Energy Consumption Level Prediction Based on Classification Approach with Machine Learning Technique

Hong-Chan Chang, Cheng-Chien Kuo, Yu-Tung Chen, Wei-Bin Wu, Eduardo Jr Piedad
Department of Electrical Engineering, National Taiwan University of Science and Technology
Taipei City, Taiwan
cckuo@mail.ntust.edu.tw

Abstract - Most of researches primarily use regression-type prediction, a method of estimating a numerical value given its historical data. In this study, a novel and practical prediction technique based on predicting identified energy consumption levels (e.g. low, average, and high levels) is proposed as an alternative on this conventional regression approach. Temperature and time features serve as predictors for short-term energy level prediction using commonly-used machine learning classifiers - artificial neural networks, support vector and random forest. The energy consumption numerical values were classified into ordinal bins created using a general percentile statistic. For verification, training subset underwent 10-times cross validation and test models are verified using the testing subset. These splitting and validation process are repeated 10 times with random permutation in each run prior to splitting into train and test subsets. The three-level energy prediction results show at least 90 % classification accuracy using any of the classifiers. However, the higher the number of desired energy levels, prediction accuracy tends to decrease. On the other hand, misclassification of energy level tends not to deviate from its true level that would still give confidence on the prediction performance when applied to any energy management system.

Keywords: Energy Consumption Prediction, Machine Learning, Energy Level Classification, Artificial Neural Network, Support Vector Machine, Random Forest.

1. Introduction

The increasing concern in energy conservation gives rise to various researches in energy consumption prediction especially in residential and commercial entities. In residential houses, data generated from sensors were used to forecast energy consumption [1]. In office buildings, an estimation of small power equipment's energy consumption and power demand can also performed [2]. Energy performance in buildings is also reviewed in [3]. Another energy prediction based on occupant behaviour was also conducted in [4]. Weather and climate variables were also considered for energy prediction [5]. A large hypermarket, a commercial center type consumer, also served as a subject for energy forecasting [6]. These researches aimed to predict either short or long term energy consumption with the ease of artificial intelligence.

In the advent of emerging algorithms, various prediction models were used in literature. Engineering estimates and statistical methods such as the classical linear regression are used in [1]–[3]. Stochastic models also reliably predict the building energy prediction and even identifies potential areas of energy waste [4], [5], [7]. Genetic programming approach can also be used [8]. Machine learning models such as the artificial neural networks, support vector machines and time series methods produce good prediction accuracy with other potential uses in [9]–[15]. Deep learning model can also play in prediction [16]. An electric load forecasting study in [17] also uses artificial neural network. These regression-type models were used to predict numerical energy value based on a given historical data.

Regression-type prediction however may not be the primary objective to address certain applications (e.g. an application of predicting which energy level a consumer consumes either low, medium or high). A conventional way to address this application is simply use regression-type models at first and later classify its predicted numerical output into desired levels or states. One may encounter however two particular issues when using this method. First, arbitrary selection of levels after the numerical prediction might incur additional accuracy error. Second, computational complexity of regression-type prediction may not be efficient to a simple energy level prediction application. This paper ought to develop a method using machine learning classifiers to provide an alternative way of predicting energy consumption levels.

Classification problems using machine learning were often used in image and speech recognition [18]. In forecasting applications, however, predicting directly the energy levels instead of undergoing regression-type prediction has not been

studied in the literature. This paper aimed to predict short-term energy levels using widely-used machine learning classifiers - artificial neural networks, random forest and support vector models.

This paper is organized as follows: Section 2 provides an overview of the energy dataset. The machine learning pipeline, implementation and classifiers are given in Section 3. The results and discussions, and conclusion are presented in Sections 4 and 5, respectively.

2. Energy Data
2.1. Data
A whole year dataset in [19] was used as test subject in this study. The dataset contains an hourly energy consumption and an hourly temperature (measured in Celsius degrees) retrieved from a smart metering device of a large hypermarket and meteorological sensors from specialized institute, respectively. The dataset originally comes from [6] and is already pre-processed by the authors to correct errors resulting from missing or abnormal values.

