การเปรียบเทียบเครื่องจักรเรียนรู้เอ็กซ์ตรีมเพื่อช่วยในการทำนายการเสียชีวิตและการเจ็บป่วยของมารดาตั้งครรภ์ที่มีความไม่สมดุลของข้อมูล

A Comparative Extreme Learning Machine to Predict the Maternal Mortality and Morbidity with Imbalance Data

กษมา ดอกดวง[1*] ศรุตติ อัศวเรืองสุข[2] และปิยภัทร โกษาพันธุ์[3]

Kasama Dokduang[1*] Sarutte Atsawaraungsuk[2] and Piyapat Kosapan[3]

[1, 3]คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี

[1, 3]Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University

[2]คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี อำเภอเมือง จังหวัดอุดรธานี 41000

[2]Faculty of Education, Udon Thani Rajabhat University, Mueang District, Udon Thani, Thailand, 41000

[*]Email: kasama.d@ubru.ac.th

Received: Mar 27, 2020

Revised: Aug 17, 2020

Accepted: Aug 20, 2020

บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อหาขั้นตอนวิธีในการทำนายการเสียชีวิตและความผิดปกติของมารดาที่ตั้งครรภ์ด้วยเทคนิคการสุ่มตัวอย่างแบบผสม กรณีของข้อมูลการเสียชีวิตและความผิดปกติของมารดาที่ตั้งครรภ์จากการสำรวจขององค์การอนามัยโลก ในปี 2007 – 2008 ซึ่งเป็นข้อมูลที่มีความไม่สมดุล ซึ่งความไม่สมดุลนี้เป็นประเด็นที่ท้าทายในการทำเหมืองข้อมูลของนักวิจัย จึงถูกเรียกว่า การเรียนรู้ความไม่สมดุลของคลาส ทำให้ยากต่อการเรียนรู้ด้วยเครื่องและลดประสิทธิภาพของตัวแบบการจำแนกข้อมูล ดังนั้น จึงได้ทำการเปรียบเทียบขั้นตอนวิธีที่หลากหลายในการแก้ปัญหาการจำแนกข้อมูลที่ไม่สมดุล แก้ปัญหาโดยใช้เทคนิคของการสุ่มตัวอย่างแบบรวม ประกอบด้วยเทคนิคการสุ่มสังเคราะห์เพิ่มตัวอย่าง (SMOTE) และเทคนิคการสุ่มลดตัวอย่าง (Tomek Link) จำแนกข้อมูลด้วยขั้นตอนวิธีการจำแนกประกอบด้วย C4.5, โครงข่ายเพอร์เซฟตรอนแบบหลายชั้น (MLP), Naïve Bayes และเครื่องจักรเรียนรู้เอ็กซ์ตรีมประเภทต่าง ๆ ได้แก่ ELM, Circular ELM (CELM), q-Gaussian Extreme Learning Machine (QELM) และ q-Gaussian CELM (QCELM) ในการจำแนกการเสียชีวิตและการเจ็บป่วยของมารดาตั้งครรภ์ วัดประสิทธิภาพโดยใช้ค่าความถูกต้อง และค่า G-mean ผลลัพธ์ที่ได้จะเห็นว่าขั้นตอนวิธี ELM ให้ประสิทธิภาพที่ดีกว่าขั้นตอนวิธีอื่น ๆ จากค่าความถูกต้อง แต่ขั้นตอนวิธี QELM จะให้ประสิทธิภาพที่ดีด้วยตัววัดประสิทธิภาพ G-mean ผลลัพธ์ที่ได้จากการทำนายจะช่วยให้แพทย์ผู้ทำการรักษาใช้เป็นแนวทางในการวางแผนการรักษาผู้ป่วยต่อไป

คำสำคัญ: เครื่องจักรเรียนรู้เอ็กซ์ตรีม ข้อมูลที่มีความไม่สมดุล การเสียชีวิตและความผิดปกติของมารดาตั้งครรภ์ สโมท โทเมค ลิงค์

