

The Prediction of Earthquake Building Structure Strength: Modified K-Nearest Neighbour Employment

Okfalisa Okfalisa¹, Septian Nugraha¹, Saktioto Saktioto², Zahidah Zulkifli³, S.S.M. Fauzi⁴

¹Department of Informatics Engineering, Faculty Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

²Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Riau, Pekanbaru, Indonesia

³Department of Information Systems, Kuliyyah of Information Communication and Technology, International Islamic University Malaysia, Malaysia

⁴Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Perlis Branch, Arau Campus 02600 Arau, Perlis, Malaysia

Article Info

Article history:

Received Apr 22, 2020

Revised Nov 10, 2020

Accepted Nov 25, 2020

Keywords:

Earthquake

Modified K-Nearest Neighbor

Classification

Prediction

Performance Measurement

ABSTRACT

The earthquake damage brings significant effects. The resilience of buildings against the earthquake and the destruction's location is not an efficient outcome from previous research. This study applied the Modified K-Nearest Neighbor (MK-NN) in predicting the concrete structures' performance despite the earthquakes. The 2-story building prediction covered earthquake history, time, concrete quality, displacement, velocity, and acceleration. The analysis of MK-NN provided the values of Euclidean, distance calculation, validity, and weight voting towards the classification of damages as "Safe" or "Immediate Occupancy" (IO). The K values exploited were 1, 3, 5, 7, 9, and 11, and simulation data training at 10:90, 20:80, 30:70. This study revealed the highest degree of accuracy at 98.85% with K=1 and a ratio of 30:70. Simultaneously, the lowest error rate was 1.15% at a similar K value and ratio. Herein, MK-NN significantly exceeds the accuracy and error rate of KNN up to 1.02% and 0.69%, respectively. To date, the automatic calculation prototyping software was then successfully developed. Ensuring the application's accuracy, the Confusion Matrix, the Black box, and User Acceptance Test (UAT) have been performed. In a nutshell, this study provides a significant contribution to planning and information analysis of earthquake-resistant construction.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Okfalisa,

Department of Informatics Engineering,

Universitas Islam Negeri Sultan Syarif Kasim Riau,

HR. Soebrantas Panam Street. Km. 15 No. 155, Riau, Indonesia, 28293.

Email: okfalisa@gmail.com

1. INTRODUCTION

Technological advancement has taken into account the control of various parts and human life fields, including the concrete building construction. Theoretically, the design planned of multi-stories construction should be equipped with structural document analysis and building Information in supporting the outlined instrument towards the physical development stage [1]. Building Information Modelling (BIM) integrates intelligent objects with the data regarding a specific component from geometric characteristics to provide the building information [1]. Previous researchers consider BIM an evolving technology that proposes excellent benefits for the architecture, engineering, and construction industry [2]. The most significant contribution of BIM in structural engineering activities is to better analyze situations through simulation, the coordination, consistency of data, and visualization [3] and [4]. The reduction of design, drafting errors, and costs were identified due to improved productivity [5-7]. BIM offers facilities in developing, executing, and managing

infrastructure projects [8]. However, big data's emergence becomes a challenge for BIM to outdated data usage and experience-based decision-making [9]. The heterogeneity information, storage complexity, and users' specialized functions tend to lead to non-intuitive data, inaccurate data, tedious and costly data [10]. Thus, inaccurate records deliver confused data analysis processes, lower data quality, and negative implications [11]. Lin et al. [11] have found that data mining processes can enhance BIM data. This data mining has the potential to find latent patterns, as well as a prediction based on the rise of information requirements [8]. Nowadays, data mining's successful application in solving extensive data analytical for many fields of study, especially in the building environment, has been pointed out with or without support by the BIM platform. It successfully classifies and predicts the big data as the purpose of analytical problems. Yu et al. [12] proposed a data mining framework, including classification, clustering, and association rule mining, to analyze building-related data more efficiently. Xiao and Fan [13] applied the clustering analysis to identify the tallest building's typical power consumption patterns in Hongkong. A big data analysis framework has been effectively retrieved and calculated the required information from dynamic BIM by Hung-Ming et al. [14]. A. Costa et al. [15] proposed a novel integrated toolkit designed to analyze performance monitoring and analysis building operation and energy performance. Herein, BIM applied as methods and technology in structured performance definitions. The above studies showed that data mining techniques are valuable knowledge discovery to obtain better building operational performance. Data mining had a tremendous potential technology to discover hidden knowledge in large data sets, and this method had significant superiority in prediction accuracy [16, 17]. Data mining techniques can accelerate the analysis to be more accurate, reliable, and computationally efficient models [12].

The impact of the earthquake is ravaging. Thus, it terminates the socio-economic activities of a region within a short period. Due to the earthquake's effect, the consideration of structure configuration, the type of material used, and the structural building system are substantial [18]. Magnitude earthquake loads on building structures depend on the horizontal force, vertical force, torque earthquake moment in the structures, weight, and stiffness of the structural material, configuration and structural system, vibration time, ground conditions, earthquake zones, and earthquake behavior. Dynamic analysis procedures are hard to predict and require classifying earthquake loads and a seismic response to the building structures [19]. Besides the structure's construction details, the uncertainty of building structures' collapse capacity during the earthquakes is also influenced by concrete compression strength, steel tension strength, earthquake energy content and frequency, and structure resistance model [20]. Thus, it triggers many efforts and risks involving uncertainty in the seismic responses.

In Indonesia, the seismic responses are arranged based on earthquake resilience planning guidelines for home and building structures in SNI 03-1726-2002. SNI was in line with the international seismic code as a seismic-resistant structure design, describing structural analysis performance despite the build-structure weight, considerable dead weight, and earthquakes' retention for sensitiveness areas. Thus, building workforce behaviors and deformations controlled are exhibited [21]. Moreover, the building's un-properly designed will cause discomfort and damage to non-structural components, including partition walls, windows, and doors, thus block the evacuation passage [22]. Unfortunately, SNI 03-1726-2002 was outdated [23] and replaced by SNI 03-1726-2012 through the increasing of peak ground acceleration (PGA) from 10% probability into 2% [20, 21]. Therefore, this research adopted SNI 03-1726-2012 to provide seismic resistant structure design as the primary data.

