

# Multi-Layer Perceptron Neural Network Model Development for Chili Pepper Disease Diagnosis Using Filter and Wrapper Feature Selection Methods

Sumitra Nuanmeesri

Faculty of Science and Technology  
Suan Sunandha Rajabhat University  
Bangkok, Thailand  
sumitra.nu@ssru.ac.th

Wongkot Sriurai

Faculty of Science  
Ubon Ratchathani University  
Ubon Ratchathani, Thailand  
wongkot.s@ubu.ac.th

**Abstract**—The goal of the current study is to develop a diagnosis model for chili pepper disease diagnosis by applying filter and wrapper feature selection methods as well as a Multi-Layer Perceptron Neural Network (MLPNN). The data used for developing the model include 1) types, 2) causative agents, 3) areas of infection, 4) growth stages of infection, 5) conditions, 6) symptoms, and 7) 14 types of chili pepper diseases. These datasets were applied to the 3 feature selection techniques, including information gain, gain ratio, and wrapper. After selecting the key features, the selected datasets were utilized to develop the diagnosis model towards the application of MLPNN. According to the model's effectiveness evaluation results, estimated by 10-fold cross-validation, it can be seen that the diagnosis model developed by applying the wrapper method along with MLPNN provided the highest level of effectiveness, with an accuracy of 98.91%, precision of 98.92%, and recall of 98.89%. The findings showed that the developed model is applicable.

**Keywords**—chili pepper diseases; feature selection; multi-layer perceptron neural network; wrapper

## I. INTRODUCTION

Chili peppers are plants used and consumed by Thai people in many forms. They are fundamental spices that can enhance the flavor, odor, and color of food. Chili peppers are considered unique due to their nutritional and medical benefits, flavors, and colors, so they cannot be replaced by other plants, hence, they are highly economically important in Thailand [1]. Chili peppers must be appropriately stored and maintained before being sold. Many farmers usually face problems in chili pepper plantations. These include inevitable natural disasters, weeds, and pests. Therefore, the farmers need to be educated about chili pepper diseases in order to protect their plants appropriately. Multi-Layer Perceptron Neural Networks (MLPNNs) have been broadly applied to diagnose diseases. For example, such models were developed in [2] and in [3] for predicting lung cancers and heart diseases respectively. These two studies developed disease diagnosis models which were more than 90% effective.

The current research aims to apply MLPNNs to diagnose chili pepper diseases. It also employs the filter and wrapper

feature selection methods to develop the diagnosis model. The feature selection methods will select only significant features for synthesizing the model rapidly and classifying data more effectively [4]. The most effective model in terms of data classification will be further applied to the development of a chili pepper disease diagnosis system that can assist the farmers in diagnosing the diseases and treat their plants timely and adequately.

## II. LITERATURE REVIEW

### A. Filter Feature Selection Methods

Filter feature selection methods are the processes that evaluate the level of effectiveness of each feature regarding its compatibility with data analysis. These processes do not rely on any learning approach. These filter methods select the data by ranking each feature in a priority queue. The priority list is based on the number specified by the user or the threshold of the selected feature. The advantages of these methods are rapid data processing and independence from learning methods [4]. This research applies two types of filter feature selection methods, namely the Information Gain (IG) and Gain Ratio (GR), as follows.

#### 1) Information Gain (IG)

IG is a measure used for classifying data by calculating the gain value of each attribute. If any attribute has the highest gain value, it will be selected as a subset with classification power, as shown in (1) [5]. Equation (1) demonstrates the calculation of entropy, while (2) explains the calculation of the gain value [5].

$$Entropy = -\sum_{i=0}^{c-1} p_i(t) \log_2 p_i(t) \quad (1)$$

$$Gain = Entropy(\text{parent}) - \sum_{i=1}^k \frac{N(v_i)}{N} Entropy(v_i) \quad (2)$$

where  $c$  refers to the number of classes,  $p_i(t)$  refers to the frequency value of class  $i$  for node  $t$ ,  $Entropy(\text{parent})$  refers to

the entropy of the parent node,  $k$  refers to the total number of feature values,  $N$  refers to the total amount of data of the parent node, and  $N(v_i)$  refers to the total amount of instances of child node  $i$ .

