

A class skew-insensitive ACO-based decision tree algorithm for imbalanced data sets

Muhamad Hasbullah Bin Mohd Razali¹, Rizauddin Bin Saian², Yap Bee Wah³,
Ku Ruhana Ku-Mahamud⁴

^{1,2,3}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis, Malaysia

⁴School of Computing, Universiti Utara Malaysia, Kedah, Malaysia

Article Info

Article history:

Received Mar 14, 2020

Revised Jun 27, 2020

Accepted Jul 27, 2020

Keywords:

Ant colony optimization

Classification

Decision tree

Hellinger distance

Imbalanced learning

ABSTRACT

Ant-tree-miner (ATM) has an advantage over the conventional decision tree algorithm in terms of feature selection. However, real world applications commonly involved imbalanced class problem where the classes have different importance. This condition impeded the entropy-based heuristic of existing ATM algorithm to develop effective decision boundaries due to its biasness towards the dominant class. Consequently, the induced decision trees are dominated by the majority class which lack in predictive ability on the rare class. This study proposed an enhanced algorithm called hellinger-ant-tree-miner (HATM) which is inspired by ant colony optimization (ACO) metaheuristic for imbalanced learning using decision tree classification algorithm. The proposed algorithm was compared to the existing algorithm, ATM in nine (9) publicly available imbalanced data sets. Simulation study reveals the superiority of HATM when the sample size increases with skewed class (Imbalanced Ratio < 50%). Experimental results demonstrate the performance of the existing algorithm measured by BACC has been improved due to the class skew-insensitiveness of hellinger distance. The statistical significance test shows that HATM has higher mean BACC score than ATM.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Muhamad Hasbullah Mohd Razali

Faculty of Computer and Mathematical Sciences

Universiti Teknologi MARA

Arau, Perlis, Malaysia

Email: hasbullah782@uitm.edu.my

1. INTRODUCTION

Imbalanced class is a condition when a data set consists of one class that has more observations than the others. The severity of imbalanced class can be measured by the imbalanced ratio (IR). For example, a data set with 5% IR indicates the majority class dominates 95% of the dataset's class label. This situation is common when dealing with data-intensive applications such as biomedical, defense and finance [1]. For instance, there were only 5.6% of recurrent disease patients out of 154 follow-up patients reported by Malaysian National Cancer Patient Registry in 2008. The other 94.4% patients are disease free thus considered as less important group. In terms of data mining perspective, it would be beneficial to the medical practitioners to have insights on what the 5.6% recurrent disease patients share in common like age, gender, race or blood type for better decision-making process. Hence, the patients can be treated effectively. This is known as supervised learning in which the machine learning algorithm learn the pattern of the training data and reveal the significant features in the model for prediction of new and unseen data.

Decision tree is one of the models in machine learning which is widely used for classification. This model is well-known for its simplicity and easy to interpret IF-THEN rules since it can be represented in graphical form. As with other machine learning models, decision tree is affected by the imbalanced class issue due to its entropy-based splitting criterion which is biased towards the majority class [2]. Developing a decision tree with significant features that can correctly classify the rare and more important group is a challenging task. Intuitively, most models are misled under imbalanced class environment since the majority group dominates the likelihood of prediction in the learning process. A heuristic approach which utilized bio-inspired behaviour for tree induction such as Ant-Tree-Miner (ATM) is also affected by the skewed class distribution. This is due to the entropy-based heuristic function which suffer from information loss since the majority class dominates the frequently occurring events. This unfavorable behaviour is illustrated in Figure 1. In this experiment setup, the Ionosphere data sets is divided into three sets of class imbalanced ratios (IR) in which the 60%:40% represents the original data sets class distribution while the synthetic domain with 5% IR is a highly imbalanced class skewness. It can be seen that as the class distribution became skewed (IR=5%), the entropy-based heuristic (gain ratio) is shrinking which indicates a significant loss of information thus justified the claim in literature that entropy-based information works well when the class is evenly distributed. Consequently, the developed decision tree with poor attribute evaluation due to biasness towards the dominant group tends to misclassify the rare class in new and unseen data sets as the majority group to achieve high accuracy rate. This is why accuracy rate is not a suitable performance measure for imbalanced class problem since higher accuracy rate does not necessarily means the classifier is good.

