International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 4
ISSN: 2088-5334

Hybrid K-Nearest Neighbour and Particle Swarm Optimization Technique for Divorce Classification

Hayder Naser Khraibet AL-Behadili[a,*], Ku Ruhana Ku-Mahamud[b]

[a]Department of Computer Science, Shatt Alarab University Colleg, Basra, Iraq
[b]School of Computing, Universiti Utara Malaysia, Sintok, Malaysia
Corresponding author: *hayderkhraibet@sa-uc.edu.iq

Abstract—**Judgment is the ability to make a considered decision by an evolution of knowledge. With the increasing trend of applications of artificial intelligence in law, divorce prediction has become the centre of research. Divorce classification and divorce factors determination are two of the most important matters in societies. Developing an effective technique is essential to prevent communities from collapsing. The traditional techniques of artificial intelligence play a major role in classifying divorce cases. Feature selection is a powerful pre-processing method used for data classification problems. Most previous studies on divorce classification focused on heuristic feature selection methods to determine the main factors behind divorce. These heuristic methods are considered the greedy strategy which does not produce an optimal solution. In this research, a new hybrid swarm intelligence technique was proposed using particle swarm optimisation for feature selection and the K-nearest neighbour algorithm for classification. Specifically, the proposed hybrid classifier can be used in real divorce applications where judges in their investigations can identify the factors that lead to the applications. For the experiment, five classifiers were used for performance analysis. The proposed technique was successfully applied and showed that the performance is better than the existing classifiers, namely naive Bayes, support vector machine, artificial neural network, repeated incremental pruning to pro-duce error reduction, and decision stump. Therefore, the proposed classification model is a more suitable technique for divorce classification than other artificial intelligence techniques.**

Keywords— **K-nearest neighbour; particle swarm optimization; data classification; machine learning; feature selection.**

Manuscript received 30 Sep. 2020; revised 8 Jan. 2021; accepted 16 Mar. 2021. Date of publication 31 Aug. 2021.

I. INTRODUCTION

Classification is the most frequently occurring task in decision- making. The use of feature selection methods for classification has been widely studied because it can improve classification accuracy and reduce computational time required in data- mining tasks [1]. Data classification is a supervised learning method involving the assigning of an instance to a group of classes, based on the features of that object. Classification algorithm consists of the training and the testing stages. In the first stage, training is performed on training data to build a classifier while in the second stage, the classifier is tested on the test data. In this manner, classification algorithms will use the input data that includes features but only some of them are suitable for the classification. Under such circumstances, it is not appropriate to select a large number of features from data as some are irrelevant and increase the classification cost [2]. Therefore, feature selection is used to automatically find a feature subset which is relevant and removes the irrelevant features. Many crucial application domains such as justice and medical have benefited from the advanced artificial intelligence approach of selecting relevant features in classification.

Law is a system of rules that are created by a society or a government to regulate crimes, economies, social relationships, business agreements and politics [3]. Judgement is the matter of going to court for adjudication in which evidence is evaluated to make decisions on the basis of law. Judgement is also the ability of carefully making decisions [4]. Law is important for people to provide a norm guideline for their behaviour and sustain normal justice in society [5].

One of the serious threats to society is divorce [6]. It primarily affects the family, which is the basic kernel in society [7]. A divorce starts between parents. A set of problems immediately occurs between children and their parents and thus weakens or damages parent-children relationships. In addition, children receive less financial help, emotional support, intellectual stimulation of academic behaviour and practical assistance from their family [8]. Various governments have reported that uncoupling has surged during the lockdown period due to the Coronavirus pandemic [9]. In reality, different reasons and factors lead to divorce [10]. Conducting an in-depth study of such causes can reduce the number of divorce cases in society. With the rise of artificial intelligence and information age, many governments use the advent of technology in the court of law and the divorce cases to first detect the main reasons and factors, and then analyse them for treatment.

The usage of artificial intelligence is progressing rapidly. Artificial intelligence can be applied to any task from self-driving cars to medical diagnosis and law applications [11-13]. Artificial intelligence aims to programmed human intelligence in computers, which refers to mimicking human actions and thinking like humans [14]. It can also be applied to other devices that show traits associated with an expert mind such as problem- solving and learning. It uses different technologies and algorithms to deal with different applications and sectors (e.g., the usage in the healthcare domain for discovering drugs and treatments for patients) [15]. Other examples of artificial intelligence include court and law systems that use artificial intelligence to detect the main reasons and factors behind the disasters threatening society and analyse them for treatment. However, most of the studies in divorce classification focus on heuristic feature selection methods to determine the main features of divorce. These heuristic methods are considered as a greedy strategy which does not produce an optimal subset of features. Thus, particle swarm optimisation (PSO)- based feature selection is used in this research to produce an optimal subset of features. This research will address the major challenges and opportunities in the literature to improve divorce data classification. Furthermore, researchers have studied divorce from different perspectives. In recent years, machine learning algorithms have been used to find various reasons that lead to divorce. Many studies have shown that the machine- learning algorithms have been successfully applied for divorce prediction, and to prevent future marriages.

