Indonesian Journal of Electrical Engineering and Computer Science
Vol. 27, No. 1, July 2022, pp. 555~562
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v27.i1.pp555-562

Improving the efficiency of machine learning models for predicting blood glucose levels and diabetes risk

Kriengsak Yothapakdee[1], Sarawoot Charoenkhum[2], Tanunchai Boonnuk[3]
[1]Department of Computer Science, Faculty of Science and Technology, Loei Rajabhat University, Loei, Thailand
[2]Hospital Director of Khok-Nong-Kae, Sub-district Health Promoting Hospital of Wangsaphung District, Loei, Thailand
[3]Department of Public Health, Faculty of Science and Technology, Loei Rajabhat University, Loei, Thailand

Article Info

Article history:

Received Feb 16, 2022
Revised May 27, 2022
Accepted Jun 8, 2022

Keywords:

Blood glucose levels
Data mining
Diabetes
Health promoting hospital
Machine learning
Prediction
Scikit-learn

ABSTRACT

Fasting blood glucose is used as an indicator in the process of predicting diabetes risk. This research aims to, i) create a model for predicting blood glucose level using data mining algorithms, ii) a selection algorithm was used to select a feature from the correlation of the data, and iii) to compare the model's performance with the classical methods. All clinical data ware recorded and compiled in a database by hospital staff from 2014-2019. In our previous research, the blood glucose prediction model had an acceptable accuracy where 18 patient features were used as input data to the data mining process. In this research, we demonstrated that the random forest classifier and extra tree classifier algorithms have an outstanding in discarding non-critical attributes. And the process of reducing the number of those features has impacted the glycemic prediction model with higher efficiency. Seventeen machine learning algorithms are used to find the best performance models. Our results clearly show that the improved prediction model is more efficient. This experiment has shown that improvements to our proposed model were able to predict blood glucose levels with 99.69% and 99.63% accuracy for random forest classifier, extra tree classifier, and Gaussian process classifier, respectively.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Kriengsak Yothapakdee
Department of Applied Science, Faculty of Science and Technology, Loei Rajabhat University
Loei 42000, Thailand
Email: kriengsak@lru.ac.th

1. INTRODUCTION

The various difficulties and limitations of rural people's access to modern information systems have been demonstrated from the past to the present [1]. The credibility of obtaining assistance from government welfare remains an issue that remains to be addressed. One of those problems, public health issues such as primary risk analysis and health monitoring systems, and have emerged as the top issues of national governments in finding solutions. In particular, a system or tool for early disease risk screening processes that would save people from having to move into the capital to diagnose the disease is severely scarce in developing countries [2]. As for this health issue, the World Health Organization (WHO) report provides very important information about the causes of death for people worldwide. In the past, 63% of the population died from non-communicable disease and more than 80% were citizens of developing countries [3]. A non-communicable disease (NCD) is a disease that is not transmissible directly from one person to another. The top five NCDs that are prevalent worldwide and the WHO are monitoring are heart disease, stroke, cancer, diabetes and chronic lung disease [4]. Diabetes is a metabolic disorder and has become one of the top 5 causes of death in the world population. According to a 2014 WHO report, there are 422 million people with diabetes worldwide and an expected increase of 629 million by 2045 [5]. In Thailand, the situation and trends of people with diabetes have continued to increase. The Ministry of Public Health has reported that over the past decade, more than 75 percent of all Thai deaths from NCD have died, or about 0.32 million people per year [6]. And the latest report, the population of Thailand over 4.8 million adults suffer from diabetes [7]. Loei Province is one of the 77 provinces of Thailand, located in the northeast of the Mekong River and on the border of the Lao people's democratic republic. It is approximately 520 kilometers from Bangkok, covering an area of approximately 11,424 square kilometers. According to the division of non-communicable diseases data report [8], the population of Loei province was diagnosed with diabetes at an increased rate from 11,406 to 12,366 in 2016 and 2017, respectively. Over the years, diabetes risk assessment has been used as a guideline for the prevention and early screening of diabetes risk levels. Today, many public health organizations have applied risk assessment processes to screen groups of people with diabetes. The results of the risk assessment enabled the volunteers to understand the trends in future diabetes risk. In addition, the results of the risk assessment were able to detect asymptomatic diabetic patients and be able to treat them at an early stage.

