# Hybrid Artificial Neural Network with Artificial Bee Colony Algorithm for Crime Classification

Syahid Anuar, Ali Selamat, and Roselina Sallehuddin

Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
syah2105@yahoo.com, {aselamat,roselina}@utm.my

**Abstract.** Crime prevention is an important roles in police system for any country. Crime classification is one of the components in crime prevention. In this study, we proposed a hybrid crime classification model by combining Artificial Neural Network (ANN) and Artificial Bee Colony (ABC) algorithm (codename ANN-ABC). The idea is by using ABC as a learning mechanism for ANN to overcome the ANN's local optima problem thus produce more significant results. The ANN-ABC is applied to Communities and Crime dataset to predict 'Crime Categories'. The dataset was collected from UCI machine learning repository. The result of ANN-ABC will be compare with other classification algorithms. The experiment results show that ANN-ABC outperform other algorithms and achieved 86.48% accuracy with average 7% improvement compare to other algorithms.

## 1 Introduction

Increasing volumes of crime had brought serious problems for community in any country. However, the increase in realization of information technology has opened up new doors for government to include crime prevention component as a strategies to reduce crime. Wilpen Gorrs and Richard Harries have introduced the crime forecasting due to the fact that forecasting crime is still new and not widely practiced by police. The main concern of crime forecasting is to study the crime patterns, to analyzing the modus operandi of serial criminals and to allocate useful resources (for tactical purpose) in the places where there are lots of crimes. With accurate forecasts, police would be able to take tactical actions such as targeting patrols to hot spots, conducting surveillance for deployment of special units, scheduling vacations and training for police officer and making crime alerts available to neighborhood watch groups [2].

Classification is one of the data mining technique which been use to analyze crime patterns. The data mining approach can help to detect the crimes patterns and speed up the process of solving crime [11]. Because of this, research on crime classification has increased because of the potential and effectiveness of classification in crime prevention programs. Some researches in crime classification have been done by several researchers [4, 5, 10]. Crime can be divided into

 
S. Phon-Amnuaisuk and T.-W. Au (eds.), *Computational Intelligence in Information Systems*,
Advances in Intelligent Systems and Computing 331, DOI: 10.1007/978-3-319-13153-5_4

several types and the most common findings at the city level are crimes against property (burglary, robbery and theft) and crime of aggression such as assaults, homicides and rape [3]. For every crime's types, the dataset can be classify into several categories such as low, medium and high. Thus, the objective of this paper is to propose a new classifier to classify the crime categories.

The artificial neural networks (ANN) is most popular computational model and usually use as a benchmark model for any classifier. Learning algorithm is an important aspect that can influence ANN model in producing better results [12]. The ANN with back-propagation (BP) algorithm is a common learning algorithm to minimize mean-squared error between the desired outputs and the actual outputs for the particular inputs to the networks [13]. However, BP has some limitations. First, the result may converge to local optima, meaning that the convergence is not guaranteed [9]. Second, the convergence of back propagation may require many iterations. The shortcomings of BP have resulted on slow learning process or sometimes produce less significance results [12, 13].

To improve the accuracy and the convergence rate of BP algorithm, a new hybrid model called as ANN-ABC is proposed. The goal of the hybridization of artificial bee colony (ABC) and back-propagation algorithm is to optimize the ANN training by avoiding the local minima solution. To validate the performance of the proposed hybrid model, ANN-ABC will be tested on crime classification problem with dataset obtained from UCI machine learning repository [1]. This data has also been used by Iqbal et al. [4] to classify crime using decision tree and Naive Bayes. This work employ machine learning technique, in contrast with work done by Iqbal et al. which was using statistical techniques. The performance of the ANN-ABC algorithm is compared with ANN using standard BP algorithm, Naive Bayes and decision tree algorithm

The rest of the paper is organize as follows. Section 2 discusses the proposed crime classification model, containing the implementation of ABC to train ANN. In section 3, the experiments and results produced by ANN-ABC, ANN, decision tree and Naive Bayes are presented and discussed. In section 4, the statistical test is presented to validate the experiment results.