Previous research has been discussed on building structures' seismic response using artificial intelligence techniques, including machine learning and intelligent data analysis or big data analysis. Maram et al. [24] predicted the seismic behavior of reinforced concrete buildings using artificial neural networks (ANN). An ANN algorithm was also studied to consider structural vibration for the uncertainties in calculating damage identification data [25]. By emphasizing the type of soil, Karbassi et al. [26] applied decision tree algorithms to predict the stable reinforced building structure and found 95% accuracy for hard soil cases and 97% for soft soil cases. Y. Zhang et al. [27] investigated the safety of building structures post-earthquake using the machine learning method, the Classification and Regression Tree (CART), and Random Forest algorithms. This assessment provided 91% and 88% accuracy for the safety state in response and damage patterns respectively. In the case of a prestressed concrete bridge subject to earthquakes, Pei and Smyth [28] have been successfully investigated a feedforward neural network. Abd-elhamed et al. [18] proposed a Logical analysis of data (LAD) to simulate and blindly predict the dynamic response behavior of building structures against the earthquake loads. Nevertheless, the number of input variables of the above studies is still relatively small and has been pre-defined based on domain knowledge. To date, this research selected the influenced variables in the big series dataset of building resistance subjected to the earthquake and then used to map into the class prediction as outputs. A vast data set mining will be analyzed using the computationally intensive method to explore the strength of databases and possible integration as valuable information for BIM. A particular prediction task requires intelligent exploratory data analysis. Thus, it leads to a reliable and accurate algorithm.

K-Nearest Neighbors (KNN) is a data mining algorithm for non-parametric data classification that classifies and predicts big data for analytical problems. KNN becomes one of the most popular neighborhood classifiers. It provides a more flexible approach, complex, ease of understands, and interpretation [29], but it delivers highly competitive results [30]. The other advantages of KNN are including robust to noise training data and more effective in extensive training data [31]. Comparing to the Bayes algorithm, SVM, Fisher's linear discriminant analysis, Partial least squares discriminant analysis, Classification tree, Random Forest, and other Euclidean distance calculations, KNN revealed better efficiency and performance [32-34] even compared to SVM [35]. The KNN algorithm, therefore, allows the whole classification features to be determined. It thus affects classification variance and lowers precision [36].

Moreover, the computation cost is relatively high due to each query instance, large memory load, low accuracy rate in multidimensional datasets, the unclear of distance-based learning [32]. As the advancement of the KNN, the modified KNN is introduced. It operates concerning the weighted data training voting values, which fail to handle the KNN calculation [36]. This method assigns the class label of the data according to k validated data points of the data train set and eliminate those instances which fail the validity test. Thus, it overcomes the low accuracy and disadvantages of KNN [37]. Hamid et al. [31] evaluated KNN and MK-KNN on five different data sets. The result indicated the improvement of accuracy in comparison with the KNN method.

This study is applying the MK-NN to predict the performance level of the building structure in 2-story offices. The proposed data mining method has never been utilized in predicting the classification of building structure resistance subjective to the earthquake. Therefore, it provokes a new challenge in this research. As a case study, 6663 data from building construction in the area of Bangkinang city, in Riau Province, Indonesia, was calculated. The earthquake history, time, concrete quality, displacement, velocity, and acceleration became consideration variables in classifying the destruction into two classes, including Safe and Immediate Occupancy (IO) classes. A software prototype has been created to calculate the data perceived in the MK-NN algorithm automatically. The software can be utilized as a consideration tool in preparing new buildings and a recommendation system for concrete construction companies, particularly against an earthquake. The advantages mining analysis method in the database provides pertinent information on the integration of BIM.

2. RESEARCH METHOD

Data mining is a branch of science that studies methods for finding useful information from big data. It utilizes statistical techniques, mathematics, artificial intelligence, and machine learning to extract and interpret useful information from various large-sized data and presents it for sophisticated, valuable information and knowledge. Data mining can solve the problems related to the classification, regression, clustering, and association rule learning of data based on the purpose of analysis. The data pattern is descriptively and predictively displayed in this approach.

Data mining in the concept of Knowledge Discovery in Database (KDD) can analyze the data by applying algorithms to generate a list of patterns, models, and data [38]. It follows an interactive stage and lets the user to directly involved in the knowledge base (See Figure 1). KDD provides some benefits to potentially useful tasks, leads to useful insight, understandable immediately or after some post-processing [38]. KDD processes iterative and interactive sequence activities, including selecting a subset variable on which discovery has to be performed; Preprocessing-clean data by removal noise modeling, handling missing data field, and miscoding and accounting time sequence information. The dirty sets of big data can lead to inaccurate analytics, uncertain outcome, and unpredictable conclusions [39]; Transformation-reducing and projecting data as a specific task performed; Datamining-extracting interesting pattern by applying some methods (e.g., summarization, classification, clustering, and regression); Interpretation/Evaluation-visualizing the pattern to interpret and extract the knowledge. Figure 1 shows the clarification.

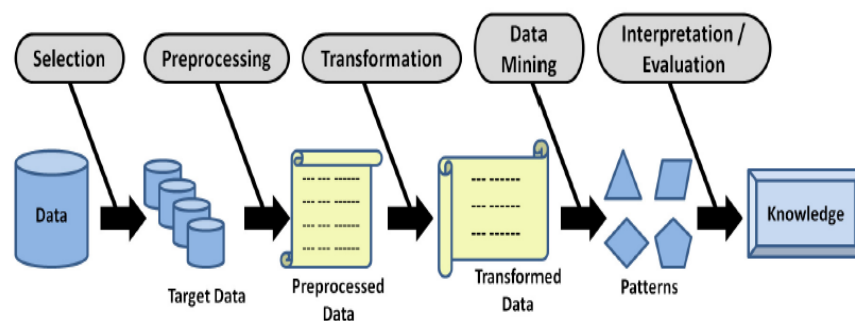


Figure 1. Flow Process of KDD [11]

Several classification algorithms, namely K-Nearest Neighbor (KNN), Modified K-Nearest Neighbor (MK-NN), Decision Tree Classifier (Random Forest, J48 Decision Tree), Rule-Based Classifier, Neural-Network, Support Vector Machine (SVM), Naive Bayes Classifier, and Artificial Neural Network (ANN), Extreme Gradient Boosting, SVM with Basis Function Kernel Approach [40]. In a nutshell, the KNN algorithm is one of the lazy learning categories commonly used to predict data. This method classifies the objects based on learning the nearest k 's value of data.

The distance of each sample of training data (x) against data tr (y) is calculated based on the following Euclidean Distance equation [41].

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

where:

- d = the distance between the points in training data x and the points in testing data y that will be classified, where $x=x_1, x_2, \dots, x_i$ and $y=y_1, y_2, \dots, y_i$
- i = the value of the attribute
- p = an attribute dimension.

