Comparison of principal component analysis and ANFIS to improve EEVE Laboratory energy use prediction performance

Desmira1,2, Norazhar Abu Bakar3, Romi Wiryadinata3, Mustofa Abi Hamid1,4, Nur Kholifah5, Muhammad Nurtanto6

1Department of Electrical Engineering Vocational Education, Faculty of Teacher Training and Education, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia
2Department of Electrical Engineering, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Malaysia
3Department of Electrical Engineering, Faculty of Engineering, Universitas Sultan Ageng Tirtayasa, Cilegon, Indonesia
4Center for Vocational and Technical Education and Training, Mahiad Research Institute, Serang, Indonesia
5Department of Fashion and Culinary Engineering Education, Faculty of Engineering, Yogyakarta State University, Yogyakarta, Indonesia
6Department of Mechanical Engineering Vocational Education, Faculty of Teacher Training and Education, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia

ABSTRACT

The energy use that is in excess of practicum students’ needs and the disturbed comfort that the practicum students experience when conducting practicums in the Electrical engineering vocational education (EEVE) laboratory. The main objective in this study was to figure out how to predict and streamline the use of electrical energy in the EEVE laboratory. The model used to achieve this research’s goal was called the adaptive neuro-fuzzy inference system (ANFIS) model, which was coupled with principal component analysis (PCA) feature selection. The use of PCA in data grouping performance aims to improve the performance of the ANFIS model when predicting energy needs in accordance with the standards set by the campus while still taking students’ confidence in conducting practicum activities during campus operating hours into consideration. After some experiments and tests, very good results were obtained in the training: R² in training; minimum RMSE=0.011900; epoch of 100 per iteration; and R²=0.37522. In conclusion, the ANFIS model coupled with PCA feature selection was excellent at predicting energy needs in the laboratory while the comfort of the students during practicums in the room remained within consideration.

1. INTRODUCTION

This study is important given that the world’s energy needs decreased in 2020 due to the global COVID-19 pandemic [1], [2]. As the IEA stated in 2021, the world’s energy needs will rebound and increase by about 1.6% per year until 2030 [3]–[7]. Energy use is predicted by several affecting things, including CO₂ and economic growth [8], [9]. Considering that the world’s supply of fossil energy sources has been depleted, it is important then to conduct an in-depth study of energy efficiency. In this regard, China's energy demand
is projected to be the largest [10], [11]. There has been a lot of research on the benefits and importance of more in-depth studies of energy utilization and consumption in major cities in the world, including those in Jordan [12], [13], China [14], Turkey [15], [16], Iran [17], Taiwan [18], South Korea [19], European [20], and Ontorio in Canada [21].

Energy consumption in colleges is heavily influenced by the climate [22], [23]. For instance, at Guang Dong University, a study has been conducted by building a conservation-oriented campus, which has a significant effect on energy saving. Various studies also found that it is difficult to measure indicators of energy consumption [24], [25]. Targeted energy efficiency strategies on university buildings in China have been developed in accordance with local conditions with different multi-campus and climate factors being among the factors of difficulty in realizing energy efficiency in the buildings [26]. Historical data on daily electricity use in two London South Bank University buildings have been regressed over normalised data from six input variables to obtain efficient energy use in a building [27]. A case study of three campus buildings in Tianjin predicted that the average electricity consumption of one occupant varies depending on the function of the building and the mode of control of electrical equipment [28]. Various developed statistical regression models are used to understand the relationship between each variable and energy consumption [29]. In another earlier research study, it was revealed that enhanced modeling could be used in other types of buildings as long as it had an energy consumption monitoring platform, not limited to campus buildings [30].