## 2 Proposed Crime Classification Model

The implementations of ABC algorithm is relatively simple but being able to produce a very good result [8] and has the advantage of not requiring a lot of parameters to be tuned [7]. Figure 1 illustrate the overall process involved in implementation of the proposed hybrid model. Detailed explanation on each process will be described in the next subsection.

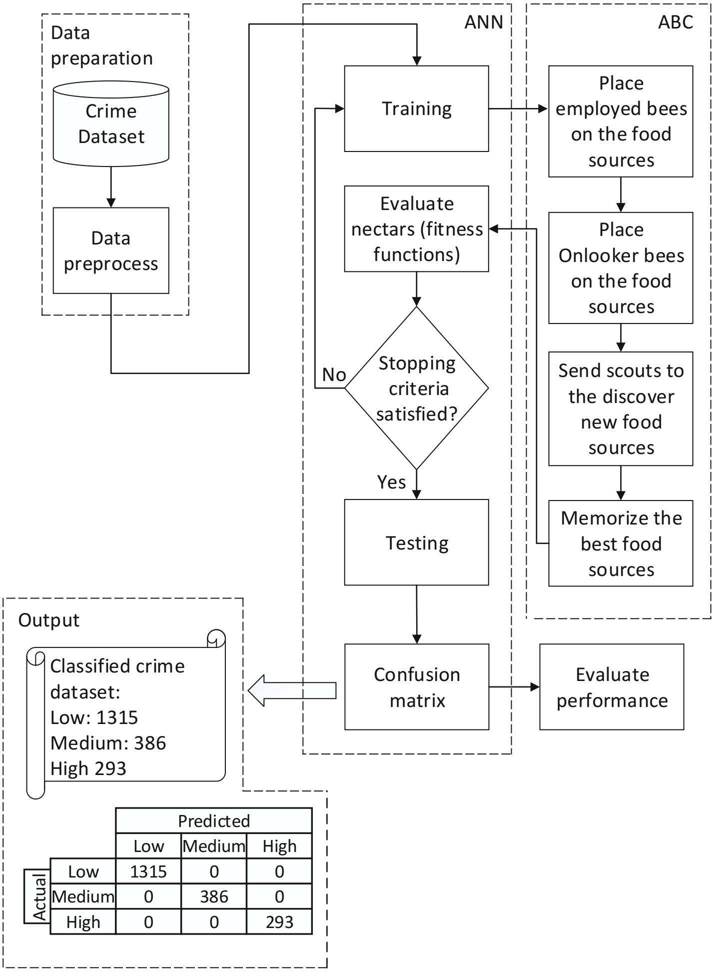


**Fig. 1.** Implementation of Proposed hybrid model ANN-ABC

### 2.1 Data Preparation

The Communities and Crime data set is obtained from the UCI Machine Learning Repository. This dataset focuses on communities in United States of America (USA). The data comprises of socio-economic data from the 90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey and crime data from the 1995 FBI UCR. The dataset consists of 128 number of attributes and 1994 number of instances with missing values [1].

Data preparation is essential for successful data classification. Poor quality data typically result in incorrect and unreliable data classification results. Thus, the following data preparation mechanisms were carried out to obtain the final set of attribute in an appropriate form for further analysis and processing. The data preparation is based on works done by Iqbal et. al [4].

