The One-Versus-One Classification Technique Based on Data Synthesis with Appropriate Distant Neighbors

Pasapitch Chujai, Kedkarn Chaiyakhan, Nittaya Kerdprasop, and Kittisak Kerdprasop

***Abstract*—Classification of unbalanced information is a problem of machine learning frameworks and learning of existing basic algorithms which will be more complicated if the data has more than two classes. Therefore, this research proposes a concept to improve the classification of unbalanced data with more than two classes with a model called One-Versus-One Classification Technique based on Data Synthesis with Appropriate Distant Neighbors (OVO-SynDN). This OVO-SynDN model will divide the problem of classification of multiclass data into binary class data classification with the learning technique "One-Versus-One". Then it will adjust the information imbalance by synthesizing the data which selects nearest neighbors with Euclidean distance techniques. For the amount of data to be synthesized of each data set, it will be selected from the number of nearest neighbors that are in the opposite group. The features of the new synthesized data, it depends on the characteristics of the original data and the nearest neighbors. Then combine the existing imbalanced data sets and synthetic data sets to construct the learning model. The standard algorithm, Support Vector Machine (SVM) with polynomial kernel function, will be selected to learn these data sets. Some parameters are adjusted so that the algorithm is suitable for learning each data set. The results show that the OVO-SynDN model has satisfactory performance and reliability with high MAvA and MFA values. In addition, the OVO-SynDN model can still classify imbalanced data better than the four techniques that are compared. That means that the proposed method can be applied to the classification of unbalanced data that has more than two classes.**

***Index Terms*—SVM with polynomial kernel function, synthesizing imbalance data, multiclass imbalanced datasets classification, one-versus-one, binary decomposition.**

I. Introduction

Classification of unbalanced information is one of the most researched machine learning frameworks [1]. The unbalanced data will be found when the amount of data in the group that is important has less data than the rest [2]. Imbalanced information can be often found in real data [3] such as medical diagnosis data [4] that has less data than the information of healthy people, production information of products with less amount of product information than good products, or credit card information that has amount of unusual customer information less than normal customers etc. When using these data to learn with the basic algorithms of machine learning tasks, especially the classification of that data, it appears that those data will affect the learning of existing basic algorithms. However, due to the classification of information, it will give equal importance to every group of information. Whereas in the case of unbalanced data classification, it is found that standard data classification algorithms are unable to predict the data that is in the minority class. At the same time, the prediction results will be biased in a group with a large number of instances. It results in the efficiency of the classification of information in the minority groups to be less accurate.

In order to solve this problem, researchers were very interested in presenting various techniques [5]-[7]. The goal of the proposed method is increasing efficiency and accuracy in classifying data for all groups. The proposed method will be divided into four levels [8]. 1) Data Level Approaches which are directly related to data in the preprocessing stage by adjusting the data that is not balanced to become balanced data with data sampling techniques [9]: Over Sampling, Under Sampling, or Hybrid Methods between Over Sampling and Under Sampling Techniques. The balance adjustment at this level is adjusted by random selection from the original data or rebuilt from existing examples [10]. 2) Algorithmic Level Approaches will solve the problem by adjusting the learning of standard algorithms for the classification of existing data to be able to learn unbalanced information by providing biased information to the class of small groups. 3) Cost Sensitive Level Approaches is a solution that brings both the solution at the data level and the algorithm level to work together [11]. At the data level, the extra cost will be added for the case of faulty classification. At the algorithm level, adjusting the learning of standard algorithms in accordance with the wrong data classification. And 4) Ensemble Learning Approach [12], [13] is a method that uses more than one algorithm to learn together to get better predictive performance than learning with only one algorithm by working with the three methods mentioned above.

In this research, the researcher adopted a method to solve both data and algorithm level problems by presenting a model called OVO-SynDN. It is a model used to improve the classification of unbalanced data that is multiclass by dividing the problem from the classification of multiclass data into a binary class data classification with techniques for learning One-Versus-One (OVO) [8], [14], [15], then adjust the information imbalance by synthesizing the data. For the synthesis of information, the nearest neighbors will be selected with Euclidean distance techniques, then select the

Manuscript received February 5, 2019; revised April 22, 2019.