For instance, related works by [3-6] focused on the ability of the classifier to detect the rare class which is known as the Minority Class Prediction (MCP). However, in this study we used the improved accuracy measure inspired by the work of [7] as the performance measure called the balance accuracy (BACC) which normalizes the correct prediction from both classes. Misclassification of important and rare class could be disastrous since for instance classifying the recurrent disease patients as disease free is a very costly decision. On the other hand, statistical distance which is based on class distribution divergence (Hellinger Distance) is found to be less affected by the class skewness since the prior distribution has no effect during the computation. Hence, Hellinger Distance is found to be a more reliable heuristic measure for attribute evaluation under imbalanced class environment. This is the main motivation of this study due to the fact that heuristic function in bio-inspired algorithms such as ATM is a crucial weighting element to bias the ants' path creations in the probabilistic tree induction process.

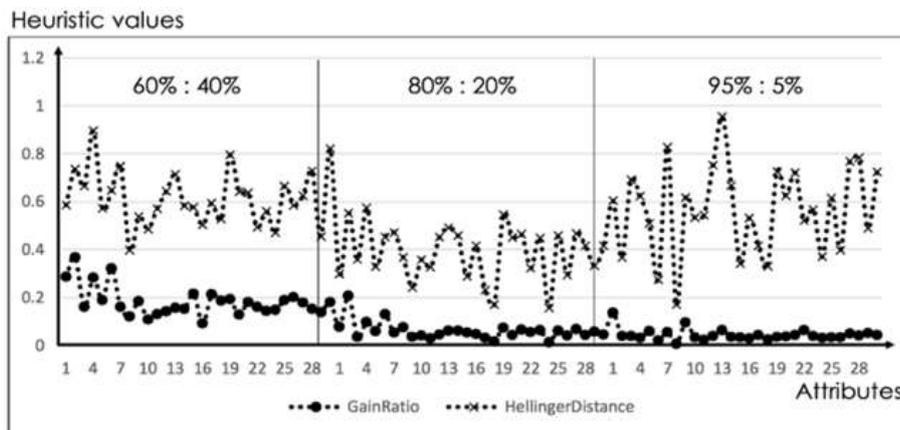


Figure 1. Synthetic ionosphere data sets

2. RELATED WORKS

2.1. Classification of imbalanced data sets

Imbalanced learning is about reducing the impact of imbalanced class either via algorithmic or resampling approach. Algorithmic approach is more promising since less effort is needed on pre-processing the data even-though the solution needed is quite complex. This approach will impose a bias on the rare class by adapting the decision threshold since most machine learning (ML) models such as support vector machine (SVM) and decision tree are affected by this condition. For instance, [8] proposed a new bias for imbalanced classification by adjusting the function learned by the SVM to improve the standardized error measures based

on sensitivity and g-means. On the other hand, resampling approach involve removing instances from majority class or duplicating instances from the minority class. Both techniques are known as under-sampling and over-sampling respectively. Latest work by [9] introduced an efficient under-sampling algorithm by calculating the membership probabilities to the majority class in order to eliminate sample with the lowest membership probability. Their work was motivated by the fact that resampling technique for imbalanced data sets need to be conducted carefully since it could potentially introduce noise or lose information during the learning phase.

The most prominent work that modify the decision tree algorithm to be skew-insensitive is by [2]. The algorithm is known as Hellinger-Distance-Decision-Tree (HDDT) which is based on the conventional decision tree algorithm but with skew-insensitive splitting criterion called the Hellinger Distance. Due to the greedy nature of conventional decision tree, [10] developed a hybrid of the swarm intelligence technique known as the Ant Colony Optimization (ACO) with the decision tree algorithm called Ant-Tree-Miner (ATM). Their work is an extension of algorithm by [11], the Ant-Miner which only extract classification rules from data sets. ATM is better than the conventional decision tree since the optimization approach has an advantage to induce a decision tree without being trapped in local optima due to its iterative way of searching for global best tree.

Meanwhile, the impact of imbalanced class can be relaxed by hybrid technique which combines the algorithmic and sampling approach simultaneously. In [12] presented a new hybrid sampling and boosting algorithm, called RUSBoost, for learning from skewed training data which provides a simpler and faster alternative to SMOTEBoost [4]. SMOTEBoost is an algorithm that combines boosting and data sampling for skewed data sets. In both techniques, the main idea is to balance the classes using a random sampling procedure before each round of boosting. Another hybrid approach by [13] proposed an algorithm called GA-EoC in handling the imbalanced class problem in Alzheimer's dataset. The algorithm utilized the genetic algorithm-based search method to find the optimum combination from a pool of base classifiers to form a heterogenous ensemble along with random sub-sampling approach to balance the class distribution. In [14] developed a cost sensitive feature selection algorithm that adds the cost-based evaluation function of a filter feature selection using a chaos genetic algorithm namely CSFSG. This approach is able to consider both feature acquiring cost (test costs) and misclassification cost which is crucial in imbalanced class learning.

