

Thai Water Buffalo Disease Analysis with the Application of Feature Selection Technique and Multi-Layer Perceptron Neural Network

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—This research aims to develop an analysis model for diseases of the water buffalo with the application of the feature selection technique along with the Multi-Layer Perceptron Neural Network (MLP-NN). The data used for analysis were collected from books and documents related to water buffalo diseases and the official website of the Department of Livestock Development. The data consist of the characteristics of 6 water buffalo diseases, including anthrax, hemorrhagic septicemia, brucellosis, foot and mouth diseases, parasitic diseases, and mastitis. Since the amount of the collected data was limited, the synthetic minority over-sampling technique was also employed to adjust the imbalance dataset. The adjusted dataset was used to select the disease characteristics towards the application of two feature selection techniques, correlation-based feature selection and information gain. Subsequently, the selected features were then used for developing the analysis model for water buffalo diseases towards the use of the MLP-NN. The evaluation results given by 10-fold cross-validation, showed that the analysis model for water buffalo diseases developed by correlation-based feature selection and MLP-NN provided the highest level of effectiveness with an accuracy of 99.71%, precision of 99.70%, and recall of 99.72%, implying that the analysis model is effectively applicable.

Keywords—water buffalo diseases; feature selection; multi-layer perceptron; neural network; synthetic minority over-sampling

I. INTRODUCTION

In Thailand, water buffaloes play an important role in the livestock economy. Thai water buffaloes are the most common farm animals in Asia and farmers typically use them for agricultural labor and as a source of food. To domesticate water buffaloes productively, it is vital to pay attention to their nutrition, habitats, sanitation, signs, behavior, and disease symptoms [1]. At present, water buffalo farmers are confronting many kinds of water buffalo diseases due to the seasonal changes, disease carriers, the lack of expertise and knowledge among farmers themselves, and the lack of experts who can analyze and diagnose these diseases. Moreover, the internet only offers basic information about the diseases gained from statistical surveys, resulting in the retrieval of incorrect or inadequate data for disease analysis and, ultimately, misunderstanding or incorrect analysis. Water buffalo farmers

who lack attention may overlook the signs and symptoms of serious infectious diseases, which could be spread to other animals and cause sickness or even casualties [2-3].

The Multi-Layer Perceptron Neural Network (MLP-NN) is one of the most popular techniques used to classify complex data. This research aims to develop an analysis model for water buffalo diseases towards the application of the feature selection technique and MLP-NN. Two feature selection techniques were employed, Correlation-based Feature Selection (CFS) and Information Gain (IG). After selecting the features, the data were then used for developing the analysis model by using the MLP-NN. The developed model can be applied to the development of a water buffalo disease analysis system, which is expected to help farmers timely analyze the diseases.

II. BACKGROUND

A. Data Imbalance Resolution

This research applies the Synthetic Minority Over-sampling Technique (SMOTE) to resolve the data imbalance problem. SMOTE helps to resynthesize data by increasing the dataset's size with a small amount of class data [4] to be compatible with the biggest dataset. This is done by randomizing a value and calculating the distance between the selected value with other values to find the nearest value [5].

B. Feature Selection Techniques

The goal of feature selection is to select the most significant features of each dataset in order to synthesize the model rapidly and increase the effectiveness of data classification. In this study, two feature selection methods are employed:

1) Correlation-based Feature Selection (CFS)

CFS is a feature selection method based on the relationship between the collections of features gained from the evaluation of feature prediction capacity used for data classification and irrelevant data management. CFS can rank the data subsets based on the data dimensions and select the data subsets based on the data dimensions with regard to high and low relationships between classes. Any irrelevant data or any data with a low level of relationship will be excluded. The same will

occur with complex data dimensions which shall be excluded from the data dimensions with a high level of relationship. The formula for evaluating the subsets of CFS data dimensions is shown in (1) [6]:

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)\bar{r}_{ff}}} \quad (1)$$

where k refers to the data dimension or features, M_s refers to the value of S data dimension subset which composes of k data dimensions, \bar{r}_{cf} refers to the average value of the relationship between the variables and classes ($f \in S$), and \bar{r}_{ff} refers to the average value of the relationship of data dimension.