P. Chujai is with the Electrical Technology Education Department, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand (e-mail: pasapitchchujai@gmail.com).

K. Chaiyakhan with Computer Engineering Department, Rajamangala University of Technology Isan, Muang, Nakhon Ratchasima, Thailand (e-mail: kedkarnc@hotmail.com).

N. Kerdprasop and K. Kerdprasop are with the School of Computer Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand (e-mail: nittaya.k@gmail.com, kittisakthailand@gmail.com).

doi: 10.18178/ijiee.2020.11.4.1

amount of data to be synthesized from the number of nearest neighbors in the opposite group. The features of the new synthesized data depend on the amount of data in the group. If there is only one sample of data in the group, the new synthesized data will have the characteristics of the data that are similar to the original data set in a variety of formats. At the same time, data groups that have more than one sample data will receive new synthesized information that has features similar to the original data set and neighbor information with the appropriate distance. Then it will bring the original unbalanced data set and the new synthesized data set to classify the data with the SVM algorithm which uses polynomial kernel function. For this section, some parameters will be adjusted so that the algorithm is suitable for learning each data set. Finally, it will test the proposed model with an unknown data set, test the performance and find the reliability of the model obtaining the accepted measurement.

II. Background

A. Imbalanced Data

The general characteristics of unbalanced data is that the data contains a large amount of data for one group rather than the amount of data of the remaining groups. The large number of group data will be called majority class or negative class and the group data with a small amount will be called minority class or positive class. This unbalanced data will affect the basic algorithm of data classification. The existing algorithm will work effectively only when the data is balanced. Whenever there is an imbalance of information, the learning of general algorithms will be biased towards most class data which is causing less incorrect predictions in the class data.

The degree of imbalanced data [16] can be expressed by the ratio between the amount of data of the majority classes and the amount of data of the minority classes in (1).

$$Imbalance\ Ratio\ (IR) = \frac{n_{majority}}{n_{minority}} \quad (1)$$

where $n_{majority}$ is the amount of data for large classes, and $n_{minority}$ is the amount of data for small classes.

The data used in this research is more than two groups, therefore, the IR of each data set is the ratio between the amount of data of the class with the most data and the amount of data of the class with the least data. An IR that has a high value means that the number of data sets have a very different sample number while the balance data will have IR value close to one.

B. Euclidean Distance

Euclidean distance [7] is a technique for measuring the normal distance between two points, $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, p_n)$, with n dimensions. Therefore, the distance between two points of Euclidean distance can be calculated in (2).

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (2)$$

C. Binary Decomposition Strategies

Algorithms for machine learning in the classification of data have been designed to address problems for information that includes both binary class and multiclass. But there are some techniques which when applied to solve the multiclass classification problem, appears to have less efficiency such as the SVM technique. The SVM technique is designed to be used to solve the problem of data with two classes. Therefore, to be able to classify multiclass data which are more complex than the binary classes, then the methods to solve the problem have been proposed [8], [15], [17]. One of those methods is to divide the complexity of the problem with more than two classes of data into a simple sub-problem with only two data classes. This technique is called *binary decomposition strategies*.

For binary decomposition strategies, there are two steps. The first step will be a step to decompose the complex problem into a simple problem by dividing the multiclass problem into binary class problems, then solve the classification problem with the standard techniques for binary classifiers, which the classifier will work independently. The second step will be a step to combine the answers obtained from each sub-problem. For the strategies used in the first step, there are two strategies: One-Versus-One (OVO) and One-Versus-All (OVA). The OVO strategy will divide the multiclass problem with *k* classes into binary class sub-problems, here the total sub-problem is *k(k-1)/*2, while the multiclass problems with OVA strategy are subdivided to binary classes problems with the *k* sub-problems.

The problem for this research is the selection of the problem of classification of multiclass data into a subclass of two class data classification using OVO technique. This means that if there are three classes (*c1, c2, c3*), it can be divided into three sub-problems with binary class (*c1, c2*), (*c1, c3*) and (*c2, c3*). Which means that it is necessary to use three independent classifiers.