2.2. Energy Levels
The dataset [19] differs from its original dataset [6] with the addition of energy levels by creating bins based on percentile measures. For example, five bins representing five energy levels (very low, low, mid, high, very high) can be created using five even percentile ranges of the energy consumption data resulting to [0.174, 0.366), [0.366, 0.634), [0.634, 0.782), [0.782, 0.874), [0.874, 1.36] energy value ranges, respectively. For prediction implementation, energy levels were converted into its respective ordinal values (1, 2, 3, 4, 5). The dataset contains three energy level cases – three, five and seven classes. With these, three prediction cases were conducted using machine learning.

3. Machine Learning
3.1. Pipeline
A typical machine learning pipeline is shown in Fig. 1. There are two main parts in machine learning: the training and testing phases. In the training phase, a train dataset is used to create a model using some learning algorithm. Three popular machine learning classifier models—artificial neural network, support vector machines and random forest—are used in this study. A training validation usually is performed to ensure that the model performed is in a generalized manner and to prevent overfitting. Then, the trained model is verified in the testing phase whether it can sufficiently predict the test dataset using its input variables. Test dataset has similar features with train dataset and usually they are usually the result of splitting the original dataset. The prediction in the testing part is evaluated using some performance metrics.
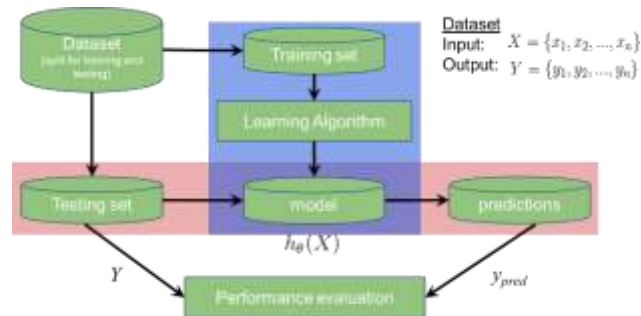

Fig. 1: A typical machine learning pipeline with training, testing and evaluation phases.

The performance is usually compared with the training performance to see whether there is overfitting, which happens when training verification is significantly higher than the testing evaluation.

3.2. Implementations
In this study, the proposed pseudocode of [20] is used as shown in Algorithm 1. The dataset was split into 70 % and 30 % for training and testing phase, respectively. In training, we verified the model by k-times cross-validation. This training

verification means that the training dataset was divided into k partitions and each partition served as test dataset to verify the model in k -times.

Algorithm 1. Pseudocode for Machine Learning Implementation
Initialization
Import libraries
Load and Split dataset into feature X and predictor y datasets
Data preprocessing
Repeat training and testing of the model n-times
for i=1: n
    Shuffle-splitting of X and y datasets into train and test datasets
    # k-times Training Cross-Validation
    for j=1:k_times
        Train the model using a ML algorithm
        Performance evaluation of the trained model
    # Testing the model
    Test the trained model
    Performance evaluation of the tested model
Display Performance Results
Calculate classification accuracy, F-score
Calculate the classification confusion matrix

We used k=10 times cross-validation and we took the average of its performance. After verification, the performance of the model was evaluated using the testing dataset. These validation and evaluation help determine whether there was overfitting of the model. There were three evaluation metrics used in F-score, classification accuracy, and confusion matrix. F-score accuracy weighs the impact of positive predictive value (also known as precision) and sensitivity (also known as recall) as measured in (1).

$$\text{F score} = \frac{2 * Precision * Recall}{Precision * Recall} \tag{1}$$

Another measure of accuracy, classification accuracy or the percent of correctly classified levels over the total number of levels was taken using (2).

$$\text{Accuracy} = \frac{No. of\ correctly\ classified\ energy\ levels}{Total\ number\ of\ classified\ energy\ levels} \tag{2}$$

In addition, a normalized confusion matrix can further visualize the result of the classification performance. Finally, the whole process was repeated 10 times, with the average of the accuracy taken. Python libraries from Scikit-learn platform [21] were used to implement the machine learning classifiers – artificial neural network, support vector, and random forest.