2.2. Bio-inspired algorithms for imbalanced learning

The idea of swarm intelligence (SI) which is a branch of artificial intelligence (AI) for machine learning is gaining interest since it combines the heuristic and pheromone measure for various applications. For example, [15] hybrid the honeybee mating optimization (HBMO) behavior with the Pi-Sigma neural network (PSNN) for classification task. An approach by [16] found that by embedding the circle chaotic search of whale optimization algorithm (WOA) in the searching iterations, they were able to effectively select relevant features of breast cancer data. An interesting application by [17] combined the fruit fly optimization algorithm (FFOA) as the features selection method with multi-kernel support vector machine (MK SVM) for classification of chronic kidney disease (CKD). Thus, it can be concluded that there are various applications of SI for ML offered in the literature since SI is able to compensate the imprecisions of local attribute evaluation by most machine learning models.

Meanwhile, imbalanced learning via bio-inspired algorithm is gaining a significant attention in data mining recently. A recent work by [18] proposed the use of entropy and information gain as a fitness function in genetic algorithm with an aim to improve the impurity and gives a more balanced result without changing the original dataset. Earlier work by [19] proposed a cost-sensitive deep neural network, which can automatically learn robust feature representations for both the majority and minority classes. The proposed approach is applicable to both binary and multiclass problems without any modification to the original data distribution, which results in a lower computational cost during the training process. The results on six major image classification data sets show that the proposed approach significantly outperforms the popular data sampling techniques. Another work by [20] introduced bio-inspired kernels in twin support vector machine (Bio-TWSVM) for prediction of imbalanced multiclass protein data sets. There is also possibility to merge SI with the sampling technique. For instance, [21] combined ACO with under-sampling for imbalanced DNA microarray data classification using SVM.

Based on the previous works, there were significant amount of efforts been taken on hybridizing SI with ML models particularly to alleviate the impact of imbalanced class problem. However, such strategy is still in its infancy. Thus, this is the main motivation of our study to improve the ACO-based decision tree by hybridizing Hellinger Distance in the algorithm. The aim is to improve the ants' heuristic measure so that the constructed decision tree is insensitive towards class skewness thus able to predict the minority class well. The proposed algorithm is called hellinger-ant-tree-miner (HATM).

3. METHODOLOGY

This study extends the work of [10] by hybridizing statistical distance called Hellinger Distance in inducing the decision tree with ACO algorithm. This new algorithm will combine Hellinger Distance and pheromone to provide a precise response of the value of a feature considering its influence in a whole decision tree under imbalanced class domains. Thus, it is inferred to compensate the skew sensitiveness of the conventional information gain measure as the ant’s bias component towards the promising paths during the tree constructions process. The proposed HATM algorithm follows the traditional structure of ATM algorithm in Figure 2. HATM algorithm starts by initializing the pheromone values and computing the Hellinger Distance for each attribute of the training set (Figure 3). Then, it enters in an iterative loop (while loop) where each ant in the colony creates a new decision tree until a maximum number of iterations is reached or the algorithm has converged. An ant creates a decision tree (for loop) in a top-down fashion by probabilistically selecting attributes to be added as decision nodes based on the amount of pheromone (τ) and heuristic information (η). The probability of an ant to visit the attribute vertex is given by (1).

The algorithm of Compute Hellinger Distance() function returns the Hellinger Distance of the attributes as in Figure 3. This algorithm will measure the features’ distributional divergence by quantifying the similarity between the majority and minority probability distribution in discriminating the class. This function also returns the Hellinger Distance for each feasible attribute in the decision tree induced during the training stage. Note that the implementation of this function is slightly different to the one applied in Hellinger-Distance-Decision-Tree (HDDT) in such a way that in HATM, the Hellinger Distance is used as the heuristic function while HDDT returns the highest Hellinger distance as the splitting criterion.