Abstract

This paper aims to find an algorithm to predict maternal mortality and morbidity with a hybrid sampling technique. Cases of maternal mortality and morbidity data from WHO global survey 2007-2008 which is imbalanced data. This imbalanced data is the subject of challenge in data mining for researchers, called class imbalanced learning (CIL), and leads to difficulties in machine learning and reduces the efficiency of classifiers. Therefore, we have compared several algorithms to handle the imbalanced data classification problem using hybrid sampling technique which consists in SMOTE and Tomek Link used to drive classification models of C4.5, MLP, Naïve Bayes, and the several kinds of Extreme Learning Machine (ELM) as follows: ELM, Circular ELM (CELM), q-Gaussian Extreme Learning Machine (QELM) and q-Gaussian CELM (QCELM) to classify the maternal mortality and morbidity results using Accuracy and G-mean ~~to~~ evaluated performance. Our results demonstrated that ELM show that the ELM algorithm outperformed the other algorithms in Accuracy but QELM outperformed other algorithms by G-mean. This may be helpful in performing clinical assessments.

Keywords: Extreme Learning Machine, Imbalanced Data, Maternal Mortality and Morbidity, SMOTE, Tomek Link

1. Introduction

More than 500,000 of 45 million pregnancy women worldwide die from pregnancy complications [1], [2]. Interestingly, approximately 90 % of maternal mortality and morbidity were found in the developing countries including Thailand. This is caused by the insufficient health service in the rural area.

The maternal mortality and morbidity data were necessary for the intelligence classification model regarding to compare the proficiency between the developed model and physician diagnosis. The data of the WHO global survey 2007-2008 were collected by the Reproductive age mortality survey (RAMOS) [3]. It is the best way to evaluate the maternal mortality and morbidity by collecting the factor of the maternal and morbidity of the woman on the reproductive age.

The dataset of maternal mortality and morbidity data is noisy as unbalanced data when contains more samples from the one class than from the other classes. The data level preprocessing approach is an important process to deal with imbalance data. However, data level approaches are independent of classifiers chosen. We focus on the balanced distribution of classes. The over-sampling approach which increasing the minority class by randomly duplicate of positive samples. The information taken from cases of patients diagnosed with maternal mortality and morbidity data was imbalanced. In supervised classification, the imbalance dataset is a challenge for the research community in data mining and is referred to as Class Imbalance Learning (CIL). The classification tends to be biased towards the majority class and the minority class is more likely to be misclassified [8].

Chawla et al. proposed Synthetic Minority Over-sampling Technique (SMOTE) which the minority class over-sampled by generating synthetic samples rather than replacement [4]. The under-sampling approach is removing the samples of majority class that too close from the decision boundary, which includes the approach of Tomek Link or TLink, One-Sided Selection (OSS), and Neighborhood Cleaning Rule [5]. The hybrid approaches combined two previous approaches as SMOTE+Tomek Link [6].

Sangaroon S. [7] has been using C4.5, MLP, and Naïve Bayes with SMOTE+Tomek Link to comparison performance for risk classification model, used G-mean to analyze the performance measurement imbalance data. It produced better results than not using hybrid approaches.

In order to improve the performance of ELM for imbalanced data learning capability of the classical ELM for imbalance data, several researches have proposed results over the past few years. Guo, S. et al. [9] propose an enhanced oversampling approach called CR-SMOTE to enhance the classification of bug reports with a realistically imbalanced severity distribution. The main idea is to interpolate new instances into the minority category that are near the center of existing samples in that category. Then, we use an *extreme learning machine* (ELM) — a feed-forward neural network with a single layer of hidden nodes — to predict the bug severity. Several experiments were conducted on three datasets from real bug repositories, and the results statistically indicate that the presented approach is robust against real data imbalance when predicting the severity of bug reports.