MK-NN is implemented by considering weight voting values in data training to pursue several merits and demerits of KNN. The validity of data training and weight voting using the following formulae is the various processes for the advancement of KNN [8].

$$Validity(x) = \frac{1}{k} \sum_{i=1}^k S(label(x)(label(N_i(x))) \quad (2)$$

K = number of closest points.

$label(x)$ = class x .

$label(N_i(x))$ = The class label nearest point x .

The S function calculates the resemblance between x point and the i^{th} data from the nearest neighbor.

$$S(a, b) = \begin{cases} 1 & a=b \\ 0 & a \neq b \end{cases} \quad (3)$$

where a is the class in the training data, and b is another one.

The following formula is used for weight voting[6].

$$W(i) = Validity(i) x \frac{1}{d_{e+0,5}} \quad (4)$$

$W(i)$ = weight voting calculation

$Validity(i)$ = validity value

Where d_e is the Euclidean distance of Equation 1.

The accuracy value calculation in the confusion matrix is required to determine the classification's success rate. The formulas in Equations 5 and 6 are given [13].

$$Accuracy = \frac{TP+TN}{P+N} x 100\% \quad (5)$$

$$Error - rate = \frac{FP+FN}{P+N} x 100\% \quad (6)$$

TP (*True Positive*) = The quantity of correctly categorized data (*Actual class (yes), Predicted class (yes)*).

TN (*True Negative*) = The quantity of correctly categorized data (*Actual class (no), Predicted class (no)*).

FN (*False Negative*) = The quantity of incorrectly categorized data (*Actual class (yes), Predicted class (no)*).

FP (*False Positive*) = The quantity of incorrectly categorized data (*Actual class (no), Predicted class (yes)*).

P = Total of TP and FN

N = Total of FP and TN

The comparative study between MK-NN and KNN was performed employing confusion matrix testing. The evaluation emphasized on the accuracy and error rate values in order to check the accomplishment

of M-KNN. Finally, the MK-NN prediction procedures mechanism was then covered by system development that adopted PHP programming languages and MySQL for the database. The MK-NN prediction system was designed as simulation tools for mining the earthquake data that can be embedded in BIM in order to provide a better analysis of situations in structural engineering activities. The architecture of the MK-NN prediction system was then explained as the outcome of this mechanism. To date, the Blackbox and UAT testing were carried out in ensuring the reliability of the MK-NN prediction system.

3. RESULTS AND DISCUSSION

As shown in Figure 2, the research flow activities were carried out by identifying previous research reviews mainly addressed in K-NN and MK-NN. The algorithm was investigated and developed as case studies for predicting the structural strength of concrete buildings against earthquakes. In the next activity, 6663 data from two-story office buildings structure in Bangkinang City, Riau Province, Indonesia, were collected and analyzed based on the SNI 03-1726-2012. A simulation tool using civil engineering software, namely SAP2000, is used to generate structural systems from earthquake data and SNI 03-1726-2012. To date, the civil engineering experts from Universitas Riau was engaged for validation.

This classification makes use of parameters input viz. the time history-time (seconds), the concrete quality ($f'c$), the displacement in point 118 (direction x, y, z), and point 124 (direction x, y, z), the velocity in point 118 (direction x, y, z), and point 124 (direction x, y, z), and acceleration in point 118 (direction x, y, z), and point 124 (direction x, y, z). Generally, the k value is determined in an odd number to avoid the appearance of similar distance values during the classification process. Herein, the values were set in 1, 3, 5, 7, 9, and 11. The cascading algorithm analysis of MK-NN was performed by providing two output classes, namely "Safe" or "Immediate Occupancy" (IO). The confusion matrix was used to measure its precision as a test of data and simulation at 10:90, 20:80, 30:70 [14]. Due to its speed, simplicity, and versatility, the above simulation was trained and tested with a high estimate of different data [14].

3.1. KDD flow process: data selection and transformation

The KDD method is followed in Figure 1. Data were cleaned, as then missing values could be eliminated. This data is conveniently complete and ready for review. The selection operation eliminated attributes, such as time history, time, and concrete quality (Fc'), which were indirectly evaluated during mining (see Figure 3). Figure 3 explained KDD's transformation process from the primary data generated by SAP2000 to execute data mining.

3.2. MK-NN process

MK-NN analysis was initiated by calculating Euclidean distance in Equations 1 and 2 between trained and tested data. The determination of S as in Equation 3 was then followed. Equation 4 has been carried out for the weight calculation. The calculation led to the exhibition of Table 1.

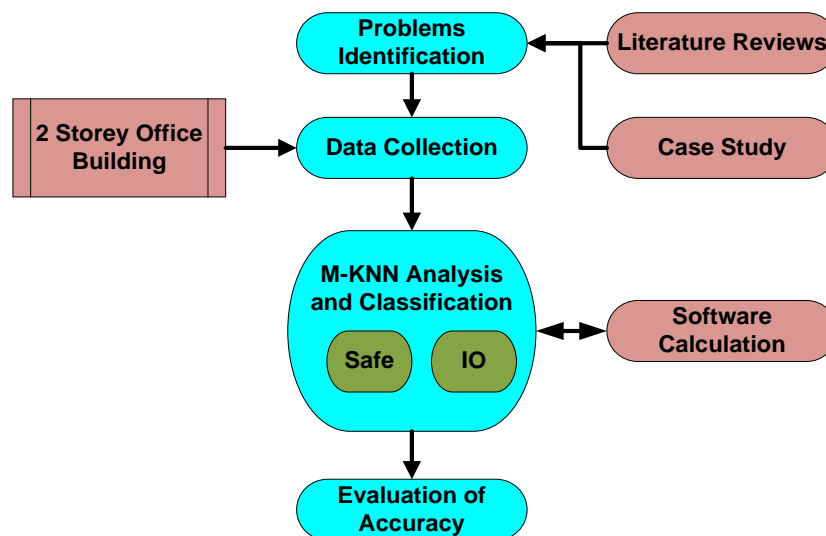
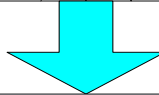


Figure 2. Research Method

No.	History Time	Time (Second)	Displacement									Velocity						Acceleration					
			Point 118			Point 124			Point 118			Point 124			Point 118			Point 124					
			Fc' (Mpa)	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z		
1	El Centro -0.024g	0.00	15	0.00001	-0.00006	-0.00017	0.00000	-0.00005	-0.00034	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
2	El Centro -0.024g	0.05	15	0.00001	-0.00006	-0.00017	0.00000	-0.00005	-0.00034	-0.00002	-0.00002	0.00000	-0.00002	-0.00002	0.00000	-0.00189	-0.00189	0.00001	-0.00188	-0.00190	-0.00001		
...		
6663	El Centro -1.00g	1.85	25	-0.04878	-0.05114	0.00325	-0.04639	-0.05473	0.00183	-0.21622	-0.14559	0.01106	-0.18588	-0.19113	0.00140	0.05919	0.66234	-0.48220	0.04281	0.67533	-0.00349		