Several studies discussing the importance of energy saving include the importance of indoor energy utilization [31], [32], including one conducted by the University of Sao Paulo using the energy plus model as an evaluation of the parameters used to predict energy consumption in an administration building [33]. A room in the United Kingdom was also used as a research object to predict the energy use in the room [34]. Indoor energy consumption models have also been widely used in several studies, one of which is used at Zhejiang University China, namely the genetic algorithm-adaptive neuro-fuzzy inference system (GA–ANFIS) comparison, which produced good accuracy from artificial neural network (ANN) [35]. Therefore, this method is excellent at predicting energy use (see Figure 1). Among the prediction models, time series analysis is used to analyze energy use in buildings, but in one study the variable occupancy rate was not covered [36]. The adaptive neuro-fuzzy inference system (ANFIS) generated heating prediction correlation coefficient of R²=0.9959 and cooling prediction correlation coefficient of R²=0.9567, which were fairly good but did not reach the correlation coefficients of R=1. Two ANFIS models can be used to predict energy consumption from renewable sources [37]. The fuzzy c-means (FCM) clustering techniques provide faster computing, but compared to genetic programming (GP) they result in longer computing times, with estimated accuracy of 73.48% and 83.11% for ANFIS modeling [38]. The ANFIS model used for modeling based on four tested model parameters resulted in root mean square error (RMSE) of 0.029817613, with generalized bell membership type and a time of 3 hours and 33 minutes [39]. Electrical energy with GA optimization is used as an ANFIS input parameter to obtain better network performance computing [26], [40]–[42]. Energy usage prediction models include relevance vector machine (RVM), group method data handling (GMDH), ANFIS-biogeography-based optimization (ANFIS-BBO), and ANFIS-improved particle swarm optimization (ANFIS-IPSO) models. Of the four, RVM produces the best accuracy with a correlation coefficient (R²) close to 1 and an average error of 0.06 [43]. The ARIMA model as an input for an ANFIS model with evaluation uses the MSE criterion of 0.026% from 0.058% [44].

To improve the performance of a network, feature selection is required to enhance the performance of the algorithmic model used [45], [46]. The purposes of input variable selection are to improve predictive performance, speed, and cost effectiveness and to provide an underlying understanding of the data [47]. Features used to accelerate support vector machine (SVM) performance in predicting energy may be irrelevant, so methods to degrade or eliminate irrelevant features were proposed [48]. Among the feature selection techniques used to reduce high-dimensional data is GA [49]. Consumption patterns were strongly influenced by the 4-hour schedule and activity of the 4 models tested in random forest, neural networks, and fuzzy inductive reasoning. Variable inputs were used to provide the best results using FSP with 24-hour predicted forecasts [50]. Predicting the use of electrical energy by using multi-output support vector regression-memetic algorithm (MSVR-MA) and selection of features using ma identify less with lower errors [51]. The implementation of binary genetic algorithm–principal component analysis (BGA-PCA) also produced excellent accuracy in getting an average error against mean absolute percentage errors (MAPE) 17% better than before [52]. The selection of features used in precision to select information based on information gain (IG) and GA has provided a decrease in text and vector dimensions [53].

In this study, a predictive model of energy consumption in the Laboratory of Electrical Engineering Vocational Education (EEVE) at Universitas Sultan Ageng Tirtayasa (UNTIRTA) was developed. Among the input parameters that affect the energy use in a building is the behavior of the room users, and it is difficult to determine how much the energy consumption is [54]. In previous energy consumption prediction models [55]–[59] the features were directly used as an input in the prediction stage. Meanwhile, in this study.
selection of features was carried out first before entering the prediction stage. Feature selection was aimed to select features that affected energy consumption the most [60]. With selection of features, it was expected that the prediction stage could be carried out more effectively and accurately. The purpose of the feature selection method was to obtain a reduction set by removing some features that were considered irrelevant for text sentiment classification in order to result in an increase in accuracy classification and a decrease in the duration of machine learning models [61]. However, the disadvantage of feature selection was that it required training of a large data set to obtain a reliable transformation [62]. One way to overcome high dimensions of features is feature selection using various feature selection techniques [63]. Information gain [64], for example, has been used to reduce vector dimensions. Another way to overcome the problem of high dimensions of features is dimension reduction.