$$P_i = \frac{\tau^\alpha(E, L, x_i) \eta_i^\beta}{\sum_{i \in F} \tau^\alpha(E, L, x_i) \eta_i^\beta}, \forall_{i \in F} \tag{1}$$

where

- a) $\tau_{(E,L,x_i)}$ is the amount of pheromone associated with the entry (E, L, x_i) – E is the attribute condition represented by the edge being followed or ‘-’ at the start of the construction procedure, L is the ant’s current level in the decision tree or 0 at the start of the construction procedure, x_i is the i -th attribute vertex of the construction graph-in the pheromone matrix.
- b) η is the heuristic information (Hellinger distance) of the i -th attribute.
- c) F is the set of available (feasible) attributes for selection.
- d) Parameter α and β used to control the influence of the pheromone and heuristic information, respectively, during the selection of vertices.

```

Input: training examples, list of predictor attributes
Output: best discovered tree
1. InitialisePheromones ();
2. ComputeHellingerDistance ();
3. treegb ← ∅;
4. m ← 0;
5. while m < maximum iterations and not CheckConvergence () do
6.   treeib ← ∅;
7.   for n ← 1 to colony size do
8.     treen ← CreateTree (Examples, Attributes, -);
9.     Prune(treen);
10.    if Q(treen) > Q(treeib) then
11.      treeib ← treen;
12.    end if
13.  end for
14.  UpdatePheromones(treeib);
15.  if Q(treeib) > Q(treegb) then
16.    treegb ← treeib;
17.  end if
18.  m ← m+1;
19. end while
20. return treegb;
    
```

Figure 2. Pseudocode of the hellinger-ant-tree-miner algorithm

```

Input: Training set T, Feature f
Output: Hellinger Distance, HD
1. Let Hellinger ← -1;
2. Let Vf be the set of values of feature f;
3. for each value v ∈ Vf do
4.   Let w ← V \ v
5.   HD ← ( √(|Bf,w+1| / |R+|) - √(|Bf,w-1| / |R-|) )2 + ( √(|Bf,w+1| / |R+|) - √(|Bf,w-1| / |R-|) )2;
6. end for
7. return √Hellinger;
    
```

Figure 3. Pseudocode of compute hellinger distance()

3.1. Simulation procedure

In this study, data X were simulated using R software and assigned to group 1 or 0 using the following logistic regression model:

$$P(Y = 1) = \frac{1}{1 + e^{-z}} \quad (2)$$

where $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ and k is the number of features. Since decision tree construction is a process of rule induction which involve determining those features important for classification, hence the predictors (or features) were set as significant features with odds-ratio greater than 1. The value of the regression coefficients was set at an incremental factor of 0.1 starting from 0.5 up to 1.4 to generate 10 features. 10 features are generated since this is to avoid poorly grown tree if too few features involved. Hence, a covariate with a coefficient of 1.2 will gives a significant odds ratio (OR) of 3.32012 for X ($OR = e^{1.2} = 3.32012$). We considered six imbalanced ratios (IR): 5%, 10%, 20%, 30%, 40% and 50% where IR of 5% reflects a highly imbalanced class in the response variable. The full multivariate logistic regression (MLR) is presented in (3):

$$z = \beta_{0k} + 0.5x_1 + 0.6x_2 + \dots + 1.4x_{10} \quad (3)$$

where β_{0k} is varied according to the IR. The distribution of the covariates (X) considered in this study is the standard normal distribution, $N(0,1)$. The sample sizes were generated for 1000, 5000 and 10000 subjects in order to permit the investigation of the learning ability of both algorithms as sample size increases. The simulated data is then split into 90% train and 10% test by using tenfold cross validation strategy which consists of dividing the data set into ten stratified partitions of examples. Each partition has a similar IR. The 90%:10% splitting strategy is practical for imbalanced class learning in order to allow the decision trees to sufficiently represent the under-represented group (rare class) to be sampled during the decision tree construction procedure [22]. Finally, based on the confusion matrix in Figure 4, the balance accuracy (BACC) as shown in (4) is averaged after 5 repetitions over various sample sizes and IRs to avoid noise.