3.3. Machine Learning Classifiers
3.3.1. Artificial Neural Network
Multilayer perceptron (MLP) is a feedforward artificial neural network that uses backpropagation for training. It has at least three layers of nodes in which, except the first layer, each node uses a nonlinear activation function. MLP minimizes an error function by learning proper weights for each pattern in training set. Using MLP as a classifier requires all output nodes to be set to zero except one that predicts the class of given input. In this study, the inputs are features such as temperature and timestamps such as time and the day of the week. An MLP classifier algorithm [22] is applied to image classification. The universal approximation theorem [23] guarantees that a single hidden layer is enough to represent any

continuous function while two hidden layers for non-continuous. Typically, the number of hidden nodes is between 5 and 100 in proportion to the increasing number of inputs and number of training cases [24]. In this study, a parameter tuning using a simple brute-force search method was performed to choose the optimum number of hidden layers and neurons. A set of one hidden layer with twenty performance evaluations, ranging from 10 to 100 neurons, was performed. Another similar set of two layers with similar range of neurons per layer was conducted.

3.3.2. Support Vector Classifier

The support vector machine (SVM) is another type of learning algorithm that made statistical learning theory a tool for creating practical algorithms, especially for estimating multidimensional functions [25]. A practical guide for classification application of SVM is studied in [26]. The radial basis function (RBF) and polynomial kernel of degree are the two commonly used SVM kernel functions. In SVM training process, the input parameters, which are the kernel parameters, sigma σ and cost C, are tuned to achieve the desired trade-off between generalization performance and training. This paper uses an RBF kernel which is a first reasonable choice for SVM [26]. We tuned the kernel parameters, σ and C, with similar brute force search method to determine which set of parameters can result to a better SVM classifier. We select the best parameters from the usual 0.1, 0.5, 1 and 5 sigma values and the range from 1 to 10 for the cost.

3.3.3. Random Forest Classifier

The ensemble learning is a method of generating a number of classifiers, and then aggregate their results. There are two well-known methods, boosting [27] and bagging [28]. In boosting, successive trees impose an extra weight to incorrect prediction of previous trees, while in bagging, successive trees are independent of the previous trees. In this study, bagging method, also known as random forest (RF), was used. This method introduced randomness to bagging. Each node in standard trees was split, using the best split among a subset of randomly chosen predictors at that node [29]. For classification using random forest, there are two tuning parameters: the number of trees n_{tree} and the depth of tree. Similar to previous machine learning models, these two parameters were tuned in 1, 4, 8 and 12 depths and the range of 10 to 100 for the number of trees, respectively, using the previous brute-force search method.
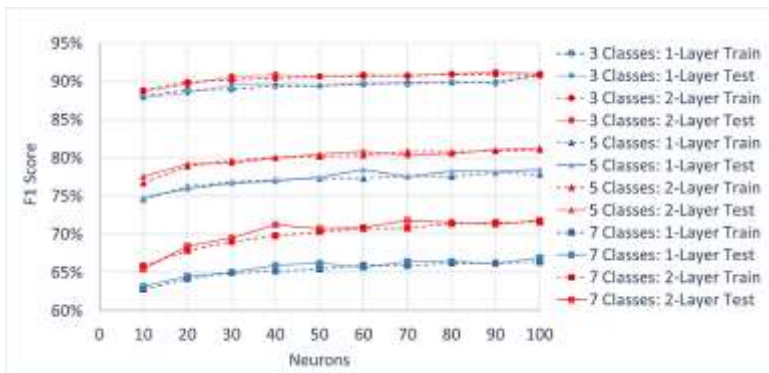
4. Results and Findings

