# Performance Evaluation of the Electrical Appliances Identification System Using the PLAID Database in Independent Mode of House

Fateh Ghazali[1][a], Abdenour Hacine-Gharbi[1][b], Khaled Rouabah[2] and Philippe Ravier[3][c]
[1]*LMSE Laboratory, University of Bordj Bou Arreridj, Bordj Bou Arréridj, Algeria*
[2]*Electronics Department, University of Mohamed Boudiaf M'sila, Algeria*
[3]*PRISME Laboratory, University of Orleans, Orléans, France*

Keywords: Non-Intrusive Load Monitoring (NILM), Electrical Appliances Identification, Statistical Feature Extraction, Discrete Wavelets Analysis, Wavelet Cepstral Coefficient (WCC), K-Nearest Neighbors (KNN), Voting Rules Method, Independent Mode of House.

Abstract: In Electrical Appliances Identification (EAI) system, Plug Load Appliance Identification Dataset (PLAID) is largely used to develop and benchmark new methods proposed for demand management in electricity networks, more particularly, automated control, non-intrusive load planning and monitoring. Particularly, this database contains electrical signals of 11 appliance electrical appliances, recorded in several houses. In state-of-the-art, the EAI systems have used this latest PLAID designed, in two parts (one for training and the other for testing). These parts can be organized on house-dependent mode or house-independent mode. In the first mode, the signals of each appliance class and house in the testing part have examples in the training part. In opposition, in the second mode, the houses in testing part have not any example in training part. In this paper, we propose a comparative study between the performance of house-dependent EAI system and those of house independent mode system. In addition, in order to more validate the results of the comparison study, we propose the use of other classifiers like Gaussian Mixture Model (GMM), Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN). The obtained results, based on the use of PLAID, have demonstrated that the performances of this system, in independent mode, are relatively low compared to those obtained in dependent mode. This shows that the house's electrical installation has a good footprint in the input current signal.

## 1 INTRODUCTION

Electrical appliance identification (EAI) systems, integrated into smart meters, are an important function in ensuring proper management of household electrical energy consumption and distribution. An EAI system is considered as a pattern recognition system containing two phases: (1) the training phase (used to learn the different class models) and (2) the testing phase (used to evaluate system performances). These last phases are used to match the data received, via a pattern recognition system, with the information stored in a specific data set. The Plug Load Appliance Identification Dataset (PLAID) dataset (Gao, et al., 2014), a public, collaborative dataset intended for load identification research, is widely used in EAI systems in house-dependent mode (Nait-Meziane, et al., 2016)- (Nait Meziane, et al., 2017)- (Hacine-Gharbi, et al., 2018)- (Ghazali, et al., 2019) (Ghazali, et al., 2020) (Ghazali, et al., 2021). In this mode, the EAI systems are designed in such a way that all houses have examples of current signals in the training and test phases.

This present work aims to study and validate the EAI systems proposed in (Ghazali, et al., 2019) (Ghazali, et al., 2020) (Ghazali, et al., 2021), based on the strategy of the voting rule, and realized in house-dependent mode. Here, both validation and study of the aforementioned works are carried out in house-independent mode using other classifiers, namely

[a] https://orcid.org/0000-0003-2839-3259
[b] https://orcid.org/0000-0002-7045-4759
[c] https://orcid.org/0000-0002-0925-6905



Ghazali, F., Hacine-Gharbi, A., Rouabah, K. and Ravier, P.
Performance Evaluation of the Electrical Appliances Identification System Using the PLAID Database in Independent Mode of House.
DOI: 10.5220/0012468900003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 880-885
ISBN: 978-989-758-684-2; ISSN: 2184-4313

