Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2020, 13(2), 600-608
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2020, 13(2), 600-608
DOI: 10.18185/erzifbed.691398
Research Article

Comparative Analysis of Regression Learning Methods for Estimation of Energy Performance of Residential Structures

Abdurrahim AKGUNDOGDU [1*]

[1] Electrical-Electronics Eng. Dep. Istanbul University-Cerrahpaşa, Istanbul, Turkey

Geliş / Received: 19/02/2020, Kabul / Accepted: 30/05/2020

Abstract

Energy efficiency is a top priority for private and commercial buildings. This study evaluates the performance of six regression learning methods, including Linear Regressor, MLP Regressor, RBF Regressor, SVM Regressor, Gaussian Processes, and ANFIS Regressor to predict the heating and cooling loads of residential buildings. 768 buildings were considered and analyzed based on the influential parameters, such as relative density, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution for predicting heating load and cooling load. Three statistical criteria such as correlation coefficient (R), mean absolute error (MAE) and root mean square error (RMSE) were used to assess the potential of the regression methods used in this study. The best estimation results were obtained with the ANFIS regression model, with R of 0.998, MAE of 0.46 and RMSE of 0.68 for HL; and with R of 0.990, MAE of 1.26 and RMSE of 1.60 for CL.

Keywords: Energy efficiency, heating and cooling loads, regression learning, ANFIS

Konut Yapılarında Enerji Performansının Tahmininde Regresyon Öğrenme Yöntemlerinin Karşılaştırmalı Analizi

Öz

Özel ve ticari binalar için enerji verimliliği birinci önceliktir. Bu çalışma, konut binalarının ısıtma ve soğutma yüklerini tahmin etmek için Lineer Regresör, MLP Regresörü, RBF Regresörü, SVM Regresörü, Gauss İşlemleri ve ANFIS Regresörü dahil olmak üzere altı regresyon öğrenme yönteminin performansını değerlendirmektedir. 768 bina, ısıtma yükü ve soğutma yükünü tahmin etmek için nispi yoğunluk, yüzey alanı, duvar alanı, çatı alanı, toplam yükseklik, yönlendirme, cam alanı ve cam alanı dağılımı gibi etkili parametrelere dayanarak düşünülmüş ve analiz edilmiştir. Bu çalışmada kullanılan regresyon yöntemlerinin potansiyelini değerlendirmek için korelasyon katsayısı (R), ortalama mutlak hata (MAE) ve kök ortalama kare hatası (RMSE) gibi üç istatistiksel kriter kullanılmıştır. En iyi tahmin sonuçları ANFIS regresyon modeli ile; HL için 0.998, MAE 0.46 ve RMSE için 0.68; ve CL için 0.990 R, MAE 1.26 ve RMSE 1.60'tır

Anahtar Kelimeler: Enerji verimliliği, ısıtma ve soğutma yükleri, regresyon öğrenme, ANFIS

*Corresponding Author: akgundog@istanbul.edu.tr

1. Introduction

The energy use of residential buildings and structures has increased over the past few decades (Sholahudin et al., 2014). This trend has increased in recent years because buildings have significant energy buyers around the world and an increase in living standards. The buildings in European countries are legitimately linked to at least prequalification conditions after the European Directive (Directive, 2002). The European Union has issued a directive that requires compliance with the pre-condition to reduce environmental impacts. It is extremely important to know the heating and cooling loads when buildings are built to reduce the total energy use of buildings (Ekici, 2016).

There is a need for some parameters to analyze architects and designers in a plan [1]. Some of these parameters are Relative Compactness (RC), Surface Area (SA), Wall Area (WA), Roof Area (RA), Overall Height (OH), Orientation (OR), Glazing Area (GA), and Glazing Area Distribution (GAD) of the building (Tsanas and Xifara, 2012). Designers should also decide which temperature and cooling load the building is having on the big impacts. It has been shown in the literature that researchers use different methods and parameters to predict the heating load (HL) and the cooling load (CL) (Pérez et al., 2008).