- All the attributes with large number of missing values were removed
- The newly added nominal attribute named Categories is created based on attribute named 'ViolentCrimesPerPop' - if the value is less than 25% than the Categories is Low. If the value is equal to or greater than 25% than the Categories is Medium. If the value is equal to or greater than 40% than the Categories is High. The final count of crime categories are low: 1315, medium: 386 and high: 293.
- All attributes are set to numeric except Categories which is set as nominal.
- The final number of attributes after data preparation are implemented is 104. This number of attribute will undergo feature selection process.
- All the data will be normalize into [0, 1] using min-max method by using Eq. (1), where $X_i$ is the normalized input data and $x_i$ is the unnormalized input data during $i$th iteration

$$X_i = \frac{x_i - min(x_i)}{max(x_i) - min(x_i)} \tag{1}$$

The dataset is divided into training and testing using 10-fold cross validation method. In this method, for each of the 10 times, 9 portion of the dataset is divided into training set and a portion of dataset is divided into testing set. The training set is used to train the algorithm for good learning capability while the testing set is applied to evaluate the generalization capability of the proposed algorithm.

### 2.2 Artificial Neural Network

Artificial neural network (ANN) consists of a set of processing elements known as neurons or nodes and connected with each other. Each node is multiplied with separate weight value and receive signal from the nodes in the previous layer. Output of the $i$th neuron can be described by (Eq. 1)

$$y_i = f_i(\sum_{j=1}^{n} w_{ij}x_j + \theta_i) \tag{2}$$

where $y_i$ is the output of the node, $x_j$ is the $j$th input to the node, $w_{ij}$ is the connection weight between the node and input $x_j$, $\theta_i$ is the threshold (or bias) of the node, and $f_i$ is the node transfer function. The goal is to minimize the mean square error (MSE) function given by (Eq. 2)

$$E(w(t)) = \frac{1}{n}\sum_{j=1}^{n}(d_k - o_k)^2 \tag{3}$$

where $E(w(t))$ is the error at the $t$th iteration; $w(t)$, the weight in the connections at the $t$th iteration; $d_k$ is the desired output node; $0_k$ is the target value for the $k$th output node.

For crime classification, the output $(y_1, ..., y_3)$, consist of three crime categories which are low, medium and high. The outputs are represent as three bit binary, for example, {1,0,0} as low, {0,1,0} as medium and {0,0,1} as high.

### 2.3 Artificial Bee Colony Algorithm

Artificial Bee Colony (ABC) algorithm consists of three groups of bees which are employed bees, onlooker bees and scouts bees. Employed bees go to the food sources and come back to hive and dance on this area. Onlookers bees watch the dances and choose food sources depending on the dances. The employed bee whose food sources has been abandoned becomes scout and starts searching a new food source. Algorithm 1 describe the detail implementation of ABC algorithm.

The position of the food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The number of employed bees or the onlooker bees is equal to the number of solutions in the population. For the first step, the ABC generates a randomly distributed initial population $P(C = 0)$ of $SN$ solutions (food source positions), where $SN$ represents the size of employed bees or onlooker bees. Each solution $x_i (i = 1, 2, ..., SN)$ is a $D$-dimensional vector where $D$ is the number of parameters to be optimize. The population of the positions (search process of the employed bees, onlooker bees and the scout bees) is repeated until reach maximum cycle number (MCN), $C = 1, 2, ...MCN$.

**Algorithm 1.** Artificial bee colony algorithm [6]

```
Initialize the population of solutions x_i, i = 1...SN
Evaluate the population
cycle = 1
repeat
    Produce new solution v_i for the employed bees by using (Eq. 4) and
    evaluate them
    Apply greedy selection process
    Calculate the probability values p_i for the solutions x_i by (Eq. 3)
    Produce the new solutions v_i for the onlookers from the solutions x_i
    selected depending on p_i and evaluate them
    Apply greedy selection process
    Determine the abandoned solution for the scout, if exists, and replace
    it with a new randomly produce solution x_i by (Eq. 5)
    Memorize the best solution achieved so far
    cycle = cycle+1
until Until cycle = MCN;
```

An employed bee produces a modification on the position using (Eq 3). If the nectar amount of the new position is higher than previous, the bee memorizes the new position and discard the old one. Otherwise, the bee keeps the position of the previous on in memory.

$$v_{ij} = x_{ij} + \emptyset_{ij}(x_{ij} - x_{kj}) \tag{4}$$

where $k \in \{1, 2, ..., SN\}$ and $j \in \{1, 2, ..., D\}$ are randomly chosen indexes. $k$ is determined randomly and should be differ from $i$. $\emptyset_{ij}$ us a random generated number between [-1,1].