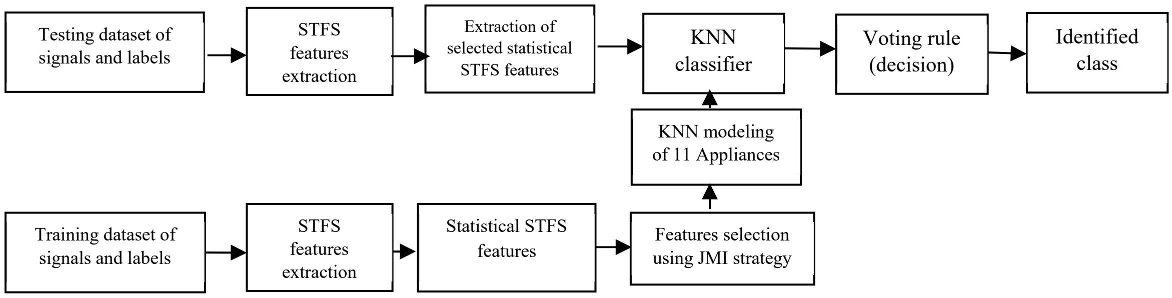


Figure 1: EAI system proposed in (Ghazali, et al., 2019).

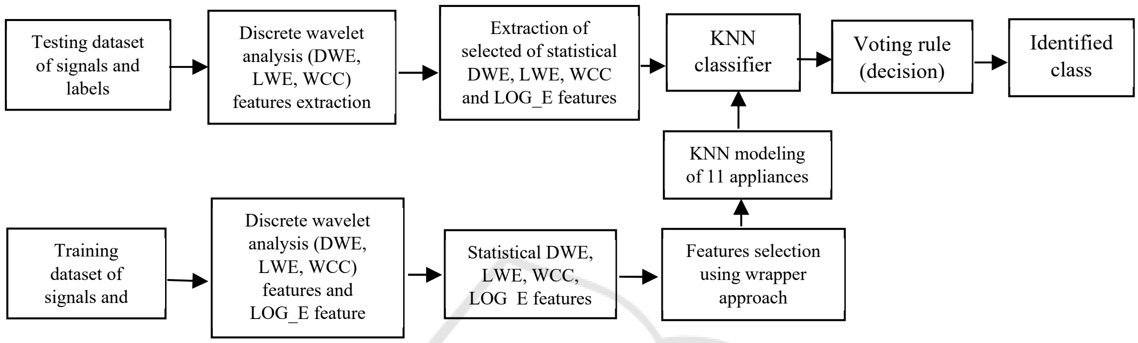


Figure 2: EAI system proposed in (Ghazali, et al., 2021).

Gaussian Mixture Model (GMM), Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN). In addition, performances of the latter EAI systems have been evaluated using the PLAID database.

## 2 RELATED WORK

In (Ghazali, et al., 2019), the authors proposed an electrical appliance identification system based on the KNN classifier combined with the voting rule method, using the statistical harmonics features from harmonic analysis. These statistical features are estimated from the sequences of the STFS coefficients vectors in order to extract a single vector representing the complete signal. Nevertheless, in practice, the measurements of the current signals are supplied continuously. This requires converting each signal into a sequence of statistical harmonic features, considering time segments of fixed duration. Figure 1 presents the system proposed in (Ghazali, et al., 2019).

The obtained results show that the combination of the mean and the standard deviation with 500 statistical features (250 for the mean and 250 for the standard deviation) give a CR classification rate of 92.63%. Also, applying the voting rule strategy improves the result to 94.97%.

In (Hacine-Gharbi, et al., 2018), the authors proposed an electrical appliance identification system based on the HMM classifier and the use of WCC coefficients as features for compact data representation to reduce dimensionality. This descriptor estimated from other descriptor called LWE (Log Wavelet decomposition-based Energy) by the application of the Discrete Cosine Transform (DCT). The last descriptor is also estimated from another descriptor called DWE (Discrete Wavelet Energy) by the application of the log10 of energy at each decomposition level from the wavelet analysis. The obtained results show that the WCC descriptor give the best performances results.