	Pred.Negative	Pred.Positive	
Act.Negative	True negative (TN)	False positive (FP)	$N = TN + FP$
Act.Positive	False negative (FN)	True positive (TP)	$P = FN + TP$

Figure 4. Confusion matrix

$$BACC = \frac{\left(\frac{TP}{P} + \frac{TN}{N}\right)}{2} \quad (4)$$

3.2. Applications to real data sets

The summary of the data sets used is presented in Table 1. The data sets have various imbalanced ratio (IR) ranging from 1.4% to 40%. The most severe class IR is the Bioassay_AID362 data set with 144 features of size 4279 while both Diabetes-Frankfurt and Diabetes-Pima have the least number of features. Hepatitis is the smallest data set which consist of 155 instances. The class labels are presented in the parentheses. Note that both Bioassay data sets are highly skewed with IR less than 15%. These data sets require a data pre-processing in order to avoid poorly structured trees since it is learnt that the minority class is not sufficient to provide effective decision boundaries for the tree algorithms to learn the rare class pattern. Thus, the synthetic minority oversampling technique (SMOTE) was applied using WEKA software in order to improve the IR to be 30% since thorough experiments show that this IR is reasonable to provide information for learning experience during the training phase. The experiment was initiated by generating a 10-folds cross-validation to split the data sets into 90% train and 10% test domains. Then, all continuous features, if there are any, were discretized based on the minimum description length (MDL) principle by [23]. The stochastic algorithms were run 5 times for all 10-folds data sets and finally the BACC was averaged and

validated. Both algorithms were run with the default parameter values as depicted in Table 2 due to the good performance than other combinations [10].

Table 1. Summary of the data sets

Data sets	Majority class freq.	Minority class freq.	Size	No of features	IR
1 Bioassay_AID1284	305 (Inactive)	57 (Active)	362	103	15%
2 Bioassay_AID362	4221 (Inactive)	60 (Active)	4279	144	1.4%
3 Diabetes_Frankfurt	1316 (Non-diabetic)	684 (Diabetic)	2000	8	30%
4 Diabetes_Pima	500 (Negative)	268 (Positive)	768	8	30%
5 Hepatitis	123 (Live)	32 (Die)	155	19	20%
6 Horse	232 (No)	136 (Yes)	368	26	40%
7 Ionosphere	225 (Good)	126 (Bad)	351	34	40%
8 Statlog-Heart	150 (Absence)	120 (Presence)	270	13	40%
9 Vote	267 (Democrate)	168 (Republican)	435	16	40%

Table 2. Parameters setting

Parameters	Values
Colony size	5
Max iterations	500
Min case	3
Stagnation	40
Evaporation rate	0.9

4. RESULTS AND ANALYSIS

4.1. Balance accuracy of simulation and real data sets

In order to compare the performance of ATM and HATM, both algorithms were fitted on the simulation data sets generated in Section 3.1 and nine (9) publicly available data sets from UCI machine learning repository [24] as summarized in Table 1. The average BACC of 10-folds cross validation for both domains were recorded in Table 3 and Table 4 respectively.

Table 3. Balance accuracy (BACC) of simulation data sets

No	Data set	ATM	HATM
1	Sim1_1000_5%	*50.5 ± 0.3	*50.5 ± 0.3
2	Sim2_1000_10%	61.6 ± 0.8	62.0 ± 0.9
3	Sim3_1000_20%	70.1 ± 0.6	72.3 ± 0.7
4	Sim4_1000_30%	73.5 ± 0.4	73.9 ± 0.5
5	Sim5_1000_40%	71.1 ± 0.4	71.5 ± 0.3
6	Sim6_1000_50%	70.8 ± 0.4	69.2 ± 0.4
7	Sim7_5000_5%	55.8 ± 0.2	55.5 ± 0.4
8	Sim8_5000_10%	58.2 ± 0.3	57.2 ± 0.3
9	Sim9_5000_20%	67.2 ± 0.2	67.0 ± 0.3
10	Sim10_5000_30%	*71.1 ± 0.3	*71.1 ± 0.3
11	Sim11_5000_40%	73.1 ± 0.2	74.0 ± 0.3
12	Sim12_5000_50%	73.9 ± 0.2	73.4 ± 0.2
13	Sim13_10000_5%	*50.0 ± 0.0	*50.0 ± 0.0
14	Sim14_10000_10%	56.1 ± 0.2	56.6 ± 0.2
15	Sim15_10000_20%	64.3 ± 0.1	64.8 ± 0.3
16	Sim16_10000_30%	69.1 ± 0.2	69.5 ± 0.2
17	Sim17_10000_40%	72.1 ± 0.2	72.2 ± 0.1
18	Sim18_10000_50%	74.3 ± 0.2	73.3 ± 0.1