Divorce data from the University of California, Irvine (UCI) repository has been used by several re-searchers in their studies related to artificial intelligence techniques for classification [16], [17]. Ranjitha and Prabhu introduced the PSO as a machine- learning technique [16]. Their study featured two stages. The first stage involved preprocessing through normalisation and binning. The second stage used PSO as the classification technique. The result was promising and showed signs of future success by using artificial intelligence in the divorce sector. How-ever, the current study does not provide a comparison with other feature selection methods and classification algorithms. Kong and Tianrui used three machine- learning algorithms for divorce classification, namely support vector machine (SVM), random forest, and natural gradient boosting (NGBoost) [17]. Output from the research highlighted nine main factors that lead to divorce. It also indicated that the NGBoost algorithm is better than the SVM and random forest. Furthermore, the research recommended future studies to collect further data with additional factors for analysis.

The divorce predictor scale was used by Mustafa et al. to predict whether the couple would continue to live together or get divorced [18]. In the first stage, the correlation-based feature selection method was used to select the most effective features leading to divorce. In the second stage, two classification algorithms-- the C4.5 decision tree classifier and the artificial neural network (ANN) -- were used. In this research, the success of the classification increased when it was applied to six significant features, rather than the 54 features that were available in the data.

Goel et al. proposed Augur Justice as a classification machine- learning algorithm to provide advice to users about their cases [19]. This algorithm can determine whether users can win or lose their cases. Data related to divorce cases were collected from India, comprising three different religions, namely Hindu, Christian and Muslim. Thus, three datasets were developed. The datasets underwent the preprocessing stage by using various methods for data preparation which were tokenization, re-moving redundancy, removing stop words and detection of religion. The comparison among the three different algorithms were through naive Bayes, decision tree, and random forest which showed that Augur Justice provided promising results.

Several machine- learning techniques were proposed by Sohail et al. to solve diverse conditions that affect divorce [20]. The research was performed at Bahria University, Pakistan, where three classifiers; namely naive Bayes, decision tree, and KNN were used to evaluate the collected divorce dataset to find the pros and cons of diverse conditions. They found that the best result was based on the naive Bayes classifier. They concluded that many changes can be made in the future to help further under-stand this social problem which can change a destroyed society to a stable society. A machine learning technique based on the random forest classifier to analyse the divorce factors for a case study in Germany was proposed by Arpino et al. [21]. They investigated divorce with the German social-economic panel data from 1984 to 2015, which had different features (e.g., age, importance, extroversion of spouse, openness, and spouse level of subjective well-being agreeableness). The machine- learning technique indicated that the random forest classifier can determine and highlight the most powerful divorce conditions according to their importance. However, this research did not use feature selection methods or evaluate the performance with other classification algorithms.

Li et al. [22] proposed the Markov logic networks to classify divorce cases related to judicial decisions in China. The data were collected online from the judgements and consisted of 695,418 cases of divorce disputes. The research extracted the legal factors in a formal way and made an induction rule, which users could easily understand. The results showed that the proposed technique provides a promising result, and the results are interpretable. However, the result of this research was not com-pared with other classification algorithms. Thus, comparison has to be performed to ascertain if this technique is better than other common techniques. Another study in China was conducted by Luo et al. [23] where a support- vector machine to predict charges for criminal cases was proposed. The data were collected from the Judgements that were available online. The researcher randomly selected 50,000 cases for training, 5,000 cases for validation and 5,000 for testing. The experiment showed a promising result with a reasonable generalisation ability on fact descriptions written by non-legal users. The system has helped legal assistants in decision making. The researchers recommended further improvements to handle multi-defendant cases. Comparison of two artificial intelligence classification algorithms (naive Bayes and KNN) on their classification performance were studied by Irfan et al. [24]. The authors used the data from the monthly divorce rate collected from the Cimahi Religious Court in Indonesia. The result revealed that naive Bayes was better than KNN in predicting the number of divorce cases.

A multidimensional online analytical processing (MOLAP) technology for the Republic of Iraq was proposed by Jabbar [25] where the divorce data was collected from various sources from the Iraqi court systems with different representation formats. The data were extracted, transformed, and loaded to become a dataset. The study provided an information system to analyse court data and provide good decisions to the authority. The system has powerful properties and gives dynamic, rapid and multiple levels of data analyses. It can produce knowledge from different perspectives. However, the system only provides information, does not have intelligent characteristics and is unable to predict the future or classify cases.

Table 1 shows a summary of the research related to divorce classification and the techniques that have been used. However, existing feature selection methods are categorized as greedy strategy. Consequently, they may only determine the local optimal subset of features, rather than the global optimal solutions (i.e., features). To overcome the gaps stated in previous studies, PSO algorithms can mitigate this drawback by using a combination of two basic ideas. Firstly, these algorithms have a stochastic capability which helps explore a large area of the search space. Secondly, indirect communication procedures are followed based on the positive feedback to determine the promising areas in the search space.