In (Ghazali, et al., 2021) the authors proposed an EAI system based on the KNN classifier combined with the voting rule strategy based on the feature extraction from wavelet analysis (DWE, LWE, WCC). As well as, the concatenation of this descriptors with the LOG_E descriptor can improve the perform of the Figure 2 presents the EAI system proposed by (Ghazali, et al., 2021). The results obtained show good performance in terms of the classification rate CR up to 98.51%.

It should be noted that this research used the PLAID database. The Database is divided into two subsets, one for learning and the other for testing, so that each house will have examples in both bases (dependent mode of the house). In this work, we are interested in the independent mode of house, ie each house will have examples in a single database (either for testing or training).

# 3 THE EAI IN THE INDEPENDENT MODE OF HOUSE

## 3.1 Presentation of the PLAID Dataset

Our system is based on the Plaid dataset (The Plug Load Appliance Identification Dataset) (Gao, et al., 2014). PLAID is a public dataset of electric signatures composed of 1074 instances recordings of currents and voltages of 11 types of electrical appliances from a variety of 55 households. These signals are sampled at a 30 kHz rate. Figure 3 summarizes the appliances found in the dataset with the different appliance types and the number of instances for each type.

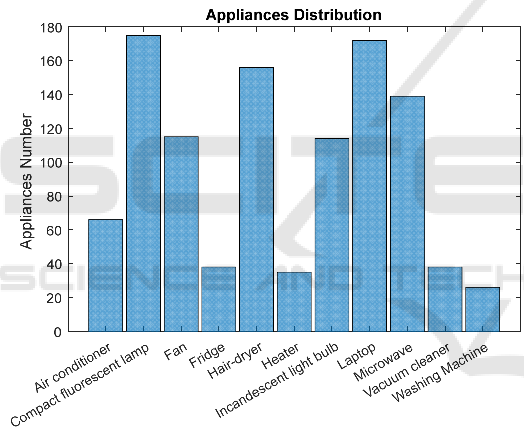


Figure 3: Distribution of appliances types and the number of theirs instances in PLAID dataset.

## 3.2 PLAID Dataset Subdivision for Training and Testing

It is clear that the design of an EAI system requires a database. This latter is divided into two datasets, one for training and the other for testing. The previous works used the PLAID database in dependent mode of house. In this work, we propose to study the case of a distribution of the database in independent mode of house. Initially the PLAID database is split into two parts almost balanced in instance number. Table 1 presents the distribution of the database in the two training and testing subsets for the different appliances.

This distribution will be applied on the same system proposed in (Ghazali, et al., 2021). the results will be compared with those obtained in the previous work. This situation is the most favorable for the real case of the electrical appliance identification.

Table 1: PLAID database distribution in independent mode of house for 50% for the training and 50% for the test.

| NB R | Appliance type | Total number of instances | Training subset | Testing subset |
|---|---|---|---|---|
| 1 | Compact fluorescent lamp | 175 | 69 | 106 |
| 2 | Vacuum cleaner | 38 | 22 | 16 |
| 3 | Hair-dryer | 156 | 59 | 97 |
| 4 | Microwave | 139 | 64 | 75 |
| 5 | Air conditioner | 66 | 51 | 15 |
| 6 | Laptop | 172 | 80 | 92 |
| 7 | Fridge | 38 | 5 | 33 |
| 8 | Incandescent light bulb | 114 | 59 | 55 |
| 9 | Fan | 115 | 85 | 30 |
| 10 | Washing Machine | 26 | 11 | 15 |
| 11 | Heater | 35 | 25 | 10 |
| Overall | | 1074 | 530 | 544 |

# 4 EXPERIENCES AND RESULTS

In this section, we present a number of experiments in which we take an EAI system applied in (Ghazali, et al., 2021). Different experiments are carried out to evaluate the performance of our EAI system. This system is based on WCC features extraction. The WCC coefficients are extracted from the current signals of the PLAID dataset (Gao, et al., 2014).This latter is divided into two groups in independent mode of house. For this study, we tested the following:

- A comparative study between different features types, like DWE, LWE and WCC in dependent and independent mode of house, taking the best configurations obtained in the previous works.
- In order to improve the performance of the system, we seeking the optimal configuration of the KNN classifier by choosing the parameters K (nearest neighbor vectors) and the optimal distance.
- We applied the wrappers selection method for studying the relevance of the concatenation of the LOG_E descriptor with different descriptors (DWE, LWE and WCC)
- In order to validate the previews works we applied other classifiers like GMM, ADL ...