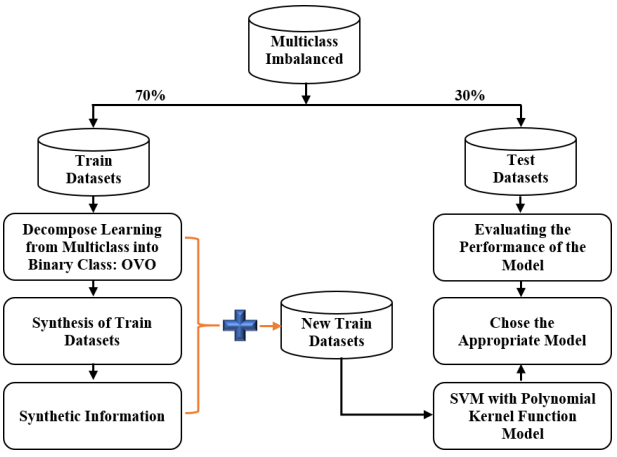

Fig. 1. Conceptual framework of the proposed model: OVO-SynDN.

III. Method

The models presented in this paper will focus on the efficiency and reliability of the multiclass imbalanced data classification. This paper will present the idea of adjusting the imbalance rate by the synthesis of data. The process of synthesizing new data is determined by the amount of data, appropriate parameters and converting the multiclass data into a binary class. The concept of the OVO-SynDN model is shown in Fig. 1.

Fig. 1 illustrates each step as follows.

A. Data Segmentation

The information used in this research is multiclass. The data were divided into two parts. The first 70% were the data set for learning by the model, while the remaining 30% were the data set for performance testing and the reliability of the model presented. Data was selected randomly.

B. Decompose Learning from Multiclass into Binary Class: OVO

In this step, the learning set is introduced. The data set for the test is a subset of learning from a multiclass dataset to a binary class. With the OVO principle, learning is *k(k-1)/2*, where *k* denotes the number of classes in the set. The pattern of the subdivision of learning is shown in Fig. 2.

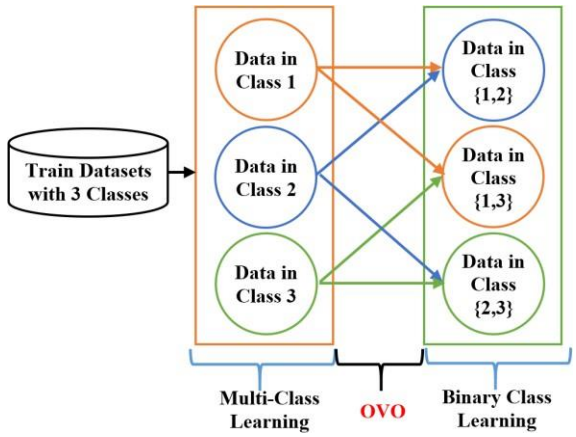

Fig. 2. Binary decompose strategy with OVO technique.

Fig. 3. Synthesis of data set in classes {a, b}.

C. Synthesis of Train Datasets

The information in Fig. 2 in the binary class learning section was used to synthesize information. In this case, *k(k-1)/2* is equal to three rounds (*k =3 classes*). Each round will have the same function as Fig. 3.

For the synthesis of data, the research is based on the amount of data in each class. The criteria are as follows.

Case 1: *Synthesize new information on a both groups*

The first case synthesizes the two new data groups. In this case, both large and small data groups have fewer than 100 samples. From modeling experiments with less than 100 samples, it was found that the model has low efficiency. Therefore, this work will synthesize a set of data with fewer than 100 samples. This work will be considered in two subsections.

***First Subcase*:** amount in minority class or majority class is equal one. In this case, it will randomly select the amount of data to be synthesized. It is not determined by the number of nearest neighbors. The amount of data to be synthesized is randomly selected from the range [*Number of minority data group, number of majority data group*].

