# Discrete Wavelet Transform and Classifiers for Appliances Recognition

El Bouazzaoui Cherraqi[1(✉)], Nadia Oukrich[1], Soufiane El Moumni[2], and Abdelilah Maach[1]

[1] Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco
elbouazzaouicherraqi@research.emi.ac.ma
[2] Laboratory of Information Processing, Hassan II-University of Casablanca, Casablanca, Morocco

**Abstract.** Recognition of appliances' signatures is an important task in energy disaggregation applications. To save and manage energy, load signatures provided by appliances can be used to detect which appliance is used. In this study, we use a low frequency database to identify appliances based on discrete wavelet transform for features extraction and data dimensionality reduction. Further that, the accuracy of several classifiers is investigated. This paper aims to prove the effectiveness of DWT in load signatures recognition. Then, the best classifier for this studied task is selected.

**Keywords:** Classifiers · Discrete Wavelet Transform (DWT) Appliances recognition

## 1 Introduction

Nowadays, electrical appliances dominate the energy consumption in residential sector. The most measurements used today are blind [1], which means that the consumption of individual units is ignored as it is described in Fig. 1. This disadvantage cannot give any indication about detailed consumption that down to the used appliances. To understand our energy consumption, appliances load monitoring [2, 3] becomes an essential element to reduce and save energy. Since its inception, Appliances Load Monitoring is based on load signatures to identify appliances, this signature can be defined as the unique electrical behavior of appliances when they turned on such as audio signals generated by human.

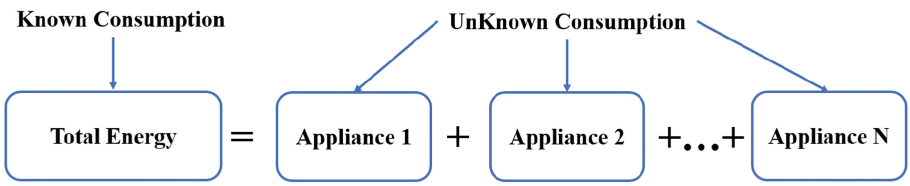


**Fig. 1.** Blind measurement of energy consumption


M. Ben Ahmed and A. A. Boudhir (Eds.): SCAMS 2017, LNNS 37, pp. 223–232, 2018.
https://doi.org/10.1007/978-3-319-74500-8_20

There are two approaches in Appliances load monitoring, the first one is Non-intrusive load monitoring [4] and the second one is intrusive load monitoring. For the first one, a unique sensor is used to collect information about energy consumption in household, this approach is proposed for the first time by Hart in 1992 [4], while, for the second approach, we use one sensor for each appliance or a group of appliances by dividing home into different zones of measurement. This method is accurate and simple. The only disadvantages are the high cost and the complexity of installation [2].

The challenge in appliance recognition is the large categories of appliances located at home, the collected databases are depended on the electrical line and transmission noise, and the preprocessing of these databases are widely different. In this paper, we use individual signatures to recognize appliances. In literatures, there are several public databases which are available for the scientific community, these datasets are divided into two groups which have low frequency (Freq $\leq$ 1 Hz): such as AMPds [5], ECO [6], GREEND [7], DRED [8], iAWE [9], REFIT [10], and Tracebase [11]. High frequency (Freq $\geq$ 50 Hz) such as WHITED [12] and COOLL [13]. According to the data acquisition frequency, the used load signatures in this work that belong to low frequency database named Tracebase, which is collected using a smart plug with a sampling frequency of 1 Hz. This appliances load signatures database contains individual signatures of 43 types of appliances from different residential devices and offices in Germany. Signatures are collected during 24 h in a separate file.

This work analyzes the accuracy of different classifiers to recognize appliances used in residential sector. The current implementation allowing to choose the most accurate classifier to solve the problem of appliances identification, two aspects are developed in the present paper: the features extracted using wavelet transform and the implementation of more than one classifier belonging to different families. In this study, the used classifiers are K Nearest Neighbors classifier (KNN), Support Vector Machine (SVM), Decision Tree classifier (DT), Random Forest classifier (RF), AdaBoost classifier (Adaboost), Gradient Boosting classifier (GB), GaussianNB (GNB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA).