Recently there has been increasing interest in developing an approach to estimating the energy performance of residential buildings (Tsanas and Xifara, 2012; Jeon et al. 2016). Many techniques have been proposed for energy performance in buildings. Some of these techniques are based on traditional regression methods (Yu et al., 2010), a statistical linear regression model (Fan et al., 2016), a least square support vector machine (LS-SVM) (Ekici, 2016) that focus on the effects of building devices on demographics, household behavior and household electricity demand (IRBFN), which designs the Linear Regression (LR) as a model around the local RBFN (Lee and Kwak, 2016). Duarte et al. (2017) used Decision Trees (DT), Multi-Layer Perceptron Neural Network (MLP), Random Forests (RF) and Support Vector Machines (SVM) for predicting energy loads in buildings.

This work points to determine the performance of the six regression models for the HL and CL estimation output variables of buildings with eight input parameters such as overall height, relative compactness, surface area, wall area, roof area, orientation, glazing area, and glazing area distribution of residential buildings.

This study is organized as follows: Section 2 shows an illustration of the set of data used to train and test regression learning methods and shows how regression learning models work. Section 3 presents the results and discussion obtained and the comparison between the proposed model and other methods. Finally, Chapter 4 summarizes the conclusion of this research.

2. Material and Method

2.1. Data set

The data set used in this work can be accessed in (Tsanas and Xifara, 2012). The data were obtained by simulation of several buildings using a computer program called Ecotect. This program is a natural review facility compatible with the building data modeling program and is used for a comprehensive preliminary building energy application review (Yang et al., 2014). The dataset contains 768 examples and eight input features aimed at predicting two real-valued responses (HL and CL). This data set performs energy analysis using 12 distinctive structures. These 12 buildings in the Greek city of Athens are composed of 18 blocks of 3.5 × 3.5 × 3.5m, each with a volume equal to 771.75 cubic meters for each simulated building. The data set consists of eight input factors and two outputs, as shown in Table 1.

Table 1. Statistical information on the input and output variables

Description	Variable	Min	Max	Mean
Relative Compactness (RC)	Input 1	0.62	0.98	0.76
Surface Area (SA)	Input 2	514.5	808.5	671.71
Wall Area (WA)	Input 3	245	416.5	318.50
Roof Area (RA)	Input 4	110.25	220.5	176.60
Overall Height (OH)	Input 5	3.5	7	5.25
Orientation (OR)	Input 6	2	5	3.50
Glazing Area (GA)	Input 7	0	0.4	0.23
Glazing Area Distribution (GAD)	Input 8	0	5	2.81
Heating Load (HL)	Output 1	6.01	43.1	22.31
Cooling Load (Cl)	Output 2	10.9	48.03	24.59

2.2. Regression Learning Methods

Regression is a machine learning algorithm that can be developed to predict output; like power, energy, current, price and so on. A continuous output variable is a real value, such as an integer or floating-point value. Regression is a method of predicting a series of incoming responses. These strategies require an educational phase, called a supervised education phase, that takes into account a set of data from selected factors in the problem domain. In this study, some successful regression models were trained and tested. WEKA 3.8 (Data Mining Program) and Matlab (MATLAB ^{TM}) were used for these predictions.

2.3. Selecting the best Regression Learning Algorithms

In this section, six regression learning algorithms are trained and tested:

1. Linear Regression (LINREG) (Wilkinson and Rogers, 1973): Linear regression is an easy and widely used method of estimating. Finding the relationship between two continuous variables is useful. If a variable can be fully expressed by the other, it is averred that the relationship between the two factors is deterministic. A variable is considered as an informative variable and the other is considered as a dependent variable.