*Ties

*Bold values correspond to algorithm with higher BACC

Table 4. Balance accuracy (BACC) of real data sets

No	Data set	ATM	HATM
1	Bioassay_AID1284 + SMOTE	84.1 ± 0.4	85.7 ± 0.3
2	Bioassay_AID362 + SMOTE	93.2 ± 0.3	93.4 ± 0.2
3	Diabetes-Frankfurt	81.4 ± 0.04	81.2 ± 0.03
4	Diabetes-Pima	73.2 ± 0.5	73.3 ± 0.5
5	Hepatitis	66.8 ± 1.3	65.9 ± 1.2
6	Horse	*86.2 ± 0.5	*86.2 ± 0.5
7	Ionosphere	88.7 ± 0.3	90.5 ± 0.4
8	Statlog Heart	82.0 ± 0.4	82.3 ± 0.5
9	Vote	95.3 ± 0.4	95.9 ± 0.4

*Ties

*Bold values correspond to algorithm with higher BACC

4.2. Statistical significance test

Based on Table 3 and Table 4, both ATM and HATM were compared for significance different using the nonparametric Sign Test [25]. The null hypothesis to be tested is $H_0 : \mu_{ATM} \geq \mu_{HATM}$ against the alternative hypothesis $H_1 : \mu_{ATM} < \mu_{HATM}$. This make the significance test a one-tailed test with significance level $\alpha = 0.05$. From both tables, we see that HATM wins 15 times over ATM, ties 4 times, and loses 8 times. Based on the table of critical values for the Sign Test, for 27 data sets, a classifier needs to win at least 11 times for the null

hypothesis to be rejected with significance level $\alpha=0.05$ for one-tailed test. Thus, since HATM wins 15 times, in this case, the null hypothesis can be rejected. We conclude that there is sufficient evidence to reject the null hypothesis thus indicating the HATM algorithm has a higher mean of BACC than ATM.

5. CONCLUSION

This study has demonstrated the hybrid of Hellinger Distance as new heuristics measure in the existing ATM algorithm is able to improve the learning experience of ATM in imbalanced class data sets. The Hellinger Distance is reliable for its skew-insensitiveness where improved trees' performance measured by BACC can be induced. The simulation study reveals that HATM performed well for imbalanced class data sets while ATM consistently induced higher BACC decision tree in all balanced data sets. This is expected since entropy-based heuristic put more weightage on the features which can separate the class evenly while suffered from information loss in imbalanced learning environment. Further applications on real data sets demonstrate HATM is statistically significance for imbalanced data sets with minimum data pre-processing efforts thus support the initial findings in our simulation procedure. Future research could used other statistical distance-based functions as the alternative heuristic measure.

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Higher Education Malaysia for funding this study under Fundamental Research Grant Scheme, FRGS/1/2018/STG06/UITM/02/3 and Research Management Centre (RMC), Universiti Teknologi MARA, Malaysia for the administration of this study.