The rest of this paper is organized as follows. Section 2 presents related works of appliances recognition. Section 3 describes, methodology of classification used in this paper. In Sect. 4 experimental results and classifiers performances are described. Finally, Sect. 5, we conclude this paper.

## 2 Related Work

In the residential sector there are several categories of appliances and this number increases accordingly to the growth of this sector [3]. According to electrical behavior of appliances, they are grouped into four groups as it described in [4] and developed in [2, 14]: appliances with two specific state ON/OFF such as toasters and lamps, appliances with a finite number of states, continuously variable appliances and permanent appliances, when the consumption maintained during a long time. The knowledge of these categories is useful to estimate the individual consumption of each appliance based on the change of event and the time between two successive states for the same device. Information down to the appliances level can reduce energy into 12% [15].

The common point between all proposed approaches in the appliance identification is the use of machine learning. Artificial neural network is used for appliances recognition based on low frequency sampling rate load signatures [24]. 10 features are used as inputs for ANN to perform the classification [15]. The heart of appliances recognition is the features extraction this stage is different from work to another according to the used method. In the paper of Reinhardt [11] 517 features are extracted from each device trace, only 15 relevant features are used in the classification and these final features are extracted using Weka toolkit to represent 33 categories of appliances. The classification is achieved using; random committee, Bayesian Network, J48, JRip, LogitBoost, Naive Bayes, Random Forest, and Random Tree the accuracy up to 95.5% obtained by Random Committee. In [16] the proposed appliances identification was built to recognize five classes of appliances, classification was performed using two classifier from two distinguished families, accuracy in this work up to 85%, This work was developed in [1] and the number of appliances categories was increased to 10 raw features that are analyzed using principal component analysis and the recognition of appliances achieved using SVM, 16 types of devices are investigated in this work given an accuracy up to 99.9%. Classification of ON/OFF appliances category was examined in [17], three classifiers namely Bayes Net, Random Forest and Hoeffding Tree are used to identify appliances through load signatures the proposed methods are implemented by using the WEKA software.

## 3 Methodology of Classification

Appliances recognition approach has been investigated through a several classifiers. In this work we use a real-world database with a large-scale modification. Table 1 shows details about the appliances used, we have chosen seven classes that are the most known in residential sector such as refrigerator, washing machine, laptop and so on. Our classification methodology is described in the Fig. 2.

**Table 1.** Number of instances prepared from Tracebase dataset for each appliance.

| Appliance | Number of load signatures instances |
|---|---|
| Refrigerator | 1026 |
| Washing machine | 56 |
| Laptop-PC | 482 |
| Desktop-PC | 1728 |
| Monitor TFT | 818 |
| Router | 455 |
| Multimedia | 111 |