The challenge of evaluating diabetes risk is the effectiveness of predicting the tendency or likelihood of developing diabetes in the future. For the above reasons, we are understood and realized their importance. The tools in Scikit-learn library will be built into predictive models of diabetes risk by researchers. In this research, the data were used as clinical data of 103,492 cases of patients receiving services from sub-district health promoting hospitals in Loei province. All data are diagnostic and collected by the process and the model of the sub-district health promotion hospital by the staff and authority of the hospital.

2. RELATED WORKS

In the past, the development of an information system for assessment and screening to predict the level of early diabetes risk has been continually improved and revised. Of course, one algorithm might be able to work or solve some problems very well. But this algorithm can be much less efficient in different environments and data formats. Therefore, there are still many opportunities to develop and improve the algorithm to be the most suitable [9]. In this research, the theories and principles involved in the development of a predictive model for screening for diabetes risk levels include:

2.1. Machine learning

Machine learning (ML) is the process by which machines (computers) can process (learn) a set of instructions and execute them on their own. The machine learning process is an attempt to enable machines to learn and understand various interrelated patterns of information. Then, when a new set of data is introduced into the system again, the machine can make predictions or make decisions based on their learning. Researcher and interested people in the field of data manipulation, known as "machine learning", were first proposed in 1959 [10]. Machine learning aims to use the special characteristics of data collected from observations or experiments in the system to automatically model, predict or explain system behavior, or to obtain derive decision rules to interact suitable for the system. ML approaches [11] can be categorized into three basic categories: supervised learning, unsupervised learning, and reinforcement learning. In particular, machine learning applied to diagnostic tasks was presented Fatima and Pasha [12] in six methods: Supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, evolutionary learning, and deep learning. Over the past decade, information systems for predicting diabetes risk have been continually developed and improved. For example, Sowjanya *et al.* [13] developed MobDBTest applications using new machine learning techniques for predicting diabetes for users. Yuvaraj and SriPreethaa [14] presented health care monitoring systems using machine learning algorithms applied in hadoop based clusters to predict diabetes risk levels. Later in 2020, an intelligent architecture for monitoring diabetic patients by using machine learning algorithms was introduced by Rghioui *et al.* [15]. This architecture consists of smart devices, sensors and smartphones in order to collect measurements from different parts of the body. Clinical data obtained from patients are categorized using machine learning to participate in the process of diagnosis. The simulation results showed that the algorithm was optimized accordingly. Most recently in our research Charoenkun *et al*. [16], a large number of patient data features such as sex, age, diastolic blood pressure, systolic blood pressure, body weight, heart rate, pulse, temperature, height, body mass index, waist, smoking, and drinking, were used for the process of machine learning in order to create a model for predicting blood sugar levels. The results of studies and developments have been demonstrated that the effectiveness of the glycemic prediction model is within the relatively high and acceptable range.

2.2. Data mining

Data mining (DM) and knowledge discovery in database (KDD) are highly relevant. Because the process of acquiring knowledge and understanding on a large database, known as knowledge discovery in database, has a data mining process under operation. Both knowledge discovery in database and data mining are among the fields of computer science that are gaining greater attention today. The principles of the implementation of the data mining model consist of two approaches [17], [18]: Description data mining and predictive data mining. i) Description data mining is the process of finding patterns for describing available data based on relevance or coherence of data. To implement the results obtained in the decision-making process, there are several methods of action such as association rule discovery, and sequence pattern discovery, clustering. ii) Predictive data mining is a method of predicting future events based on a process of learning from those historical data or situations. The results (experiences) obtained from this learning process are then used to construct a model to predict data with similar patterns. The most popular methods such as classification and prediction. The application of predictive data mining techniques, for example, Islam *et al.* [19] has presented a tool for predicting the likelihood of early diabetes risk using symptomatic diagnostics and techniques of data mining. Likewise, a study by Yang *et al.* [20] presented a computational design for predicting diabetes risk based on a combination of data from physical measurements. And developed a prediction model using xtreme gradient boosting (XGBoost). Khatun *et al*. [21] have presented the concept of how crime investigation agencies should use data mining techniques to find anti-crime predictions. They used a supervised classification algorithm, decision tree, K-nearest neighbors (KNN), and random forest in their knowledge discovery process.

