was being performed, several features were extracted from the audio clips. As shown in [20, 21], 10-second audio clips seem to be adequate for this task. However, some activities lasted longer than ten seconds, yielding longer than needed record samples. Such recordings were trimmed into as many 10-second audio clips as possible. To avoid issues between Mono and Stereo recordings, only the information from the left channel of the latter kind of recordings was used. Each 10-second clip was then converted into an integer array, where each integer represented the magnitude of the sound wave in that instant of time. Even though all clips had the same duration, the length of the arrays that represented them varied from 80,000 to 441,000 samples depending on the sample rate of the original recording.

Features that statistically describe sound waves have previously been found to be of importance in solving similar problems [22]. Thus, the 16 statistical features listed in the following list were extracted from each sample. Mascia et al. showed that Mel-frequency cepstral coefficients (MFCC) [23] can be used to identify acoustic descriptors from the environmental sound [24] and thus were also extracted from the audio samples. To do so, each 10-second audio clip was split into ten 1-second audio clips, from which 12 cepstral coefficients were calculated, resulting in 120 MFCC per sample. In order to avoid the matrix that is generated

during the extraction of the MFCC, the vectorization process showcased by Mascia et al. [24] was performed.

#### *Statistical Features That Were Extracted from Each Sample*

kurtosis of the probability distribution of the integer array,

skewness of the probability distribution of the integer array,

mean of the integer array,

median of the integer array,

standard deviation of the integer array,

variance of the integer array,

coefficient of variation (CV) of the probability distribution of the integer array

inverse CV,

1st, 5th, 25th, 50th, 75th, 95th, and 99th percentile of the probability distribution of the integer array,

mean of the integer array after trimming the bottom and top 5% elements.

In order to avoid any outlier-related problems, features were rank-normalized as described in (1), where  $x_i$  is the  $i$ th value of the feature  $x$  and  $x'_i$  is the  $i$ th value of the rank-normalized feature  $x'$ . As a result, all features range between 0 and 1, with equidistant steps between each element of the array:

$$x'_i = \frac{\text{trunc}(\text{rank}(x_i))}{\text{length}(x)}, \quad (1)$$

TABLE 3: Audio clips metadata per activity.

Activity	Sample rate	Encoding format	Channels
Brew coffee	8,000 Hz–44,100 Hz	m4a, amr	Stereo, Mono
Cooking	44,100 Hz	m4a	Stereo
Use microwave oven	44,100 Hz	m4a	Stereo
Take a shower	44,100 Hz	m4a, mp3	Stereo
Dish washing	44,100 Hz	m4a	Stereo
Hand washing	8,000 Hz–44,100 Hz	m4a, amr	Stereo, Mono
Brushing teeth	44,100 Hz	m4a	Stereo
No activity	8,000 Hz–44,100 Hz	m4a, amr	Stereo, Mono

*2.3. Feature Selection.* In the first step of the feature selection process, a genetic algorithm called Galgo [25] reduced the size of the database by determining which features had more chances of being useful. To do so, a set of random five-feature models were evolved throughout 200 generations, during which the models mutated, reproduced, and recombined, eventually yielding a highly accurate model. Fitness was defined as the accuracy of the model to classify the eight previously defined activities using a nearest-centroid approach and following a 3-fold train-test methodology. This whole process was repeated 300 times, resulting in 300 highly accurate five-feature models. The number of times each feature was found in these models was used to determine a feature rank, which described the potential classification capabilities of each feature.

Based on this rank, forward selection and backward elimination were performed, defining which features were to be used in the next stage of the feature selection procedure. Forward selection is a well-known methodology used to build models at low computational costs. From the list of ranked features, this approach added one feature at a time and evaluated the performance of the models. Once the last feature was added and the model with all features was evaluated, the features from the model that achieved the highest accuracy were kept, and the rest were disregarded. Backward elimination was then performed to avoid redundant information and to further reduce the amount of features to be used. This process consisted in removing one feature at a time and evaluating the performance of the model, starting with the final model of the forward selection procedure and removing first the least frequent feature in the 300 Galgo-generated models. If eliminating a feature did not decrease the accuracy of the model, then such a feature was removed from the final model. This process was repeated until model stability was achieved. Accuracy in both the forward selection and the backward elimination procedures was measured following a 3-fold train-test methodology, using the same folds as the ones used in the genetic algorithm.

The features selected by the backward elimination algorithm were then used to generate two HAR models, one through a random forest (RF) implementation [26] and the other through a neural network (NN) one. RF is a robust machine learning technique that can handle classification problems based on bagging and random feature selection [27, 28]. Moreover, it allows for the calculation of the error

TABLE 4: Instances per activity.