The purpose of dimension reduction is to obtain new data representations in lower dimensions [65]. The linear model of dimension reduction consists of the singular value decomposition (SVD) model and the PCA model [66]. However, the linear model of dimension reduction has advantages in that it produces a linear combination of all the features that may be contaminated by noise and that it decreases the performance of the classification model. When the linear model of dimension reduction is used, it will become difficult to deal with non-linear data [67]. PCA is used to improve calculation accuracy and reduce training time [68]. In this study, feature selection was conducted using the PCA model, while the prediction stage was carried out using the ANFIS model. The features used to predict energy consumption included pressure, temperature (DHT22), current, voltage, power, light intensity, temperature (BMP180), altitude, and humidity. Meanwhile, the prediction model resulted in comparison of actual data of calculations in a traditional way using mathematical formulations with the proposed model, the ANFIS model. Computer computing was used in making predictions with the ANFIS model [69]–[72].

2. RESEARCH METHOD

A flowchart of the energy consumption prediction model using PCA as a feature selection method is shown in Figure 1. It can be seen that historical data were the input for the PCA model. The input variables were nine-fold, reduced so that the data would be fewer, easier, and faster to display using the ANN model into 4 pairs of variables. Another goal of the use of the PCA model here was transformation. Therefore, the data that were initially correlated with each other became uncorrelated, and they also became easier and faster to display using the ANN model into the next stage. Normalization is the initial stage in PCA to obtain the covarian matrix used to obtain eigenvectors and eigenvalues.

![Figure 1. Flowchart of the PCA+ANFIS-based energy consumption prediction method](image)

The research flow can be seen from 2 historical data images based on real data retrieval using 9 sensors as inputs, namely pressure sensor, temperature sensor (DHT22), electric current sensor, voltage sensor, power sensor, luminous intensity sensor, temperature sensor (BMT 180), altitude sensor, and humidity sensor. By using the PCA model for feature selection in accordance with (1), PC1 to PC9 were obtained, in which case features with a value greater than 1 were selected. Figure 2 shows the stages performed in the grouping of data using PCA.
Comparison of PCA and ANFIS to improve EEVE Laboratory energy use prediction performance (Desmira)

The covariance with n number of attributes after data standardization is given by (1).

$$cov(A_1, ..., A_n) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

(1)

If the covariance is positive, the two dimensions increase together. Contrarily, if the covariance is negative, one dimension will decrease when the other one increases. If the covariance equals zero, the two dimensions are independent of each other. The largest number (m) of eigenvalues is selected, then the selected eigenvectors are stored as \(v_1, ..., v_m\). If M is the mxm matrix, then each \(\lambda\) with the (2).

$$Mv = \lambda$$

(2)

\(mx1, \ \text{vector} \ x \neq 0\), is called eigenvalue of A. Vector x is called eigenvector from A, which relates to eigenvalue \(\lambda\). The equation determinant in (2) must be met. The contribution of each feature is calculated using (3).

$$c_j = \sum_{p=1}^{m}|v_{pj}|$$

(3)

The largest value of \(c_j\) is obtained according to the number of features one would want to maintain, from which the \(j^{th}\) feature as a significant feature comes. A number of selected features are then used as input in the stage of energy consumption prediction. Energy consumption prediction is performed using two different models, namely the PCA and ANFIS models.

The accuracy of the data collected in real time at the EEVE laboratory were tested and analyzed by comparing the actual/original data against the prediction data using the ANFIS method and the root mean square error (RMSE) formula. The RMSE formula was used to search for the accuracy of the prediction data and original data of measurements in the laboratory. The forecasting results and MSE values would range from 0 to infinity, with 0 being the best value [45], [72]. The RMSE formula is as (4).

$$RSME = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

(4)

Where \(\hat{y}_i\) is the predictive value of the PCA-ANFIS model, \(y_i\) is the actual value, and \(n\) is the data size.