TABLE I
A SUMMARY OF RELATED STUDIES USED IN DIVORCE CLASSIFICATION

No	References	Application domain	Classification techniques
1	Ranjitha & Prabhu [16]	The divorce sector	1-Preprocessing (normalization and binning) 2- PSO -based classification
2	Kong & Tianrui [17]	The divorce sector	1- SVM 2- Random forest 3- NGBoost
3	Mustafa et al. [18]	Divorce predictors scale	1- Correlation-based feature selection 2- C4.5 3- ANN
4	Goel et al. [19]	Divorce cases	1- Preprocessing using NLTK functions 2-Augur Justice
5	Sohail et al.[20]	Divorce Information	1- Naive Bayes 2-Decision tree 3-KNN
6	Arpino et al. [21]	German social-economic panel divorce data	Random forest
7	Li et al. [22]	Divorce cases based on China judicial decisions	Random forest
8	Irfan et al. [24]	Monthly divorce rate	1- Naive Bayes 2-KNN
9	Luo et al. [23]	Criminal cases in China	SVM
10	Jabbar [25]	Republic of Iraq court data	Information system based on MOLAP

The rest of this paper is arranged as follows. Section II explain the proposed hybridization technique presented the K-nearest neighbour (KNN) algorithm for classification combined with PSO for feature selection. This section also provides an explanation on the divorce dataset and the experimental design.. Section III illustrates the classification results of the proposed hybrid technique. Finally, in Section IV the conclusion and the future research direction on divorce classification are drawn.

II. MATERIAL AND METHOD

A. Implementation of Hybrid KNN & PSO Technique for Divorce Classification

In this subsection, a new technique is explained, which works effectively and efficiently with irrelevant and redundant features. The proposed technique consists of two main steps: (a) KNN- based classification of the divorce data, and (b) search for the optimal divorce subset of features based on PSO. The architecture of the proposed technique is presented in Fig 1.

The classifier uses the popular 10-fold cross-validation method to divide the dataset into 10 parts. All parts are equal. Nine parts are used for the learning process (i.e., divorce training dataset), while the remaining part is used for the testing of the model (i.e., divorce testing dataset). This method is repeated 10 times in the same way but with different parts of the data for learning and testing to guarantee that all parts are used in both stages (learning and testing).

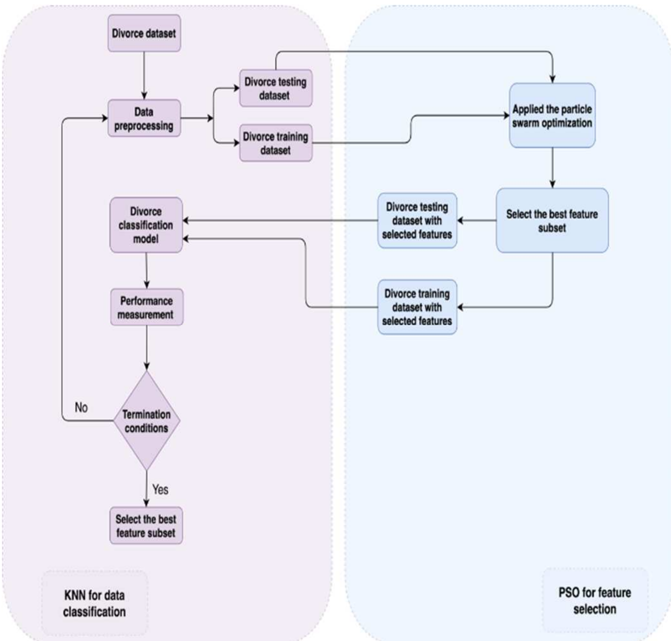

Fig. 1 The Architecture of the PSO-based Attributes Selection Technique for KNN.

The search principle of the PSO algorithm uses the initial population called particles (i.e., solution). This population is improved from iteration to iteration in the PSO algorithm. To find the optimal subset of the feature, each particle in the PSO modifies its search direction based on two aspects. The first aspect is the optimal previous experience for an individual particle (*pbest*) and the global best solution of all other particles (*gbest*).

Each particle in the population represents a candidate's position in the search space (i.e., the candidate's subset of features). Thus, each particle (solution) is considered a position in a *D*-dimension space, and its status is described basedon its position and velocity. Fig. 2 shows the population of particles for the divorce dataset at iteration *t*. The possible position for each particle *i* at iteration *t* is represented as $x_i^t = \{x_{i0}^t, x_{i1}^t, x_{i2}^t, \ldots\ldots, x_{iD}^t\}$. Meanwhile, the distance change (i.e., velocity), which is the *D*-dimension vector for particle i at iteration t is characterised as follows: $V_i^t = \{V_{i0}^t, V_{i1}^t, V_{i2}^t, \ldots\ldots, V_{iD}^t\}$. In PSO, each particle is defined as a position in the search space. For each particle in the iteration, the position, velocity and the *pbest* point found in the hyperspace are recorded.

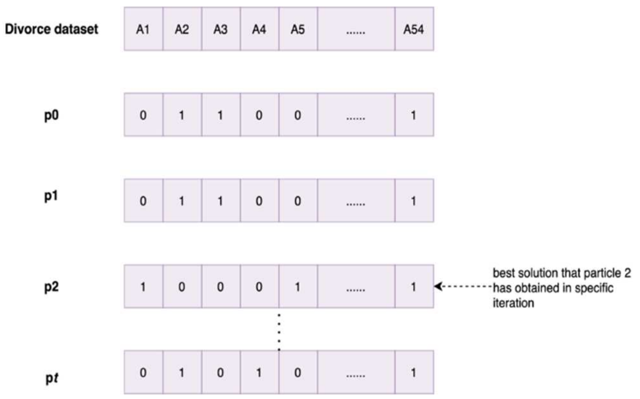

Fig. 2 PSO- based Feature Selection Technique.