2.3. Scikit-learn library

Scikit-learn or Scikit-learn library is a modern open-source machine learning library on a wide variety of python modules supporting data mining techniques both supervised and unsupervised learning. This package focuses on the early learning and implementation of machine learning algorithms for students, researchers, and non-specialists using a high-level language for general purposes [22]. And it also provides a wide range of tools for model fitting, data processing, model selection and evaluation, and many other utilities [23]. Source code, binaries, and documentation can be downloaded from https://scikit-learn.org/stable/. Within these libraries, there are examples of applications and algorithms for working with enormous data. Examples of applications for performing classification of data, such as spam detection, and image recognition. An application for predicting a continuous value attribute in connection with an object or target of interest, such as stock prices, and drug response, is often referred to as the regression method. Many scikit-learn algorithms are provided to students and researchers as a tool in the process of mining data for very large databases. For scikit-learn applications in the field of prediction, for example Marcelino *et al*. [24] used a Python machine learning library, scikit-learn, to predict asphalt pavement friction. Their friction prediction machine learning model was built using two algorithms: linear regression and regularized regression with lasso.

3. MATERIALS AND METHODS

3.1. Proposed method

The format of this study is as show in Figure 1. We present the four steps of the conceptual process in this study. Clinical data of 103,492 cases from a patient database at a Tambon health promoting hospital were performed through three processes of data preprocessing includes data cleaning, data integration, data selection, and data transformation. Subsequently, we used 3 algorithms to select the most important and suitable features to create a learning model for predicting blood sugar levels. After that, the clinical data were divided into two sections for training and testing data of the glycemic prediction model. In step 2, seventeen algorithms from scikit-learn library are applied and developed into a learning model for predicting blood sugar levels. And in step 3, we will evaluate the performance of the model obtained from the previous step through comparing the prediction accuracy. Then, the accuracy of those models is compiled to be proven and compared for each model to find the best model. Subsequently, the top five precision machine learning models are tuned to various parameters in order to make the model more suitable for the characteristics or patterns of the data and obtain high accuracy values. And finally, a predictive model of blood glucose levels of service participants in sub-district health promoting hospital was presented, and this model will be developed as an application for screening and assessing diabetes risk levels in the near future. In this study, the research process and methodology were validated by Loei Rajabhat University's Research Ethics Committee (LRUREC No. H 010/2564) details of the trial will be presented in a later section.

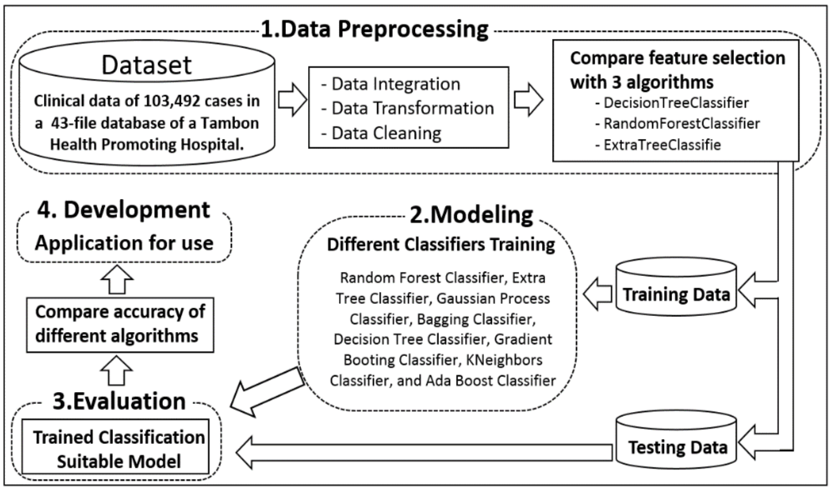

Figure 1. The process of conceptual framework

3.2. Data collection

In this study, the database was used for the development and improvement of blood glucose prediction models. We have the support of the right authority and responsibility. These datasets are clinical data that were collected and recorded as a database of service attendants in Tambon health promoting hospital in Loei province between 2014 and 2019. The dataset initially included 103,492 records of service participants with age groups in the 23-103 range. This dataset has 18 features as shown in Figure 2.

```
df = pd.read_csv('DB_diabetesRisk.csv', sep=',')
df.head()
```