## 2) Gain Ratio (GR)

GR is a measure that uses the gain ratio as the indicator for dividing datasets into sub-datasets based on the IG value. However, when the IG value is used for classifying datasets, there would be the occurrence of bias in favor of attributes with large numbers of values. Thus, the GR can be derived from the gain value divided by the *SplitInfo* value that is calculated in (3) [6], where  $k$  refers to the total number of split data. Therefore, the GR is calculated in (4) [5]:

$$\text{SplitInfo} = -\sum_{i=1}^k \frac{N(v_i)}{N} \log_2 \frac{N(v_i)}{N} \quad (3)$$

$$\text{GR} = \frac{\text{Gain}}{\text{SplitInfo}} \quad (4)$$

## B. Wrapper Feature Selection

Wrapper feature selection methods are processes that select subsets from all features. They explore the feature subsets which specifically match a learning approach. Therefore, these methods can increase the effectiveness of the learning process to the greatest extent. This research applies an evolutionary wrapper for selecting features. The features, which are predictor variables, are randomized into the equation each by each. Then, their effectiveness in prediction is tested. If the level of effectiveness in prediction increases, the effective features will be maintained. In contrast, if the level of effectiveness in prediction decreases, the ineffective features will be removed [7].

## C. Multi-Layer Perceptron Neural Network

MLPNN is a form of perceptron neural network with multiple layers. It is suitable for complex computational tasks. It consists of 3 layer levels, namely the Input layer, the Hidden layer, and the Output layer. The Hidden layer could be composed of many layers, but there must be at least one layer in it [6, 8]. The data are the input of the Input layer. The output is sent out from the Output layer. The summation function of MLP is calculated in (5):

$$n = \sum_{i=1}^k p_i w_i + b \quad (5)$$

where  $n$  is the total sum gained from the summation function,  $p_i$  is the input of neuron  $i$ ,  $w_i$  is the weight of neuron  $I$ ,  $k$  is the number of Input layer neurons, and  $b$  is the bias value. The MLPNN model is illustrated in Figure 1.

## D. Model's Effectiveness Evaluation

The model's effectiveness was evaluated by the confusion matrix, which can calculate the precision, recall, and accuracy of the model [9-12].

## E. Related Works

Authors in [2] developed a model to predict lung cancer

using the MLPNN, comparing the efficacy of disease classification with the K-Nearest Neighbor (KNN) technique. The results showed that the disease classification model using the MLPNN was more effective. Authors in [3] applied the MLPNN to predict heart diseases. The efficacy of the heart disease classification model was compared with 9 classification methods and the results showed that the MLPNN was the most effective.

Authors in [13] applied the Wrapper method to select feature data to classify cardiac arrhythmias. The model's effectiveness was tested using 10-fold cross-validation, comparing 5 modeling techniques. The results showed that the model used to classify cardiac arrhythmias using Wrapper combined with MLP had the best performance. Authors in [14] developed an MLP model by applying the Correlation-based Feature Selection (CFS) and IG to analyze Thai water buffalo diseases. The experimental results showed that the developed model by CFS and MLP was efficient with accuracy, precision, and recall greater than 99.0%.

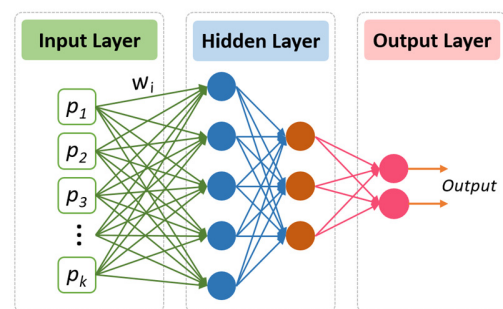


Fig. 1. Multi-Layer Perceptron Neural Network model.

## III. METHODS

In this work, a MLPNN model for chili pepper disease diagnosis was developed using filter and wrapper feature selection methods. Six 6 processes were involved: data collection, data preparation, feature selection, modeling, evaluation, and deployment.

### A. Data Collection

Primary data were collected from 5 agricultural professionals and 33 chili pepper farmers in Ubon Ratchathani Province of Thailand with the use of questionnaires. This is a significant chili pepper cultivation area in 3 districts, namely Muang Sam Sip District, Mueang District, and Khueang Nai. The secondary data were collected from the review in [15]. An analysis of the data revealed 14 diseases in chili peppers in those plantations, as shown in Figure 2. The details of the data are illustrated in Table I.