Fig. 3 depicts the above PSO concept of the modulation of the search position. x_i^t: current position, x_i^{t+1}: modified position, V_i^t: current velocity, V_i^{t+1}: modified velocity, V_i^{pbest}: *pbest* velocity and V_i^{gbest}: *gbest* velocity. The best solution obtained by each particle is denoted as $p_i^t = \{p_{i0}^t, p_{i1}^t, p_{i2}^t, \ldots\ldots, p_{iD}^t\}$. Meanwhile, $p_g^t = \{p_{g0}^t, p_{g1}^t, p_{g2}^t, \ldots\ldots, p_{gD}^t\}$ indicates the global best solution obtained from p_i^t population at iteration *t*.

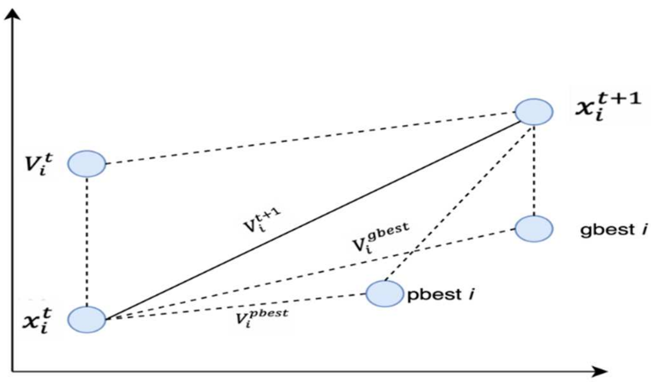

Fig. 3 PSO Searching Concept.

Fig. In addition, each particle searches for the optimal subset of features by changing its velocity on the basis of the following Equation (1):

V_id^(t+1)= V_id^t+c_1^ r_1^ (P_id^t- X_id^t)+c_2^ r_2^ (P_gd^t- X_id^t) (1)

where c1 denotes the individual learning rate; c2 represents the social learning rate; r1 and r2 are real numbers distributed between (0, 1) and generated randomly; d indicates the dimension of the search space; t denotes a pointer of iterations. Therefore, the following Equation (2) allows the particle to move to a new position:

X_id^(t+1)= X_id^t+V_id^(t+1) (2)

The main steps of the PSO-based feature selection algorithm is presented as follows:

- Step One: Generate initial particles randomly by using the initial population.
- Step Two: Measure the quality (i.e., fitness) of each individual solution in the population.
- Step Three: Calculate the velocity value for each particle by using Equation (1).
- Step Four: Construct solution (each particle moves to the next position) on the basis of Equation (2).
- Step Five: Stop the movement if the termination condition is met (number of iterations); otherwise, return to Step Two.

The main idea behind feature selection is finding the optimal set of features that represents the overall data and determines the main factors behind the divorce between spouses. Therefore, the selected feature from the dataset can only be considered in the training and testing datasets for the KNN classifier.

The KNN classifier is a popular nonparametric machine-learning and data- classification task. It is considered as one of the 10 best techniques in data- mining [26]. In supervised learning, KNN assigns unseen instances from the divorce-testing dataset to the nearest class of the divorce- training instance according to the Euclidean distance metric. As illustrated in Fig. 4, each x instance is classified; its KNNs are determined, and x is classified under the class label to which the voting of its neighbours belongs. The selection of k can affect the classifier performance. Experimental results in this research show that k = 3 is the best value used for the divorce data classification.

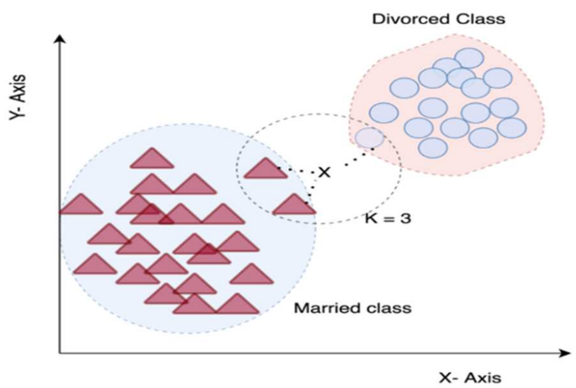

Fig. 4 K-NN Classifier for Divorce Dataset.

B. Divorce Database

The divorce dataset from the University of California, Irvine (UCI) repository were used in this study. The dataset consisted of divorce case information collected by Mustafa et al. [18]. This dataset included 170 instances, and 54 features, which were divided into two classes: 84 instances related to the cases that ended with divorce between the spouses, and 86 instances related to the cases that did not end with divorce between the spouses. This dataset contained no missing data. Table 2 summarizes the characteristics of the divorce datasets.