2) Information Gain (IG)

IG is a feature selection method in which the gain value of each node is evaluated. If a node has the highest gain value, it will be chosen as the root node, and the rest of the data will be reassessed in order to find the gain value of the next node. The formula for finding the IG value is [7]:

$$Gain(Y; X) = H(Y) - H(Y | X) \quad (2)$$

where Y refers to the feature value, which is a data class belonging to the $\{Y_1, Y_2, \dots, Y_n\}$ set where n is the number of features, X refers to the value of other features that are not classes ranging between $\{X_1, X_2, \dots, X_n\}$, $Gain(Y; X)$ refers to the score value gained from sample randomization ranging between 0 and 1, $H(Y)$ refers to the probability value gained from the randomization of Y samples, and $H(Y | X)$ refers to the probability value gained from the randomization of Y samples when compared to X . $H(Y)$ and $H(Y | X)$ are calculated in (3) and (4), respectively:

$$H(Y) = -\sum_{i=1}^{i=k} P(Y = y_i) \log_2 P(Y = y_i) \quad (3)$$

$$H(Y | X) = -\sum_{i=1}^{i=k} P(X = x_i) H(Y | X = x_i) \quad (4)$$

where $P(Y = y_i)$ refers to the probability value from y_1 to y_k , $P(X = x_i)$ refers to the probability value from x_1 to x_k , and k refers to the number of features.

C. Multi-Layer Perceptron Neural Network

The MLP-NN, illustrated in Figure 1, consists of an input layer, hidden layers, and an output layer [8]. In each layer, there are a collection of nodes. Possibly, there are more than one hidden layers [9]. The MLP-NN operates by inserting data into the input layer to estimate and deliver the results to the output layer. The estimation requires the sum total of input data multiplied by weight values, as shown in (5). After that, the output is used for the calculation with the sigmoid function, as shown in (6).

$$n = \sum_{i=1}^k P_i W_i \quad (5)$$

where n refers to the sum total of input P_i multiplied by the weight W_i and i refers to the number of inputs or weight value. In (6), x refers to the input value.

$$f_x = \frac{1}{1 + e^{-x}} \quad (6)$$

The output of the hidden layer is delivered to the output layer, where there is a comparison between the estimated and the target outputs. If there are different values that cannot be accepted, the outputs will get into the backpropagation process and go back to the hidden and input layer. Simultaneously, there is the weight adjustment process, which will find the most acceptable value after testing with the data. Subsequently, the output is estimated with the sigmoid function once again [8].

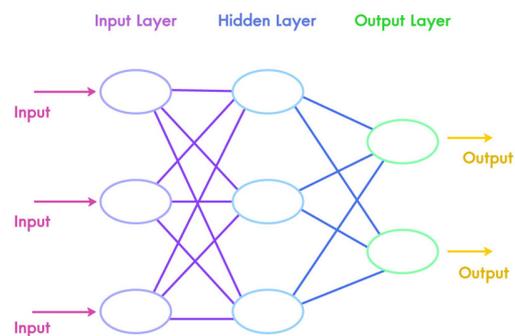


Fig. 1. MLP-NN.

D. Similar Studies

Authors in [13] compared the effectiveness of data imbalance resolution techniques by using diabetes patients' data. The research team compared four different methods, which include oversampling, undersampling, hybrid method, and SMOTE. Two data classification techniques were applied, Multinomial Logistic Regression Analysis and Decision Tree, to classify diabetes patients. The research findings showed that the combination of the data adjusted by SMOTE and the data classification with the decision tree technique provided the best results for the classification of the diabetes patients. Authors in [14] studied the classification of heart diseases using MLP-NN and IG as a feature selection method. The findings showed that the number of features could be reduced from 13 to 8, while the accuracy of the training dataset increased by 1.1% and the accuracy of the trial dataset increased by 0.82%. Authors in [15] studied the classification of ovarian cancer towards the application of SMOTE and MLP-NN. The findings indicated that the SMOTE technique could adjust the data balance, and after using the adjusted data to construct the model with MLP-NN, the model's effectiveness increased. The experiment results showed that the application of SMOTE+MLP provided a data classification accuracy of 96%, which was higher than the one gained from the application of SMOTE+RBF. Authors in [16] applied feature selection along with MLP to predict chronic diseases. The research findings showed that applying these two methods provided higher effectiveness in terms of chronic disease prediction than the application of Support Vector Machine (SVM) and Decision Tree.