Extreme learning machine (ELM) proposed by Huang [10], [11] provides better generalization performance at a much faster learning speed than traditional learning algorithms, ELM is not designed to handle the class imbalance problem as it minimizes least-squares error along with regularization to obtain the output layer weights [12]. The distribution of the original data and new data with SMOTE+TL are used to train learning models. Therefore, we propose ELM algorithms with SMOTE+Tomek Link for the imbalance problem in the content of predictive analysis of maternal mortality and morbidity data to compare the performance of a classification algorithm that was proposed previously.

The rest of this paper is organized as follows: section two describes the detail of ELM, CELM, QELM, QCELM, and SMOTE+Tomek Link; section three explains the experimental setup and the experimental results. And the final section is the conclusion.

2. Maternal mortality and morbidity

This research and analytical study is aimed to create and test a model for predicting the maternal mortality and morbidity. The maternal mortality and morbidity data used in the model were collected from WHO global survey 2007-2008. This dataset is nominal attributes and considered two-class maternal mortality and morbidity (class yes) with 176 samples and no maternal mortality and morbidity (class no) with 9,569 samples that cause imbalanced data. This data has imbalance ratio = 48.83

The 9,745 records from the WHO global survey 2007-2008 is selected only Thailand cases. In the study consisted of 18 attributes (the details in table 1).

Table 1 The dataset detail

Order	Attribute name	The values	The number of attributes
1	Maternal age	16-35	8,205
		<=16	1,307
		>=35	233
2	Year of education	<7	2,597
		7-12	5,187
		>12	1,961
3	Primiparaous	Yes	4,698
		No	5,047
4	Birth weight	2,500-4,000	7,074
		<2,500	991
		>4,000	207
5	History of neonatal death or still birth	Yes	96
		No	9,649
6	HIV	Yes	116
		No	9,629
7	Chronic hypertension	Yes	60
		No	9,685
8	Cardiac/Renal diseases	Yes	28
		No	9,717
9	Sickle cell anemia	Yes	30
		No	9,715
10	Other medical conditions	Yes	190
		No	9,715
11	Prelabour rupture of membranes	Yes	726
		No	9,555
12	Pregnancy included hypertension	Yes	234
		No	9,511
13	Pre-eclampsia	Yes	210
		No	9,535
14	Eclampsia	Yes	12
		No	9,733
15	Vaginal bleeding in 2^{nd} half of pregnancy	Yes	53
		No	9,692
16	Any antenatal antibiotic treatment	Yes	196
		No	9,549
17	Referred for complication related to pregnancy or delivery	Yes	648
		No	9,097
18	Maternal mortality and morbidity (answer)	Yes	176
		No	9,569

3. Evaluation matrices

In the two-class case with classes yes and no, a single prediction has the four different possible outcomes shown in Table 2. The overall success rate is the number of correct classifications divided by the total number of classifications [13] (equation 1).

A classifier achieves a good prediction score when it gains high values for both sensitivities or TP rate is divided by the total number of positives, which is TP+FN (equation 1); and specificity or TN rate is divided by the total number of negative, which is TN+FP (equation 4), in which *TN* (true negative), *TP* (true positive), *FP* (false positive), and *FN* (false negative) are defined in table 2. [14].

In this paper, we use Accuracy (equation 1), and G-mean (equation 2) as performance evaluation measures.