TABLE II
THE CHARACTERISTICS OF DIVORCE DATASETS

No	Name of feature	Attribute value	Type of attribute
1	A spouse will apologize to end a bad discussion.	5	Discrete
2	Ignorant of differences	5	Discrete
3	Sometimes, the spouse will continue with the discussions and amend the mistake	5	Discrete
4	Discussion with my spouse will eventually work.	5	Discrete
5	Time spent with spouse considered as distinctive time.	5	Discrete
6	Spouses do not spend time at home.	5	Discrete
7	Feel like strangers rather than a family at home.	5	Discrete
8	I am happy to spend my holidays with my spouse.	5	Discrete
9	I am happy to travel with my spouse.	5	Discrete
10	My partner and I have mutual goals.	5	Discrete
11	If I look back, I see a harmonious relationship with my spouse.	5	Discrete
12	Spouse have equal values in terms of freedom of choice.	5	Discrete
13	Spouses have the same sense of entertainment.	5	Discrete
14	Our goals (friends, children, etc.) are similar.	5	Discrete
15	Spouse has similar harmonious dreams.	5	Discrete
16	Spouse are compatible about love.	5	Discrete
17	Spouse are compatible about happiness in life.	5	Discrete
18	Spouse have the same ideas about marriage.	5	Discrete
19	Spouse have the same ideas about marriage roles.	5	Discrete
20	Spouse has equal trust values.	5	Discrete
21	Spouse knows exactly what partner likes.	5	Discrete
22	Spouses take care of each other when one is unwell.	5	Discrete
23	Knowing my spouse's favorite food.	5	Discrete
24	Aware that my spouse is feeling stressed.	5	Discrete
25	Spouse knows the inner world of the partner.	5	Discrete
26	Knowing my spouse's anxieties.	5	Discrete
27	Knowing my spouse's sources of stress.	5	Discrete
28	Knowing my spouse's hopes and wishes.	5	Discrete
29	Knowing my spouse very well.	5	Discrete
30	Knowing my partner's friends and the social relationships between them.	5	Discrete
31	Feeling aggressive during arguements with my partner.	5	Discrete
32	Using emotional expressions such as 'you never' or 'you always' during discussions.	5	Discrete
33	Using negative statements during arguements with my spouse.	5	Discrete
34	Using deeply offensive expressions during discussions.	5	Discrete
35	Insulting my partner during discussions.	5	Discrete
36	Spouse can be humiliated during conversation.	5	Discrete
37	The discussion is always deafening.	5	Discrete
38	Spouse has an unpleasant way of opening a subject for discussion.	5	Discrete
39	Discussions frequently occur suddenly.	5	Discrete
40	Discussion starts before we understand what is going on.	5	Discrete
41	I lose my patience quickly in any discussion.	5	Discrete
42	Argument will end when a spouse leaves .	5	Discrete
43	Usually my partner stays silent in any argument.	5	Discrete
44	Spouse leaves home for a while to calm down after any heated situation.	5	Discrete
45	I like to be silent rather than discuss with my partner.	5	Discrete
46	I like to be silent to hurt my spouse.	5	Discrete
47	I stay silent to control my anger.	5	Discrete
48	I feel I am right during discussions.	5	Discrete
49	I have nothing to do with what I have been accused of.	5	Discrete
50	Spouse feels not guilty about what he/she is accused of.	5	Discrete
51	Spouse is not guilty about problems.	5	Discrete
52	Spouses do not hesitate to tell each other about their inaptitudes.	5	Discrete
53	Spouse reminds partner about her/his inaptitude during discussions.	5	Discrete
54	Spouses are not afraid to tell each other about their ineptitudes.	5	Discrete

C. Experiments

In the experiment, the famous 10-folds cross-validation method was used in all classifiers [26]. The experiment was carried out using the MacBook Pro, with the macos mojave operating system, the 2.9 GHz Intel Core i5 processor, and the 8 GB 2133 MHz memory. The Java object-oriented programming language in the Eclipse development environment was used to develop the proposed hybrid technique.

D. Classification Algorithms

The performance of the proposed KNN+PSO technique for classification on the divorce dataset was compared with five most-popular classifiers which were naïve Bayes, SVM, ANN, RIPPER and decision stump [27], [28]

The naive Bayes classifier is considered as one of the simple probabilistic classifiers according to the Bayes principle together with strong independence assumptions among the attributes (naive) [29]. The support vector machine classifier is considered as one of the best supervised learning-classifiers used for regression and data classification. The principle of the SVM is based on finding the hyperplane in a multidimensional space (i.e., multifeatured) to evidently classify the data instances [30]. Artificial neural network is one of the machine learning algorithms that is used for classification and regression. It is designed on the basis of the human brain processes and analyses the information. Artificial neural network constructs an important element called neuron , which represents the base for a classifier to categorise data accurately [31].

The repeated incremental pruning to produce error reduction (RIPPER) algorithm is the decision tree classifier, which makes classification rules based on the minimum descriptive length of the heuristic function [32]. The classifier iteratively adds construction rules to the rule list until all positive instances are covered [33]. A decision stump is a supervised- learning classifier that comprises a one-level decision tree. It is called a one-rule classifier because it makes classification according to only one feature. This is a decision tree with one internal node [34].

E. Evaluation Analysis

This study used the most popular evaluation criteria in the literature of data classification. These evaluation criteria are correctly classified instances (accuracy), precision, recall, F-measure, Kappa statistic and relative absolute error [26], [35]. The classification accuracy of any given classification algorithm represents the percentage of the correctly classified instances on the test set. Classification accuracy is calculated as shown in Equation (3). Precision evaluates the rate of positive classification. A high value of precision indicates low false positives, whereas a low value of precision means high false positives. Precision can be calculated using Equation (4). Recall evaluates positive observations that are correct; high recall value means low false negatives, whereas low recall value means additional false negatives. Recall is calculated as shown in Equation (5). The F-measure or F-score is measured according to recall and precision. It measures the performance balance between multiple classes. The calculation of F-measure is done as in Equation (6). Kappa statistic is a measurement used to evaluate the intra/inter reliability of a classifier. It evaluates the agreement of classification with the true class. The Kappa equation is given in (7). The relative absolute error (RAE) in (8) is measured the difference between a measured or inferred value and the actual value of a quantity. The relative absolute error value is normalised by the total squared error of the evaluation classification model and divides it by the value of the total squared error of the simple classifier. The best value of this criterion is near 0.