***Second Subcase*:** number of minority class or majority class is greater than one. This case is determined by the number of nearest neighbors. The number of closest neighbors is five (*KNN=5*), but the amount of data that will be synthesized will depend on the amount of information in the opposite group. For example, the number of nearby five neighboring data samples. There are three nearby neighbors in the same class and there are two different classes. This work will synthesize the data by the number of nearest neighbors with different classes, which is two. For finding nearby neighbors, the Euclidean distance is used for measurement.

Case 2: *Synthesize new information on a small group*

In this case, most groups will have more than 100 samples. Therefore, new data will be synthesized only in small groups. The amount of information to be synthesized by a minority group is determined by the first and second sub-cases.

For the synthesis of new data, there are two criteria for synthesis.

Case One: *Number of members in the group is one*

In this case, the data that will be synthesized will be similar to the original data. Therefore, the new data obtained is shown in (3).

$$X_i * Threshold_A \tag{3}$$

where $X_i$is the original data that needs to be synthesized.

$Threshold_A$ is a random value which is in the [0,1].

Case Two: Number of members in group more than one

In case of new synthesized data, it will look similar to the original data and neighbor information. Therefore, the new data obtained is shown in (4).

$$X_i * Threshold_B + (\hat{X}_i - X_i) \tag{4}$$

where $\hat{X}_i$ is neighbor's information of X_i

$Threshold_B$ is a random value which is in the range of [-2,2].

D. Creating the OVO-SynDN Model

In this step, the synthesized information is merged into the new training set (from training set: *class{a,b}*). The new

information is called "*New Train Dataset for class{a,b}*". This set of information is learned with the SVM algorithm and uses the polynomial kernel function. In this research, it has chosen the *fitcsvm* algorithm in MATLAB R2018, which has an appropriate parameter adjustment *BoxConstraint*. In this step, it obtains the model of *k(k-1)/2*.

E. Evaluating the Performance of the Model

For model testing, it tests with the unknown test datasets. Because of the unbalanced series, there will be a total number of *k(k-1)/2* models. The researcher chose the right model with the statistical mode function for measuring performance and the reliability of the model is used only for two measures.

IV. Experimental Evaluation

A. Data Sets

The data set used in this study is a set of data from KEEL repository [18], a set of data that has more than two classes. Each data set has an unbalanced appearance, and still has the characteristics of overlapping space. Thirteen sets of unbalanced data will have different details. As in Table I, *#att* is the number of attributes. *#IR* represents imbalance, and *#c1, #c2, ..., #c10* means number of instances in each class.

B. Performance Evaluation

For measuring performance and reliability of the proposed model, using two measurements

1) Marco Average Arithmetic (MAvA):

$$= \frac{1}{m}\sum_{i=1}^{m} ACC_i \tag{5}$$

where ACC_i is the accuracy rate of the class i

2) Mean F-Measure (MFM)

$$= \frac{1}{m}\sum_{i=1}^{m} F-Measure_i \tag{6}$$

where $F-Measure_i$ is in (7)

$$F-Measure_i = \frac{2 * recall_i * precision_i}{recall_i + precision_i} \tag{7}$$

TABLE I: Displays Details of Unbalanced Data Sorted by Unbalance Rate

Dataset	#att	#IR	(#c1, #c2, #c3, #c4, #c5, #c6, #c7, #c8, #c9, #c10)
Pen-Based	16	1.10	(115, 114, 114, 106, 114, 106, 105, 115, 105, 106)
Wine	13	1.48	(59, 71, 48)
Hayes-Roth	4	1.70	(51, 51, 30)
Contraceptive	9	1.89	(629, 333, 511)
New Thyroid	5	5.00	(35, 30, 150)
Dermatology	34	5.60	(112, 61, 72, 49, 52, 20)
Balance	4	5.88	(49, 288, 288)
Glass	9	8.44	(70, 76, 17, 13, 9, 29)
Thyroid	21	39.18	(17, 37, 666)
Ecoli	7	71.50	(143, 77, 2, 2, 35, 20, 5, 52)
Yeast	8	92.60	(463, 5, 35, 44, 51, 163, 244, 429, 20, 30)
Page Blocks	10	164.00	(492, 33, 3, 8, 12)
Shuttle	9	853.00	(1706, 2, 6, 338, 123)