The developed model of the current research can be applied to the development of a water buffalo disease analysis system. Two feature selection techniques were employed, CFS and IG.

E. Effectiveness Evaluation

The effectiveness evaluated by the confusion matrix is an evaluation method of discriminants' accuracy, which means the discriminants can be classified in accordance with their genuine value. The accuracy can be calculated by (7) [10], while precision and recall can be calculated in (8) and (9) [11-12], with the values represented in Figure 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (9)$$

where TP refers to when the target class is "Yes," and the model predicts it as "Yes" (True Positive), FP refers to when the target class is "Yes," but the model predicts it as "No" (False Positive), TN refers to when the target class is "No," and the model predicts it as "No" (True Negative), and FN refers to when the target class is "No," but the model predicts it as "Yes" (False Negative).

		Predicted	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	TP

Fig. 2. Confusion matrix.

III. RESEARCH METHODOLOGY

The research methodology for the analysis of Thai water buffalo diseases consists of 1) data collection and preparation, 2) data imbalance adjustment using SMOTE, 3) feature selection by CFS and IG, 4) model development using MLP-NN, and 5) model effectiveness evaluation (Figure 3).

A. Data Collection and Preparation

This research collected data from books and documents related to water buffalo diseases [3] and the Department of Livestock Development's official website. There are totally 480 records of data. The data involves information about six water buffalo disease classes, namely anthrax, hemorrhagic septicemia, brucellosis, foot and mouth diseases, parasitic diseases, and mastitis. These data were used for developing the disease analysis model. There are 33 attributes and six classes, as illustrated in Table I.

TABLE I. DATA CHARACTERISTICS USED FOR DEVELOPING THE ANALYSIS MODEL OF THAI WATER BUFFALO DISEASES

Symptoms	Data symbol	Data value
Fever	Symptom 1	High: AA1 Moderate: AA2 No Fever: AA3
Drooling	Symptom 2	Normal: Ab1 Excessive: AB2
Breathing characteristics	Symptom 3	Fast and frequent breathing: AC1 Difficulty in breathing: AC2 Normal breathing: AC3
Bad breath	Symptom 4	Bad breath: AD1 Normal: AD2
Swollen breasts	Symptom 5	Little swollen: AE1 Red, hard, and swollen: AE2 Normal: AE3
Endometritis	Symptom 6	Inflammatory: AF1 Not inflammatory: AF2
Drowsiness	Symptom 7	Drowsy: AG1 Not drowsy: AG2
Eating habits	Symptom 8	Appetite loss: AH1 Normal: AH2
Swelling characteristics	Symptom 9	Body or neck swelling: AI1 Joint swelling: AI2 Not swollen: AI3
Urine characteristics	Symptom 10	Blood in urine: AJ1 Normal urine: AJ2
Testicle characteristics	Symptom 11	Swollen: AK 1 Normal: AK2
Stool characteristics	Symptom 12	Stools contained blood: AL1 Melena: AL2 Smelly stools: AL3 Normal: AL4
Walking	Symptom 13	Stiff: AM1 Normal: AM2
Milk characteristics	Symptom 14	Sticky or purulent: AN1 Normal: AN2
Retained placenta	Symptom 15	Retained: AO1 Not: AO2
Emaciation	Symptom 16	Emaciated: AP1 Not: AP2
Blisters inside the mouth and on the feet	Symptom 17	Has: AQ1 Has not: AQ2
Vesicular state (within mouth, tongue, beneath tongue, cheeks, hoofs, ankles, nipples, and breast)	Symptom 18	Has: AR1 Has not: AR2
Wounds around the hoofs and/or ankles	Symptom 19	Has: AS1 Has not: AS2
Weight	Symptom 20	Weight loss: AT1 Normal weight: AT2
Hard and rough hair	Symptom 21	Hard and rough: AU1 Normal: AU2
Sleep habits	Symptom 22	Excessive sleep: AV1 Normal: AV2
Muscle characteristics	Symptom 23	Shaky or twitching: AW1 Normal: AW1
Bleeding from nose, mouth, ears, eyes	Symptom 24	Has: AX1 Has not: AX2
Lip characteristics	Symptom 25	Pale: AY1 Normal: AY2
Oral mucositis and red eye sockets	Symptom 26	Has: AZ1 Has not: AZ2
Forward head posture	Symptom 27	Has: BA1 Has not: BA2
Tongue-tie	Symptom 28	Has: BB1 Has not: BB2
Cough	Symptom 29	Has: BC 1 Has not: BC2
Vaginal discharge	Symptom 30	Has: BD 1 Has not: BD2
Amount of milk	Symptom 31	Stopped milk production: BE1 Lower amount of milk: BE2 Normal: BE3
Excretion characteristics	Symptom 32	Frequent: BF1 Normal: BF2
Recurrent seizures	Symptom 33	Has: BG 1 Has not: BG2
Water buffalo diseases	Class	Anthrax Hemorrhagic septicemia Brucellosis Foot and mouth disease Parasitic disease Mastitis