$$\text{Accurate} = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{F-measure} = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (6)$$

$$\text{Kappa Statistics} = \frac{Accurate-RandomAccurate}{1-RandomAccurate} \quad (7)$$

$$\text{RAE} = \frac{(TN+FP)*(TN+FN)+(FN+TP)*(FP+TP)}{Total*Total} \quad (8)$$

where *TN* is true negative (refers to negative instances that were correctly classified), *TP* is true positive (refers to positive instances correctly classified), *FN* is false negative (refers to positive instances incorrectly classified), and *FP* is false positive (refers to negative instances incorrectly classified). These variables are often used to test the performance of the classifier (classification algorithms) on a given set of data for which the true values for these instances in the data are known [36-38].

III. Results and Discussion

From the experiment, different classification results of divorce predication were obtained. These classification results represent a model, which can be applied to predict divorce classification in the future. The experiment was conducted with the classification accuracy criterion to test the classification performance. The result obtained by the proposed technique (99.41%) is the best among the other techniques (refer to Fig. 5). This is very close to perfect classification. The measurement of classification was compared with the state-of-art classifiers that are used in the literature of divorce prediction. Figs. 6, 7 and 8 illustrate the average performance in terms of precision, recall and F-measure, respectively. Among all the evaluation criteria, the proposed technique shows the best result compared to other classifiers. The results in Fig. 9 show that KNN+PSO obtained the best result among the five other classifiers in the Kappa statistic measurement that evaluated the intra/inter reliability of the classifier. Moreover, Fig. 10 displays that KNN+PSO is lower than other classifiers with a relative absolute error of 1.907 for the 10-folds cross-validation.

Therefore, the proposed technique is the best classifier among all the evaluation criteria because this technique can find the appropriate number of features which represent the divorce dataset. It uses the power of the artificial swarm intelligence (i.e., PSO algorithms) that can mitigate such drawbacks.

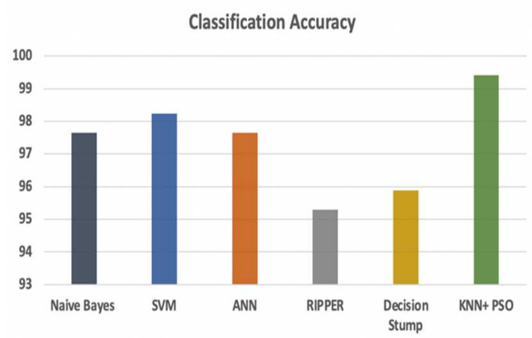

Fig. 5 Classification Accuracy.

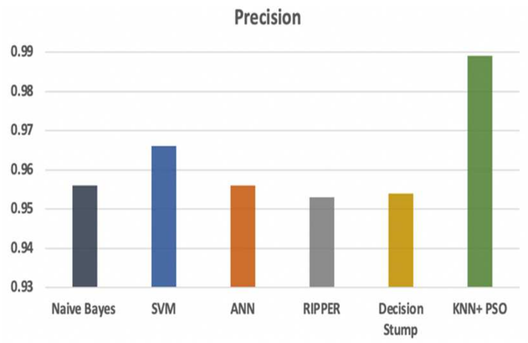

Fig. 6 Precision Performance.

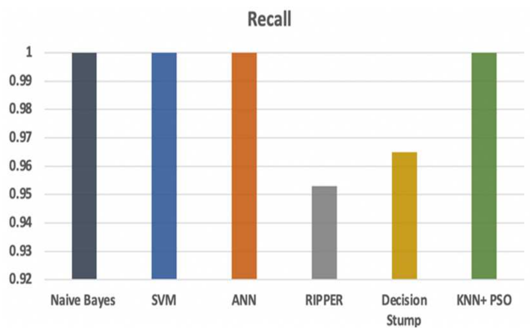

Fig. 7 Recall Performance.

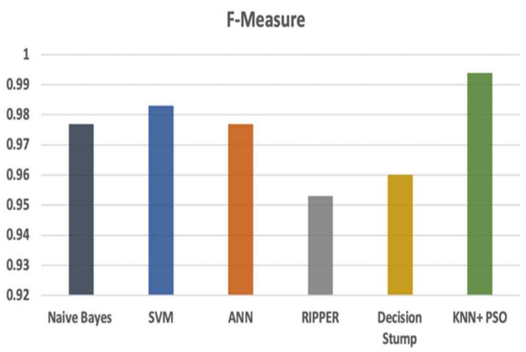

Fig. 8 F-Measure Result.

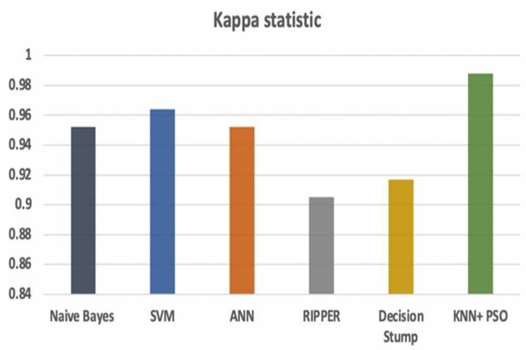

Fig. 9 Kappa Statistic Result.