Fig. 2: Parameter tuning of the number of hidden layers and the number of neurons per layer for MLP classifier with 3, 5 and 7 classes.
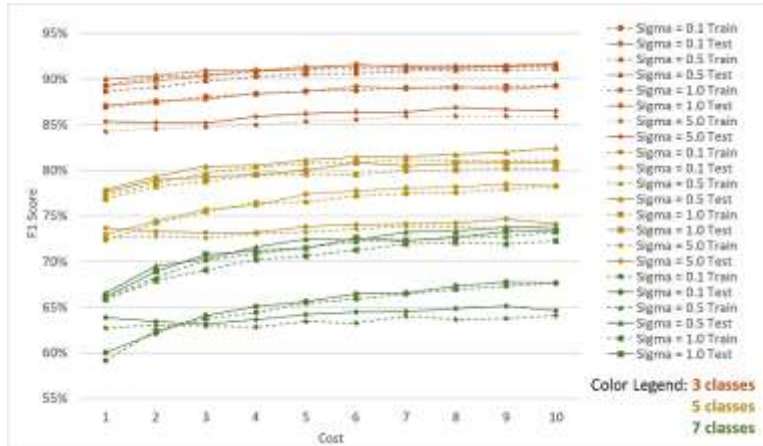
Fig. 3: Parameter tuning of sigma σ and cost C for support vector classifier with 3, 5 and 7 classes.
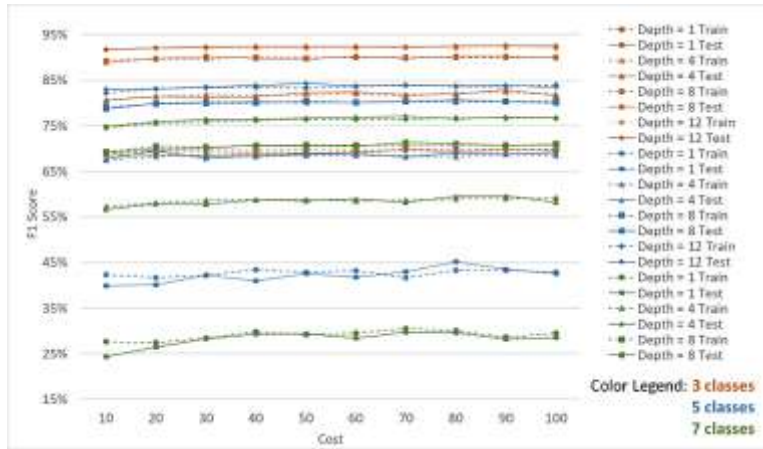

Fig. 4: Parameter tuning of the number of trees n_{tree} and the maximum depth of tree for random forest classifier with 3, 5 and 7 classes.

Fig. 2 shows that choosing 30 neurons for single and two-layer MLP classifiers can already predict sufficiently without overfitting. Figs. 3 and 4 show that selecting the parameters σ = 0.5 and cost = 5, and depth = 12 and trees = 30 for SVM and RF classifiers, respectively, can also generalize with greater F score and without overfitting. In all the classifiers, the lower the number of classes the better the F-score. The three classes represent low, mid, high energy levels, respectively.

Table 1: Performance Comparisons of Three Tuned Classifiers.

No. of Class	Classification Accuracy (%)			
	ANN		SVC	RF
	1-hidden layer neurons = 30	*2-hidden layer neurons = 30*	*σ = 0.5 cost = 5*	*depth = 12 trees = 30*
3	89.89	91.19	91.38	92.28
5	78.07	81.04	81.07	83.40
7	66.51	71.43	72.57	76.39