No.	Displacement						Velocity						Acceleration						
	Point 118			Point 124			Point 118			Point 124			Point 118			Point 124			
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	
1	0.00001	-0.00006	-0.00017	0.00000	-0.00005	-0.00034	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.00001	-0.00006	-0.00017	0.00000	-0.00005	-0.00034	-0.00002	-0.00002	0.00000	-0.00002	-0.00002	0.00000	-0.00189	-0.00189	0.00001	-0.00188	-0.00190	-0.00001	
...
6663	-0.04878	-0.05114	0.00325	-0.04639	-0.05473	0.00183	-0.21622	-0.14559	0.01106	-0.18588	-0.19113	0.00140	0.05919	0.66234	-0.48220	0.04281	0.67533	-0.00349	

Figure 3. KDD Process Transformations

Table 1. Euclidean and Weight Voting Calculation

Euclidean Distance between Trained and Tested Data							
No.	de Tested 1	de Tested 2	de Tested 3	de Tested 4	de Tested 5	de Tested 6	...
1	0.070	1.964	12.441	1.426	1.265	3.275	...
2	0.235	2.203	12.141	1.585	0.986	2.985	...
3	0.070	1.964	12.441	1.426	1.265	3.275	...
...
110	1.088	2.509	11.747	1.954	1.144	2.769	...
...
5973	3.670	1.823	15.608	3.367	4.814	6.706	...
...
6663	5.458	4.054	17.809	4.807	6.609	8.633	...

Weight Voting							
No.	Wv Tested 1	Wv Tested 2	Wv Tested 3	Wv Tested 4	Wv Tested 5	Wv Tested 6	...
1	1.753	0.406	0.077	0.519	0.566	0.265	...
2	1.360	0.370	0.079	0.480	0.673	0.287	...
3	1.753	0.406	0.077	0.519	0.566	0.265	...
...
110	0.630	0.332	0.082	0.407	0.608	0.306	...
...
5973	0.240	0.430	0.062	0.259	0.188	0.139	...
...
6663	0.168	0.220	0.055	0.188	0.141	0.109	...

The dominant class was cross-checked between the actual and predicted class and the majority-weight voting class’s effectiveness. Equations 5 and 6 have been measured regarding the confusion matrix, with the class prediction accuracy. For example, Table 2 shown only nine data (1.35 percent) were found in the “False” from the prediction of 666 tested ratios 10:90.

3.3. Comparative analysis of testing results

The confusion matrix evaluated the accuracy and error rate measurement values based on Equation 5 and 6, as presented in Table 3. It explained the calculation values for simulation 90:10 with 5973 trained dan 666 tested data, for k=1. The formula determined the values of TP=651, FN=6, FP=3, and TN=6. Thus, the accuracy was figured at 98.65%, and the error-rate at 1.35%. The comparative category applied for 90:10, 80:20, and 70:30, and the value of k=1, 3, 5, 7, 9, and 11, respectively. The recapitulation appraisal has contributed to the presentation of Table 4 and Table 5. Table 4 exhibits the maximum precision found in k=1 (98.85%) in either ratio 30:70 and k=11(98.30%) in even a ratio of 30:70 for the lowest sensitivity. Table 5 clarified that the lower error rate for k=1 (1.15 percent) was discovered at 30:70, and the highest error rate of 30:70 for k=11 (1.70 percent) is unveiled. This finding has shown that MK-NN is an improved predictive process with minimum accuracy at 98.30 percent.

Table 6 and 7 construed the testing result of the confusion matrix for KNN. The highest accuracy initiated at k=9 and ratio 30:70 (97.83%). Meanwhile, the lowest error rate got going on 1.84% with a similar k value and ratio. To support this finding, Okfalisa et al. [42] have successfully conducted a comparative study between KNN and M-KNN for classifying 7395 records data of the conditional cash transfer implementation unit. The testing result using the confusion matrix identified the highest accuracy of KNN achieved into 94.95%. Meanwhile, the highest accuracy of M-KNN brought about 99.51%. Hamid et al. [31] experimented that the M-KNN method significantly outperforms the KNN method up to 3.2%, using the different choices of value k over five different datasets. The case study was arranged with two-class data set, 34 features, and 351 sample points. In a nutshell, M-KNN succeeded in proposing a very significant increase in accuracy by applying weighted KNN [16].

In contrast with the previous investigation, ANN has achieved only 95% accuracy with comparable primary data resources [43]. In a nutshell, the reliable prediction for the earthquake cases significantly impends above the shape of the loading vector used in the pushover analysis for such buildings condition such as numbers of story, PGA values, earthquake region, earthquake surrounding environment, and others earthquake parameters [26] [27]. Meanwhile, machine learning reconnoiters algorithm prediction accuracy related to different training values for each attribute. Consequently, the algorithms can be easily updated and improved [27]. Herein, this study showed how M-KNN reveals the likelihood values that could be enhancing the accuracy of reinforcement concrete buildings prediction, subjected to the earthquakes [31][37].

Table 2. Classes Prediction

No.	Actual Class	Predicted Class	Result
1	SAFE	SAFE	True
2	SAFE	SAFE	True
3	IO	IO	True
4	IO	SAFE	False
5	SAFE	SAFE	True
6	SAFE	SAFE	True
...
113	IO	SAFE	False
114	SAFE	SAFE	True
...
223	IO	SAFE	False
224	SAFE	SAFE	True
...
663	SAFE	SAFE	True
664	SAFE	SAFE	True
665	SAFE	SAFE	True
666	SAFE	SAFE	True

Table 3. Confusion Matrix Evaluation for simulation 90:10

Actual	Prediction		Total
	SAFE	IO	
SAFE	651	7	658
IO	2	6	8
Total	653	13	666

Table 4. M-KNN Confusion Matrix Evaluation for Accuracy

Neighbors	10: 90 Ratio	20: 80 Ratio	30: 70 Ratio
K=1	98,65%	98,80%	98,85%
K=3	98,65%	98,65%	98,65%
K=5	98,50%	98,50%	98,70%
K=7	98,50%	98,50%	98,60%
K=9	98,50%	98,57%	98,50%
K=11	98,35%	98,42%	98,30%