3. RESULTS AND ANALYSIS

Data training was conducted based on the data taken from the data sensors as variable input after feature selection using the PCA model (see Table 1). Eigenvectors and eigenvalues obtained using (1) are presented. The data were grouped from the PCA results obtained according to Table 1.

| No | PCA | Results of dimensional reduction transformation= \(c_j = \sum_{p=1}^{m}|v_{pj}|\) |
|----|-----|-----------------------------------------------|
| 1  | PC1 | 3.183383339858511                           |
| 2  | PC2 | 1.65908166416605                             |
| 3  | PC3 | 1.50879133963896                             |
| 4  | PC4 | 1.072473976011161                            |
| 5  | PC5 | 0.736835812172469                            |
| 6  | PC6 | 0.511552879372422                            |
| 7  | PC7 | 0.251043706402677                            |
| 8  | PC8 | 0.0463160415988100                           |
| 9  | PC9 | 0.0305211820519049                           |
Data standardization in the first step would affect the covariance matrix based on the input variables and historical data as shown in Figure 3. In the second step, the covariance matrix served the input value to calculate the eigenvalues and eigenvectors according to (2). The eigenvalues would inform about how much diversity a PC variable would be able to explain by the formation of $\lambda_1 \ldots \lambda_n$ based on (3). In the fourth step, as the eigenvalues and eigenvectors were known, the value of each PC could be calculated according to (4). In the fifth step, only PCs with an eigenvalue of >1 would be selected. Based on Table 1, the results for the dimension reduction process were that PC1 to PC9 that had an eigenvector and an eigenvalue greater than one would be selected for the ANFIS model, in this case PC1 to PC4. Thus, PC1 to PC4 were used in the next stage as the input for the ANFIS model to predict the energy needs of the EEVE laboratory. The next process was to conduct training using the input parameters of the ANFIS model using PC1 to PC4, from which coefficient correlation $R=1$ was obtained, with the average error calculated using the RMSE formula in (4) being 0.011900. From Figure 4 it can be seen that the distribution of data between actual data and prediction data using the ANFIS model was very good, with a correlation coefficient value of $R=0.37522$ and the target MSE value reaching 1,000 iterations.

An average RMSE value of 0.011900 was resulted. The results of the comparison of training data against actual energy calculation data under the ANFIS model can be seen from Figure 5. From Table 2 and Figure 5 it can be seen that the comparison of actual energy data and testing data using PCA and ANFIS
Comparison of PCA and ANFIS to improve EEVE Laboratory energy use prediction performance (Desmira)

shows similarity to the actual energy data of the EEVE laboratory. The training data from predictions 1 to 50 of actual energy usage in the EEVE laboratory are almost the same as the prediction data collected using PCA+ANFIS. A significant difference occurred in prediction 55 between PCA and ANFIS, amounting 802.6744347 Wh.