### 4.1 Comparative Study Between Dependent and Independent Mode of House

In this experiment, we present the CR obtained in dependent and independent mode of house for a PLAID database distribution. taking the best configurations obtained in the previews work (Ghazali, et al., 2021), which are the mean and the standard deviation statistical features extracted from the window durations of 8 cycle length using DB5 mother wavelet with a decomposition level equal to 5, and 15 voting vectors for WCC, LWE and DWE statistical features.

Table 2: CR obtained in dependent and independent mode of house (NBF is the number of features).

| | Dependent mode of house | | | Independent mode of house | | |
|---|---|---|---|---|---|---|
| Descriptors NBF | DWE 12 | LWE 12 | WCC 12 | DWE 12 | LWE 12 | WCC 12 |
| CR (%) | 94.04 | 97.57 | 98.13 | 66.54 | 75.55 | 75.55 |

From this table, we show that the classification rate (CR) in independent mode of house is relatively low compared to that in dependent mode of house for all descriptors type. These results show that the electrical installation of the house has an influence on the performance of our EAI system, the objective then to look for another configuration of our system to improve the identification results in independent mode of the installation. In this context, the next experiment consists in finding the right configuration of the KNN classifier in terms of the optimal number of nearest neighbor vectors as well as the corresponding optimal distance.

### 4.2 Optimal Configuration of the KNN Classifier

The objective of this experiment is to seek the optimal configuration of the KNN classifier by choosing the number of nearest neighbor vectors K as well as the optimal distance in terms of the classification rate of the TCS signals. Table III.3 presents the TCS classification rates for the different values of k varying from 1 to 100 and for the different distances (Euclidean, Cosine, Correlation, Cityblock) using the LWE descriptor. Figure 4 shows the evolution of the classification rates of the different distances for the different values of the variable k.

From this figure we show that, the maximum classification rate of 78.49% is obtained by choosing 'Euclidean' distances with k equal to 80. The

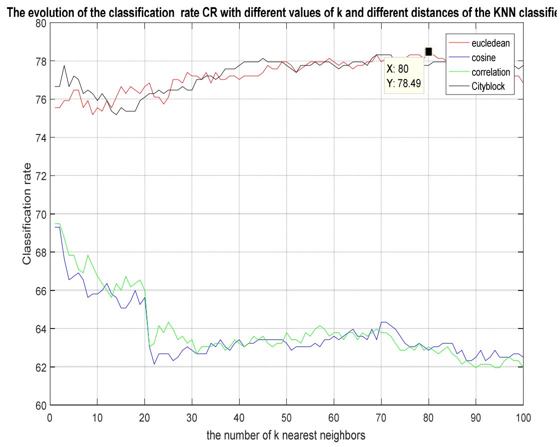


Figure 4: CR as a function of k for KNN classifier using 12 LWE features.

'Cityblock' distance gives an equally acceptable classification rate of 78.30% with the value of k equal 69. On the other hand, the 'Cosine', 'Correlation' distances, give classification rates of 69.30%, 69.48% respectively with the value of k equal 1. This configuration (Euclidean, k=80) is taken for the rest of the study.