2. Multilayer Perceptron (MLP) (Haykin, 1999): Multilayer Perceptron is a feedforward network that maps sets of input data onto an appropriate output pattern. A supervised learning methodology, called backpropagation, is used to train a net that links a large number of simple perceptron models, especially those that can recognize non-specific data.

3. RBF Network (RBFN) (Haykin, 1999): RBFN is a specific member of the feed-forward neural networks and has supervised and supervised stages. There are input points and hidden nodes to characterize the activation of each node. An RBFN performs classification by measuring the similarity of the input to the samples in the training set.

4. SMOreg (SMO) (Smola and Schölkopf, 1998): SMO is an algorithm for training Support Vector Machine (SVM). SMOreg is the execution of a consecutive minimum optimization calculation to train an SVM regression method. SMO divides large quadratic programming (QP) problems into a set of smallest QP problems, which are then solved analytically.

5. Gaussian Processes (GP) (Williams, 1998): The Gaussian Process is a type of supervised learning that is a generalization of the Gaussian likelihood distribution. This procedure is represented by an average and a covariance function, and the output function in any data modeling problem can be considered as a single example from this Gaussian distribution.

6. ANFIS Regression (Jang, 1993): ANFIS is derived from the Adaptive Neuro-Fuzzy Inference System. ANFIS is an adaptive network class that relates both neural networks and fuzzy logic inference systems. When fuzzy logic and neural networks in fuzzy clusters are combined with cluster values, membership functions are evaluated in training and neural networks are used to estimate weights.

2.4. Performance Evaluation Indices

In this study, three evaluation criteria are employed to evaluate the performance of each of the regression models. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Correlation Coefficient (R) values are calculated. These performance values can be formulated as follows.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (2)$$

$$R = \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum(y_i-\bar{y})^2\sum(x_i-\bar{x})^2}} \quad (3)$$

where y_i and x_i are the desired output and estimated output respectively; $\bar{y}$ and $\bar{x}$ represent averages values, and n represents each sample in the data set.

3. Results and Discussion

At this stage, 70% of the whole data set were randomly selected to develop the HL and CL of building forecasting models; the remaining 30% of the data set was used to test and re-evaluate the accuracy, as well as the carrying out of the developed regression models.

After training the system, a test operation was carried out to verify the success of the models. In this work, the available data sets were split into two subsets randomly, namely, a training set and a test set. The total number of samples used was 768, out of which the first 538 were for training and the remaining 230 for testing. The best accuracy was obtained by using ANFIS-Fuzzy C-Means (FCM) clustering method. FCM clustering as recommended by Bezdek is a data clustering procedure in which every data point has a place with at least two clusters (Bezdek, 1973). FCM is an iterative calculation, which needs to find cluster centers based on the minimization of a goal function. The target function is the whole of squares distance between every data point and the cluster centers and is weighted by its membership (Mehrabi and Sharifpur, 2012).

Hybrid learning calculation, a combination of least squares and back-propagation, has been connected to recognize the membership function parameters of ANFIS. After simulations, the ANFIS parameters can be pictured in Table 2.

Table 2. The optimal values of the ANFIS algorithm in this study.

ANFIS FCM Parameters	Value
Input MF type	Gaussian
Number of Input MF	16
Output Function Type	Linear
Total Number of Input MFs	8*16
Number of Clusters	16
Number of Rules	16
Number of Epoch	200

In this work, the genfis3 function of Matlab's Fuzzy Toolbox was used to generate a Fuzzy Inference System (FIS) using the fuzzy c-means (FCM) clustering model of data behavior. The rule extraction strategy first uses the FCM function to decide the number of rules and membership functions for the forerunners and consequents. The number of clusters decides the number of membership functions and rules in the generated fuzzy system.

For this work, the number of clusters was chosen automatically by the command. The input membership function was chosen to be 'gaussmf', and the output membership function was chosen to be 'linear'. The input and output were given to genfis3 utilizing the database created in the first stage of the approach. The number of cycles for genfis3 was chosen to be 200. Along these lines, the training process of ANFIS was executed.