After all employed bees complete the search process, the sharing information begins where the food sources and their position information shared with the onlooker bees. An onlooker bee evaluates the nectar information and choose a food source with a probability, $p_i$ related to its nectar amount following (Eq. 4).

$$p_i = \frac{fit_i}{\sum\limits_{n=1}^{SN} fit_n} \tag{5}$$

where $fit_i$ is the fitness value of the solution $i$ and $SN$ is the number of food sources. The employed bee produces a modification of the position and checks the nectar amount of the candidate source. If the nectar is higher than previous one, the onlooker bee memorizes the new position and discards the old one. The food source of which the nectar is abandoned by the bees is replaced with a new food source by the scouts by (Eq. 10) in case if the position cannot be improved further. The parameter "$l$imit" is the control parameter to determine the abandonment of the food sources within predetermined number of cycles.

$$x_i^j = x_{min}^j + rand(0,1)(x_{max}^j - x_{min}^j) \tag{6}$$

## 3 Experiment Results and Discussions

The performance of ANN-ABC is compare with ANN using standard BP algorithm. The result from other two algorithm, decision tree and Naive Bayes are taken from the work done by Iqbal et. al [4]. The aim of this comparison is to evaluate the performance capability of the proposed ANN-ABC in crime classification. Or in other words, to investigate whether the application of ABC has successfully optimize ANN and increase the classification accuracy.

Figure 2 shows the experiment result of correctly classified (or accuracy) and incorrectly classified crime dataset. The blue column indicate the dataset is correctly classified and the red column indicate the dataset is incorrectly classified by tested algorithm. The result show that ANN-ABC have correctly classified 86.48% of the dataset outperformed other. ANN itself score 83.49%, decision tree score 83.95% and Naive Bayes score 70.81%. Moreover, ANN-ABC reduced the amount of incorrectly classified and this shows that ANN-ABC has improved in term of accuracy compared with others.

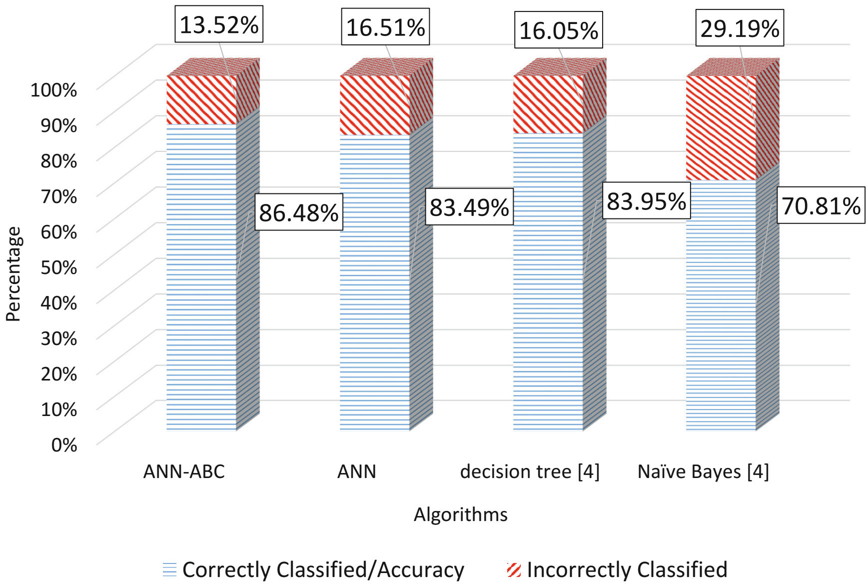


**Fig. 2.** Comparison of correctly classified and incorrectly classified

Figure 3 illustrate the experiment result for classification. The performance measurements consists of precision, recall and F-measure. The result for precision shows that ANN-ABC outperforms the other algorithms with 86.48%. Meanwhile, decision tree is the second algorithm with 83.50% followed by ANN with 83.49%. The last algorithm is Naive Bayes with 66.40%. ANN-ABC shows 8.6% of improvement over other algorithms for precision.