Table 2 Confusion Matrices for two-class

		Predictive	
		Positive (class yes)	Negative (class no)
Actual	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

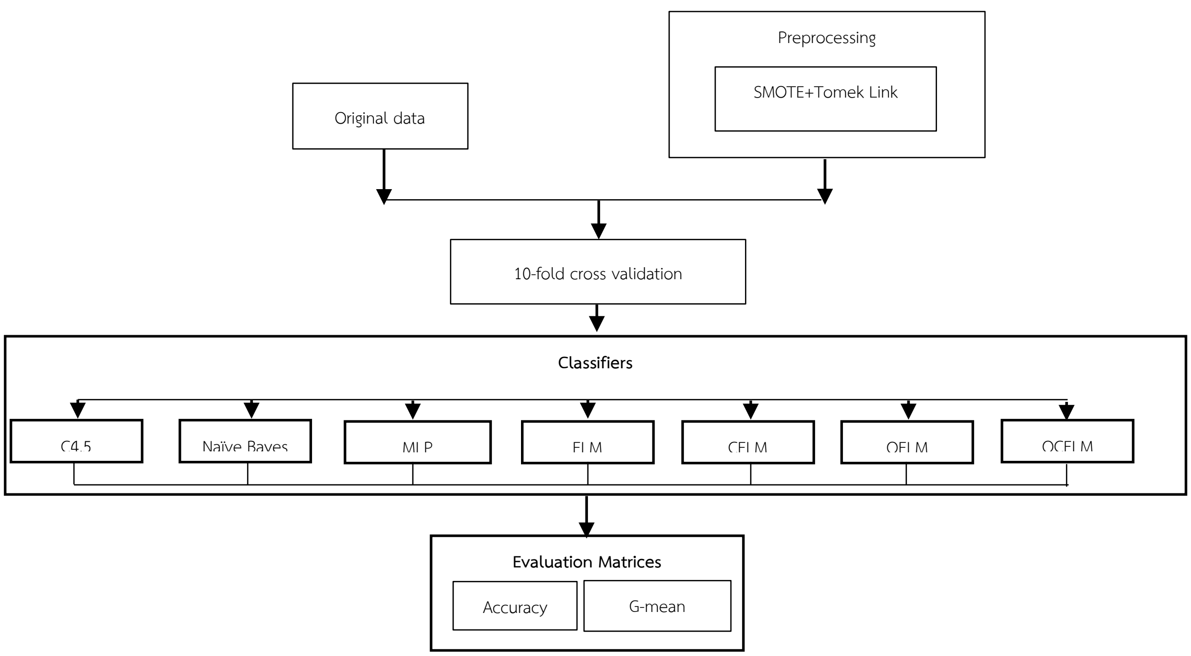


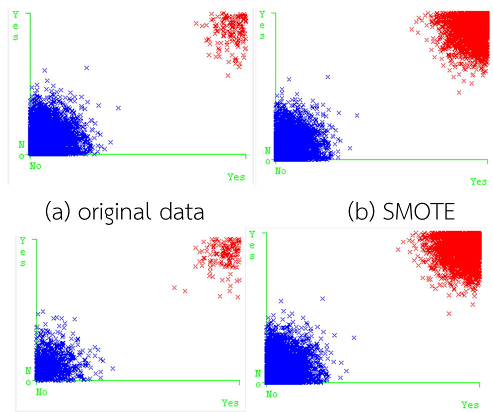
Figure 1 Overview of the Methodology

4. Methodology

Figure 1 shows an overview of classification algorithms with SMOTE+Tomek Link technique.

4.1 Hybrid Over-Sampling and Under-Sampling

SMOTE+Tomek-Link is a hybrid technique, through using SMOTE, removes only those samples that compose the Tomek link. Tomek link consists of two samples that are the nearest neighbors, however, do not belong to the same class. SMOTE technique oversamples the original data set and then detects and removes Tomek links. The result is a balanced data set with well-defined class clusters [15] as demonstrated in Figure 2



(a) original data (b) SMOTE

(c) Tomek Link (d) SMOTE+Tomek Link

Figure 2 Data in borderline of SMOTE+Tomek Link

$$Accuracy=\frac{(TP+TN)}{TP+FN+TN+FP} \qquad (1)$$

$$G\text{-}mean=\sqrt{Sensitivity \times Specificcity} \qquad (2)$$

By $$Sensitivity\ (TP\ rate) = \frac{TP}{TP+FN} \qquad (3)$$

$$Specificity\ (TN\ rate) = \frac{TN}{TN+FP} \qquad (4)$$

4.2 Algorithms for Comparison and Parameter

4.2.1 C4.5

C4.5 is one of statistical classifiers which can be used to generate decision tree. C4.5 developed by Ross Quinlan [16]. It's a top-down decision tree learner extension of the ID3 algorithm. At each node predictive and split node based on information gain as well as ID3.Prune tree; Confidence factor = 0.25, Minimum of instances per leaf = 2.