In this work, the system has eight inputs and one output. HL and CL were estimated respectively. The ANFIS-FCM Regression structure is seen in Figure 1. At the input of this structure, 128 Gaussian membership functions with 16 clusters are used for 8 input.

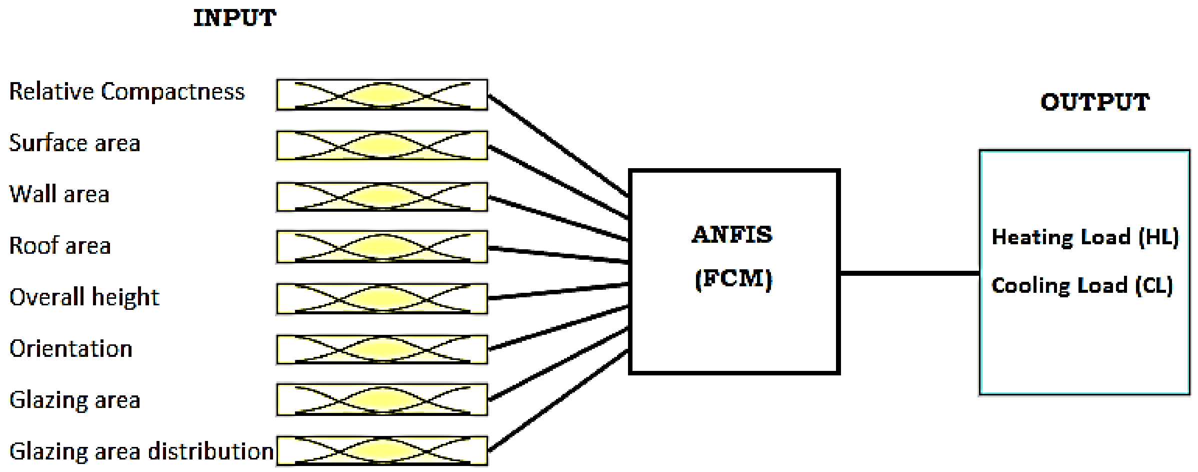

Figure 1. ANFIS Structure

After the training of the used data set was over, 230 data sets that were not used in training before were used to test the success rates of the system. As a result of these test operations for the HL and CL estimates, the graphs of the target values and estimated values can be seen in Figure 2 and Figure 3. The actual values are plotted in black, and the predicted values are shown in red.