TABLE II: Details the Conversion of a Multiclass to A Binary Class with OVO Learning Principles: An Example Is the Shuttle Data Set

Group No.	Amount Data	Group No.	Amount Data
(#c1,#c2)	(1194,1)	(#c2,#c4)	(1,237)
(#c1,#c3)	(1194,4)	(#c2,#c5)	(1,86)
(#c1,#c4)	(1194,237)	(#c3,#c4)	(4,237)
(#c1,#c5)	(1194,86)	(#c3,#c5)	(4,86)
(#c2,#c3)	(1,4)	(#c4,#c5)	(237,86)

C. Results and Analysis

The detailed results obtained from the proposed model are as follows.

The First Part: *Results from the selection of data to be synthesized.*

Here are some examples of shuttle data. This information will have five classes. In order to find the amount of data to be synthesized, it is necessary to convert the learning from the multiclass to the binary class with OVO Learning Principles. Therefore, this dataset can be grouped into ten new learning groups. The details are shown in Table II.

From Table II *(#c1,#c2)(1194,1)* is the data in classes *#c1* and *#c2* by class *#c1* has 1194 samples, and *#c2* has only one sample. Here, *#c1* is a majority class and *#c2* is a minority class. Considering both data classes, *#c1* does not synthesize data. It is because there are more than 100 samples, while *#c2* contains less than 100 samples, additional data needs to be synthesized, *#c2* has only one instance. In this research, if the data in the class contains only one sample the researcher will randomly select the amount of data to rebuild. The amount of data is in range [*Number of minority class, Number of majority class*]. In this section, it will synthesize the new data by random number [*1,1194*]. For new synthesized data, it has the same attributes as the original data. The new synthesized data is derived from (3).

In the case of *(#c3,#c5)(4,86)*, it is found that both classes have more than one but not more than 100 samples. This results in the synthesis of both groups. Each information will have up to five nearest neighbors. However, the amount of new information that will be synthesized depends on the number of nearest neighbors in different groups. The new synthesized data is derived from (4).

The case of *(#c4,#c5)(237,86)* found that *#c4* which is majority class has more than 100 instances and *#c5* is a minority class. It will only synthesize minority class data. The new synthesized data is derived from (4).

TABLE III: Shows the Appropriate Threshold Values for Each New Synthesis Data

Dataset	Appropriate Threshold
Balance	0.80
Contraceptive	-0.60
Dermatology	-1.40
Ecoli	-1.00
Glass	1.00
Hayesroth	1.91
Newthyroid	1.60
Pageblocks	1.86
Penbased	1.00
Shuttle	1.90
Thyroid	2.00
Wine	0.80
Yeast	1.56

The Second Part: *The results of the selection of the appropriate threshold*

There are two values of threshold used in this research. The first threshold used in (4) is the appropriate threshold for generating new synthesized data. This value is used to synthesize a minority class or a majority class with more than one instance. The appropriate threshold is in the range [-2,2]. The details are shown in Table III.

The second Threshold for parameters BoxConstraint of fitcsvm algorithm in the MATLAB R2018 program is shown in Table IV.

The Third Part: *The results from the proposed model*

For the classification of data, the imbalance is multiclass with the proposed model. In each data set, there is a model with all binary learning *k(k-1)/2*. The model is best chosen with a statistical value of mode. In addition, the results from the proposed model are compared with the other four methods. The first two models do not synthesize the data, but there will be learning algorithms with *fitcecoc* type: *onevsone* (OVO_Easy) and *onevsall* (OVA_Easy). The other two models are OVO_SMOTE1 and OVO_SMOTE2. There will be a similar approach to the proposed model. The difference is that the OVO_SMOTE1 model will only synthesize the information contained in the minority. While the OVO_SMOTE2 model will synthesize data, it will only synthesize the information contained in the minority class. That means only one group is a majority class and all the others are the minority class. When synthesizing new data, we will have a set of data for a balanced learning, which is multiclass.