Activity	Recordings	10-second audio clips
Brew coffee	9	245
Cooking	6	132
Use microwave oven	3	42
No activity	5	16
Take a shower	11	428
Washing dishes	6	134
Washing hands	15	70
Brushing teeth	9	92
No activity	5	16

during the model generation without having to split the data into train and test sets. This algorithm uses an out-of-bag (OOB) error, an unbiased estimate of the true prediction error, in which as the forest is built; each tree can be tested on the samples not used in building that tree. Breiman [29] demonstrated that estimating the OOB error has the same result as estimating the error using a test set of the same size as the training set. The NN model was obtained using 70% of the data to train the model and the rest to test it. This algorithm was analyzed because it can be easily implemented in a mobile phone, giving ubiquity to the proposed HAR solution. These models were compared with an RF- and an NN-based model that included all original features, an RF- and an NN-based model that included all MFCC features but no statistical features, and a model developed by Kabir et al. [30] was evaluated in three different setups.

### 3. Results

Audio collection resulted in a total of 64 recordings and 1,159 10-second audio clips. Table 4 details the number of recordings and audio clips obtained for each activity. Since each audio clip had 16 statistical and 120 MFCC features extracted from it, the final database had a size of  $1,159 \times 136$  elements.

Figure 2 shows how each one of the 300 Galgo-generated models evolved throughout 200 generations, yielding an average accuracy of 0.68. It can also be seen that the models had achieved stability; that is, no more generations were needed. Similarly, Figure 3 shows that the frequency in which

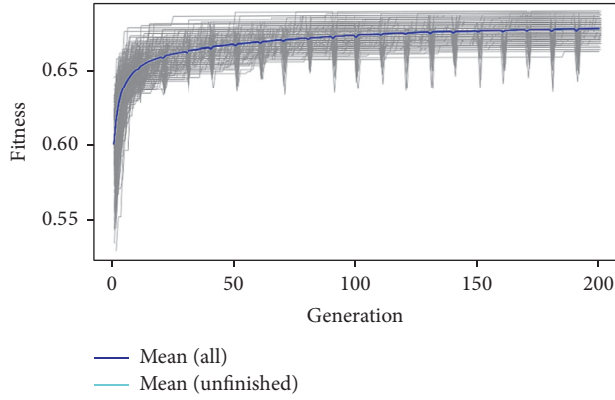


FIGURE 2: Evolution of the Galgo-generated models.

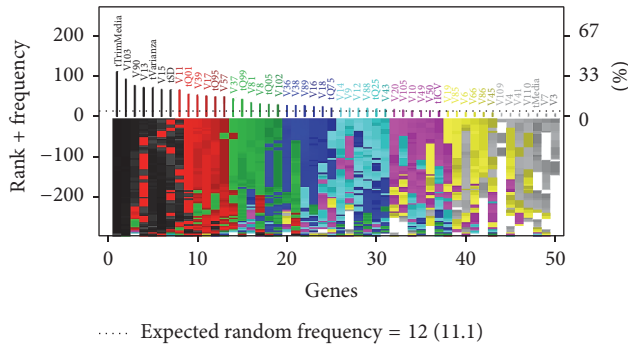


FIGURE 3: Evolution of the rank for the most frequent features.

features appeared in the 300 models had stabilized, at least for the 30 most frequent features. This means that even though more models had been generated, the rank of the most frequent features would not have changed.

The forward selection procedure selected the 35 most frequent features, and the backward elimination strategy removed 25 of them. This resulted in only 9 features with potential classification power: the trimmed mean, the standard deviation, the 95th percentile, and 6 MFCC. These features, whose heat map is shown in Figure 4, were used to generate an RF- and an NN-based HAR model. The weights of the NN-based HAR model were adjusted throughout 100 iterations and the RF-based HAR model was adjusted using 5,000 trees.

Table 5 shows the classification accuracy achieved by each model. It can be noted that by adding more features to the RF-based HAR model the accuracy increased. That is, the model with all features had the highest accuracy, followed by the model with all MFCC features and then by the model with 9 features. Conversely, the NN-based HAR model decreased its accuracy when more features were included, having its best performance when only the 9 selected features were used. Nonetheless, both 9-feature models were able to outperform the model proposed by Kabir et al. Confusion matrices that describe how each model classified each sample are shown in Tables 6, 7, 8, 9, 10, and 11. In addition, Table 5 shows that independently from the scenario, models based in RF

approach outperform all the others scenarios, including Kabir et al. ones.