Result for recall performance measurement shows that ANN-ABC outperforms other with 86.60%. The second algorithm is decision tree with 84%. The third algorithm is Naive Bayes followed by ANN both with 70.80% and 69.34% respectively. Result for recall indicates that ANN-ABC 11.9% improvement over other algorithms.

From F-measure results, ANN-ABC is the best classification algorithms for crime classification with 86.53%. The second best algorithm is decision tree with 82.6% followed by ANN with 72.09% and Naive Bayes produces the worst performance with 67.5%. This result indicate 12.5% improvement of ANN-ABC in term of F-measure performance measurement.

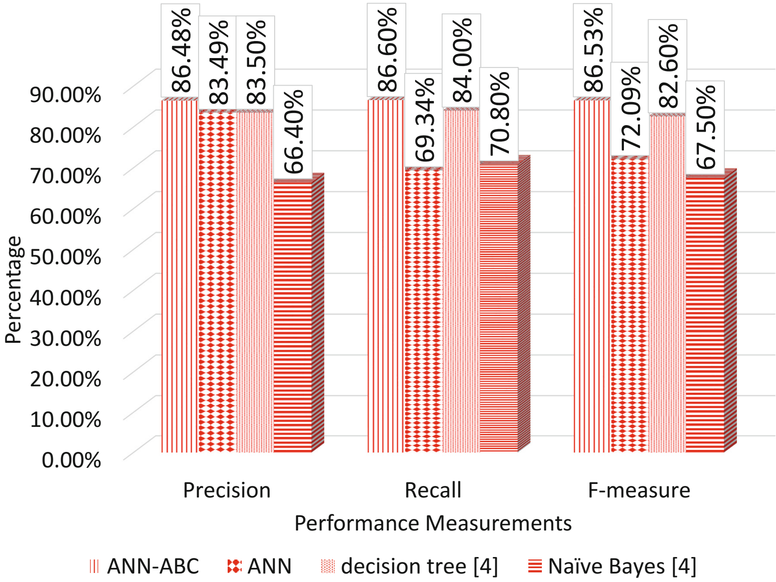


**Fig. 3.** Comparison of performance measurements

The experiment results indicate that the ANN-ABC outperforms other algorithms on classification of crime dataset. The exploration and exploitation process of ABC to optimize ANNs weights has succesfully minimize the learning error for better convergence results. Despite of using the gradient stepest descent that prone to trap in local minima, the ANN-ABC used search strategy that analyze a wider region of search space has efficiently optimize the search strategies

## 4 Statistical Test

The experiment results shown in previous section indicate that ANN-ABC outperformed other algorithms. To further validate whether the result is significant, independent sample t-test are carried out. This test is performs to examine whether the ANN-ABC is performs better the ANN at the level of confidence 0.05. This test is perform on test dataset and the test variable are the performance metrics such as accuracy, precision, recall and F-measure.

The null hypothesis ($H_0$) is define as there is no statistically significant difference between the performance of ANN-ABC and ANN, whereas the alternative hypothesis ($H_1$) has been defined as, the performance of ANN-ABC is better than the ANN. Table 1 shows the result of statistical test.

**Table 1.** Statistical test for ANN-ABC versus ANN

| | df | t | Sig. (2-tailed) |
|---|---|---|---|
| Accuracy | 18 | 12.81 | .000 |
| Presicion | 18 | 2.23 | .039 |
| Recall | 18 | 8.00 | .000 |
| F-measure | 18 | 7.29 | .000 |

By observing Table 1, it can be seen that the $p$-value of the performance metrics are smaller than 0.05. This indicate that there is a significant difference for the performance of both algorithms. Hence, the evidence support the claim on $H_1$ which conclude that the performance of ANN-ABC is better than the ANN in classification of crime dataset.