### B. Data Preparation

At the data preparation stage, the selected data were prepared before being applied to the model. The data were prepared in the following steps: 1) data selection, 2) data cleaning, and 3) data transformation. Ultimately, 863 questionnaires were processed, and were divided into 80% for training and 20% for testing. The data used for developing the chili pepper disease diagnosis model are illustrated in Table II.

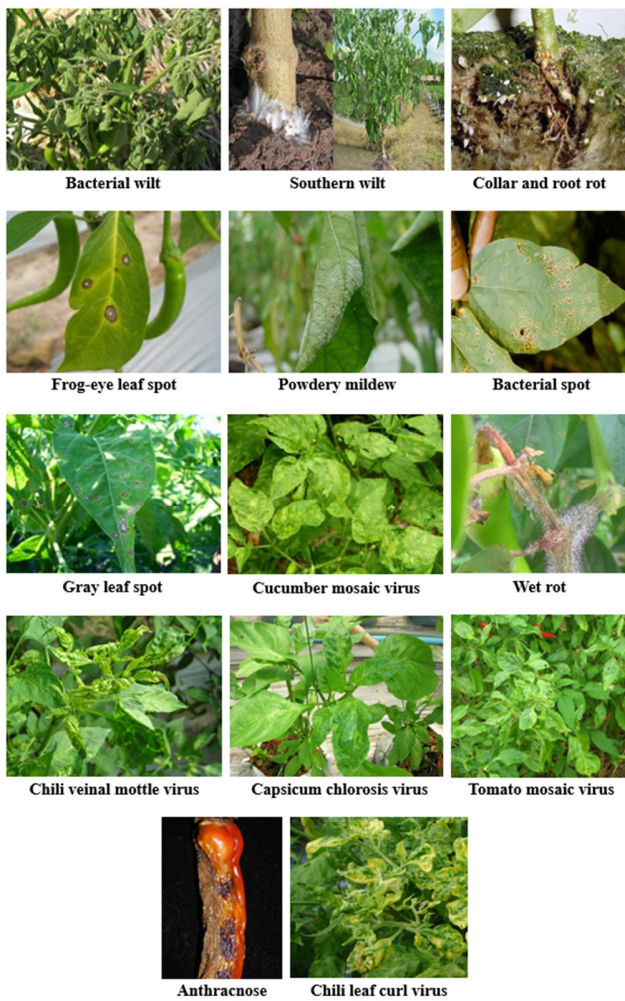


Fig. 2. Sample image of 14 chili peppers diseases.

TABLE I. DATA USED IN THIS RESEARCH

No.	Data
1	Types of chili pepper diseases
2	Causative agents
3	Infected areas
4	Growth stages of infection
5	Conditions for diseases
6	Symptoms
7	Fourteen types of chili pepper diseases: 1. Bacterial wilt 2. Southern wilt 3. Collar and root rot 4. Frog-eye leaf spot 5. Powdery mildew 6. Bacterial spot 7. Wet rot 8. Gray leaf spot 9. Anthracnose 10. Cucumber mosaic virus 11. Chili veinal mottle virus 12. Chili leaf curl virus 13. Capsicum chlorosis virus 14. Tomato mosaic virus