### 4.3 Features Selection Results

The feature selection-based wrapper method is applied to select the most relevant features among the 12 statistical features extracted from the wavelet analysis. The three descriptors DWE, LWE and WCC are concatenated with the two statistical features of LOG energy. The initial 14 features subset is given as follows:

F_DWE = $\{M_{ap}, M_{d5}, M_{d4}, M_{d3}, M_{d2}, M_{d1}, Std_{ap}, Std_{d5}, Std_{d4}, Std_{d3}, Std_{d2}, Std_{d1}, M_E, Std_E\}$.

F_LWE = $\{LM_{ap}, LM_{d5}, LM_{d4}, LM_{d3}, LM_{d2}, LM_{d1}, LStd_{d5}, LStd_{d4}, LStd_{d3}, LStd_{d2}, LStd_{d1}, M_E, Std_E\}$.

F_WCC = $\{wcc_1, wcc_2, wcc_3, wcc_4, wcc_5, wcc_6, wcc_7, wcc_8, wcc_9, wcc_{10}, wcc_{11}, wcc_{12}, M_E, Std_E\}$.

Table 4 gives the CR and the selected features at each iteration j following the same selection procedure as in (Ghazali, et al., 2021).

From the table 3 we can observe the following points:

- The concatenation of the LOG energy descriptor adds an improvement in the classification rate for all the descriptors.

with the WCC descriptor the classification rate CR reaches the value obtained with all 14 features with only 6 features, and exceeds this value by a maximum of 79.77% with 8 features only.

Table 3: CR% as a function of the selected features, j is the iteration number, Sel is the selected features number and Feat is the feature name.

| J | DWE + LOG_E (14 features) | | | LWE + LOG_E (14 features) | | | WCC + LOG_E (14 features) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sel | Feat | CR% | Sel | Feat | CR% | Sel | Feat | CR% |
| 1 | 5 | $M_{d2}$ | 53.86 | 5 | $LM_{d2}$ | 51.65 | 1 | $wcc_1$ | 55.69 |
| 2 | 6 | $M_{d1}$ | 73.52 | 6 | $LM_{d1}$ | 70.22 | 2 | $wcc_2$ | 70.03 |
| 3 | 11 | $Std_{d2}$ | 74.44 | 3 | $LM_{d4}$ | 75.55 | 3 | $wcc_3$ | 75.73 |
| 4 | 12 | $Std_{d1}$ | 74.44 | 11 | $LStd_{d2}$ | 76.65 | 14 | $Std_E$ | 77.20 |
| 5 | 10 | $Std_{d3}$ | 74.44 | 2 | $LM_{d5}$ | 76.83 | 6 | $wcc_6$ | 77.94 |
| 6 | 9 | $Std_{d4}$ | 72.42 | 13 | $LM_E$ | 78.12 | 13 | $M_E$ | 78.49 |
| 7 | 14 | $Std_E$ | 72.79 | 4 | $LM_{d3}$ | 79.41 | 5 | $wcc_5$ | 79.77 |
| 8 | 8 | $Std_{d5}$ | 73.71 | 1 | $LM_{ap}$ | 79.41 | 9 | $wcc_9$ | 79.96 |
| 9 | 2 | $M_{d5}$ | 73.16 | 7 | $LStd_{ap}$ | 79.41 | 10 | $wcc_{10}$ | 79.96 |
| 10 | 4 | $M_{d3}$ | 72.79 | 8 | $LStd_{d5}$ | 79.41 | 11 | $wcc_{11}$ | 79.96 |
| 11 | 7 | $Std_{ap}$ | 72.42 | 9 | $LStd_{d4}$ | 79.41 | 7 | $wcc_7$ | 79.77 |
| 12 | 3 | $M_{d4}$ | 69.85 | 14 | $LStd_E$ | 79.59 | 8 | $wcc_8$ | 79.77 |
| 13 | 1 | $M_{ap}$ | 63.97 | 10 | $LStd_{d3}$ | 79.59 | 12 | $wcc_{12}$ | 79.77 |
| 14 | 13 | $M_E$ | 58.08 | 12 | $LStd_{d1}$ | 79.59 | 4 | $wcc_4$ | 79.41 |

- The LWE descriptor gave the CR of 79.41 % at only with 7 features. It reaches a maximum of 79.59% with 12 parameters.

with the DWE descriptor the CR classification rate reaches its maximum value of 74.44% with only 3 features, this value exceeds that obtained with all 14 features of 58.08%.