TABLE V: EXPERIMENTAL RESULTS FROM THE PROPOSED MODEL COMPARED WITH OTHER METHODS: *MFM*

Dataset	OVO_ Easy	OVA_ Easy	OVO_ SMOTE1	OVO_ SMOTE2	OVO-SynDN
Balance	96.85	94.11	98.73	98.73	***100.00***
Contraceptive	50.93	51.71	54.51	54.43	***58.35***
Dermatology	95.46	95.61	98.99	***100.00***	***100.00***
Ecoli	59.14	57.77	66.50	61.03	***67.42***
Glass	67.24	68.53	64.54	70.25	***80.46***
Hayes-Roth	84.29	78.42	79.09	82.87	***93.33***
New Thyroid	88.25	93.89	95.71	92.06	***100.00***
Page Blocks	63.15	82.66	91.80	65.01	***93.44***
Pen-Based	98.15	97.28	98.78	98.15	***99.38***
Shuttle	45.99	79.73	99.27	54.78	***100.00***
Thyroid	84.52	68.02	81.47	80.16	***93.61***
Wine	98.29	96.28	92.51	96.28	***100.00***
Yeast	51.89	52.29	54.59	57.19	***60.48***

TABLE IV: SHOWS THE APPROPRIATE THRESHOLD VALUES FOR THE PARAMETERS BOXCONSTRAINT OF FITCSVM ALGORITHM

Dataset	Appropriate BoxConstraint
Balance	0.30
Contraceptive	5.00
Dermatology	0.70
Ecoli	1.10
Glass	1.90
Hayesroth	0.12
Newthyroid	0.90
Pageblocks	4.40
Penbased	0.50
Shuttle	2.50
Thyroid	1.70
Wine	0.40
Yeast	0.42

Regarding the performance and reliability of the classification of unbalance data model as multiclass, the method is presented and compared with the other four methods shown in Table V and VI, respectively.

From Table V and VI, when considering the effective MFM and MAvA of the model in the range of 100%, 90%, 80%, and below 80%, as shown in Fig. 4 and 5, the OVO-SynDN model can distinguish all sets of data as unbalanced and the efficiency is higher than other methods. Details are as follows.

TABLE VI: EXPERIMENTAL RESULTS FROM THE PROPOSED MODEL COMPARED WITH OTHER METHODS: *MAVA*

Dataset	OVO_ Easy	OVA_ Easy	OVO_ SMOTE1	OVO_ SMOTE2	OVO-SynDN
Balance	95.17	91.11	99.61	99.61	***100.00***
Contraceptive	51.18	51.63	54.51	54.80	***58.07***
Dermatology	94.58	95.69	99.07	***100.00***	***100.00***
Ecoli	59.91	62.35	67.62	62.58	***68.42***
Glass	70.11	66.33	66.42	73.10	***84.65***
Hayes-Roth	84.45	77.04	78.52	82.96	***93.33***
New Thyroid	88.15	94.82	95.56	91.85	***100.00***
Page Blocks	68.73	88.59	92.59	62.73	***95.59***
Pen-Based	98.15	97.29	98.79	98.17	***99.38***
Shuttle	60.27	80.00	98.87	69.56	***100.00***
Thyroid	80.71	66.98	80.71	80.55	***93.61***
Wine	98.41	96.83	93.65	96.83	***100.00***
Yeast	57.89	55.72	57.91	57.45	***61.57***

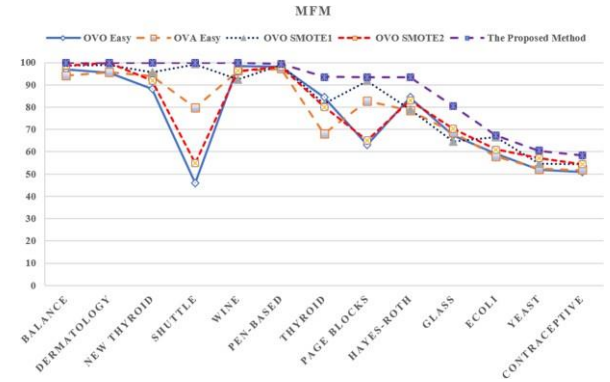

Fig. 4. Performance of the OVO-SynDN model compared to other models: MFM measurement.