4.2.2 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a type of neural network with a multi-layered structure. There is a supervised training strategies and a Backpropagation process, which maps given input data to expected target data. MLP consists of multiple layers of nodes in a directed graph, each layer connecting fully to the next [17]. Parameters used in the classifiers include: Learning rate = 0.3, Momentum = 0.2, training time = 500, threshold = 20.

4.2.3 Naïve Bayes

Naïve Bayes is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set, which makes it fast and can handle discrete and continuous attributes. Naive Bayes classification is based on the assumption that the principle of maximum posteriori hypothesis to identify the object that most likely to be classified under the category. Bayes theorem shows the relation between one conditional probability and its inverse [18].

4.2.4 Extreme Learning Machine (ELM)

ELM has been reported by Huang et al. [10][11]. In his report, it has very fast training data because of its structure. The SLFNs is the structure that processes the input data to the result on straight toward only and expand the size of the network or the number of hidden nodes to get the best performance. Let N is the number of samples that come into the ELM, the samples can be written as follows (x_i, t_i), i = 1, 2, ..., N where $x_i=(x_{i1},x_{i2},...,x_{in}) \in R^N$ is training samples and $t = (t_{i1},t_{i2},...,t_{im}) \in R^C$ is the sample target (C is the number of classes). Input layer of ELM contains the input weights and biases that are generated by the random numbers in the range [-1, 1] and [0, 1], respectively and ELM can be written to the least square form as follows:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T} \tag{5}$$

where $\beta = \left[\beta_{j1}, \beta_{j2}, ..., \beta_{jC}\right]^T, j = 1,2,...,K$ is output weights metric.

$\mathbf{H}^{\dagger}$ is the inverse of $\mathbf{H}$ from the Moore–Penrose pseudo inverse by

$$\mathbf{H} = h_{ij} = \mathrm{g}(\mathrm{w} \times x + b) \tag{6}$$

where $w_j = \left[\mathrm{w}_{j1}, \mathrm{w}_{j2}, ..., \mathrm{w}_{jD}\right]^T$ is the input weight and b_j denote the input bias.

Parameters used in the classifiers include: Number of hidden nodes = 500, Activation function = 'sig' (sigmoidal).

4.2.5 Circular Extreme Learning Machine

The Circular Extreme Learning Machine (CELM) [18] is the extended version of ELM which as the same structure as ELM and uses Circular Back Propagation (CBP) [19], [20] architecture. This allows CELM can map both linear and circular decision boundaries. CELM has the same calculation as ELM, but the input formulation is different. The CELM input formulation can be computed as follows:

$$\mathbf{H} = h_{ij} = z_j (\left\| x_i - c_j \right\|^2 - b_j), \ j = 1, 2, ..., K \tag{7}$$

where $$z_j = w_{j,M+1} \tag{8}$$

$$c_j = \left[-\frac{w_{j,1}}{2w_{j,M+1}}, \ldots, -\frac{w_{j,M}}{2w_{j,M+1}} \right] \tag{9}$$

$$b_j = \frac{1}{w_{j,M+1}} \left(\sum_{k=1}^{K} \frac{w_{j,M}^2}{4w_{j,M+1}} - br_j \right) \tag{10}$$

$$w_j = \left[w_{j1}, \mathrm{w}_{j2}, ..., \mathrm{w}_{jK} \right]^T$$

and $w_{j,M+1}$ are the weight vectors.

b_j and br_j refers to the bias vectors.