REFERENCES

- [1] H. He and Y. Ma, "Imbalanced learning: foundations, algorithms, and applications," Hoboken, NJ: IEEE Press, Wiley, 2013.
- [2] D. A. Cieslak, et al., "Hellinger distance decision trees are robust and skew-insensitive," *Data Mining and Knowledge Discovery*, vol. 24, no. 1, pp. 136-158, 2012.
- [3] C. Cardie, "Improving minority class prediction using case-specific feature weights," in *Proceedings of the Fourteenth International Conference on Machine Learning*, pp. 57-65, 1997.
- [4] N. V. Chawla, et al., "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," in *Knowledge Discovery in Databases: PKDD 2003*, Berlin, Heidelberg, pp. 107-119, 2003.
- [5] T. M. Padmaja, et al., "Majority filter-based minority prediction (MFMP): An approach for unbalanced datasets," in *TENCON 2008 - 2008 IEEE Region 10 Conference*, pp. 1-6, 2008.
- [6] B. Krawczyk, et al., "Improving minority class prediction using cost-sensitive ensembles," 2011.
- [7] V. García, et al., "Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions," in *Pattern Recognition and Image Analysis*, Berlin, Heidelberg, pp. 441-448, 2009.
- [8] H. Núñez, et al., "Improving SVM Classification on Imbalanced Datasets by Introducing a New Bias," *Journal of Classification*, vol. 34, no. 3, pp. 427-443, 2017.
- [9] G. Ahn, et al., "A Membership Probability-Based Undersampling Algorithm for Imbalanced Data," *Journal of Classification*, 2020.
- [10] F. B. Otero, et al., "Inducing decision trees with an ant colony optimization algorithm," *Applied Soft Computing*, vol. 12, pp. 3615-3626, 2012.
- [11] A. Freitas, et al., "Ant colony algorithms for data classification," *Encyclopedia of Information Science and Technology*, vol. 1, pp. 154-159, 2009.
- [12] C. Seiffert, et al., "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 1, pp. 185-197, 2010.
- [13] M. N. Haque, et al., "Heterogeneous Ensemble Combination Search Using Genetic Algorithm for Class Imbalanced Data Classification," *PLOS ONE*, vol. 11, no. 1, p. e0146116, 2016.
- [14] J. Bian, et al., "An Efficient Cost-Sensitive Feature Selection Using Chaos Genetic Algorithm for Class Imbalance Problem," *Mathematical Problems in Engineering*, vol. 2016, pp. 1-9, 2016.
- [15] J. Nayak and B. Naik, "A Novel Honey-Bees Mating Optimization Approach with Higher order Neural Network for Classification," *Journal of Classification*, vol. 35, no. 3, pp. 511-548, 2018.
- [16] G. I. Sayed, et al., "A New Chaotic Whale Optimization Algorithm for Features Selection," *Journal of Classification*, vol. 35, no. 2, pp. 300-344, 2018.
- [17] L. J. Rubini and E. Perumal, "Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm," *International Journal of Imaging Systems and Technology*, 2020.
- [18] M. A. U. H. Tahir, et al., "A Classification Model For Class Imbalance Dataset Using Genetic Programming," *IEEE Access*, vol. 7, pp. 71013-71037, 2019.
- [19] S. H. Khan, et al., "Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data," *arXiv:1508.03422*, 2017.

- [20] S. K. Guramand, et al., "Optimized bio-inspired kernels with twin support vector machine using low identity sequences to solve imbalance multiclass classification," *Journal of Environmental Biology*, vol. 40, no. 3(SI), pp. 563-576, 2019.
- [21] H. Yu, et al., "ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data," *Neurocomputing*, vol. 101, pp. 309-318, 2013.
- [22] G. M. Weiss and F. Provost, "Learning when training data are costly: the effect of class distribution on tree induction," *Journal of Artificial Intelligence Research*, vol. 19, no. 1, pp. 315-354, 2003.
- [23] U. M. Fayyad and K. B. Irani, "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," Jet Propulsion Laboratory, California Institute of Technology, 1993.
- [24] D. Dua and C. Graff, "UCI Machine Learning Repository," Irvine, CA: University of California, School of Information and Computer Science, 2019. Available: <http://archive.ics.uci.edu/ml>.
- [25] N. Japkowicz and M. Shah, "Evaluating Learning Algorithms: A Classification Perspective," USA, Cambridge University Press, 2011.

BIOGRAPHIES OF AUTHORS



Muhamad Hasbullah Bin Mohd Razali hold a Bachelor in Statistics (2008) and Master of Applied Statistics (2011) from Universiti Teknologi MARA, Malaysia. Currently he is pursuing PhD in Statistics in Universiti Teknologi MARA, Malaysia.



Rizauddin Saian is currently a Senior Lecturer at Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (Perlis), Malaysia and has been teaching there since October 1999. He received his BSc Honours in Mathematics from Universiti Malaya in 1998, MSc (Information Technology) (Title: Stemming Algorithm in Searching Malay Text) in 2005, and PhD (Title: A Hybrid of Ant Colony Optimization Algorithm and Simulated Annealing for Classification Rules) in 2013 from Universiti Utara Malaysia.



Yap Bee Wah holds a Bachelor in Mathematics from Universiti Sains Malaysia, Malaysia in 1985 and Masters degree in Statistics from University of California, Riverside, USA in 1993. Her PhD in Statistics was obtained from Universiti Malaya, Malaysia in 2001. Her research interests are in Data Mining, Multivariate Analysis, Statistical Modeling, Computational Statistics



Ku Ruhana Ku-Mahamud holds a Bachelor in Mathematical Sciences and a Masters degree in Computing, both from Bradford University, United Kingdom in 1983 and 1986 respectively. Her PhD in Computer Science was obtained from Universiti Pertanian Malaysia in 1994. As an academic, her research interests include ant colony optimization, pattern classification and vehicle routing problem.