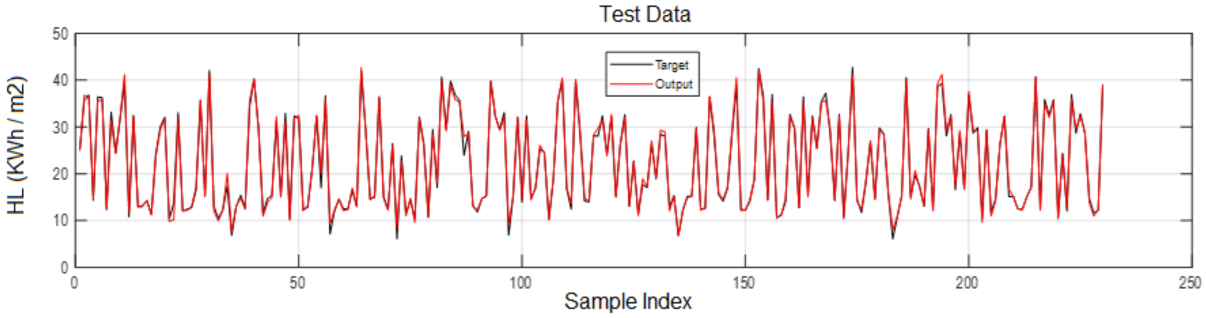

Figure 2. The testing results of ANFIS for HL

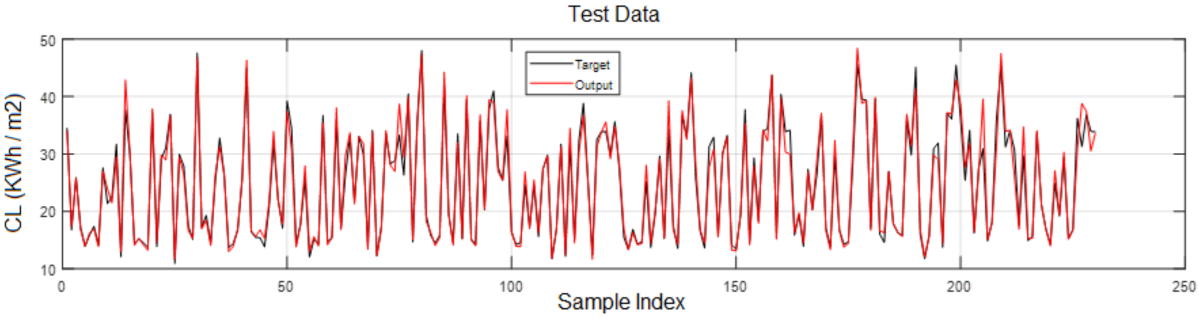

Figure 3. The testing results of ANFIS for CL

It is seen that correlation of training, testing and all predicted results have a good correspondence with the experimental data in Figure 4 and Figure 5. The value of the coefficient of determination R^2 is obtained as 0.997 for HL and 0.990 for CL, which is very close to 1 and they indicate the development of a good correlation between estimated output and target output. Regression results of HL and CL predictions can be seen in Figure 4 and Figure 5 respectively.

Table 3 and Table 4 present the comparison results of six regression models for the prediction of heating and cooling loads.