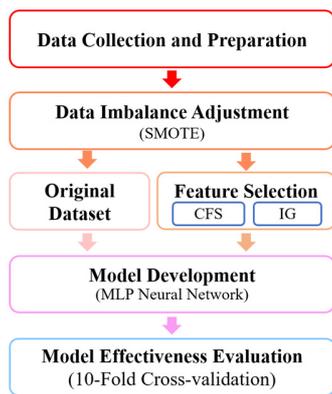


Fig. 3. The model framework.

After collecting the data illustrated in Table I, the research team rechecked their reliability and accuracy to ensure that no attribute was incorrect or missing. For example, there are no data out of range for each attribute. Then, the data were converted into .CSV file format in order to be operated with the Weka version 3.9, as shown in Figure 4.

#	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
1	Symptom	class																		
2	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	antiradiase	
3	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	antiradiase	
4	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	antiradiase	
5	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	antiradiase	
6	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	antiradiase	
7	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	hemorrhagiceptemia	
8	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	hemorrhagiceptemia	
9	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	hemorrhagiceptemia	
10	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	hemorrhagiceptemia	
11	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	hemorrhagiceptemia	
12	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	brucellosis	
13	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	brucellosis	
14	AR2	AR2	AS2	AT2	AV2	AW2	AX2	AY2	AZ2	BA2	BB2	BC2	BD2	BE2	BF2	BG2	BH2	BI2	brucellosis	
15	AR1	AR1	AS1	AT1	AV1	AW1	AX1	AY1	AZ1	BA1	BB1	BC1	BD1	BE1	BF1	BG1	BH1	BI1	footandmouthdisease	
16	AR1	AR1	AS1	AT1	AV1	AW1	AX1	AY1	AZ1	BA1	BB1	BC1	BD1	BE1	BF1	BG1	BH1	BI1	footandmouthdisease	
17	AR1	AR1	AS1	AT1	AV1	AW1	AX1	AY1	AZ1	BA1	BB1	BC1	BD1	BE1	BF1	BG1	BH1	BI1	footandmouthdisease	
18	AR1	AR1	AS1	AT1	AV1	AW1	AX1	AY1	AZ1	BA1	BB1	BC1	BD1	BE1	BF1	BG1	BH1	BI1	footandmouthdisease	
19	AR1	AR1	AS1	AT1	AV1	AW1	AX1	AY1	AZ1	BA1	BB1	BC1	BD1	BE1	BF1	BG1	BH1	BI1	footandmouthdisease	

Fig. 4. Data used for model construction.

B. Data Imbalance Adjustment using SMOTE

Since the prepared data were found to be imbalanced in class/label, the research team decided to adjust the data imbalance of the datasets by increasing the number of datasets with a small size of classes. The best result was provided after increasing the *k*-nearest neighbor value from 1 to 5. It was experimentally found that *k* = 5 and *randomSeed* = 1 give the best result. The data size was then increased from 100% until the highest level of effectiveness could be gained (as evaluated by 10-fold cross-validation). The experiment results showed that the data could be balancing and upsizing to 300%. Thus, the new dataset increased to 528, 688, and 768 records for data balancing using SMOTE at 100%, 200%, and 300% of its original size respectively.