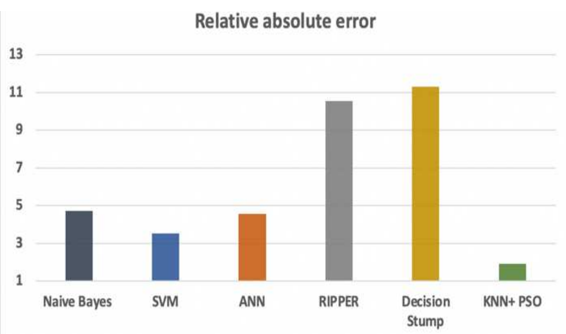

Fig. 10 Relative Absolute Error.

In addition, the proposed technique discovered 13 significant features out of the 54 features that were available in the dataset. These features :are apologies when a discussion becomes progressively worse, the discussion with spouse eventually works, the type of house- sharing (strangers and family), the goal of life, similar views about being happy in life, similar trust values, spouse's favourite food, spouse's worries, insulting spouse during a discussion, discussion starts without planning, the action after an argument with the spouse, keeping silent during a discussion to hurt spouse and paying no attention to what one is accused of. Furthermore, the proposed classification model gives the highest classification accuracy, precision, recall and F-measure evaluation measurements. The Kappa value is close to 1,1, which indicates the measure of agreement between the classification made by the experts and the classifiers. KNN+PSO has a lower absolute error than other classifiers used in the literature. Thus, the KNN+PSO technique is more appropriate for divorce prediction than the other classifiers.

IV. Conclusion

This research has devised a KNN+PSO classifier for feature selection and classification on divorce data. The aim was to determine the main factors that lead to divorce. Therefore, the proposed technique discovered the features with the use of artificial intelligence and 'expert manner' in the context of the dataset. The proposed classifier has a strong search ability in the problem space and can effectively find significant features from the dataset. In order to evaluate the performance of the proposed classifier, comparisons were performed using the baseline classification algorithms. Furthermore, the proposed classifier gives the highest classification accuracy, precision, recall and F-measure evaluation measurements. The obtained Kappa value indicates the measure of agreement between the classification made by the experts and classifiers. KNN+PSO, has a lower absolute error than other classifiers. Thus, the KNN+PSO technique is more appropriate for divorce prediction than other classifiers. This research can be extended in the future with various bio-inspired algorithms such as ant colony optimization, bat algorithm, and gray wolf which can be hybridized with other feature selection methods such as genetic algorithms.

Acknowledgement

The researchers thank the Malaysian Ministry of Higher Education for financially supporting this research under the Grant Scheme (TRGS /1/2018/UUM/02/3/3 (S/O code 14163).

References

[1] Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature selection algorithm based on bare bones particle swarm optimization," *Neurocomputing*, vol. 148, pp. 150–157, 2015.

[2] H. B. Alwan and K. R. Ku-Mahamud, "Mixed-variable ant colony optimisation algorithm for feature subset selection and tuning support vector machine parameter," *Int. J. Bio-Inspired Comput.*, vol. 9, no. 1, pp. 53–63, 2017.

[3] J. L. Entrikin, "The Death of Common Law," *Harv. J. Law Public Policy*, vol. 42, no. 351, pp. 1–137, 2019.

[4] G. Hinchliffe and H. Walkington, "Cultivating the art of Judgement in Students," *Grad. Employab. Context*, pp. 213–235, 2017.

[5] Y. Feldman, “The law of good people: Challenging states ability to regulate human behavior,” Cambridge University, 2018.

[6] K. Oladotun, “Divorce, a Debacle or a Panacea? A Rethink from Biblical Evaluation,” Olabisi Onabanjo University, 2018.

[7] A. Heidari, H., Kimiaei, S. A., & Mashhadi, “Discovering the Factors Affecting Divorce in Early Marriages: A Systematic Qualitative Study,” *J. Res. Behav. Sci.*, vol. 17, no. 2, pp. 314–323, 2019.

[8] K. Donahey, “Effects of Divorce on Children: The Importance of Intervention,” *BYU Undergrad. J. Psychol.*, vol. 13, no. 1, pp. 21–33, 2018.

[9] H. Zhang, “The Influence of the Ongoing COVID-19 Pandemic on Family Violence in China,” *J. Fam. Violence*, pp. 1–11, 2020.

[10] U. Eyo, “Divorce: Causes and Effects on Children,” *Asian J. Humanit. Soc. Stud.*, vol. 6, no. 5, pp. 172–177, 2018.

[11] B. Alarie, A. Niblett, and A. Yoon, “How artificial intelligence will affect the practice of law,” *Univ. Tor. Law J.*, vol. 68, no. 1, pp. 106–124, 2018.

[12] P. Bouletreau, M. Makaremi, B. Ibrahim, A. Louvrier, and N. Sigaux, “Artificial Intelligence: Applications in orthognathic surgery,” *J. Stomatol. Oral Maxillofac. Surg.*, vol. 120, no. 4, pp. 347–354, 2019.

[13] P. Sharma, H. Liu, W. Honggang, and Z. Shelley, “Securing wireless communications of connected vehicles with artificial intelligence,” in *2017 IEEE International Symposium on Technologies for Homeland Security, HST 2017*, 2017, pp. 1–7.