vn	hn	sex	age	marry	bpd	bps	bw	hr	pulse	temp	Resp	height	bmi	waist	smoking	drinking	fbs
3140323	22020.0	2.0	47	2.0	70.0	134.0	51.0	130.0	130.0	37.0	20.0	154.0	21.504	74.0	1.0	1.0	117.0
3141029	22438.0	2.0	43	2.0	70.0	105.0	66.0	68.0	68.0	37.0	20.0	160.0	25.781	85.0	1.0	1.0	NaN
3141152	20805.0	2.0	74	3.0	85.0	139.0	75.0	78.0	78.0	37.0	20.0	162.0	28.578	82.0	1.0	1.0	143.0
3142118	21187.0	2.0	47	2.0	82.0	151.0	85.0	66.0	66.0	37.0	20.0	163.0	31.992	80.0	1.0	1.0	113.0
3143640	19915.0	2.0	57	2.0	76.0	132.0	65.0	72.0	72.0	37.0	20.0	160.0	25.391	94.0	1.0	1.0	160.0

Figure 2. Examples of patient data sets and features

3.3. Data preprocessing

At this stage, we verify the completeness and accuracy of the patient's clinical data. The results of the operation showed that some of those patient data were missing and the data types were not in the standard format. In particular, the clinical database of patients contains a number of features that are both very important and inconsistent in the process of predicting blood glucose levels. Therefore, 3 popular features selection algorithms were used to select attributes suitable for machine learning processes and to develop glycemic predictive models. Experimentally, the random forest classifier algorithm has shown outstanding performance in the process of classifying important features of patient data. The results of the selection feature very important and the ability to classify the dataset well as shown in Figure 3. In Figure 3, the most important features obtained from the random forest classifier algorithm with a score of>10 are bmi (18.40%), age (17.80%), bps (15.76%), bpd (14.63%), waist (14.19%), and hr (10.80%), respectively. Next, these features will be used in the learning process through 20 algorithms to develop a model to predict blood sugar levels in step 2 (Modeling).

3.4. Constructing models

In this process, we used seventeen algorithms in scikit-learn library to find a highly efficient and suitable algorithm to develop a model to predict blood glucose levels in sub-district health promoting hospital. Those algorithms include: Ada boost classifier, bagging classifier, decision tree classifier, extra tree classifier, Gaussian process classifier, Gaussian Naive Bayes (NB), gradient boosting classifier, KNN classifier, Linear support vector classification (SVC), logistic regression, multilayer perceptron (MLP) classifier, nearest centroid (NC), passive aggressive classifier, Perceptron, random forest classifier, ridge classifier, and stochastic gradient descent (SGD) Classifier.

3.5. Performance measurement

After the model was obtained, the performance measurements used in our experiment were the accuracy values with mean square error (MSE), R-squared (R2), and explained variance score (EVS) [25], [26]. MSE is computed using (1):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (1)$$

R-Squared (R^2) is given by (2):

$$R^2 = 1 - \left(\frac{\sum(y-\hat{y})^2}{\sum(y-\hat{y})^2}\right) \quad (2)$$

And EVS is given by (3):

$$EVS = 1 - \frac{Var(y-\hat{y})}{Var(y)} \quad (3)$$

where y represents the estimated value, $\hat{y}$ is the actual value, and n is the number of observations.

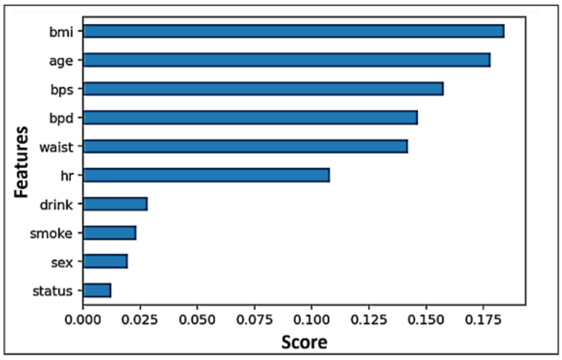

Figure 3. The graph shows the score for each feature

4. RESULTS AND DISCUSSION

The data set of this study was clinical data of 103,492 patients in sub-district health promoting hospitals. All data were collected in 2014-2019 from experts and hospital staff. We used the clinical data of those patients to study and develop a model predicting blood glucose level. In this research, the process of analyzing patient data accurately and completely in the process from the early stages is the main guideline. Then the feature selection of the clinical data of the patients to find important and appropriate features was performed. The random forest classifier algorithm has been used to find the key features that are very important by comparing the scores for each feature as shown in Figure 3. Subsequently, our results and investigations showed that six features, including body mass index (BMI), age, diastolic blood pressure (BPD), systolic blood pressure (BPS), waist, and heart rate (HR) were able to classify patients with high accuracy. Then, we modified and developed a blood glucose level prediction model from 17 algorithms in the Scikit-learn library. In the evaluation phase, we compared the efficacy of blood glucose level models using three approaches: MSE, R-Squared , and EVS. The results of comparing the predictive performance of the models with the mean square error and explained variance score showed trends in outcomes with similar precision and accuracy as shown in Figure 4.