Table 5. M-KNN Confusion Matrix Evaluation for Error rate

Neighbors	10: 90 Ratio	20: 80 Ratio	30: 70 Ratio
K=1	1,35%	1,20%	1,15%
K=3	1,35%	1,35%	1,35%
K=5	1,50%	1,50%	1,30%
K=7	1,50%	1,50%	1,40%
K=9	1,50%	1,43%	1,50%
K=11	1,65%	1,58%	1,70%

Table 6. KNN Confusion Matrix Evaluation for Accuracy

Neighbors	10: 90 Ratio	20: 80 Ratio	30: 70 Ratio
K=1	33,62%	17,11%	10,47%
K=3	55,92%	36,02%	24,64%
K=5	66,43%	56,09%	47,20%
K=7	81,80%	72,98%	66,37%
K=9	96,85%	96,85%	97,83%
K=11	97,48%	97,68%	97,68%

Table 7. KNN Confusion Matrix Evaluation for Error rate

Neighbors	10: 90 Ratio	20: 80 Ratio	30: 70 Ratio
K=1	66,38%	82,89%	89,53%
K=3	44,08%	63,98%	75,36%
K=5	33,57%	43,91%	52,80%
K=7	18,20%	27,02%	33,63%
K=9	3,15%	3,15%	1,84%
K=11	2,52%	2,32%	2,17%

3.4. Method testing through software MK-NN prediction

The software application was designed developed by following detailed analysis. It begins with the KDD analysis, MK-NN calculation, and confusion matrix assessment phase (see Figure 4). The application can be used to measure and present the knowledge analysis in the data mining platform automatically. The end-user takes in the generated earthquake data and is then processed by following the KDD steps activities inclusive of Euclidean and weighted voting calculation. The MK-NN algorithm will inaugurate the readily used data and classify it as a parameter description. The MK-NN system interface displays the classification calculation results together with the explanation of accuracy and error rate level. This MK-NN prediction system is adequate to accomplish 6663 data subjected to this case study. However, the MK-NN prediction system was constructed dynamically to enclose the various data and earthquake parameters. Figure 5 demonstrated the example of the MK-NN prediction system interface for Euclidean calculation.

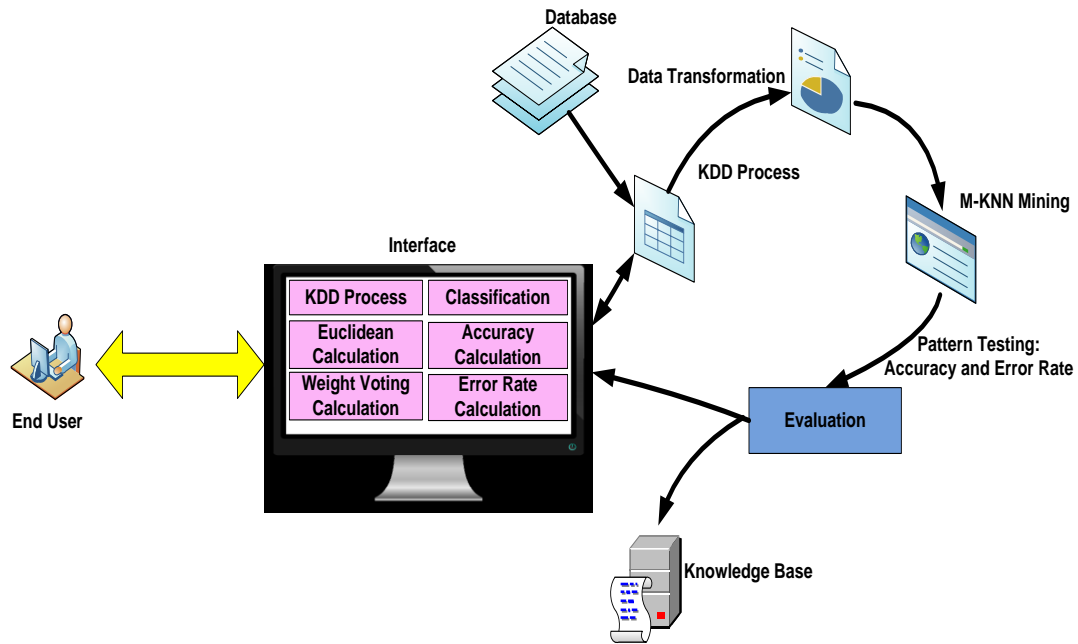


Figure 4. Software Architecture of MK-NN Prediction

The screenshot shows the 'Euclidean calculation' interface. It features a sidebar with navigation options: Home, Data, Evaluation (highlighted), Exit, and Main Page. The main area displays a table with 18 columns labeled 'Trained 1' through 'Trained 18' and 18 rows labeled 'Trained 1' through 'Trained 18'. Each cell in the table contains a numerical value representing the Euclidean distance between two samples. The diagonal elements are all 0, indicating zero distance between a sample and itself.

	Trained 1	Trained 2	Trained 3	Trained 4	Trained 5	Trained 6	Trained 7	Trained 8	Trained 9	Trained 10	Trained 11	Trained 12	Trained 13	Trained 14	Trained 15	Trained 16	Trained 17	Trained 18
Trained 1	0	0.30437	0	3.13141	0.02908	8.24272	0.19366	57.31571	0.00024	3.7332	0.00023	0.29317	0.0041	3.01445	0.00023	8.31548	1.06-6	12.37776
Trained 2	0.30437	0	0.30437	2.82704	0.30549	8.54479	0.35987	57.2425	0.30437	3.4605	0.30437	0.0614	0.30437	2.717	0.30437	8.61753	0.30437	12.07904
Trained 3	0	0.30437	0	3.13141	0.02908	8.24272	0.19366	57.31571	0.00024	3.7332	0.00023	0.29317	0.0041	3.01445	0.00023	8.31548	1.06-6	12.37776
Trained 4	3.13141	2.82704	3.13141	0	3.13127	11.35634	3.13633	56.63628	3.13141	1.61454	3.13141	2.84885	3.13138	0.63287	3.13141	11.4289	3.13141	9.32222
Trained 5	0.02908	0.30549	0.02908	3.13127	0	8.24277	0.17794	57.31596	0.02908	3.72844	0.02908	0.29426	0.02557	3.01424	0.02908	8.31507	0.02908	12.37774
Trained 6	8.24272	8.54479	8.24272	11.35634	8.24277	0	8.24499	59.80358	8.24272	11.75227	8.24272	8.5344	8.24272	11.24552	8.24272	0.4432	8.24272	20.58848
Trained 7	0.19366	0.35987	0.19366	3.13633	0.17794	8.24499	0	57.31871	0.19366	3.68661	0.19366	0.35089	0.19284	3.02011	0.19366	8.31249	0.19366	12.37973
Trained 8	57.31571	57.2425	57.31571	56.63628	57.31596	59.80358	57.31871	0	57.31571	57.7187	57.31571	57.24227	57.31572	56.6286	57.31571	59.93661	57.31571	54.80028
Trained 9	0.00024	0.30437	0.00024	3.13141	0.02908	8.24272	0.19366	57.31571	0	3.7332	1.06-6	0.29317	0.0041	3.01445	1.06-6	8.31548	0.00023	12.37776
Trained 10	3.7332	3.4605	3.7332	1.61454	3.72844	11.75227	3.68661	57.7187	3.7332	0	3.7332	3.46855	3.73286	1.59465	3.7332	11.79752	3.7332	9.21838