#### 4. Discussion and Conclusions

The focus of this research is to find features that describe efficiently the behavior of audio signal that represents activities performed by humans in order to develop a HAR model using well-known machine learning techniques that can be used in low power consumption and mobile devices, to provide context information. The results, presented in Section 3, allowed us to identify the following aspects to answer questions presented in Section 1:

- (i) *Mel-Frequency Cepstral Coefficients describe better the behavior of the audio signal:* we identified that MFCC describe accurately the behavior of an audio signal used to generate a HAR model. Even after the feature selection process, the model was composed of 9 features, 6 of them are MFCC descriptors.
- (ii) *Statistical features are important describing audio signals:* we propose the use of signal time evolution and the first and second statistical features extracted that can be computed at low computational cost. This means that can be used in low computational processors, as new portable embedded systems (i.e., Arduino and Galileo among others) or low-cost smart phones and in high-end smart phones at low battery cost. In results section we can see that even though several feature selection procedures were carried on, three of them survived until the final features set, meaning that can be used to HAR at low computational cost.
- (iii) *Selected features can describe the behavior of the audio signal losing some fitness than all the features together:* the aim of feature selection is to reduce the computational cost and to maximize fitness; nevertheless, the selected features cannot describe the behavior as when using all features. Yet, we consider that reduction of features needed to be extracted from the signal is more important to allow ubiquity in future works, given the lower computational cost and lower battery consumption processing the audio signal.

One of the strongest points presented in this research is the use of random forest for human activity recognition targeting quasi-real time mobile applications. The presented methodology builds a Random forest for said classification; the procedure is consistent and adapts to sparsity; its rate of convergence depends only on the number of strong features and not on how many noise variables are present [31]; the complexity of the classification methodology is  $\approx O(M(mn \log n))$ , where  $m$  is the number of features,  $n$  is the number of instances, and  $M$  is the number of trees. Nevertheless, once the model is trained and refined, the model could be transferred to a mobile application; said application will only be using the random forest structure and values, and it will not need to retrain the model inside the mobile device avoiding said computation complexity.

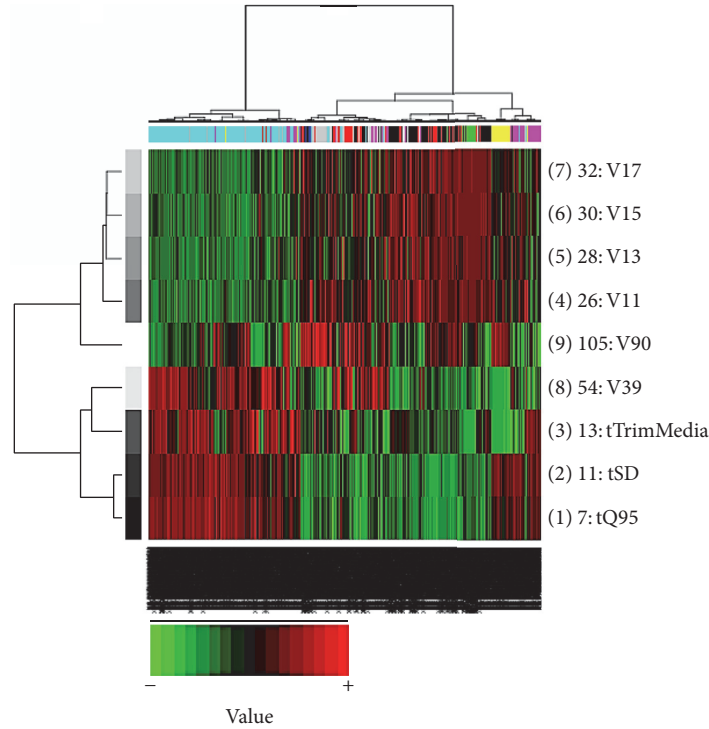


FIGURE 4: Heat map of the 9 features with potential classification capabilities.

TABLE 5: Evaluation of each model.

HAR model	Accuracy	Number of samples	Number of activities
RF with 9 features	0.814	1,191	8
NN with 9 features	0.777	319	8
RF with all features	0.856	1,191	8
NN with all features	0.646	319	8
RF with all MFCC features	0.842	1,191	8
NN with all MFCC features	0.110	319	8
Kabir et al. scenario A	0.748	1,000	10
Kabir et al. scenario B	0.748	1,300	13
Kabir et al. scenario C	0.733	1,600	16

Additionally we found particular behavior of neural networks (NN) when only MFCC are used to describe the audio signal; in this scenario a NN with a single hidden layer cannot adjust weights efficiently leading us to worst misclassification case, this given by the fact that MFCC after percentile rank normalization tends to have low standard deviation meaning a high similarity which tends to overfit the NN; nevertheless, the use of MFCC in combination with selected statistical features produces an efficient model. These arguments allow us to conclude that best results are obtained using a clever selection of statistical and MFCC features.