4.2.6 QELM

QELM [21] is ELM using a different Gaussian activation function called q-Gaussian activation function. This activation function is in the hidden layer of QCELM which can be written as follows:

$$e_q(-\left\| x - a \right\|^2 / b^2) \equiv [1 + (1-q)(-\left\| x - a \right\|^2 / b^2)]^{\frac{1}{1-q}} \tag{11}$$

The success key of QCELM is to vary the parameter q by training many CELMs. Then select the parameter q that takes the CELM get to the highest training accuracy. The parameter q range is divided (follows the experimental in [17]) with heavy-tailed Gaussians for $1 \leq q < 3$ and low-order polynomial functions for q = 0, 0.5.

4.2.7 QCELM

QCELM [22] is QELM in the CELM version that changes the ELM input formulation to CELM input formulation. The formulation can be computed as follows:

$$e_q(z_j \left\| x_i - c_j \right\|^2 - b_j) \equiv [1 + (1-q)(z_j \left\| x_i - c_j \right\|^2 - b_j)]^{\frac{1}{1-q}} \tag{12}$$

4.3 Experimental setup

This section will describe the experimental details. The dataset is maternal mortality and morbidity dataset obtained from the WHO global survey 2007-2008 (the details in table 1) that are used to test the performance of methods. The 9,745 records from WHO global survey 2007-2008 is selected only Thailand cases. MATLAB version R2014a and keel implement the method code. The 10-fold cross-validation method is used to shuffle the train and test data. To finding the best accuracy for each ELM, the optimal number of hidden nodes is selected in the range 1-200.

5. Result and Discussion

The experimental results obtained from the Hybrid Classification method, which is using the method of solving imbalanced data and to analyze the performance of methods: C4.5, Naive Bayes, MLP, ELM, QELM, CELM, and QCELM with SMOTE + Tomek Link and without SMOTE + Tomek Link in the maternal mortality and morbidity dataset. Measured by TP rate or sensitivity, TN rate, G-mean, and Accuracy are shown in table 3 and 4.

The resulting in all datasets have 16,998 samples consisting of two-class maternal mortality and morbidity (class yes) with 8,594 samples and no maternal mortality and morbidity (class no) with 8,404 samples, which make the dataset balanced. The imbalanced ratio is 1.94.

Table 3 Summary of 10 fold cross validation performances of simple Classification

Methods	Non-SMOTE+TL				
	Accuracy	G-mean	TP rate	TN rate	Avg.
C4.5	**98.19**	**0**	**0**	**100**	**49.10**
Naïve Bayes	98.03	0	0	100	49.02
MLP	94.74	0	0	100	47.37
ELM (sigmoid)	97.98	0	0	100	48.99
QELM	97.98	0	0	100	48.99
CELM (sigmoid)	97.98	0	0	100	48.99
QCELM	97.98	0	0	100	48.99

From table 3, showed that all classifiers unable to predict maternal mortality and morbidity, it has a sensitivity or TP rate is 0% but the TN rate is 100%. Despite the higher accuracy, but the G-mean = 0%, indicating that the two classes are very unbalanced.

Table 4 Summary of 10 fold cross validation performances with SMOTE+TL

Methods	SMOTE+TL				
	Accuracy	G-mean	TP rate	TN rate	Avg.
C4.5	86.08	50.85	38.13	67.81	68.47
Naïve Bayes	76.29	59.32	43.55	80.80	67.81
MLP	82.15	49.46	27.26	89.08	65.81
ELM (sigmoid)	**91.68**	58.46	36.84	92.77	**75.07**
QELM	81.06	**65.52**	52.63	81.57	73.29
CELM (sigmoid)	80.90	65.39	52.61	81.46	73.15
QCELM	84.00	61.77	45.00	84.81	72.89