Figure 5. The MK-NN Prediction Interface for Euclidean calculation

The equivalence class dividing technique was performed in BlackBox testing. The Blackbox verified three primary interfaces in conjunction with the login page, data management page, and evaluation page. The testing reveals that the entire functions in the application can run as well as expected. Via the distribution of questionnaires to 30 interviewees in the civil engineering field, the user acceptance test (UAT) was carried. The questionnaire was designed in five Linkert scales from strongly agreed and strongly not agreed to ask the interface, contribution, and system functionality. As a result, eighty-seven percent of respondents strongly agreed that this application helped them employ the MK-NN method to predict building structure concrete, and eighty-eight percent of respondents on the device interface were extremely user friendly.

4. CONCLUSION

This study was successfully applied to the MK-NN approach for preventing earthquakes from the quality of concrete structures. Three significant parameters, meaning displacement, velocity, and acceleration of data set buildings, were analyzed and rated at 98.85 percent and 1.15 percent, respectively, with supreme precision and error rates. The confusion matrix calculation indicated that the comparative ratio of 30:70 in k=1 of the dataset revealed promising results in prediction. A comparative analysis revealed that MK-NN could improve the accuracy and error rate of KNN up to 1.02% and 0.69%, respectively. A prototype application has been successfully developed and tested to measure the enormous size of data using MK-NN. This tool gave the prediction of the concrete construction structure. It may thus play an essential role in the preparation and

study of construction. The employment of data mining techniques in this knowledge base offers significant analytical value for future inclusion in the BIM platform.

ACKNOWLEDGMENTS

The authors acknowledged the generous assistance provided to the Faculty of Science and Technology of the University of Islam Negeri Sultan Syarif Kasim Riau and Dr. Reni Suryanita of the Department of Civil Engineering Riau University. They contribute significantly to the feasibility and efficacy of this study. Besides, the authors also would like to thank the Ministry of Higher Education Malaysia for the research collaboration and sponsoring us with Fundamental Research Grant Scheme (FRGS) project ID number 19-037-0645, International Islamic University Malaysia, and Islamic Science University of Malaysia.

REFERENCES

- [1] McCuen, Tamera., Suermann, Patrick C., Krogulecki, and Matthew J, "Evaluating award-winning BIM projects using the Nation Building Information Model standard capability maturity model," *Journal of Management in Engineering ASCE*, vol. 28, no. 2, pp. 224-230, 2012.
- [2] Migilinskas, Darius., Popov, Vladimir., Juocevicius, Virgaudas. Ustinovichius, and Leonas, "The benefits, obstacles, and problems of practical BIM implementation," *Procedia Engineering*, vol. 57, pp. 767-774, 2013.
- [3] Azhar, Salman, "Building information modeling (BIM): trends, benefits, risks, and challenges for the AEC industry," *Leadership and Management in Engineering ASCE*, vol. 11, no. 3, pp. 241-252, 2011.
- [4] Ikerd, and William F, "The importance of BIM in structural engineering," *Structure Magazine*, pp. 37-40, 2008.
- [5] Shim, C.S., Yun, N.R., and Song, HH, "Application of 3D bridge information modeling to design and construction of bridges," *Procedia Engineering*, vol. 14, pp. 95-99, 2011.
- [6] Bynum, Patrick, Issa, Raja R. A., Olbina, and Svetlana, "Building information modeling in support of sustainable design and construction," *Journal of Construction Engineering and Management ASCE*, vol. 139, no 1, pp. 24-34, 2013.
- [7] Badiru Yunusa Y, Kherun N A, and Muhammad Rashid E, "Building Information Modeling (BIM): A Potential for effective Building Industry Practice in Malaysia," *Jurnal Teknologi (Sciences & Engineering)*, vol. 77, no. 15, pp. 55-61, 2015.
- [8] M. Bilal, L.O. Oyedele, O.O. Akinade, S.O. Ajayi, H.A. Alaka, H.A. Owolabi, J. Qadir, M. Pasha, and S.A. Bello, "Big data architecture for construction waste analytics (CWA): A conceptual framework," *Journal of Building Engineering*, vol. 6, pp. 144-156, 2016.
- [9] Peng Y, Lin J-R, Zhang J-P, and Hu Z-Z, "A hybrid data mining approach on BIM-based building operation and maintenance," *Building and Environment*, vol. 126, pp. 483-495, 2017.
- [10] J. Han, M. Kamber, J. Pei, "Data Mining: Concepts and Techniques", *Morgan Kaufmann Publishers Inc.*, 2011.
- [11] Lin, J.R., ZZ. Hu, J.P. Zhang, and FQ. Yu, "A Natural Language Based Approach to Intelligent Data Retrieval and Representation for Cloud BIM," *Computer-Aided Civil and Infrastructure Engineering*, vol. 31, no. 1, pp. 18-33, 2015.
- [12] Yu, Zhun., Fung, B. C. M., and Haghghat, F, "Extracting knowledge from building-related data — A data mining framework," *Building Simulation*, vol. 6, no. 2, pp. 207–222, 2013.
- [13] Xiao, Fu., and Fan, Cheng, "Data mining in building automation system for improving building operational performance," *Energy and Buildings*, vol. 75, pp. 109–111, 2014.
- [14] Hung-Ming Chen, Kai-Chuan Chang, and Tsung-His Lin, "A Cloud based system framework for performing online viewing, storage, and analysis on big data of massive BIMs," *Automation in Construction*, vol. 71, pp. 34-48, 2016.
- [15] Costa, A., Keane, M. M., Torrens, J. I., and Corry, E, "Building operation and energy performance: Monitoring, analysis and optimization toolkit," *Applied Energy*, vol. 101, pp. 310–316, 2013.
- [16] Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Applied Energy*, vol. 87, no. 3, pp. 925-933, 2010.
- [17] A. Ahmed, M. Otreba, N.E. Korres, H. Elhadi, and K. Menzel, "Assessing the performance of naturally day-lit buildings using data mining," *Advanced Engineering Informatics*, vol. 25, no. 2, pp. 364-379, 2011.
- [18] Abd-elhamed, A., Shaban, Y., and Mahmoud, S, "Predicting Dynamic Response of Structures under Earthquake Loads Using Logical Analysis of Data," *Journal of Building Engineering*, vol. 8, no. 61, pp. 1–12, 2018.
- [19] Sengara, I. W., Sidhi, I. D., Mulia, A., Asrurifak, M., and Hutabarat, D, "Development of Risk Coefficient for Input to New Indonesian Seismic Building Codes," *Journal of Engineering and Technological Sciences*, vol. 48, no. 1, pp. 49–65, 2016.
- [20] Luna Ngerjaratan,P.K and Nagesh R. Iyer, "A critical review on earthquake resistant design provisions of SNI 03-1726-2002 of Indonesia and IS 1893 (part 1)-2002 of India," *Journal of Structural Engineering*, vol. 38, no. 3, pp. 285-296, 2011.
- [21] S. Chopra and K. J. Marfurt, "Seismic attribute expression of differential compaction," *Lead Edge*, vol. 31, no. 12, pp. 1418–1422, 2012.
- [22] Indonesian Standard Code, *Earthquake Resistance Design for Buildings (SNI 03-1726-2002)*, National Standardization Agency, 2002.