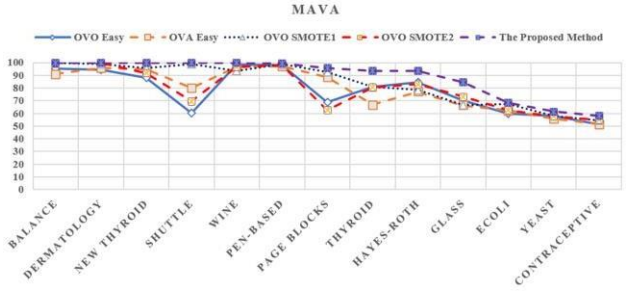

Fig. 5. Performance of the OVO-SynDN model compared to other models: MAvA measurement.

For 100% range of MFM and MAvA it appears that the OVO-SynDN model can best classify in five datasets: Balance, Dermatology, New Thyroid, Shuttle, and Wine data, followed by OVO_SMOTE1 model, OVO_SMOTE2 model, OVO_Easy model and OVA_Easy model, respectively.

For 90% range of MFM and MAvA it appears that the OVO-SynDN model can best classify in four datasets: Pen-Based, Page Blocks, Thyroid and Hayes-Roth data,

followed by OVO_SMOTE1 model, OVO_SMOTE2 model, OVO_Easy model, and OVA_Easy models, respectively.

For 80% range of MFM and MAvA it appears that the OVO-SynDN model can only classify one set of data which is Glass Information, followed by OVO_SMOTE2 model, OVO_Easy model, OVO_SMOTE1 model and OVA_Easy models, respectively.

Lower than 80% range of MFM and MAvA it appears that the OVO-SynDN model could best classify in three datasets: Ecoli, Yeast and Contraceptive data, followed by OVO_SMOTE1 model, OVO_SMOTE2 model, OVA_Easy model and OVO_Easy model, respectively.

In this experiment, there were three datasets: Ecoli, Yeast, and Contraceptive data, the efficiency of the OVO-SynDN model has an MFM and MAvA of less than 70%. However, compared to other methods, it is demonstrated that the proposed method performs better.

V. Conclusion

In this research, the researcher has proposed a model called OVO-SynDN that can effectively classify imbalances in a multiclass. In the model, the learning curve of the multiclass is learned in the form of a binary class. Then, the data in the form of binary class is used to adjust again the rate of imbalance with the synthesis of information. The synthesis of new data is based on the number of members of each class. If both classes are less than 100, both groups will be synthesized. At the same time, if the majority class data is larger than 100, it will only synthesize information on a minority group. In the synthesis, if the data in each group is only one, it will synthesize new data by random number. The synthesized data will be similar to the original data with distance [0,1]. If the data in each group is more than one, it will be synthesized by the number of nearest neighbors in the different class. The synthesized data is same as the original data and the nearest neighbor with the appropriate range [-2,2]. The researcher has applied the SVM algorithm. The polynomial kernel function is used for data classification. The results showed adjusting the imbalance rate with the proposed method can effectively classify multiclass data in the majority and minority groups with high MFM and MAvA.

References

[1] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Prog. Artif. Intell*, vol. 5, no. 4, pp. 221–232, 2016.

[2] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Trans. Syst. Man Cybern. –Part B: Cybern*., vol. 42, no. 4, pp. 1119–1130, 2012.

[3] N. V. Chawla, "Data mining for imbalanced datasets: an overview," *In: Data Mining and Knowledge Discovery Handbook*, Springer US, pp. 875–886, 2012.

[4] M. Mazurowski, P. Habas, and J. Zurada, "Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance," *Neural Netw*, vol. 21, no. 2, pp. 427–436, 2008.