C. Feature Selection by CFS and IG

The data with 33 attributes and 6 classes were brought into the feature selection process using the CFS and the IG in Weka. In this work, there are 8 groups of data applied in this process, namely 1) the original dataset through the CFS, 2) the original dataset through IG, 3) 100% of SMOTE through CFS, 4) 200% of SMOTE through CFS, 5) 300% of SMOTE through CFS, 6) 100% of SMOTE through IG, 7) 200% of SMOTE through IG,

and 8) 300% of SMOTE through IG. These resulting datasets will be used at the next step.

D. Model Development Using MLP neural network

After the data imbalance had been adjusted, the data were transferred to the learning process to construct the model by which the research team applied two feature selection techniques, CFS and IG, along with the MLP-NN. In this research, the input layer consisted of 33 neurons. The output layer consisted of 6 neurons. Therefore, the optimum parameters for the MLP-NN model set in Weka were: Hidden Layer = 4, Training Time = 500, Learning Rate = 0.3, Momentum = 0.2, and 21 epochs= 500. These values provided the highest level of effectiveness after being evaluated by 10-fold cross-validation. Besides, the model was also generated from the original dataset that SMOTE, or CFS, or IG had not processed in any way, for effectiveness comparison with the other models that have undergone balancing and or feature selection. Thus, 9 models were built from the original dataset and 8 feature selection datasets.

IV. RESULTS

According to the experiment results, the most appropriate data size adjusted by SMOTE was 200%. Then, the features of the complete dataset were selected by CFS and IG. Afterwards, the MLP-NN was developed and its effectiveness was evaluated by 10-fold cross-validation, as illustrated in Table II and Figure 5.

TABLE II. MODEL'S EFFECTIVENESS EVALUATION

Method	Data (rows)	Precision (%)	Recall (%)	Accuracy (%)
MLP (Original data)	480	98.77	98.78	98.79
CFS+MLP	480	98.80	98.82	98.81
IG+MLP	480	91.59	91.54	91.52
SMOTE (100%) +CFS+MLP	528	98.85	98.87	98.86
SMOTE (200%) +CFS+MLP	688	99.70	99.72	99.71
SMOTE (300%) +CFS+MLP	768	99.60	99.61	99.62
SMOTE (100%) +IG+MLP	528	92.48	92.52	92.53
SMOTE (200%) +IG+MLP	688	93.62	93.59	92.61
SMOTE (300%) +IG+MLP	768	93.05	93.08	93.06

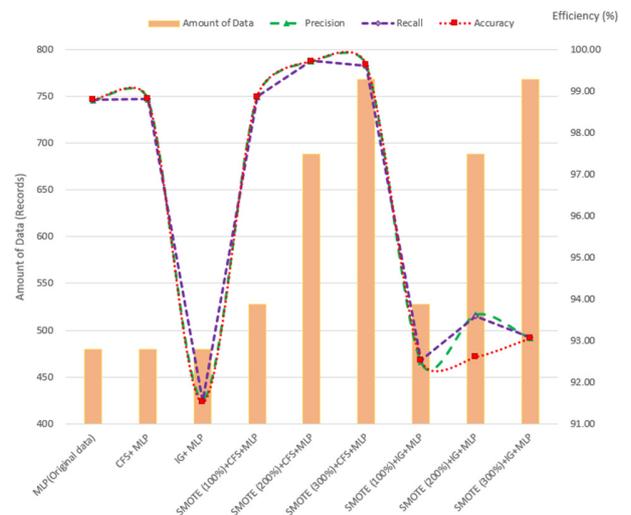


Fig. 5. The comparison of the model's effectiveness evaluation.

According to Table II and Figure 5, the 10-fold validation data classification method of the SMOTE (200%)+CFS+MLP dataset provided an accuracy of 99.71%, a precision of 99.70%, and a recall of 99.72%. These were the highest values. The results of the effectiveness comparison between the outputs of CFS and IG given by the MLP-NN showed that after being adjusted by SMOTE, the CFS method provided better feature selection than the IG method.