- [23] Indonesian Standard Code, *Earthquake Resistance Design for Buildings (SNI 03-1726-2012)*, National Standardization Agency, 2012.
- [24] Maram, M.P., Rao, K.R.M., and Poursalehi, A, "An artificial Neural Network for Prediction of Seismic Behavior in RC Buildings with and Without Infill Walls," *Int. J. Mod. Eng. Res. 1*, pp. 3071–3078, 2013.
- [25] Norhisham Bakhary, "Statistical Vibration Based Damage Identification Using Artificial Neural Network," *Jurnal Teknologi*, vol. 52, no.1, pp. 49-60, 2010.
- [26] Karbassi, A., B. Mohebi, S. Rezaee, and P. Lestuzzi, "Damage prediction for regular reinforced concrete buildings using the decision tree algorithm," *Comput. Struct. Jan.*, Vol. 130, pp. 46–56, 2014.
- [27] Zhang, Y., H. V. Burton, H. Sun, and M. Shokrabadi, "A machine learning framework for assessing post-earthquake structural safety," *Struct. Saf. May.* Vol. 72, pp. 1–16, 2018.
- [28] Pei, J.-S., and Smyth, A.W, "New approach to designing multilayer feedforward neural network architecture for modeling nonlinear restoring forces. II: Applications," *J. Eng. Mech.*, vol. 132, pp. 1301–1312, 2006.
- [29] Badhiye, S. S., Sambhe, N. U., and Chatur, P, "KNN Technique for Analysis and Prediction of Temperature and Humidity Data," *International Journal of Computer Application*, vol. 61, pp. 7–13, 2013.
- [30] Max and Kjell, *Applied Predictive Modeling*, Springer, 2018.
- [31] Hamid P, *et al.*, "MKNN: Modified K-Nearest Neighbor," *Proceeding of the World Congress on Engineering and Computer Science*, vol. 1, 2008.
- [32] Maryam Kuhkan, "A Method to Improve the Accuracy of K-Nearest Neighbor Algorithm," *International Journal of Computer Engineering and Information Technology*, Vol. 9, No. 6, 2016.
- [33] Okfalisa, R. Fitriani, and Y. Vitriani, "The Comparison of Linear Regression Method and K-Nearest Neighbors in Scholarship Recipient," *ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 194–199, 2018.
- [34] Jadhav, S. D., and Channe, H. P, "Comparative Study of K-NN , Naive Bayes and Decision Tree C lassification Techniques," *International Journal of Science and Research*, vol. 5, pp. 2014–2017, 2016.
- [35] Tesfamariam, S., and Liu, Z, "Earthquake induced damage classification for reinforced concrete buildings," *Structural Safety*, vol. 32, no. 2, pp. 154–164, 2010.
- [36] H. Parvin, H. Alizadeh, and B. Minaei, "A Modification on K-Nearest Neighbor Classifier," *Glob. J. Comput. Sci. Technol. Jan.*, vol. 10, 2010.
- [37] M. Abrar, A. Tze Hiang Sim, D. Shah, S. Khusro, and Abdusalam, "Weather Prediction using Classification," *Sci. Int. Dec.* vol. 26, pp. 2217–2223, 2014.
- [38] F. Gullo, "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do," *Phys. Procedia*, vol. 62, pp. 18–22, 2015.
- [39] M.Z.H. Jesmeen, J.Hossen, S. Sayeed, C.K.Ho, K. Tawsif, Armanur Rahman, E.M.H.Arif, "A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics," *Indonesian Journal of Electrical Engineering and Computer Science*. June.vol.10, no.3, pp.1234-1243.2018.
- [40] M.Swami D, A.Govardhan, and D.Vijaya L, "Classification of web services using data mining algorithms and improved learning model," *Telekomnika*, vol. 17, no. 6, 2019.
- [41] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *Int. J. Emerg. Technol. Adv. Eng.* vol. 3, no. 6, pp. 354–360, 2013.
- [42] Okfalisa, Mustakim, Ikbal Gazalba, Nurul Gayatri Indah Reza, "Comparative analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor algorithm for data classification," *International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. Nov, Yogyakarta, pp. 294-298, 2017.
- [43] R. Suryanita, H. Maizir, E. Yuniarto, M. Zulfakar, and H. Jingga, "Damage Level Prediction of Reinforced Concrete Building Based on Earthquake Time History Using Artificial Neural Network," *MATEC Web Conf.* Dec, vol. 138, pp. 02024, 2017.