[5] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid- based approaches," *IEEE Trans. Syst. Man, Cybern. C: Appl. Rev*, vol. 42, pp. 463–484, 2012.

[6] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] P. Chujai, K. Chomboon, K. Chaiyakhan, K. Kerdprasop, and N. Kerdprasop, "A cluster based classification of imbalanced data with overlapping regions between classes," in *Proc. the International MultiConference of Engineers and Computer Scientists*, vol. 1, 2017.

[8] Z. L. Zhang, X. G. Luo, S. González, S. García, and F. Herrera, "DRCW-ASEG: One-versus-One distance-based relative competence weighting with adaptive synthetic example generation for multi-class imbalanced datasets," *Neurocomputing*, vol. 285, pp. 176-187, 2018.

[9] S. Piri, D. Delen, and T. Liu, "A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets," *Decision Support Systems*, vol. 106, pp. 15-29, 2018.

[10] J. A. Sáez, J. Luengo, J. Stefanowski, and F. Herrera, "SMOTE–IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," *Information Sciences*, vol. 291, pp. 184-203, 2015.

[11] F. Li, X. Zhang, X. Zhang, C. Du, Y. Xu, and Y. C. Tian, "Cost-sensitive and hybrid-attribute measure multi-decision tree over imbalanced data sets," *Information Sciences*, vol. 422, pp. 242-256, 2018.

[12] Z. Zhang, B. Krawczyk, S. Garcìa, A. Rosales-Pérez, and F. Herrera, "Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced data," *Knowledge-Based Systems*, vol. 106, pp. 251-263, 2016.

[13] G. Collell, D. Prelec, and K. R. Patil, "A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data," *Neurocomputing*, vol. 275, pp. 330-340, 2018.

[14] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, pp. 1761–1776, 2011.

[15] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera, "DRCW-OVO: Distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems," *Pattern Recognition*, vol. 48, no. 1, pp. 28-42, 2015.

[16] S. L. Phung, A. Bouzerdoum, and G. H. Nguyen, "Learning pattern classification tasks with imbalanced data sets," in *P . Yin (Eds.), Pattern recognition, Vukovar, Croatia: In-Teh*, pp. 193-208, 2009.

[17] M. Galar, A. Fernndez, E. Barrenechea, H. Bustince, and F. Herrera, "Aggregation schemes for binarization techniques," *Methods Description, Research Group on Soft Computing and Intelligent Information Systems*, 2011.

[18] Keel Datasets. [Online]. Available: http://www.keel.es/datasets.php

Pasapitch Chujai is a lecturer at the Electrical Technology Education Department, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Thailand. She received her bachelor degree in Computer Science from Ramkhamhaeng University, Thailand, in 2000, master degree in computer and information technology from King Mongkut's University of Technology Thonburi, Thailand, in 2004 and doctoral degree in computer engineering, Suranaree University of Technology, Thailand, in 2015. Her current research includes ontology, recommendation system, time series, machine learning, and imbalanced data classification.

Kedkarn Chaiyakhan is currently a faculty member of the Computer Engineering Department, Rajamangala University of Technology Isan, Thailand. She received her bachelor degree in computer engineering from Rajamangala University of Technology Thanya-buri, Thailand, in 1998, master degree in computer engineering from King Mongkut's University of Technology Thonburi, Thailand, in 2007, and doctoral degree in computer engineering from Suranaree University of Technology, Thailand, in 2016. Her current research includes data mining, machine learning, image classification and image clustering.

Nittaya Kerdprasop is an associate professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. She received her bachelor degree in radiation techniques from Mahidol University, Thailand, in 1985, master degree in computer science from the Prince of Songkla University, Thailand, in 1991, and doctoral

degree in computer science from Nova Southeastern University, U.S.A, in 1999. Her research of interest includes knowledge discovery in databases, artificial intelligence, logic and intelligent databases.

Kittisak Kerdprasop is an associate professor and chair of the School of Computer Engineering, Suranaree University of Technology, Thailand. He received bachelor degree in mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991, and doctoral degree in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes data mining and data science, artificial intelligence, and computational statistics.