V. CONCLUSION

This research aimed to analyze the water buffalo diseases towards the application of feature selection techniques along with the MLP-NN. The data imbalance was adjusted by the SMOTE method. Two feature selection methods were employed: CFS and IG. After that, the data were classified by MLP-NN, and the model's effectiveness was evaluated by 10-fold cross-validation. The research findings showed that the most suitable data size after the data imbalance adjustment was 200%. After using the obtained data to construct the model, it was found that the model whose data size was adjusted by SMOTE and developed by CFS and MLP-NN provided the highest level of effectiveness in data classification with an accuracy of 99.71%. So, the developed model can be applied to the development of an analysis system for water buffalo diseases. The results of this study conform to the research conducted in [15], in which the SMOTE+MLP method was applied for data classification and a high level of effectiveness was reached with an accuracy of 90%, and also are in accordance with [16], in applied feature selection techniques were which alongside the MLP for data classification and gained a higher level of effectiveness.

ACKNOWLEDGMENT

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

REFERENCES

- [1] *Thailand's livestock statistics*. Bangkok, Thailand: Information and Communication Technology Center, Department of Livestock Development, 2015.
- [2] N. Hongboonmee and P. Sornrung, "Applying decision tree classification techniques for diagnose the disease in cow on mobile phone," *Journal of Science and Technology*, vol. 20, pp. 44–58, 2018.
- [3] P. Booranamanas, *Water buffaloes and treatments*, Bangkok, Thailand: Thaiwattanapanich, 1988.
- [4] M. G. Tsiouras, "Uterine EMG Signals Spectral Analysis for Pre-Term Birth Prediction," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3310–3315, Oct. 2018, <https://doi.org/10.48084/etasr.2146>.
- [5] P. Paranya, "Improving decision tree technique in imbalanced data sets using SMOTE for internet addiction disorder data," *Information Technology Journal*, vol. 12, no. 1, pp. 54–62, Jun. 2016.
- [6] A. H. Mark, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, Hamilton, New Zealand, 1999.
- [7] T. Puripat, "Ensemble algorithm for feature selection", M.S. thesis, Thammasat University, Bangkok, Thailand, 2016.
- [8] S. Boubaker, S. Kamel, and M. Kchaou, "Prediction of Daily Global Solar Radiation using Resilient-propagation Artificial Neural Network and Historical Data: A Case Study of Hail, Saudi Arabia," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5228–5232, Feb. 2020, <https://doi.org/10.48084/etasr.3278>.
- [9] A. Montaphan, "Comparison of feature selection methods to improve breast cancer prediction," *Royal Thai Air Force Medical Gazette*, vol. 65, no. 2, pp. 49–56, 2019.
- [10] 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>.
- [11] S. Nuanmeesri, "Mobile application for the purpose of marketing, product distribution and location-based logistics for elderly farmers," *Applied Computing and Informatics*, 2019, <https://doi.org/10.1016/j.aci.2019.11.001>.
- [12] 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>.
- [13] W. Kesornsit, V. Lorchirachoonkul, and J. Jitthavech, "Imbalanced data problem solving in classification of diabetes patients," *Khon Kaen University Research Journal*, vol. 18, no. 3, pp. 11–21, Jul. 2018.
- [14] A. Khempila and V. Boonjing, "Heart Disease Classification Using Neural Network and Feature Selection," presented at the Proceedings - ICSEng 2011: International Conference on Systems Engineering, Las Vegas, NV, USA, Sep. 2011, vol. 64, pp. 406–409, <https://doi.org/10.1109/ICSEng.2011.80>.
- [15] M. A. Hambali and M. D. Gbolagade, "Ovarian cancer classification using hybrid synthetic minority over-sampling technique and neural network," *Journal of Advances in Computer Research*, vol. 7, no. 4, pp. 109–124, Nov. 2016.
- [16] S. K. Hegde and R. Hedge, "Symmetry Based Feature Selection with Multi layer Perceptron for the prediction of Chronic Disease," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 3316–3322, Jul. 2019, <https://doi.org/10.35940/ijrte.B2658.078219>.

AUTHOR PROFILES



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.