In Figures 4 and 5, we demonstrated the very high performance of five models for predicting blood glucose levels from four algorithms: random forest regressor, extra tree classifier, Gaussian process classifier, bagging classifier and decision tree classifier. Based on previous studies and trials, we believe that these four models are the most appropriate models for predicting blood glucose levels in the clinical data of these patients.

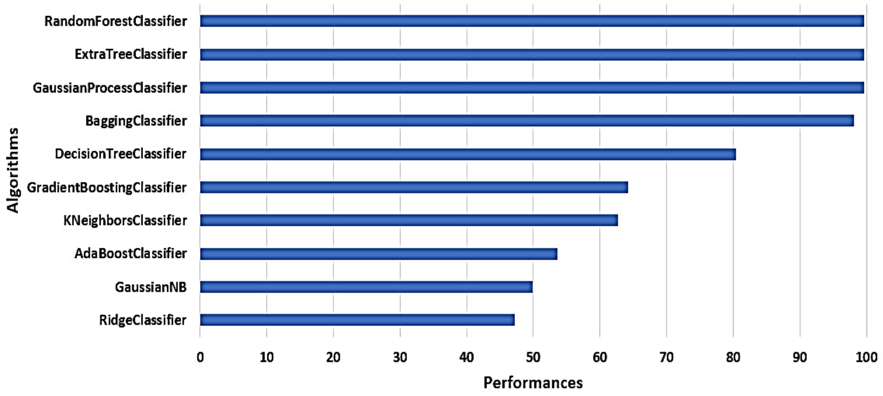

Figure 4. Performance graph (EVS) of the top ten algorithms

Algorithms	Performances
RandomForestClassifier	0.9969135802469140
ExtraTreeClassifier	0.9969135802469130
GaussianProcessClassifier	0.9963524130190790
BaggingClassifier	0.9806397306397300
DecisionTreeClassifier	0.8044332210998870
GradientBoostingClassifier	0.6425364758698090
KNeighborsClassifier	0.6271043771043770
AdaBoostClassifier	0.5356341189674520
GaussianNB	0.4983164983164980
RidgeClassifier	0.4719416386083050
NearestCentroid	0.4579124579124570
LogisticRegression	0.4523007856341180
MLPClassifier	0.4480920314253640
PassiveAggressiveClassifier	0.4447250280583610
SGDClassifier	0.3916947250280580
LinearSVC	0.3689674523007850
Perceptron	0.3672839506172830

Figure 5. Detailed accuracy of the model in each algorithm

5. CONCLUSION

Our study aims to build and improve a predictive model of blood glucose level trends of service participants in a sub-district health promotion hospital in Loei Province. A wide variety of machine learning algorithms and data mining methods have been used to modify models with accuracy and precision to suit the dataset. Therefore, the random forest classifier algorithm was used to select the very important and essential features for predicting blood glucose levels. The results of the features selection operation demonstrated that only six features were able to accurately and accurately identify the patient dataset. Thereafter, seventeen machine learning algorithms obtained from the scikit-learn library have been applied to a database of service participants in sub-district health promoting hospital. We have been trying to find and develop suitable models in order to obtain a predictive model of blood glucose trends. At the beginning of the operation, all algorithms were tested with a data mining technique in the database in order to select the models that received the accuracy within acceptable criteria. After that, the selected algorithms will be tuned to the parameters for the best benefit of further use. Our experiment showed that the top ten models with the best accuracy were the random forest classifier (99.69%), extra tree classifier (99.69%), Gaussian process classifier (99.63%), bagging classifier (98.06%), decision tree classifier (80.44%), gradient booting classifier (64.25%), KNeighbors classifier (62.71%), AdaBoost classifier (53.56%), GaussianNB (49.83%), and ridge classifier (47.19%), respectively.