TABLE II. FEATURES USED FOR DEVELOPING THE CHILI PEPPER DISEASE DIAGNOSIS MODEL

No.	Data	Features
1	Types of chili pepper diseases	Non: nonliving (abiotic) agents Liv: living (biotic) agents
2	Causative agents	Bac: bacterial agents Fung: fungal agents Vir: viral agents
3	Infected areas	S1: stems F2: fruits R3: roots L4: leaves St5: stems and roots Lf6: leaves and fruits Ls7: leaves and stems Fs8: fruits and stems Sf9: stems, leaves, and fruits
4	Growth stages of infection	G1: early stage of growth G2: flowering G3: harvest G4: at all stages
5	Conditions for diseases	C1: poorly drained soil C2: humidity and rain consecutive days C3: rains and wet soil C4: strong and dry winds C5: frost and strong winds C6: overcrowded sprouts; poor ventilation C7: thrips C8: tobacco whiteflies
6	Symptoms	S1: browned vessels S2: white powdery coating or brown pellets appearing on the lower surface S3: water soaked lesions on stems and stalks S4: midget growth S5: severely spotted leaves turn yellow and drop S6: black narrow elongated lesions or streaks developing throughout the stems to the top of the plant S7: bleached, pale veinlets S8: wet, macerated stalks with black-ended silvery hair covering the lesions S9: distorted, irregular, curly leaves S10: yellowing of lower leaves at the front and back S11: a dark ring and a yellowish halo around the ring, forming a "frog-eye" appearance on the leaves
7	Types of chili pepper diseases (classes)	Bw1: bacterial wilt Sw2: southern wilt Cr3: collar and root rot Fl4: frog-eye leaf spot Pm5: powdery mildew Bs6: bacterial spot Wr7: wet rot Gr8: gray leaf spot An9: anthracnose Cm10: cucumber mosaic virus Cv11: chili veinal mottle virus Cl12: chili leaf curl virus Cc13: capsicum chlorosis virus Tm14: tomato mosaic virus

Once the data in Table II had been collected, the research team verified their accuracy. Then, they were converted into a csv file to be computed in Weka 3.9, as illustrated in Figure 3. After converting the data, the authors input the data into Weka to select features by applying IG, GR, and Wrapper.

Fig. 3. A sample of the data used in model development.

IV. RESULTS

The chili pepper disease diagnosis model was developed by applying filter and wrapper methods along with MLPNN. The number of features remaining after the feature selection process is shown in Table III.

TABLE III. REMAINING FEATURES AFTER FEATURE SELECTION

Feature selection method	Number of features remaining
IG	7
GR	7
Wrapper	6

The effectiveness of the models was evaluated by 10-fold cross-validation. The research results showed that the model developed by Wrapper and MLPNN (Wrapper+MLP) provided the highest results with accuracy of 98.91%, precision of 98.92%, and recall of 98.89%. Next in the rankings were IG+MLP, GR+MLP, and MLP (original data), as shown in Table IV and Figure 5. Furthermore, compared to other studies that detected or classified chili pepper diseases (Table V), the developed model by the Wrapper+MLP method gave the highest effectiveness result.

TABLE IV. EFFECTIVENESS EVALUATION RESULTS

Method	Precision (%)	Recall (%)	Accuracy (%)
MLP (original data)	92.60	92.54	92.52
IG+MLP	96.50	96.42	96.45
GR+MLP	93.34	93.20	93.26
Wrapper+MLP	98.92	98.89	98.91

C. Modeling

The data were divided into 4 datasets, namely 3 datasets undergoing the feature selection process and 1 original dataset used for MLPNN modeling. The research team specified the parameters for the MLPNN modeling as follows: Hidden layer = 2, Training time = 500, Learning rate = 0.3, and Momentum = 0.2. These parameters provided the best results.

D. Evaluation

The effectiveness of the chili pepper disease diagnosis model was evaluated by 10-fold cross-validation with the test dataset (20% of the total data). In addition, it was measured by examining the precision, recall, and accuracy of the developed models [6].

E. Deployment

After constructing the model, the most effective sample was applied to develop the chili pepper disease diagnosis prototype system. This system can diagnose chili pepper diseases timely and suggest proper treatments for each disease rapidly. The research model framework is illustrated in Figure 4.

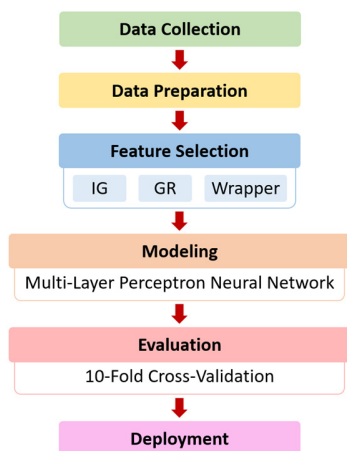


Fig. 4. The research model framework.