Table 1 summarizes the performance of the classifiers using their respective tuned parameter values. With only three classes, we can observe classification accuracy greater than 90 % except for a single-layer ANN. RF classifier achieves the highest classification accuracy 92.28 % among the three classifiers.
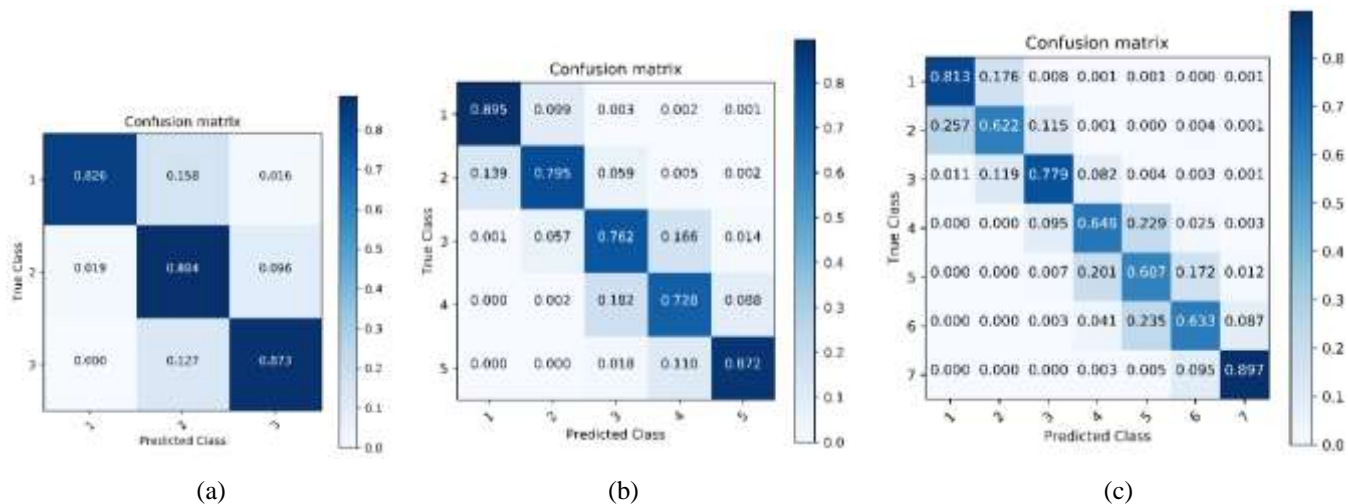
(a)   (b)   (c)

Fig. 5: Normalized confusion matrix of (a) three, (b) five, and (c) seven energy levels using multilayer perceptron classifier.
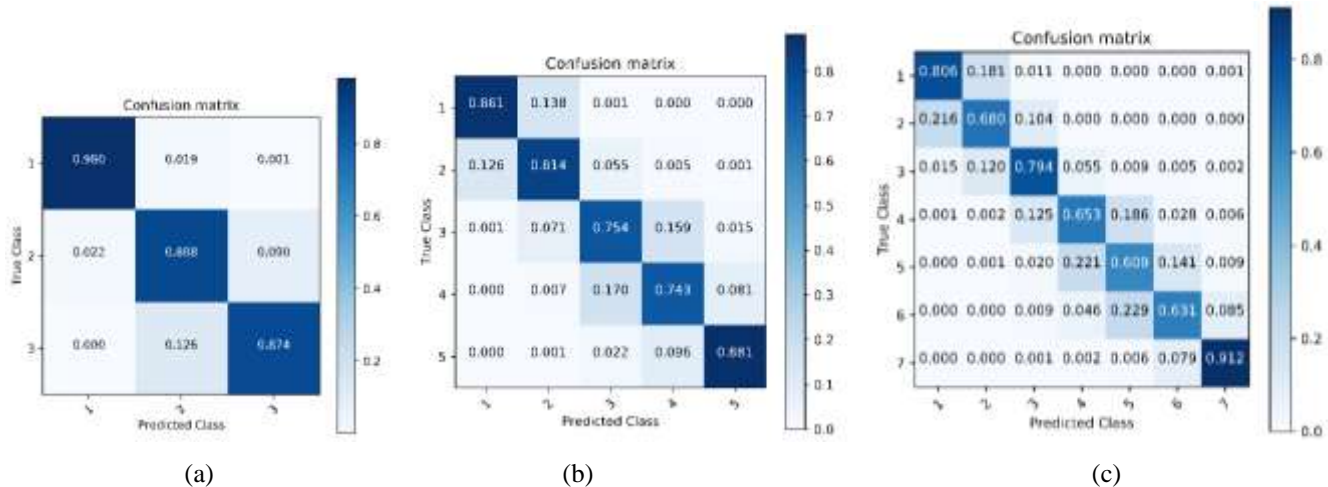

(a)   (b)   (c)