Our future work; The unique characteristics of this dataset will be further studied for predicting the trend of diabetes in the local population of a health promoting hospital in Wang Saphung District, Loei province. We will study and develop models of diabetes prognosis based on this set of data through scikit-learn libraries in order to find the most suitable and accurate model. Thereafter, those models will be further studied and researched in order to be used for a preliminary investigation or analysis of service participants in the Sub-district Health Promoting Hospital in order to be used as a primary diagnostic tool for diabetes. In the end, we will use a model that is most suitable for developing applications in both web applications and mobile applications in the near future.

ACKNOWLEDGEMENTS
This work is partially supported by the Research and Development Institute of Loei Rajabhat University and the Department of Applied Science, Faculty of Science and Technology, Loei Rajabhat University, Thailand.

REFERENCES
[1] B. Sikhakhane and S. Lubbe, "Preliminaries into problems to access information – the digital divide and rural communities," *SA Journal of Information Management*, vol. 7, no. 3, Dec. 2005, doi: 10.4102/sajim.v7i3.273.
[2] A. Z. Chowdhury and K. S. Jomo, "Responding to the COVID-19 pandemic in developing countries: Lessons from selected countries of the global south," *Development (Basingstoke)*, vol. 63, no. 2–4, pp. 162–171, Dec. 2020, doi: 10.1057/s41301-020-00256-y.
[3] O. Simpson and S. G. Camorlinga, "A framework to study the emergence of non-communicable diseases," *Procedia Computer Science*, vol. 114, pp. 116–125, 2017, doi: 10.1016/j.procs.2017.09.026.
[4] World Health Organization, "NCDs | Noncommunicable diseases and their risk factors," *Word Health Organization*, 2018. https://www.who.int/ncds/en/%0Ahttp://www.who.int/ncds/en/ (accessed Dec. 26, 2019).
[5] Word Health Organization, "Classification of diabetes mellitus 2019," *Word Health Organization*, 2019. https://apps.who.int/iris/bitstream/handle/%0A10665/325182/9789241515702-eng.pdf%0A (accessed Jan. 12, 2020).
[6] Strategy and Planning Group, "Bureau of non-communicable diseases*," 5-Year National NCDs Prevention and Control Strategic Plan. Department of Disease Control, Ministry of Public Health, Thailand*. https://www.iccp-portal.org/system/files/plans/Thailand National NCD plan 2017-2021.pdf (accessed Mar. 26, 2019).
[7] P. Payuha, N. Mongkonchoo, and S. Siwina, "Health behavioral changes of patients with uncontrolled diabetes mellitus by a nutritional surveillance and resistance exercise program," *Journal of Preventive Medicine Association of Thailand*, vol. 10, no. 1, pp. 95–107, 2020.
[8] Force Interagency Task, Diseases Noncommunicable, and Oleg Chestnov, "Non Communicable Diseases Newsletter," *Department of Disease Control, Ministry of Public Health, Thailand*, 2016. https://www.who.int/news-room/fact-sheets/detail/noncommunicablediseases%0Ahttps://www.moh.gov.om/documents/272928/0/Department+of+Non+Communicable+Diseases+Newsletter+2016/31774b6d-88dc-4b85-a46c-8349e6c15f89 (accessed Feb. 25, 2019).
[9] N. Sureja, B. Chawda, and A. Vasant, "A novel salp swarm clustering algorithm for prediction of the heart diseases," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, pp. 265–272, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp265-272.
[10] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development,* vol. 44, no. 1–2, pp. 207–219, Jul. 2000, doi: 10.1147/rd.441.0206.
[11] Y. Yang and L. Wu, "Machine learning approaches to the unit commitment problem: Current trends, emerging challenges, and new strategies," *Electricity Journal*, vol. 34, no. 1, p. 106889, Jan. 2021, doi: 10.1016/j.tej.2020.106889.
[12] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *Journal of Intelligent Learning Systems and Applications*, vol. 09, no. 01, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.
[13] K. Sowjanya, A. Singhal, and C. Choudhary, "MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices," in *Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015*, Jun. 2015, pp. 397–402, doi: 10.1109/IADCC.2015.7154738.
[14] N. Yuvaraj and K. R. SriPreethaa, "Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster," *Cluster Computing*, vol. 22, no. S1, pp. 1–9, Jan. 2019, doi: 10.1007/s10586-017-1532-x.
[15] A. Rghioui, J. Lloret, S. Sendra, and A. Oumnad, "A smart architecture for diabetic patient monitoring using machine learning algorithms," *Healthcare (Switzerland)*, vol. 8, no. 3, p. 348, Sep. 2020, doi: 10.3390/healthcare8030348.
[16] S. Charoenkun, T. Klinkaew, K. Yothapakdee, N. Kaewsuwan, and T. Boonnuk, "Comparison of Machine learning models in predicting blood glucose levels of service participants at Khok Nong Kae subdistrict health promoting hospital Wang Saphung district, Loei Province," in *Proceedings of the 13th Conference on Application Research and Development (ECTI-CARD 2021)*, 2021, pp. 118–121.
[17] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12547–12553, Dec. 2009, doi: 10.1016/j.eswa.2009.05.032.
[18] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*. Wiley-IEEE Press, 2003.
[19] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, "Likelihood prediction of diabetes at early stage using data mining techniques," in *Advances in Intelligent Systems and Computing*, vol. 992, 2020, pp. 113–125, doi: 10.1007/978-981-13-8798-2_12.
[20] H. Yang *et al*., "Risk prediction of diabetes: Big data mining with fusion of multifarious physical examination indicators," *Information Fusion*, vol. 75, pp. 140–149, Nov. 2021, doi: 10.1016/j.inffus.2021.02.015.
[21] M. R. Khatun, S. I. Ayon, M. R. Hossain, and M. J. Alam, "Data mining technique to analyse and predict crime using crime categories and arrest records," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, p. 1052, May 2021, doi: 10.11591/ijeecs.v22.i2.pp1052-1060.
[22] P. Fabian *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
[23] Scikit-Learn, "scikit-learn Machine Learning in Python," *scikit-learn*, 2022. https://scikit-learn.org/stable/ (accessed Jan. 16, 2021).
[24] P. Marcelino, M. de Lurdes Antunes, E. Fortunato, and M. C. Gomes, "Machine learning for pavement friction prediction using scikit-learn," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10423 LNAI, 2017, pp. 331–342, doi: 10.1007/978-3-319-65340-2_28.
[25] M. S. Hossain Lipu, M. A. Hannan, A. Hussain, M. H. Saad, A. Ayob, and M. N. Uddin, "Extreme learning machine model for state-of-charge estimation of lithium-ion battery using gravitational search algorithm," *IEEE Transactions on Industry Applications*, vol. 55, no. 4, pp. 4225–4234, Jul. 2019, doi: 10.1109/TIA.2019.2902532.
[26] M. Solas, N. Cepeda, and J. L. Viegas, "Convolutional Neural Network for Short-term Wind Power Forecasting," in *Proceedings of 2019 IEEE PES Innovative Smart Grid Technologies Europe, ISGT-Europe 2019*, Sep. 2019, pp. 1–5, doi: 10.1109/ISGTEurope.2019.8905432.