In table 4. The result of the methods with SMOTE+TL, ELM is the highest average accuracy = 75.07 (Accuracy = 91.68, G-mean = 58.46, TP rate = 36.84 and TN rate = 92.77), QELM has average = 73.29 (Accuracy = 81.06, G-mean= 65.52, TP rate = 52.63 and TN rate = 81.57), CELM has average = 73.15 (Accuracy = 80.90, G-mean= 65.39, TP rate = 52.61 and TN rate = 81.46), QCELM has average = 72.89 (Accuracy = 84.00, G-mean= 61.77, TP rate = 45.00 and TN rate = 84.81), C4.5 has average = 68.47 (Accuracy = 86.08, G-mean= 50.85, TP rate = 38.13 and TN rate = 67.81), Naïve Bayes has average = 67.81 (Accuracy = 76.29, G-mean = 59.32, TP rate = 43.55 and TN rate = 80.80) MLP has average = 65.81 (Accuracy = 82.15, G-mean= 49.46, TP rate = 27.26 and TN rate = 89.08).

In general, we tend to expect confidence in the detection of sensitivity and specificity, but when we increase sensitivity, the tests tend to have less specificity, but when we add, the more sensitive detection is often a lower specificity. On the other hand, higher specific tests tend to have the lowest sensitivity. Sensitivity and specificity measurements in this research show that high rates of TP indicate that there is a chance of maternal mortality, while a high TN rate indicates that there is no chance of maternal mortality.

This research studies the classification model to predict the maternal mortality and morbidity. It's consist of C4.5, MLP, Naïve Bayes, and the several kinds of Extreme Learning Machine (ELM) as follows: ELM, Circular ELM (CELM), q-Gaussian Extreme Learning Machine (QELM) and q-Gaussian CELM (QCELM) with hybrid sampling and choose the performance measurement that is suitable for evaluating the performance of the predictive model which is G-mean in order to compare with the measurement of the accuracy. The most suitable model is ELM has the highest average of accuracy and G-mean than Naive Bayes (the method on the previous research proposed [7]) highly to 14.53 percentages.

6. Conclusion

This paper aims to improve the intelligence classification model in the machine learning part. That can predict the maternal mortality and morbidity risk by training the WHO global survey 2007-2008 data. In the machine learning part, we compared algorithm to deal with the data imbalance classification problem by combining technique is the synthetic minority over-sampling (SMOTE) and under-sampling (Tomek Link) technique used to drive classifiers in performance of the several kinds of Extreme Learning Machine (ELM) as follows: ELM, Circular ELM (CELM), q-Gaussian Extreme Learning Machine (QELM) and q-Gaussian CELM (QCELM) using Accuracy and G-mean for evaluates performance. The results of this paper show that the ELM (Sigmoid) algorithm outperformed the other algorithms by average accuracy, but QELM outperformed other algorithms by G-mean.

However, the highest average of accuracy and G-mean indicates that ELM is higher the average of accuracy and G-mean than previous research [7] highly to 14.53 percentages and suggested to use in the intelligence classification model.

7. Acknowledgement. We deeply thank you Professor Dr. Malini Raophaiboon and WHO for maternal mortality and morbidity dataset that make this reseach are successfully.

8. References

[1] World Health Organization (WHO). 1993. **International Statistical Classification of Diseases and Related Health Problem**: 10th Revision.

[2] World Health Organization (WHO). 1996. **Maternal Health and Safe Motherhood Program**. Revised 1990 estimates of Maternal Mortality: A New Approach by WHO and UNICEF. Geneva: WHO.

[3] Walker, G.J and et al. 1986. Maternal mortality in Jamaica. **The Lancet**. 327(8479): 486-488.

[4] Feng, W. and et al. 2019. Dynamic Synthetic Minority Over-Sampling Technique-Based Rotation Forest for the Classification of Imbalanced Hyper-spectral Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.** 12(7): 2159-2169.