Table 2. Data prediction of energy (actual) and PCA+ANFIS

<table>
<thead>
<tr>
<th>No</th>
<th>Day/Date</th>
<th>Humidity (%)</th>
<th>Energy (actual)</th>
<th>PCA+ANFIS</th>
<th>Error=energy (actual)-(PCA+ANFIS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monday, 6/01/2020</td>
<td>75.5979592</td>
<td>135.111837</td>
<td>135.1122</td>
<td>-0.000363265</td>
</tr>
<tr>
<td>2</td>
<td>Tuesday, 7/01/2020</td>
<td>74.0591837</td>
<td>99.6604082</td>
<td>99.6594</td>
<td>0.001008163</td>
</tr>
<tr>
<td>3</td>
<td>Wednesday, 8/01/2020</td>
<td>76.2469388</td>
<td>119.275102</td>
<td>119.2766</td>
<td>-0.001497959</td>
</tr>
<tr>
<td>4</td>
<td>Thursday, 9/01/2020</td>
<td>63.3408163</td>
<td>312.561633</td>
<td>312.5597</td>
<td>0.001932653</td>
</tr>
<tr>
<td>5</td>
<td>Friday, 10/01/2020</td>
<td>76.4755102</td>
<td>132.153469</td>
<td>132.153</td>
<td>0.000469388</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>52</td>
<td>Friday, 20/03/2020</td>
<td>50.4918367</td>
<td>215.000816</td>
<td>147.604</td>
<td>67.3968163</td>
</tr>
<tr>
<td>53</td>
<td>Monday, 23/03/2020</td>
<td>57.6612245</td>
<td>225.668571</td>
<td>-78.213</td>
<td>303.8815714</td>
</tr>
<tr>
<td>54</td>
<td>Tuesday, 24/03/2020</td>
<td>62.2346939</td>
<td>190.089796</td>
<td>445.111</td>
<td>-255.0216041</td>
</tr>
<tr>
<td>55</td>
<td>Wednesday, 25/03/2020</td>
<td>56.9857143</td>
<td>326.236735</td>
<td>-476.4377</td>
<td>802.6744347</td>
</tr>
<tr>
<td>56</td>
<td>Thursday, 26/03/2020</td>
<td>60.7918367</td>
<td>237.327347</td>
<td>498.3804469</td>
<td>54.5147939</td>
</tr>
<tr>
<td>57</td>
<td>Friday, 27/03/2020</td>
<td>77.3875511</td>
<td>223.441633</td>
<td>224.4397</td>
<td>-1.018067347</td>
</tr>
<tr>
<td>58</td>
<td>Monday, 30/03/2020</td>
<td>45.4061224</td>
<td>223.813878</td>
<td>-203.1022</td>
<td>426.9160776</td>
</tr>
<tr>
<td>59</td>
<td>Tuesday, 31/03/2020</td>
<td>51.9102041</td>
<td>230.292245</td>
<td>-334.9615</td>
<td>565.2537449</td>
</tr>
<tr>
<td>60</td>
<td>Wednesday, 1/04/2020</td>
<td>47.2346939</td>
<td>317.438367</td>
<td>-459.4389</td>
<td>776.8772673</td>
</tr>
<tr>
<td>61</td>
<td>Thursday, 2/04/2020</td>
<td>45.2591837</td>
<td>363.2489</td>
<td>-296.0683</td>
<td>67.1806795</td>
</tr>
<tr>
<td>62</td>
<td>Friday, 3/04/2020</td>
<td>74.1367347</td>
<td>432.040816</td>
<td>-498.006</td>
<td>-65.96518367</td>
</tr>
<tr>
<td>63</td>
<td>Monday, 6/04/2020</td>
<td>44.422449</td>
<td>289.204898</td>
<td>-64.04</td>
<td>353.248498</td>
</tr>
<tr>
<td>64</td>
<td>Tuesday, 7/04/2020</td>
<td>53.2142857</td>
<td>387.998367</td>
<td>295.0779</td>
<td>92.2046735</td>
</tr>
<tr>
<td>65</td>
<td>Wednesday, 8/04/2020</td>
<td>53.6918367</td>
<td>363.2489</td>
<td>296.0683</td>
<td>67.1806795</td>
</tr>
<tr>
<td>66</td>
<td>Thursday, 9/04/2020</td>
<td>68.8387755</td>
<td>432.040816</td>
<td>-498.006</td>
<td>-65.96518367</td>
</tr>
<tr>
<td>67</td>
<td>Friday, 10/04/2020</td>
<td>61.2571429</td>
<td>391.766531</td>
<td>560.6681</td>
<td>-168.9015694</td>
</tr>
</tbody>
</table>

Figure 5. Prediction energy (actual) vs PCA+ANFIS training and testing

4. CONCLUSION

The result of the research conducted at EEVE laboratory by comparison energy data (actual) and prediction data using the ANFIS model with feature selection in PCA was very good, with training correlation coefficient of $R=1$, $R=0.37522$, RMSE=0.011900, and epoch=100. The result of the research that was conducted for 3 months showed that the actual data or traditional calculation data were very similar when using the ANFIS model with PCA feature selection. Thus, the ANFIS model with feature selection in PCA is well-suited to predict the energy needs of a laboratory while still paying attention to the comfort the students feel when conducting practicums in the room.