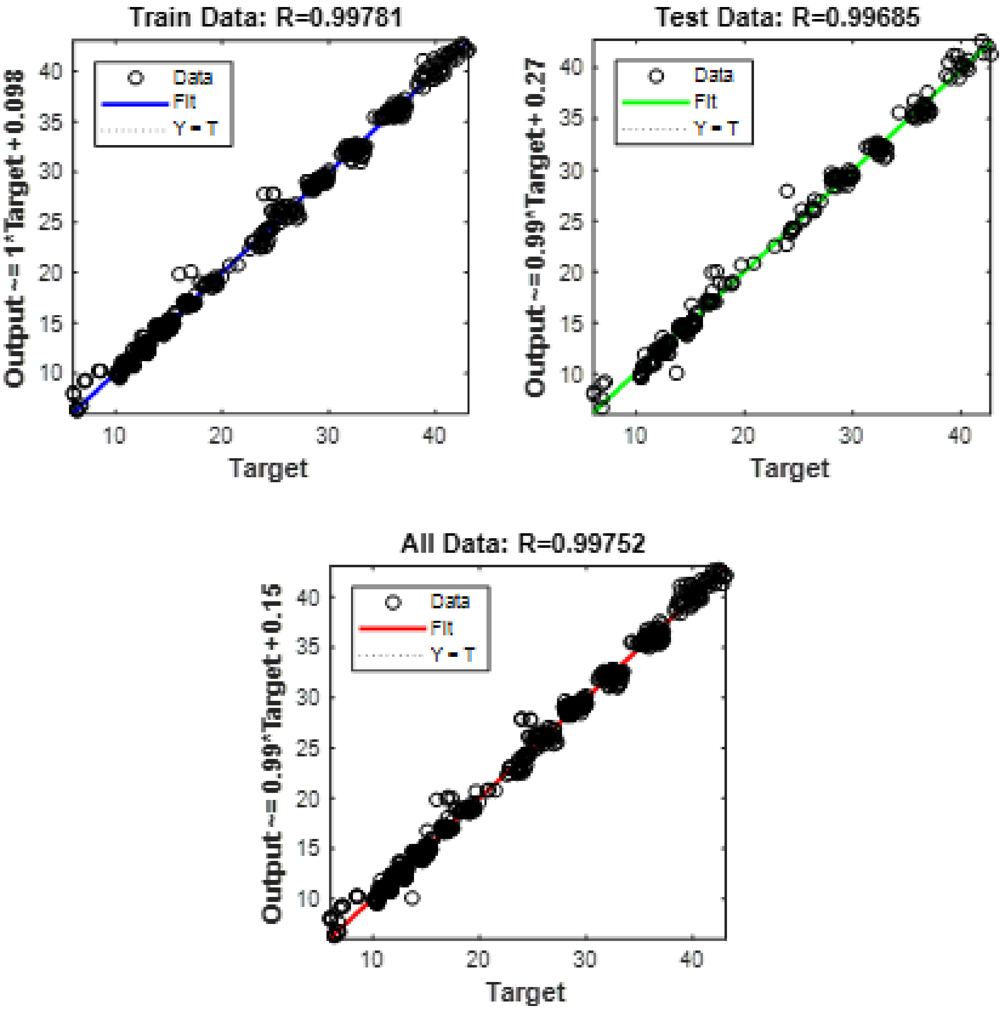

Figure 4. Regression results of HL estimation's scattering plots

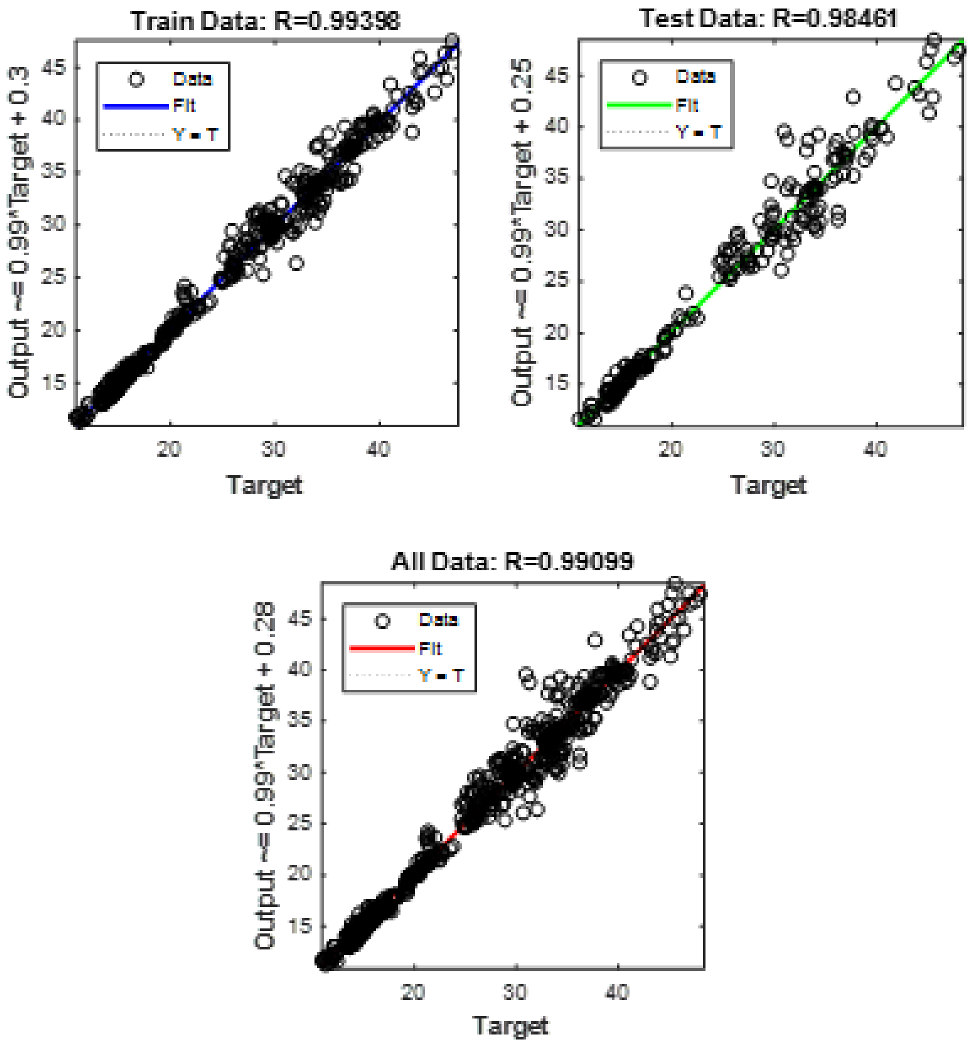

Figure 5. Regression results of CL estimation's scattering plots

RMSE, MAE and R values can be seen in Table 3 and Table 4. It is seen that the ANFIS Regression approach is a good candidate for regression learning.

The comparison of previous studies for the predictions of CL and HL and the results obtained with the ANFIS model used in this study are shown in Table 5.

Table 3. Results of the developed regression models for HL

Regression Models	R	MAE	RMSE
Linear	0.95	1,97	2,81
MLP	0.9815	1.4062	1.9119
RBF Regressor	0.9673	1.7944	2.5437
SVM	0.96	1.8922	2.7939
Gaussian	0.9603	1.9588	2.7924
Proposed ANFIS Regressor*	**0,99752**	**0.46**	**0.68**

Table 4. Results of the developed regression models for CL

Regression Models	Correlation Coefficient	MAE	RMSE
Linear	0.948	2.1469	2.959
MLP	0.9736	1.6353	2.0974
RBF	0.9542	2.0018	2.7763
SVM	0.9438	2.0666	3.102
Gaussian	0.949	2.1509	2.9224
Proposed ANFIS Regressor*	**0,99099**	**1.26**	**1.60**

Table 5. MAE and RMSE obtained the regression methodologies, including the proposed approaches.

Study (Model)	Cooling Load		Heating Load	
	MAE	RMSE	MAE	RMSE
Tsanas (2012) (IRLS)	2.21	3.38	2.14	3.14
Tsanas (2012) (RF)	1.42	2.57	0.51	1.01
Castelli (2015) (GSGP)	1.47	2.36	1.31	1.06
Castelli (2015) (GSGP-LS)	1.37	2.36	1.26	1.04
Cheng (2014) (RBFNN)””	1.30	1.69	0.51	0.67
Cheng (2014) (MARS)	1.12	1.65	0.53	0.68
Le (2019) (GA-ANN)	-	-	0.79	1.625
This study (ANFIS)	1.26	1.60	0.46	0.68

4. Conclusions

In this study, building energy application evaluation was carried out using various Regression Student models to wait for heating and cooling loads. After making comparisons with the regression learning strategies found in the literature, the gains obtained in this study show that there is an alternative to early estimates of building cooling and heating loads. Relative compactness, overall height surface area, wall area, roof area orientation, glazing area, and glazing area distribution of different structure shapes are utilized as the input of system and HL and CL of the buildings were used as the output of the proposed ANFIS Regression model. The data set consists of 768 examples of simulated buildings. Test errors and regression results show that the proposed model gives very convincing results with R = 0.998 for the HL estimation and R = 0.991 for the CL estimation and it gives us reasonable accuracy. In future studies, the new data set will be evaluated with the expectation to improve the results presented here.

5. Acknowledgment

This study was funded by Scientific Research Projects Coordination Unit of Istanbul University-Cerrahpasa. Project numbers: 23444

6. References

Bezdek, J.C. 1973. "Fuzzy Mathematics in Pattern Classification. Ph.D. dissertation", *Cornell University*, Ithaca, NY.