[5] Bach, M., Werner, A. and Palt, M. 2019. The Proposal of Undersampling Method for Learning from Imbalanced Datasets. **Procedia Computer Science.** 159: 125-134

[6] Fotouhi, S., Asadi, S. and Kattan, M. W. 2019. A comprehensive data level analysis for cancer diagnosis on imbalanced data. **Journal of Biomedical Informatics.** 90: 103089.

[7] Sangaroon, S. and Wattana, M. 2015. A comparison of data mining technique for risk classification model of maternal mortality and morbidity from WHO global survey 2007-2008. **National Computer Science and Engineering Conference 2015**, Chiang Mai University, Thailand, 23-26 Nov 2015. Chiang Mai, Thailand.

[8] Dokduang, K. and et al. 2014. A Comparative Machine Learning Algorithm to Predict the Bone Metastasis Cervical Cancer with Imbalance Data Problem. In: Boonkrong S., Unger H., Meesad P. (eds) **Recent Advances in Information and Communication Technology. Advances in Intelligent Systems and Computing**, vol 265. Springer, Cham.

[9] Guo, S. and et al. 2019. Identify severity bug report with distribution imbalance by CR-SMOTE and ELM. **International Journal of Software Engineering and Knowledge Engineering.** 29(2): 139-175.

[10] Huang, G. B., Wang, D. H. and Lan, Y. 2011. Extreme learning machines: a survey. **International Journal of Machine Learning and Cybernetics.** 2(2): 107-122.

[11] Huang, G.-B. and et al**.** 2012. Extreme learning machine for regression and multiclass classification. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).** 42(2): 513-529

[12] Raghuwanshi, B. S. and Shukla, S. 2018. Class-specific extreme learning machine for handling binary class imbalance problem. **Neural Networks.** 105: 206-217.

[13] Ri, J., and Kim, H. 2020. G-mean based extreme learning machine for imbalance learning. **Digital Signal Processing.** 98: 102637.

[14] Darzi, M. R. K., Niaki, S. T. A. and Khedmati, M. 2019. Binary classification of imbalanced datasets: The case of CoIL challenge 2000. **Expert Systems with Applications.** 128: 169-186.

[15] Fotouhi, S., Asadi, S. and Kattan, M. W. 2019. A comprehensive data level analysis for cancer diagnosis on imbalanced data. **Journal of Biomedical Informatics.** 90: 103089

[16] Mohammad, A. H., Al-Momani, O., and Alwada'n, T. 2016. Arabic text categorization using k-nearest neighbour, Decision Trees (C4. 5) and Rocchio classifier: a comparative study. **International Journal of Current Engineering and Technology.** 6(2): 477-482.

[17] Zhang, Y. and et al. (2016). A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. **Journal of Medical Systems.** 40(7): 173.

[18] An, Y., Sun, S. and Wang, S., 2017. Naive Bayes classifiers for music emotion classification based on lyrics. **IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)**, pp. 635-638. Wuhan.

[19] Decherchi, S and et al. 2013. Circular-ELM for the reduced-reference assessment of perceived image quality**. Neurocomputing**. (102): 78–89.

[20] Gastaldo, P.and et al. 2002. Circular Back propagation Networks for Measuring Displayed Image Quality. In: **Proceeding of ICANN 2002. LNCS**, vol.2415, pp. 1219-1224. Springer, Heidelberg.

[21] Stosic, Dusan, et al. 2.16. QRNN: $ q $-Generalized Random Neural Network. **IEEE Transactions on Neural Networks and Learning Systems** 28(2): 383-390.

[22] Atsawaraungsuk S. 2016. q-Gaussian activation function Circular Extreme Learning Machine for classification problems. In: **International Conference on Information Technology and Electrical Engineering (ICITEE) 2016**, pp. 125-129, IEEE, Yogyakarta, Indonesia.