ACKNOWLEDGEMENTS

Our gratitude is due to the Department of Electrical Engineering Vocational Education, Faculty of Teacher Training and Education, Sultan Ageng Tirtayasa University, for having given permission to conduct this research at the laboratory as well as to all the supervisors involved.
REFERENCES


Comparison of PCA and ANFIS to improve EEVE Laboratory energy use prediction performance (Desmira)


BIographies of AUTHORS

Desmira is currently a lecturer with Universitas Sultan Ageng Tirtayasa (Untirta), Indonesia. She is also an Assistant Professor with the Department of Electrical Engineering, Vocational Education, Untirta. She has published more than 40 research articles, with more than 80 citations received in the Google Scholar and H-index 3. His research interests are in the areas of electrical engineering, sensors, and electronics. She can be contacted at email: desmira@unirta.ac.id.

Norazhar Abu Bakar received his BEng. (Hons.) degree in electronics and electrical engineering from Leeds University, UK, MSc. (Eng.) in control systems from Sheffield University, UK, and PhD degree in electrical and computer engineering from Curtin University, Australia. He started his career as an assembly engineer at Toshiba Electronics, and then moved to TNB as a SCADA/DA Project Engineer. Currently, he is the Deputy Dean (Research & Postgraduate Studies) with the Faculty of Electrical Engineering, UTeM. He is a member of the High Voltage Engineering Research Laboratory in UTeM. His research interests are condition monitoring, asset management, and artificial intelligence. Since 2013, he has published several articles related to transformer condition monitoring and asset management in IEEE Transactions on Dielectrics and Electrical Insulation and IEEE Electrical Insulation Magazine. He can be contacted at email: norazhar@utem.edu.my.
Comparison of PCA and ANFIS to improve EEVE Laboratory energy use prediction performance (Desmira)

Romi Wiryadinata received his B.Eng. degree in electrical engineering from Universitas Islam Indonesia, Yogyakarta, Indonesia. M.Eng. in control systems from Universitas Gadjah Mada, Yogyakarta, Indonesia and Dr. degree in electrical engineering from Universitas Gadjah Mada, Yogyakarta, Indonesia. He started his career as a lecturer at Department of Electrical Engineering, Universitas Sultan Ageng Tirtayasa. Currently, he is the Head of Department of Electrical Engineering, Faculty of Engineering, Universitas Sultan Ageng Tirtayasa. He has published more than 80 research articles, with more than 258 citations received in the Google Scholar and H-index 9. His research interests are sensor and instrumentation, robotics, telecontrol, and disaster and mitigation. Hen can be contacted at email: wiryadinata@untirta.ac.id.

Mustofa Abi Hamid (Member, IEEE) has been the secretary of Department of Electrical Engineering Vocational Education, Universitas Sultan Ageng Tirtayasa (Untirta), Indonesia, since 2021. He is currently an Assistant Professor with Untirta. He is also an Assistant Professor with the Department of Electrical Engineering Vocational Education, Faculty of Teacher Training and Education, Untirta. His research interests include in the areas of electrical engineering, electronics, technical and vocational education, curriculum design, learning strategies, learning media, ICT for learning. He has written several papers in the areas of technical and vocational education and training (TVET), learning strategies, learning media, technology enhanced learning, engineering education, curriculum development, education for sustainability development, professional learning. He is on the Editor-in-Chief of the journal VOLT: Jurnal Ilmiah Pendidikan Teknik Elektro, Journal of Vocational and Technical Education and Training; and reviewer more than international journal. He can be contacted at email: abi.mustofa@untirta.ac.id

Nur Kholifah is a lecturer at the Department of Fashion and Culinary Engineering Education, Yogyakarta State University, Yogyakarta, Indonesia. Research interest in the field of empowering vocational communities, learning methods, vocational teachers, and gamification. Email: nur.kholifah@uny.ac.id.

Muhammad Nurtanto is an assistant professor in Department of Mechanical Engineering Education, Universitas Sultan Ageng Tirtayasa, Banten, Indonesia. Research interests in the fields of professional learning, Teacher Emotion, Teacher Identity Philosophy of education, STEM education, gamification, and teacher quality in vocational education. Email: mnurtanto23@untirta.ac.id