## Future Work

As part of the future work, we propose adding more activities that are commonly performed in residential homes; additionally we propose adding more features to be extracted and a

clever feature selection to reduce even further the amount of data needed. The proposed future work is as follows:

- (i) To study other human activities in residential homes
- (ii) The use of different features with efficient computational cost
- (iii) The implementation of the predictive models in a mobile application, for a real world deployment
- (iv) Implement other feature selection techniques to do a comparison and acquire descriptive features

Also, compare the obtained models with a second group of models obtained using a clusterization technique to evaluate which can offer better results in a mobile application implementation. Once the model is built and optimized, we plan to implement the methodology in an application for mobile phone. Also, it aims to check whether these applications can

TABLE 6: Confusion matrix of the RF-based HAR model with 9 features.

	Brew coffee	Cooking	Use microwave oven	No activity	Take a bath	Dish washing	Hand washing	Teeth brushing	Type II error
Brew coffee	218	12	0	1	0	8	1	5	0.110
Cooking	44	74	0	0	8	1	0	5	0.439
Use microwave oven	6	2	34	0	0	0	0	0	0.190
Take a bath	6	3	0	1	399	9	0	10	0.068
Dish washing	3	7	0	0	15	99	1	9	0.261
Hand washing	1	0	0	0	6	4	56	3	0.200
Teeth brushing	12	8	0	0	17	8	1	78	0.371
No activity	2	0	0	11	1	0	0	2	0.313
Type I error	0.253	0.302	0.000	0.154	0.105	0.233	0.051	0.304	

TABLE 7: Confusion matrix of the NN-based HAR model with 9 features.

	Brew coffee	Cooking	Use microwave oven	No activity	Take a bath	Dish washing	Hand washing	Teeth brushing	Type II error
Brew coffee	45	10	2	0	3	1	0	3	0.297
Cooking	9	22	0	0	2	2	0	1	0.389
Use microwave oven	1	1	10	0	0	0	0	1	0.231
Take a bath	0	1	0	0	119	9	2	3	0.112
Dish washing	3	0	0	0	3	24	0	1	0.226
Hand washing	0	0	0	0	0	1	12	2	0.200
Teeth brushing	0	1	0	0	3	0	2	13	0.316
No activity	2	0	1	3	0	0	0	1	0.571
Type I error	0.250	0.371	0.231	0.000	0.085	0.351	0.250	0.480	

TABLE 8: Confusion matrix of the RF-based HAR model with all features.

	Brew coffee	Cooking	Use microwave oven	No activity	Take a bath	Dish washing	Hand washing	Teeth brushing	Type II error
Brew coffee	232	0	0	1	0	8	2	2	0.053
Cooking	50	77	0	0	5	0	0	0	0.417
Use microwave oven	9	0	33	0	0	0	0	0	0.214
Take a bath	4	5	0	1	399	9	0	10	0.068
Dish washing	4	7	0	0	11	106	2	4	0.209
Hand washing	0	0	0	1	3	3	60	3	0.143
Teeth brushing	4	5	0	0	15	0	1	99	0.202
No activity	1	0	0	13	0	0	0	2	0.188
Type I error	0.237	0.181	0.000	0.188	0.079	0.159	0.077	0.175	





have a beneficial use in various areas of everyday life, from personal security to medical application.

## Competing Interests

The authors declare that there are no competing interests regarding the publication of this manuscript.