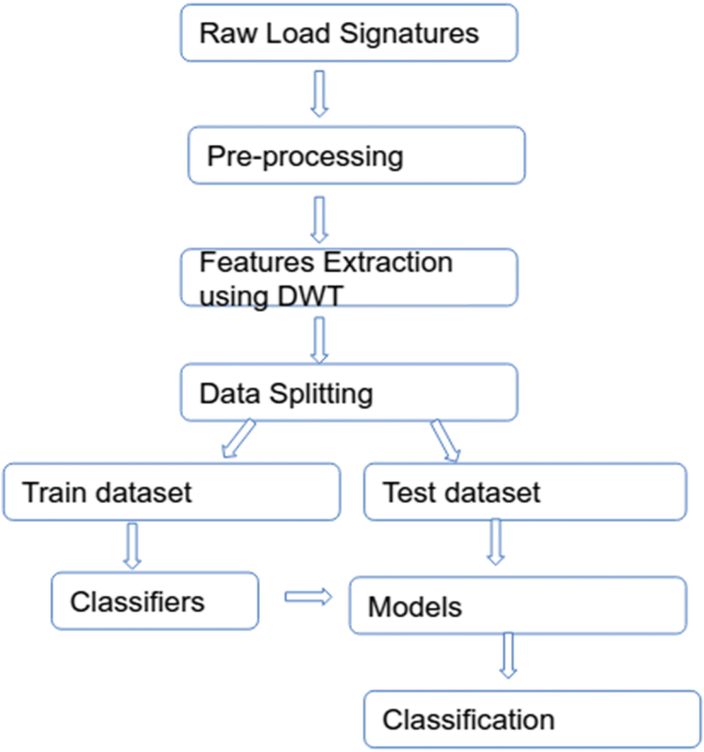


**Fig. 2.** Block diagram of the proposed study

### 3.1 Pre-processing

In the chosen database, we noticed that the collected data contains the raw time series of appliances which means the existence of duration when appliances are turned on and when they are inactive. In our application we are only interested on the tuned-on state, especially when the appliance generates its electrical behavior through a load signature. The tuned off periods of time are ignored here. Further to this, time series are subdivided into a several sub-sequences that describe each device.

### 3.2 Features Extraction Using DWT

Load signature is a sequence of values that varies during time, in this case values are real power. formally a load signature can be represented by Ls = {ls1, ls2, ls3,…,lsn} where Ls is the whole load signature of appliance and ls is the recorded value of real power and n is total number of values. While the number of observation, high dimensionality and multivariate property makes the classification difficult. To reduce the size of data and denoised it Discrete Wavelet Transform is performed to extract approximate coefficients that represents the whole signal. Decomposition of signal is generated as a tree known as Mallat's decomposition tree that shown in Fig. 3, where Ls (n) is the load signature and the high and low filters are represented respectively by h and g, the first and second wavelet details are d1 and d2. The second level approximation is represented by a2 which represents the input of classifiers.

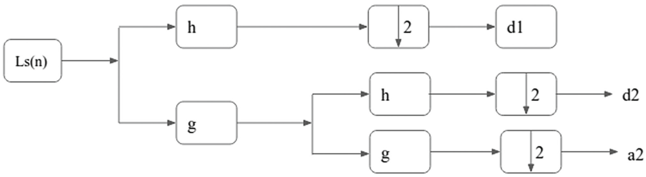


**Fig. 3.** Discrete wavelet transform with two levels

## 3.3 Splitting Database

The whole dataset that prepared for classification is splited randomly into two disjoint set, training and test. Different combinations are valuated to smooth results. The function used here to split database is implemented using the Sklearn library [9]. It allows a simple manipulation of database.

## 3.4 Classification

In our work, nine classifiers are implemented belonging to eight distinct families, namely, nearest-neighbors, support vector machines, decision trees, random forests, Boosting, Bayesian, discriminant analysis; Above, we briefly introduce the implemented classification algorithms.

**K-Nearest Neighbors (KNN)**
KNN is one of the most simple and fundamental classifier, based on the minimal distance between the training dataset and the testing dataset. This algorithm is widely used to solve the classification problems.

**Support vector machine (SVM)**
An SVM classifier is based on representation of outputs as points in space, every output category is separated by a clear gap. In test, new examples are conducted to one category based on which side of the gap they fall [18].

**Decision Trees Classifier (DT)**
DT is composed of nodes structured like a tree [19]. Nodes relate to direct edge from starting root node. Internal nodes are created with one incoming edge and producing two or more than two edges in this level. Values are compared to choose the right decision according to the feature. DT ending with terminal nodes.

**Random Forest Classifier (RF)**
Random forests use multiple of tree. Each tree depends on random vector values that tested independently with the same distribution for all trees in the forest. Each tree makes its own decision and the final decision is completed based on voting. Where the decision of a significant part of trees is the decision of overall outcome.

**Adaboost Classifier (ADA)**
An ADA classifier is a popular boosting technique [20]. This classifier helps to combine multiple classifiers that perform poorly into a single strong classifier. Beginning by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases.

**Gradient Boosting classifier (GB)**
GB is one of the most powerful machine-learning techniques for building predictive models. GB approach is based on construction of new base-learners to be excellently linked with the negative gradient of the loss function that connected with the whole ensemble. The choice of the loss function may be arbitrary, with both of rich variety of loss functions derived distant and the possibility of applying one's own task-specific loss [21].

**Gaussian NB (GNB)**
GNB subdivided inputs into continuous variables and output as discrete variable variables. GNB is characterized by the conditional probability between features that given the label [22].

Supervised classification with conditional Gaussian networks increasing the structure complexity from naive Bayes [23].

**Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)**
LDA is widely used for dimensionality reduction and classification. This technique maximizes the ratio between classes. This approach is known as class dependent transformation. Also, it can be used to create independence between different classes [23], QDA has the same property as LDA but the observations in QDA are separated by a quadratic hyperplane.

## 4 Experimental Results

### 4.1 Implementation

Classifiers are implemented in python based on Sklearn modules [9]. This library contains a wide range of machine learning algorithms. Our implementation consists on bringing all classifiers in the same script and run them in the same python successively. The machine used here is a personal notebook with 2.4 GHz in processor and 4 GB of RAM, all this work has been set up under GNU/Linux environment.

### 4.2 Results

This study is based on preprocessing of Tracebase database to enhance classification. As it is described in the previous sections the main step in classification is feature extraction in order to simplify learning for classifiers. In our work features are extracted and dimensionality reduced using the DWT. The performance of classifiers is evaluated based on the reached accuracy by each classifier and the duration of providing results.

DT with accuracy 100% of and takes 1.07 s to map all database and create the model that used for the test, GB reaches 100% of precision but it takes a long time exactly 115.70 s compared with KNN which reaches 98.93% in only 2.27 s. Random Forest Classifier is the third best classifier, RF classifier is the faster classifier with 0.86 s in this application, but the accuracy is less than KNN, Gradient Boosting and DT. QDA reach less than 20% that showing the lowest accuracy among the nine classifiers performed in this work. Table 2 gives more details.

**Table 2.** Classifiers accuracies and training times

| Classifier | Accuracy (%) | Time (S) |
|---|---|---|
| KNN | 98.93 | 2.27 |
| SVM | 64.93 | 370.98 |
| DT | 100 | 1.07 |
| RF | 95.33 | 0.61 |
| Adaboost | 61.20 | 23.16 |
| GBoosting | 100 | 117.99 |
| GaussianNB | 66.53 | 0.91 |
| LDA | 69.06 | 50.62 |
| QDA | 19.06 | 27.33 |

To quantify accuracy of each classifier, penalization of false classification is performed by log loss function. Classifier with the maximum accuracy is minimizes log loss function. Mathematically this metric can be defined as negative log-likelihood of the desired labels given an observed probabilistic classifier's predictions. Figure 4 shows the variation of log loss function for all used classifiers. According to the negative log-likelihood GB is more accurate the only weakness of this algorithm is the running time compared to DT. The poor classification detected by Log loss function is obtained by the algorithm QDA.

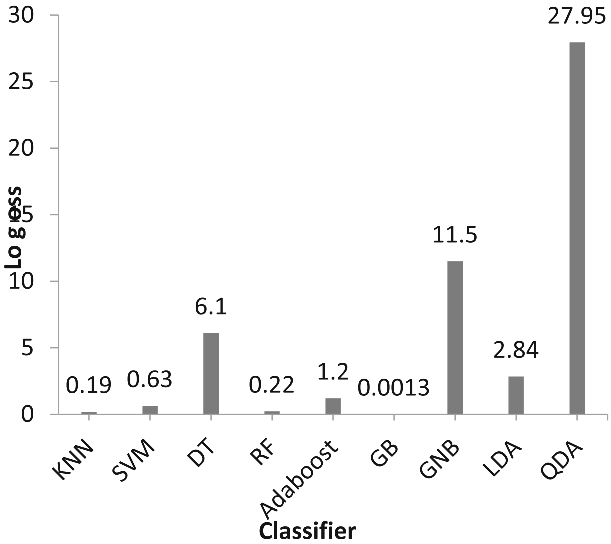


**Fig. 4.** Variations of Log loss function for the nine classifiers used

## 5 Conclusion

In this paper, load signatures recognition was presented using an automatic preprocessing based on Discrete Wavelet Transform. The aim of this work is to identify the most accurate classifier for appliances recognition using a low frequency sampling rate dataset. The methodology applied showed promising results proved by the well-known classifiers such as DT, KNN, LDA and so on. Training time and accuracy indicated that Decision Tree classifier was faster and more accurate for such prepared database. In the future we will analysis more load signatures dataset, and compare the powerful of classifiers in such complex problem. The improvement of appliances identification stage Non-intrusive load monitoring will help to save energy in the future.