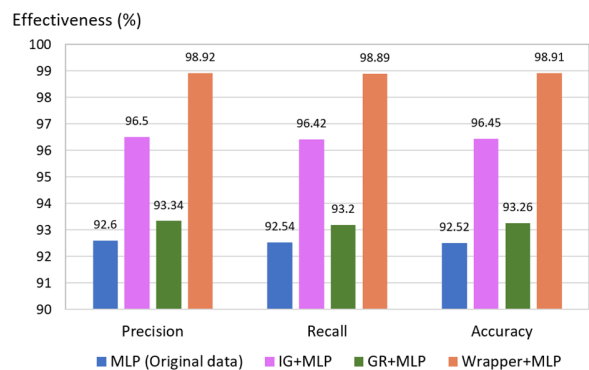


Fig. 5. Comparison between the models' effectiveness evaluation results.

TABLE V. COMPARISON WITH OTHER RESEARCH STUDIES

Method	No. of classes	Accuracy (%)
HSV model feature extraction [16]	2	80.00
Hyperspectral image + MLP [17]	2	83.26
Deep Belief Network [18]	2	91.96
Deep Learning + Support Vector Machine [19]	7	92.10
Fuzzy C-Means segmentation [20]	5	97.56
Wrapper + MLP (This study)	14	98.91

V. CONCLUSION AND DISCUSSION

This research aimed to develop a chili pepper disease diagnosis model by applying filter and wrapper feature

selection methods along with the MLPNN. The data used for classifying the 14 considered chili pepper diseases were processed by 3 feature selection techniques, namely IG, GR, and Wrapper.

Once the feature selection process was completed, the selected data were applied to develop the chili pepper disease diagnosis model with the use of the MLPNN. The experimental results indicated that the diagnosis model developed by the Wrapper method and the MLPNN provided the highest level of effectiveness, with an accuracy of 98.91%. This result means that the developed model can be used in a chili pepper disease diagnosis system, suggesting chili pepper disease prevention and treatment. The results of this research conform to the study conducted in [3], in which the authors applied the MLP method to develop a disease diagnosis model which proved to perform better than other known techniques. The research findings also conform to the findings in [13]. The authors employed the wrapper method and the MLP to classify the symptoms of cardiac arrhythmia. Their research results showed that the wrapper feature selection method could increase the effectiveness of the classification process.

In general, a disease on chili peppers can be identified by its symptoms. Most research is focused on image processing [16, 17, 20] and deep learning [18, 19] disease analysis from chili pepper leaves that require large image datasets to train the model. However, some symptoms that appear on chili pepper leaves may be similar, resulting in discrepancies in the efficiency and accuracy of classification. In addition, the disease identification of chili peppers from leaves is limited to some diseases with distinctive features appearing on the leaves. It is often necessary to analyze symptoms in other parts of the chili pepper plant, such as roots, stems, leaves, and fruits. Therefore, this study used the information presented on plants throughout the chili pepper life cycle and considered the environmental conditions to help identify 14 occurring diseases. Besides, the Wrapper feature selection process proved to be more efficient than the IG, and GR, and the classification without feature selection.

#### ACKNOWLEDGMENT

The authors are grateful to the Institute for Research and Development, Suan Sunandha Rajabhat University, and the Faculty of Science, Ubon Ratchathani University for supporting this research.