Fig. 6: Normalized confusion matrix of (a) three, (b) five, and (c) seven energy levels using support vector classifier.
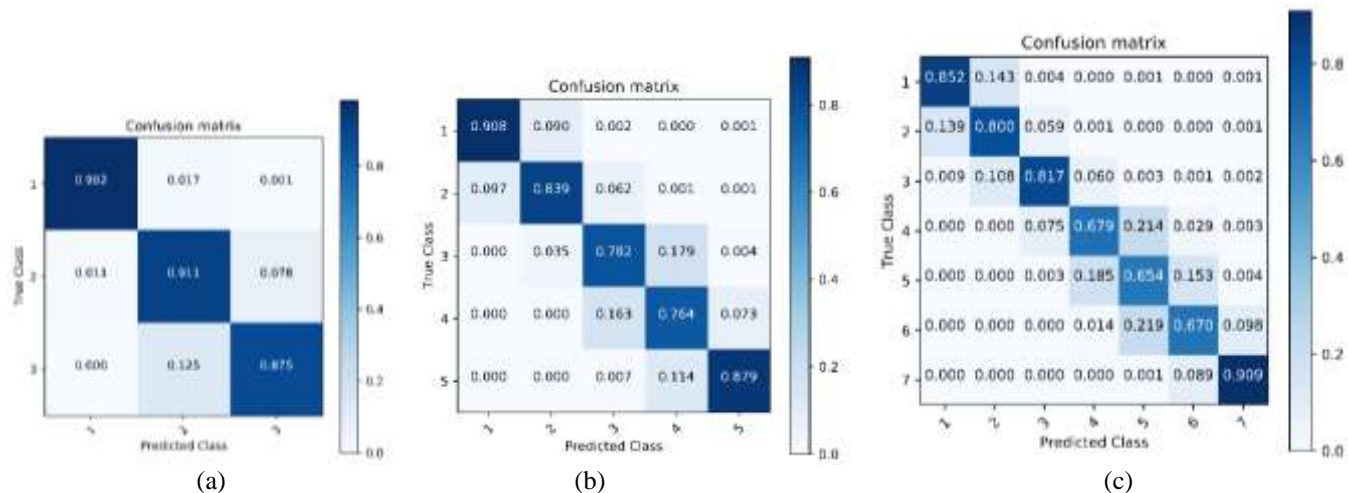

(a)   (b)   (c)

Fig. 7: Normalized confusion matrix of (a) three, (b) five, and (c) seven energy levels using random forest classifier

Fig. 5 present the normalized confusion matrix of the classification accuracy performance of three, five and seven energy levels, respectively, using MLP classifier. It can be observed that as the number of energy levels increases, the prediction decreases. For example, Fig. 5 (a) has at least 82.6 % accuracy while Fig. 5 (b) has at least 72.7 % and Fig. 5 (c) has 60.7 %. Similar observation can be found in the normalized confusion matrices of support vector and random forest classifiers as shown in Figs. 6 (a)-(c) and Figs. 7 (a)-(c), respectively. The confusion matrices of higher energy levels (e.g. five and seven classes in Figs. 5-7.) show that the predicted value does not deviate from its true value. As example, class 3 is predicted correctly in Fig. 7 (b) as class 3 with 78.2 % accuracy while it is misclassified as class 2 and class 4 with 3.5 % and 17.9 % accuracy, respectively.

5. Conclusion

A prediction of energy levels was performed using machine learning classifiers. All three machine learning models – artificial neural network, support vector and random forest classifiers, performed satisfactorily reaching at least 90 % classification accuracy for three energy level prediction. On the other hand, it was observed that as the number of energy levels increases, the classification accuracy decreases. However, the misclassification of energy levels tends not to deviate from its true level even with low classification accuracy predicted in higher energy levels such as five and seven energy levels.

References

[1]   L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy Build.*, vol. 140, pp. 81–97, 2017.

[2]   A. C. Menezes, A. Cripps, R. A. Buswell, J. Wright, and D. Bouchlaghem, "Estimating the energy consumption and power demand of small power equipment in office buildings," *Energy Build.*, vol. 75, pp. 199–209, 2014.