Castelli, M., Trujillo, L., Vanneschi, L. and Popoviˇc, A. 2015. “Prediction of energy performance of residential buildings: A genetic programming approach” *Energy Build*, 102, 67–74.

Cheng, M.Y. and Cao, M.T. 2014. “Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines” *Appl. Soft Comput*, 22, 178–188.

Directive 2002/91/EC of The European Parliament and of The Council of 16 December 2002 on the energy performance of buildings.

Duarte, G.R. Fonseca, L.G., Goliatt, P.V.Z.C. and Lemonge, A.C.C. 2017. “Uma comparação de técnicas de aprendizado de máquina para a previsão de cargas energéticas em edifícios”, *Ambiente Construído*, 17(3), 103-115.

Ekici, B.B. 2016. “Building energy load prediction by using LS-SVM”, *International Journal of Advances in Mechanical and Civil Engineering*, vol. 3, No. 3, p.p. 163-166.

Fan, H., MacGill, I.F. and Sproul, A.B. 2016. “Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia", *Energy and Buildings* , vol.105, p.p.: 9–25.

Haykin, S. 1999. "Neural Networks A Comprehensive Foundation, second ed", *Prentice Hall,* New Jersey.

Jang, J.S. 1993. "ANFIS: adaptive-network-based fuzzy inference system", *Systems, Man and Cybernetics, IEEE Transactions on* 23: 665–685.

Jeon, Y.K., Kim, T., Nam, H.S. and Lee, II.W. 2016. "Implementation of energy performance assessment system for existing building", *International Conference on Information and Communication Technology Convergence (ICTC),* Jeju, South Korea, p.p.:393-395.

Le, L.T., Nguyen, H., Dou, J. and Zhou, J. 2019. "A Comparative Study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in Estimating the Heating Load of Buildings' Energy Efficiency for Smart City Planning" *Appl. Sci,* 9, 2630.

Lee, M-W. and Kwak, K-Ch. 2016. "An Incremental Radial Basis Function Network Based on Information Granules and Its Application", *Computational Intelligence and Neuroscience* . Vol. 16, p.p.:1-6.

Mehrabi, M., Sharifpur, M. and Meyer, J.P. 2012. "Application of the FCM-Based Neuro-Fuzzy Inference System and Genetic Algorithm-Polynomial Neural Network Approaches to Modelling the Thermal Conductivity of Alumina–Water Nanofluids", *International Communications in Heat and Mass Transfer*, vol. 39, pp. 971–997.

Pérez-Lombard, L., Ortiz, J., and Pout, C. 2008. "A review on buildings energy consumption information", *Energy and Buildings*, vol. 40, no. 3, pp. 394–398.

Sholahudin, S., Alam, A.G., Baek, C. and Han, H. 2014. "Prediction and analysis of building energy efficiency using artificial neural networks and design of experiments", *Jurnal Mekanikal*, vol. 37, no. 2, pp. 37– 41.

Smola, A.J. and Schölkopf, B. 1998. "On a kernel-based method for pattern recognition, regression, approximation and operator inversion", *Algorithmica* 22, 211–231.

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

Wilkinson, G.N. and Rogers, C.E. 1973. "Symbolic descriptions of factorial models for analysis of variance", *Applied Statistics* 22, 392–399.

Williams, C.K.I. 1998. "Prediction with Gaussian processes: From linear regression to linear prediction and beyond", In: M.I. Jordan, editor, *Learning in Graphical Models,* Kluwer, pp. 599–621.

Yang, L., He, B.J. and Ye, M. 2014. "Application Research of Ecotect in Residential Estate Planning", *Energy and Buildings*, v. 72, p. 195–202.

Yu, Z., Haghighat, F., Fung, B.C., and Yoshino H. 2010. "A decision tree method for building energy demand modeling*", Energy Build*, vol. 42, no. 10,p.p.: 1637–1646.