## References

1. Ridi, A., Gisler, C., Hennebert, J.: Automatic identification of electrical appliances using smart plugs. In: 2013 8th International Workshop on Systems, Signal Processing and Their Applications (WoSSPA), pp. 301–305 (2013)
2. Zoha, A., Gluhak, A., Imran, M.A., Rajasegarar, S.: Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. Sensors **12**(12), 16838–16866 (2012)
3. Ridi, A., Gisler, C., Hennebert, J.: A survey on intrusive load monitoring for appliance recognition. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 3702–3707 (2014)
4. Hart, G.W.: Nonintrusive appliance load monitoring. Proc. IEEE **80**(12), 1870–1891 (1992)

5. Makonin, S., Popowich, F., Bartram, L., Gill, B., Bajic, I.V.: AMPds: a public dataset for load disaggregation and eco-feedback research. In: 2013 IEEE Electrical Power & Energy Conference (EPEC), pp. 1–6 (2013)
6. Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., Santini, S.: The ECO data set and the performance of non-intrusive load monitoring algorithms. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, pp. 80–89 (2014)
7. Monacchi, A., Egarter, D., Elmenreich, W., D'Alessandro, S., Tonello, A.M.: GREEND: an energy consumption dataset of households in Italy and Austria. In: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 511–516 (2014)
8. Uttama Nambi, A.S., Reyes Lua, A., Prasad, V.R.: Loced: location-aware energy disaggregation framework. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, pp. 45–54 (2015)
9. Batra, N., Gulati, M., Singh, A., Srivastava, M.B.: It's different: insights into home energy consumption in India. In: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, pp. 1–8 (2013)
10. Murray, D., Stankovic, L., Stankovic, V.: An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. Sci. Data **4**, 160122 (2017)
11. Reinhardt, A., et al.: On the accuracy of appliance identification based on distributed load metering data. In: Sustainable Internet and ICT for Sustainability (SustainIT 2012), pp. 1–9 (2012)
12. Kahl, M., Haq, A.U., Kriechbaumer, T., Jacobsen, H.-A.: Whited-a worldwide household and industry transient energy data set. In: 2016 Proceedings of the 3rd International Workshop on Non-Intrusive Load Monitoring (NILM) (2016)
13. Picon, T., et al.: COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification, arXiv Preprint arXiv:161105803 (2016)
14. Baranski, M., Voss, J.: Nonintrusive appliance load monitoring based on an optical sensor. In: 2003 IEEE Bologna Power Tech Conference Proceedings, vol. 4, p. 8 (2003)
15. Paradiso, F., Paganelli, F., Luchetta, A., Giuli, D., Castrogiovanni, P.: ANN-based appliance recognition from low-frequency energy monitoring data. In: Proceedings of the 14th IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1–6, June 2013
16. Zufferey, D., Gisler, C., Khaled, O.A., Hennebert, J.: Machine learning approaches for electric appliance classification. In: 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), pp. 740–745 (2012)
17. Salihagić, E., Kevric, J., Doğru, N.: Classification of ON-OFF states of appliance consumption signatures. In: 2016 XI International Symposium on Telecommunications (BIHTEL), pp. 1–6 (2016)
18. Vapnik, V.: The support vector method of function estimation. In: Nonlinear Modeling, pp. 55–85. Springer (1998)
19. Nguyen, N.-T., Lee, H.-H.: Decision tree with optimal feature selection for bearing fault detection. J. Power Electron. **8**(1), 101–107 (2008)
20. Freund, Y., Schapire, R.E.: A desicion-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory, pp. 23–37 (1995)
21. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front. Neurorobotics **7**, 21 (2013)
22. Perez, A., Larranaga, P., Inza, I.: Supervised classification with conditional gaussian networks: increasing the structure complexity from naive bayes. Int. J. Approx. Reason. **43**(1), 1–25 (2006)

23. Balakrishnama, S., Ganapathiraju, A.: Linear discriminant analysis-a brief tutorial. Inst. Signal Inf. Process. **18** (1998)
24. Cherraqi, E.B., Maach, A.: Load signatures identification based on real power fluctuations. In: Noreddine, G., Kacprzyk, J. (eds.) International Conference on Information Technology and Communication Systems, ITCS 2017. AISC, vol. 640 (2018)