[3]   H. X. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012.

[4]   J. Virote and R. Neves-Silva, "Stochastic models for building energy prediction based on occupant behavior assessment," *Energy Build.*, vol. 53, pp. 183–193, 2012.

[5]   F. Oldewurtel *et al.*, "Energy efficient building climate control using Stochastic Model Predictive Control and weather predictions," *Am. Control Conf.*, pp. 5100–5105, 2010.

[6]   A. Pîrjan, S. V. Oprea, G. Carutasu, D. M. Petrosanu, A. Bâra, and C. Coculescu, "Devising hourly forecasting solutions regarding electricity consumption in the case of commercial center type consumers," *Energies*, vol. 10, no. 11, 2017.

[7]   N. Arghira, L. Hawarah, S. Ploix, and M. Jacomino, "Prediction of appliances energy use in smart homes," *Energy*, vol. 48, no. 1, pp. 128–134, 2012.

[8]   M. Castelli, L. Trujillo, and L. Vanneschi, "Prediction of energy performance of residential buildings: A genetic programming approach," vol. 102, pp. 67–74, 2015.

[9]   K. Li, H. Su, and J. Chu, "Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study," *Energy Build.*, vol. 43, no. 10, pp. 2893–2899, 2011.

[10]   S. A. Kalogirou and M. Bojic, "Artificial neural networks for the prediction of the energy consumption of a passive solar building," *Energy*, vol. 25, no. 5, pp. 479–491, 2000.

[11]   B. Dong, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy Build.*, vol. 37, no. 5, pp. 545–553, 2005.

[12]   P. A. González and J. M. Zamarreño, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy Build.*, vol. 37, no. 6, pp. 595–601, 2005.

[13]   J. Yang, H. Rivard, and R. Zmeureanu, "On-line building energy prediction using adaptive artificial neural networks," *Energy Build.*, vol. 37, no. 12, pp. 1250–1259, 2005.

[14]   G. Regina and P. Capriles, "Prediction of energy load of buildings using machine learning methods Database and Machine Learnig Methods," 2016.

[15]   A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy Build.*, vol. 49, pp. 560–567, 2012.

[16] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Sustainable Energy , Grids and Networks Deep learning for estimating building energy consumption ☆," *Sustain. Energy, Grids Networks*, vol. 6, pp. 91–99, 2016.

[17] D. C. Park, M. a. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, 1991.

[18] B. Javidi, *Image recognition and classification: algorithms, systems, and applications.* Javidi, Bahram, 2002.

[19] E. J. Piedad and C.-C. Kuo, "A 12-month Data of Hourly Energy Consumption Levels from a Commercial-type Consumer," *Mendeley Data*, 2018. [Online]. Available: https://data.mendeley.com/datasets/n85kwcgt7t/1/files/6cfc7434-315c-4a2d-8d8c-ce6a2bb80a01/energy_consumption_levels.csv?dl=1 [25 June 2018]

[20] E. J. Piedad, J. I. Larada, G. J. Pojas, and L. V. V. Ferrer, "Postharvest classification of banana (Musa acuminata) using tier-based machine learning," *Postharvest Biol. Technol.*, vol. 145, pp. 93–100, Nov. 2018.

[21] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[22] A. Khotanzad and J. H. Lu, "Classification of Invariant Image Representations Using a Neural Network," *IEEE Trans. Acoust.*, vol. 38, no. 6, pp. 1028–1038, 1990.

[23] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control. Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[24] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, USA, Springer, 2009.

[25] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[26] C.-J. L. Chih-Wei Hsu, Chih-Chung Chang, "A Practical Guide to Support Vector Classification," *BJU Int.*, vol. 101, no. 1, pp. 1396–400, 2008.

[27] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Stat.*, vol. 26, no. 5, pp. 1651–1686, 1998.

[28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[29] A. Liaw, A. Liaw, and M. Wiener, "Classification and Regression by RandomForest Classification and Regression by randomForest," no. November 2001, 2014.