#### REFERENCES

- [1] K. Lertrat, "Production, planting, processing, marketing, and chili pepper products in Thailand," *Research Community*, vol. 73, pp. 15–20, May 2007.
- [2] S. Potghan, R. Rajamenakshi, and A. Bhise, "Multi-Layer Perceptron Based Lung Tumor Classification," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, Mar. 2018, pp. 499–502, <https://doi.org/10.1109/ICECA.2018.8474864>.
- [3] K. Subhadra and B. Vikas, "Neural Network Based Intelligent System for Predicting Heart Disease," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 5, pp. 484–487, 2019.
- [4] K. Sutha and J. J. Tamilselvi, "A review of feature selection algorithms for data mining techniques" *International Journal on Computer Science and Engineering*, vol. 7, no. 6, pp. 63–67, Jun. 2015.
- [5] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, 2nd ed. New York, USA: Pearson Education, 2019.
- [6] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 1st ed. Boston, Massachusetts, USA: Addison-Wesley, 2005.
- [7] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika Journal of Science & Technology*, vol. 26, no. 1, pp. 329–340, Jan. 2018.
- [8] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *International Journal of Artificial Intelligence and Expert Systems*, vol. 1, no. 1, pp. 111–122, 2011.
- [9] S. Nuanmeesri, S. Chopvitayakun, P. Kadmateekarun, and L. Poomhiran, "Marigold flower disease prediction through deep neural network with multimodal image," *International Journal of Engineering Trends and Technology*, vol. 69, no. 7, pp. 174–180, Jul. 2021, <https://doi.org/10.14445/22315381/IJETT-V69I7P224>.
- [10] S. Nuanmeesri, L. Poomhiran, and K. Ploydanai, "Improving the prediction of rotten fruit using convolutional neural network," *International Journal of Engineering Trends and Technology*, vol. 69, no. 7, pp. 51–55, Jul. 2021, <https://doi.org/10.14445/22315381/IJETT-V69I7P207>.
- [11] A. N. Saeed, "A Machine Learning based Approach for Segmenting Retinal Nerve Images using Artificial Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5986–5991, Aug. 2020, <https://doi.org/10.48084/etasr.3666>.
- [12] M. B. Ayed, "Balanced Communication-Avoiding Support Vector Machine when Detecting Epilepsy based on EEG Signals," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6462–6468, Dec. 2020, <https://doi.org/10.48084/etasr.3878>.
- [13] A. Mustaqeem, S. M. Anwar, M. Majid, and A. R. Khan, "Wrapper method for feature selection to classify cardiac arrhythmia," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Jeju, South Korea, 2017*, pp. 3656–3659, <https://doi.org/10.1109/EMBC.2017.8037650>.
- [14] S. Nuanmeesri and W. Sriurai, "Thai Water Buffalo Disease Analysis with the Application of Feature Selection Technique and Multi-Layer Perceptron Neural Network," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 6907–6911, Apr. 2021, <https://doi.org/10.48084/etasr.4049>.
- [15] S. Sudhi-Aromna *et al.*, *A guide to chili pepper pests*, Nonthaburi, Thailand: The Agricultural Co-operative Federation of Thailand, Ltd., 2014.
- [16] D. P. Patil, S. R. Kurkute, P. S. Sonar, and S. I. Antonov, "An advanced method for chilli plant disease detection using image processing," in *52nd International Scientific Conference On Information, Communication and Energy Systems and Technologies*, Niš, Serbia, 2017, pp. 309–313.
- [17] M. Ataş, Y. Yardimci, and A. Temizel, "A new approach to aflatoxin detection in chili pepper by machine vision," *Computers and Electronics in Agriculture*, vol. 87, pp. 129–141, 2012, <https://doi.org/10.1016/j.compag.2012.06.001>.
- [18] S. Jana, A. R. Begum, and S. Selvagesan, "Design and analysis of pepper leaf disease detection using Deep Belief Network," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 9, pp. 1724–1731, 2020.
- [19] N. N. Ahmad Loti, M. R. Mohd Noor, and S.-W. Chang, "Integrated analysis of machine learning and deep learning in chili pest and disease identification," *Journal of the Science of Food and Agriculture*, vol. 101, no. 9, pp. 3582–3594, 2021, <https://doi.org/10.1002/jsfa.10987>.
- [20] S. Das Chagas Silva Araujo, V. S. Malemath, and K. M. Sundaram, "Symptom-Based Identification of G-4 Chili Leaf Diseases Based on Rotation Invariant," *Frontiers in Robotics and AI*, vol. 8, 2021, Art. no. 650134, <https://doi.org/10.3389/frobt.2021.650134>.

## AUTHORS PROFILE



**Sumitra Nuanmeesri** received her Ph.D. in Information Technology at the King Mongkut's University of Technology North Bangkok, Thailand. She is an Assistant Professor in the Information Technology Department, Faculty of Science and Technology at Suan Sunandha Rajabhat University, Thailand. Her research interests include speech recognition, data mining, deep learning, image processing, mobile application, supply chain management system, internet of things, robotics, augmented reality, and virtual reality.



**Wongkot Sriurai** received her Ph.D. in Information Technology at the King Mongkut's University of Technology North Bangkok, Thailand. She is an Assistant Professor in the Mathematics, Statistics and Computer Department, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani Province, Thailand. Her research interests include data mining, text mining, web mining, recommender system, information filtering, information retrieval, decision support systems, expert systems, multimedia technology, and computer education.