## 5 Conclusions

This paper present the usage of ABC algorithm in learning ANN for hybrid crime classification model. The hybrid model, named ANN-ABC was compared with other classification algorithm such as ANN with back propagation algorithm, Naive Bayes and decision tree algorithm. The performance of the classification algorithms was tested on real world Communities and Crime dataset, collected from UCI machine learning repository. Overall specific performance measurement for classification such as accuracy, precision, recall and F-measure are compared with each other. The results from the experiment show that ANN-ABC has produce higher accuracy and improve average 7% compare to other algorithms. Furthermore, the statistical test has been conducted and the results show that the performance of ANN-ABC is significantly better than the ANN. Overall results indicate that ANN-ABC is better than the other algorithms in learning and classifying crime dataset. For crime classification purpose, the improvement is important for ANN-ABC to establish reliable classification model thus provide constant result for crime analyst to analyze the crime data. On another point of view, suggestion of using machine learning techniques to analyze crime data efficiently can be highlighted after the success of this experiment. The future direction include the consideration of using feature selection to select the significant attributes of the crime dataset.

**Acknowledgement.** The Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia (UTM) under research grant 00M19 and ERGS Grant Vote number :R.J130000.7828.4L107, are hereby acknowledged for the research funding support and some of the facilities utilized during the course of this research work.

## References


1. Bache, K., Lichman, M.: UCI machine learning repository (2013), http://archive.ics.uci.edu/ml
2. Gorr, W., Harries, R.: Introduction to crime forecasting. International Journal of Forecasting 19(4), 551–555 (2003)
3. Gorr, W., Olligschlaeger, A., Thompson, Y.: Assessment of crime forecasting accuracy for deployment of police. International Journal of Forecasting (2000)
4. Iqbal, R., Azmi Murad, M.A., Mustapha, A., Panahy, S., Hassany, P., Khanahmadliravi, N.: An experimental study of classification algorithms for crime prediction. Indian Journal of Science & Technology 6(3) (2013)
5. Janeela Theresa, M., Joseph Raj, V.: Fuzzy based genetic neural networks for the classification of murder cases using trapezoidal and lagrange interpolation membership functions. Applied Soft Computing 13(1), 743–754 (2013)
6. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Techn. Rep. TR06, Erciyes Univ. Press, Erciyes (2005)
7. Karaboga, D., Akay, B., Ozturk, C.: Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks. In: Torra, V., Narukawa, Y., Yoshida, Y. (eds.) MDAI 2007. LNCS (LNAI), vol. 4617, pp. 318–329. Springer, Heidelberg (2007)
8. Karaboga, D., Gorkemli, B., Ozturk, C., Karaboga, N.: A comprehensive survey: artificial bee colony (abc) algorithm and applications. Artificial Intelligence Review, 1–37 (2012)
9. Kattan, A., Abdullah, R., Salam, R.A.: Harmony search based supervised training of artificial neural networks. In: 2010 International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 105–110. IEEE (2010)
10. Nasridinov, A., Ihm, S.-Y., Park, Y.-H.: A decision tree-based classification model for crime prediction. In: Park, J.J., Barolli, L., Xhafa, F., Jeong, H.-Y. (eds.) Information Technology Convergence. LNEE, vol. 253, pp. 531–538. Springer, Heidelberg (2012)
11. Nath, S.V.: Crime pattern detection using data mining. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, WI-IAT 2006 Workshops, pp. 41–44. IEEE (2006)
12. Ozturk, C., Karaboga, D.: Hybrid artificial bee colony algorithm for neural network training. In: 2011 IEEE Congress on Evolutionary Computation (CEC), pp. 84–88. IEEE (2011)
13. Yaghini, M., Khoshraftar, M.M., Fallahi, M.: A hybrid algorithm for artificial neural network training. Engineering Applications of Artificial Intelligence 26(1), 293–301 (2013)