## References

- [1] N. Oliver, E. Horvitz, and A. Garg, "Layered representations for human activity recognition," in *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, pp. 3–8, IEEE, Pittsburgh, Pa, USA, 2002.
- [2] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48–53, 2010.
- [3] H. Kataoka, K. Hashimoto, and Y. Aoki, "Feature integration with random forests for real-time human activity recognition," in *Proceedings of the 7th International Conference on Machine Vision (ICMV '14)*, International Society for Optics and Photonics, Milan, Italy, November 2014.
- [4] A. Masuda, K. Zhang, and T. Maekawa, "Sonic home: environmental sound collection game for human activity recognition," *Journal of Information Processing*, vol. 24, no. 2, pp. 203–210, 2016.
- [5] M. A. M. Shaikh, K. Hirose, and M. Ishizuka, *Recognition of Real-World Activities from Environmental Sound Cues to Create Life-Log*, INTECH Open Access, 2011.
- [6] E. A. Mosabbeeb, R. Cabral, F. De la Torre, and M. Fathy, "Multi-label discriminative weakly-supervised human activity recognition and localization," in *Computer Vision—ACCV 2014*, D. Cremers, I. Reid, H. Saito, and M.-H. Yang, Eds., vol. 9007 of *Lecture Notes in Computer Science*, pp. 241–258, Springer, Berlin, Germany, 2015.
- [7] K. Zhan, S. Faux, and F. Ramos, "Multi-scale Conditional Random Fields for first-person activity recognition on elders and disabled patients," *Pervasive and Mobile Computing*, vol. 16, pp. 251–267, 2015.
- [8] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: a survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.
- [9] Y. Zhan and T. Kuroda, "Wearable sensor-based human activity recognition from environmental background sounds," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 1, pp. 77–89, 2014.
- [10] J. M. Sim, Y. Lee, and O. Kwon, "Acoustic sensor based recognition of human activity in everyday life for smart home services," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 679123, 24 pages, 2015.
- [11] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "Acoustic detection of human activities in natural environments," *Journal of the Audio Engineering Society*, vol. 60, no. 9, pp. 686–695, 2012.
- [12] D. J. Moore, I. A. Essa, and M. H. Hayes III, "Exploiting human actions and object context for recognition tasks," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 1, pp. 80–86, IEEE, September 1999.
- [13] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI '05)*, pp. 766–772, Edinburgh, UK, August 2005.
- [14] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [15] M. Z. Uddin, D.-H. Kim, and T.-S. Kim, "A human activity recognition system using HMMs with GDA on enhanced independent component features," *International Arab Journal of Information Technology*, vol. 12, no. 3, pp. 304–310, 2015.
- [16] I. T. Jolliffe, *Principal Component Analysis*, Wiley Online Library, Chichester, UK, 2002.
- [17] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on  $\mathfrak{R}$  transform," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, IEEE, June 2007.
- [18] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete cosine transform and SVM," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 5041–5044, IEEE, San Antonio, Tex, USA, October 2009.
- [19] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 46, John Wiley & Sons, New York, NY, USA, 2004.
- [20] J. R. Delgado-Contreras, J. P. Garcá-Vázquez, R. F. Brena, C. E. Galván-Tejada, and J. I. Galván-Tejada, "Feature selection for place classification through environmental sounds," *Procedia Computer Science*, vol. 37, pp. 40–47, 2014.
- [21] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Demo: indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*, pp. 385–386, ACM, July 2011.
- [22] A. Martínez-Torteya, V. Treviño-Alvarado, and J. Tamez-Peña, "Improved multimodal biomarkers for Alzheimer's disease and mild cognitive impairment diagnosis: data from ADNI," in *Proceedings of the Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670 of *Proceedings of SPIE*, pp. 86700S–86700S, International Society for Optics and Photonics, Photonics, Orlando, Fla, USA, February 2013.
- [23] U. Ligges, S. Krey, O. Mersmann, and S. Schnackenberg, *tuneR: analysis of music*, 2011.
- [24] M. Mascia, A. Canclini, F. Antonacci, M. Tagliasacchi, A. Sarti, and S. Tubaro, "Forensic and anti-forensic analysis of indoor/outdoor classifiers based on acoustic clues," in *Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO '15)*, pp. 2072–2076, IEEE, Nice, France, August 2015.
- [25] V. Trevino and F. Falciani, "GALGO: an R package for multivariate variable selection using genetic algorithms," *Bioinformatics*, vol. 22, no. 9, pp. 1154–1156, 2006.
- [26] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [27] V. Y. Kulkarni and P. K. Sinha, "Pruning of Random Forest classifiers: a survey and future directions," in *Proceedings of the International Conference on Data Science and Engineering (ICDSE '12)*, pp. 64–68, Cochin, India, July 2012.
- [28] C. E. Galván-Tejada, J. P. García-Vázquez, E. García-Ceja, J. C. Carrasco-Jiménez, and R. F. Brena, "Evaluation of four classifiers as cost function for indoor location systems," *Procedia Computer Science*, vol. 32, pp. 453–460, 2014.
- [29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

- [30] M. H. Kabir, M. R. Hoque, K. Thapa, and S.-H. Yang, "Two-layer hidden markov model for human activity recognition in home environments," *International Journal of Distributed Sensor Networks*, vol. 12, Article ID 4560365, 2016.
- [31] G. Biau, "Analysis of a random forests model," *Journal of Machine Learning Research*, vol. 13, pp. 1063–1095, 2012.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