[14] L. Huimin, L. Yujie, C. Min, H. Kim, and S. Seiichi, “Brain Intelligence: Go Beyond Artificial Intelligence,” *Mob. Networks Appl.*, vol. 23, no. 2, pp. 368–375, 2018.

[15] M. Komorowski, L. Celi, O. Badawi, A. Gordon, and A. Faisal, “The intensive care AI clinician learns optimal treatment strategies for sepsis,” *Nat. Med.*, vol. 24, no. 11, pp. 716–1720, 2018.

[16] Ranjitha and A. Prabhu, “Improved Divorce Prediction Using Machine learning- Particle Swarm Optimization (PSO),” in *2020 International Conference for Emerging Technology (INCET)*, 2020, pp. 1–5.

[17] J. Kong and C. Tianrui, “Is Your Marriage Reliable ? Divorce Analysis with Machine Learning Algorithms,” in *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 2020, pp. 1–4.

[18] Y. Mustafa, A. Kemal, T. İLHAN, and K. Serhat, “Divorce Prediction Using Correlation Based Feature Selection and Artificial Neural Networks,” *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Derg.*, vol. 9, no. 1, pp. 259–273, 2019.

[19] S. Goel, S. Roshan, R. Tyagi, and S. Agarwal, “Augur Justice: A Supervised Machine Learning Technique To Predict Outcomes Of Divorce Court Cases,” in *Proceedings of the IEEE International Conference Image Information Processing*, 2019, vol. 2019-Novem, pp. 280–285.

[20] S. Sohail, S. Aziz, F. Tahir, S. Haqqui, and A. Hussain, “Implementation of machine learning algorithm on factors effecting divorce rate,” in *2018 International Conference on Engineering and Emerging Technologies (ICEET)*, 2018, pp. 1–5.

[21] B. Arpino, M. Le Moglie, and L. Mencarini, “Machine-Learning techniques for family demography : An application of random forests to the analysis of divorce determinants in Germany,” *Annu. Meet. PAA*, no. 56, p. 32, 2018.

[22] J. Li, G. Zhang, H. Yan, L. Yu, and T. Meng, “A Markov logic networks based method to predict judicial decisions of divorce cases,” in *Proceedings - 3rd IEEE International Conference on Smart Cloud*, 2018, no. 1, pp. 129–132.

[23] B. Luo, Y. Feng, J. Xu, X. Zhang, and D. Zhao, “Learning to predict charges for criminal cases with legal basis,” in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2727–2736.

[24] M. Irfan, W. Uriawan, O. T. Kurahman, M. A. Ramdhani, and I. A. Dahlia, “Comparison of Naive Bayes and K-Nearest Neighbor methods to predict divorce issues,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 1.

[25] A. M. Jabbar, “Design and Develop an Information system for Court Data in the Republic of Iraq by using SSRS Reports with SSAS Cubes,” *Iraq J. Electr. Electron. Eng.*, vol. 11, no. 1, pp. 105–109, 2015.

[26] I. E. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, no. San Francisco, CA. Elsevier, 2005.

[27] S. Kamruzzaman, “Rule Extraction using Artificial Neural Networks,” *Arxiv Prepr. arXiv1009.4984*, 2010.

[28] X. Wu *et al.*, “Top 10 algorithms in data mining,” *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

[29] G. M. Bressan, V. A. Oliveira, E. R. Hruschka, and M. C. Nicoletti, “Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop,” *Eng. Appl. Artif. Intell.*, vol. 22, no. 4–5, pp. 579–592, 2009.

[30] A. A. Abdoos, P. K. Mianaei, and M. R. Ghadikolaei, “Combined VMD-SVM based feature selection method for classification of power quality events,” *Appl. Soft Comput. J.*, vol. 38, pp. 637–646, 2016.

[31] H. N. K. Al-behadili, “Classification Algorithms for Determining Handwritten Digit,” *Iraq J. Electr. Electron. Eng.*, vol. 12, no. 1, pp. 96–102, 2016.

[32] W. Cohen, “Fast effective rule induction”, In Machine Learning Proceedings, 2435, 115–123, 1995.

[33] B. Seerat and U. Qamar, “Rule Induction Using Enhanced RIPPER Algorithm for Clinical Decision Support,” in *Sixth International Conference on Intelligent Control and Information Processing*, 2015, vol. 33, no. November, pp. 83–91.

[34] P. N. Tan, M. Steinbach, and Vipin Kumar, *Introduction to data mining*. 2006.

[35] I. Faniqul, F. Rahatara, R. Sadikur, and B. Humayra, “Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques,” in *International Symposium, ISCMM 2019*, Springer, 2019, p. 154.

[36] H. N. K. Al-Behadili, R. Sagban, and K. R. Ku-Mahamud, “Adaptive parameter control strategy for ant-miner classification algorithm,” *Indones. J. Electr. Eng. Informatics*, vol. 8, no. 1, pp. 149–162, 2020.

[37] A. Gupta, A. Mohammad, A. Syed, and M. N., “A Comparative Study of Classification Algorithms using Data Mining: Crime and Accidents in Denver City the USA,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 7, pp. 374–381, 2016.

[38] J. Wahid and H. F. A. Al-Mazini, “Classification of Cervical Cancer Using Ant-Miner for Medical Expertise Knowledge Management,” in *Knowledge Management International Conference (KMICe)*, 2018.