BIOGRAPHIES OF AUTHORS

Dr. Kriengsak Yothapakdee received the Higher Dip. Tech. (Business Computer) from Mahasarakham Vocational College, Thailand, in May 2001, the M.S. (Computer Science) degree in June 2006 and the Ph.D. degree in computer science from Nuresuan University, Thailand, in July 2016. He is currently a lecturer in faculty of science and technology at Loei Rajabhat University. His current interests include pattern recognition, data mining, the Internet of Things, and Smart technology. He can be contacted at email: kriengsak@lru.ac.th.

Sarawoot Charoenkhum received B.S. (Public Health) degree from Mahasarakam University, Thailand. He is currently a hospital director in Khok-Nong-Kae Sub-district health promoting hospital. He can be contacted at email: sarawoot39283@gmail.com.

Asst. Prof. Dr. Tanunchai Boonnuk received the B.Sc. (Environmental Health) degree from Burapha University, Thailand, in 1997, the M.Sc. (Biomedical Instrumentation) degree from Mahidol University, Thailand, in 2003, B. Eng. (Computer Engineering) degree from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2006, the D.Eng. (Electrical Engineering) degree from Mahanakorn University of Technology, Thailand, in 2016. The Dr.P.H (Public Health) degree from Burapha University, Thailand, in 2017. the M.P.H. (Biostatistics) degree from Khon Kaen University, Thailand, in 2019. He is currently a lecture in faculty of science and technology at Loei Rajabhat University. His current interests include natural algorithm, spatio-temporal data engineering, and multilevel analysis. He can be contacted at email: tanunchai.boo@lru.ac.th.