### 4.4 Validation of the System with Other Classifiers

Table 4: CR obtained in independent mode of house for the different classifiers and the different descriptors.

| Classifiers | DWE 12 features | LWE 12 features | WCC 12 features |
|---|---|---|---|
| KNN (Euclidean, k=80) | 63.97 | 78.49 | 78.49 |
| GMM (GN=5) | 42.09 | 78.49 | 65.25 |
| ANN (HLS = 100) | 49.26 | 76.47 | 78.86 |
| ADL (DFT= quadratic) | 27.20 | 66.36 | 67.09 |

In order to validate the electrical appliances identification system architecture, we applied several classifiers such as GMM, LDA, ANN. Table 4 presents the classification rate results for the different classifiers and for the different descriptors (with HLS: hidden Layer Size; GN: Gaussian Number; DFT= Discriminant Function Type).

From the results of Table 4, we can observe that the performance of the system may vary between 27.20% and 78.49%, depending on the classifier and the type of features. This shows the important role of both factors.

## 5 CONCLUSIONS

In this work, we have investigated a comparative study between the performances of EAI system in house-dependent and house-independent modes. The comparative study is carried out firstly using EAI system based on the KNN applied on statistical wavelet features vectors, and combined with the voting rule strategy. The performances of the EAI system are evaluated in two previous modes using the PLAID database. The comparative study is extended using other classifier such as GMM, LDA and ANN classifiers.

The obtained results in independent mode which represent the real case are relatively low with those obtained in dependent mode. These results demonstrate that the electrical installation in the house will have an imprint in the input current signal, and has an influence on the performance of our EAI system.

## REFERENCES

Gao J [et al.] Plaid: A public dataset of high resolution electrical appliance measurements for load identification research: demo abstract. [Conférence] // In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy Efcient Buildings. - New York, NY, USA : [s.n.], 2014. - pp. pp. 198–199..

Ghazali F. [et al.] Extraction and selection of statistical harmonics features for electrical appliances identification using kNN classifier, combined with voting rules method [Revue] // Turkish Journal of

Electrical Engineering & Computer Sciences. - 2019. - No. 4 : Vol. Vol. 27. - pp. pp. 2980-2997.

Ghazali Fateh, Hacine-Gharbi Abdenour et Ravier Philippe Selection of statistical wavelet features using a wrapper approach for electrical appliances identification based on a KNN classifier combined with voting rules method [Revue] // Int. J. Computational Systems Engineering. - 2021. - No. 5 : Vol. Vol. 6. - pp. pp. 220–230.

Ghazali Fateh, Hacine-Gharbi Abdenour et Ravier Philippe Statistical features extraction based on the discrete wavelet transform for electrical appliances identification [Conférence] // International conference of intelligent systems and pattern recognition .. - 16-18 October 2020, Hammamet , Tunisia : [s.n.], 2020.

Hacine-Gharbi Abdenour et Ravier Philippe Wavelet Cepstral Coefficients for Electrical Appliances Identification using Hidden Markov Models [Conférence] // 7th International Conference on Pattern Recognition Applications and Methods. - Funchal, Portugal : [s.n.], 2018.

Nait Meziane M. [et al.] Electrical Appliances Identification and Clustering using Novel Turn-on Transient Features [Conférence] // 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM). - Porto, Portugal: [s.n.], 2017. - pp. pp. 647-652.

Nait-Meziane M. [et al.] HMM-based transient and steady-state current signals modeling for electrical appliances identifcation [Conférence] // 5th International Conference on Pattern Recognition Applications and Methods. - Rome, Italy : [s.n.], 2016. - pp. 670-677.