BIOGRAPHY OF AUTHORS



Dr. Okfalisa, ST, MSc. completed her Ph.D. in Information System and Computer Science at Universiti Teknologi Malaysia in 2012. Her research interests are performance measurement, strategy execution, management information system, knowledge management, knowledge management system, decision support system, data science, machine learning, and data mining. She has been involved in several research projects and publication, national (cooperation with UIN Jakarta; UIN Bandung; UIN Malang; UGM) and international (cooperation with Universiti Teknologi Malaysia-Johor Bahru Malaysia; Price of Songkla University-Hat Yai; Universiti Teknologi MARA-Perlis, Malaysia; International Islamic University Malaysia (IIUM)-Kuala Lumpur, Malaysia, University De Haute-Alsace-France; Ostbayerische Technische Hochschule Amberg-Weiden (OTH-AW)-German; University of Antwerp-Belgium; and Sheffield Hallam University) from UIN Suska Riau and Indonesia Islamic Higher Education Ministry funds. Some of the locals and international journals have been reviewed by her. Two silvers and one bronze medals are awarded by her during the paper presentation and exhibition at 2nd International Indonesia, Malaysia, Thailand Symposium on Innovation and Creativity 2018. Currently, she is Associate Professors in Informatics Engineering Department in UIN Suska Riau and licensee as Software Development Assessors in Badan Nasional Sertifikasi Profesi (BNSP). Okfalisa is a life

member of Association Higher Education Informatics and Computers (Aptikom) and Institute of Electrical and Electronics Engineers (IEEE).



Septian Nugraha, ST completed his undergraduate study at Informatics Engineering Department, Universitas Islam Negeri Sultan Syarif Kasim Riau in 2018 with excellence achievement. His expertise in machine learning and data mining led him to collaborate with the University of Abdurrab in developing an academic information system (Since 2016-2018) and data mining project with Dr. Okfalisa. Currently, he works in one of the private companies in Bangkinang, Riau, and engaged as manager of development and services.



Ts. Dr. Zahidah Zulkifli started her career as an IT executive in Ambank (M) Berhad in 2007 after graduated her Master degree (IT Management) in 2007. She then being awarded a Fellowship Programme under SKIM LATIHAN AKADEMIK BUMIPUTRA (SLAB), Ministry of Higher Education (MOHE), 2008 to further her study in PhD. Starting her PhD in 2009 and finished in 2013. Her PhD research was selected as one of 5 recipients in Malaysia for MIMOS Prestigious Award 2013. She also won 2 medals for her PhD research (bronze & silver) in IRIIE 2014 organized by IIUM. Several journal papers for her PhD research were accepted to publish and 2 of them were indexed by ISI. One of the papers was selected as the Department Best Indexed Journal Article in IRIIE 2015. She returned to KICT in April 2013 an appointed as an assistant professor. She started teaching System Analysis and Design followed by Management Information System. She is a member of Malaysian National Computer Confederation (MNCC), Membership, since 2013, (National) IEEE, Membership, since 2013, (International) and International Association of Computer Science and Information Technology (IACSIT), Member, 2013, (International). Dr Zahidah secured several numbers of grants as principal investigator and co-investigator. Starting with UMRG (University Malaya Research Grant) for her PhD research. After graduating, she joined TRGS and Myra Grant as coinvestigator and the latest in 2014, she managed to secure a grant (RAGS) and became the principal investigator. She is now doing research under Semantic Body of Knowledge and Technology, and Information Systems, Internet, & Governance Research unit. Her area of research interest covers digital library, research methodology and design, online database system, system technology and development and online education. Presently Dr. Zahidah is an Assistant Professor at the Department of information system at the International Islamic University Malaysia (IIUM) since April 2013.



Prof. Dr. Sc. Saktioto, S.Si M.Phil CPhys, M.InstP, is an alumnus of the University of Riau who completed his bachelor's degree in Physics at the Department of Physics FMIPA 1993, He pursued to Master Degree at the University of Manchester Institute of Science and Technology (UMIST) in 2000 in Plasma Physics, and Doctoral Degree at Universiti Teknologi Malaysia in 2009 in Photonic Physics. Besides a lecturer at the University of Riau, he is also heavily involved in teaching and serving in the institutions of the Riau Province and National Government as a consultant at the Bappeda of Riau Province since 2003 as well as at District Education Offices in Riau. At the National and International level, he was involved as the OSN Team, Asian Physics Olympiad and also the Educational Expert Team under the Ministry of National Defense. He is a Physicist and his research is carried out well at the National, Regional and International levels. In the past 10 years, he has owned more than 87 scientific articles recorded at SCOPUS, as the Chair of Nanophotonics Research on the International Photonic Joint Team from KMITL Universities in Thailand, Optiwave Companies in Canada, NUS in Singapore, Abex Companies and Universities in Malaysia, Universities in Yemen, NASA, America and Universities in England. He is also an active member of the International Professional agency in IEEE, OSA, APS, IOP, WASET, HFI, SPIE. Prof. Saktioto is also active as a reviewer/editorial board of more than 6 international journals in Poland, Malaysia, China, Croatia, Bulgaria, and Spain. 4 books have been published in Germany and Croatia and 3 books in Indonesia. He has won many international awards. In 2015, he was awarded as the first rank UNRI lecturer and became a national finalist lecturer finalist. In 2016, he returned as an outstanding lecturer at UNRI and in 2017 he established research as the head of Optoelectronic research with Ton Duc Thang University in Vietnam, until 2018. Until now he is a Chair of the Indonesian Physics Association in Pekanbaru and he leads the establishment of a Global Positioning System (GPS) research with the Institute National Aeronautics and Space (LAPAN) at the Photonic Laboratory.



Associate Professor Ts. Dr. Shukor Sanim Mohd Fauzi is a Deputy Rector (Research and Industrial Linkage) at UiTM Perlis Branch, and also a faculty member of Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis Branch, Malaysia. Previously, he has held Visiting Fellow position at Universitas Airlangga (2018). He received Diploma in Computer Science and Bachelor Science (Hons) specializing in Information Systems Engineering from the Universiti Teknologi MARA, and Master of Science (Computer Science - Real Time Software Engineering) from the Centre for Advanced Software Engineering, Universiti Teknologi Malaysia. He then obtained his PhD in Software Engineering from the University of New South Wales (UNSW), Australia and in the working group of Data61 (formerly known as Software Systems Research Group), Commonwealth Scientific and Industrial Research Organisation (CSIRO) which is based at the National ICT Research Centre Australia (NICTA), Australian Technology Park (ATP). His area of specialization includes software engineering, empirical software engineering, mining software repository, social network analysis, socio-technical congruence, computer supportive collaborative work, and software process. He has received more than 16 research grants (internal and external grants) totaling over RM1,000,000. He has published more than 100 scientific publications and has won more than 50 medals at national and international exhibition. He also is the editor for 5 books including “Software Process Improvement and Management: Approaches and Tools for Practical Development”.