Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression

https://doi.org/10.3991/ijim.v15i05.17173

Ameer Saleh Hussein, Rihab Salah Khairy,
Directorate General of Education in Babylon, Babylon, Iraq

Shaima Miqdad Mohamed Najeeb
Northern Technical University, Mosul, Iraq

Haider Th. Salim ALRikabi (✉)
Wasit University, Wasit, Iraq
hdhiyab@uowasit.edu.iq

Abstract—The global online communication channel made possible with the internet has increased credit card fraud leading to huge loss of monetary fund in their billions annually for consumers and financial institutions. The fraudsters constantly devise new strategy to perpetrate illegal transactions. As such, innovative detection systems in combating fraud are imperative to curb these losses. This paper presents the combination of multiple classifiers through stacking ensemble technique for credit card fraud detection. The fuzzy-rough nearest neighbor and sequential minimal optimization are employed as base classifiers. Their combined prediction becomes data input for the meta-classifier, which is logistic regression resulting in a final predictive outcome for improved detection. Simulation results compared with seven other algorithms affirms that ensemble model can adequately detect credit card fraud with detection rates of 84.90% and 76.30%.

Keywords—Fraud detection, Credit card, Ensemble technique, Stacking, Machine learning

1 Introduction

The motive that drives fraud is for criminal purposes. This act of pursuing committing fraud is basically for siphoning money illegally that leads to loss of financial or personal gain [1]. According to definition, credit card fraud is the usage of information assigned to credit card without the users' knowledge for purchases [2]. The transactions performed with the credit card are orchestrated physically or virtually. Physical in the sense, transactions involve an exchange of the card in person by the user during point of purchase. Virtual transactions encompass online operations via the World Wide Web [2], [3]. While credit card usage paved the way for easy, convenient, and proficient online transaction through e-commerce, it also created a loophole for criminal activities thereby inflating rate of fraud [4]. Online transactions for goods and services over the years have skyrocketed. It is reported that an estimate of US$15 billion was the overall orders executed in 2009, with online payment of 84% [2]. In Malaysia, credit card transactions accounted for 320 million in 2011, and rose to 360 million in 2015 [5]. Fraud hiked from US$ 23 billion in 2013 to US$32 billion in 2014 [6]. Another source stated that in 2015 [7], the global credit card fraud was US$ 21.84 billion which is decrease to the report in [6]. There are different numbers of techniques that have been proposed and developed for tackling fraud detection. They comprise of Bayesian network, Markov model, decision tree, support vector machines, and a host of algorithms that are nature-inspired [8]–[12]. In this paper, an alternative method for the detection of credit card fraud is proposed based on stacking ensemble technique. It adopted machine learning algorithms of fuzzy-rough nearest neighbor (FRNN), sequential minimal optimization (SMO), and logistic regression (LR). By combining the predictions of these algorithm results in a classification outcome for effective detection. Datasets from well-known database were retrieved for experimentations and evaluated with standard metrics for fair comparison. The organization of the paper is as follows: Section 2 summarizes relevant literatures in relation to credit card fraud detection. Section 3 discusses the algorithms of fuzzy-rough nearest neighbor, sequential minimal optimization, and logistic regression. The proposed ensemble model is formulated in Section 4. Experimentations are analyzed in Section 5. In Section 6 occupies the conclusion and future works.

2 Related Works

There has been lots of research conducted for the detection of credit card fraud in literature. This section reviews the various work carried out to solve the problem of fraud detection. Ref. [13] proposed a number of different modifications of artificial neural network (ANN) totaling five new ANNs for the classification of fraud in credit card as well as identification of customers. Dependent on real-life data, experimental outcomes show the developed models measured up and, in some cases,, performed better in comparison to other algorithms. A bagging ensemble based on decision tree was constructed to adequately predict credit card fraud [14]. The authors made use of real-world data to investigate the performance of the devised model, and after undergoing experimental rudiments, the bagging ensemble outperformed support vector machine, naïve bayes and *k*-nearest neighbor. The combination of random forest (RF) and rough set theory (RST) proved efficient for the detection of fraud as put forward by Ref. [15]. RF serves the purpose of selecting relevant attributes, which is passed onto RST for classification. The decision tree and neural network were also drafted for proper comparison. Final results places that RST was able to give a better classification performance. By adapting the algorithmic methods of AdaBoost and majority voting, a selection of twelve stand-alone algorithms have been incorporated for ascertain criminalities by fraudsters on credit card [5]. Employing a collated data over a period of three months and a benchmark data, the empirical results confirms majority voting show superiority with the inclusion of noise. According to [11], fisher discriminant analysis was adjusted by injecting a weighted average that promotes linear discriminant to suite the profitable projections as conceptualized by the authors. The classification and regression tree was used for streamlining important attributes. The selected ones are thus applied by proposed fisher discriminant analysis, and edged the decision tree, naïve bayes, ANN, and original fisher discriminant analysis in terms of detecting fraud. A comparative analysis of algorithms used mostly for credit card fraud detection was conducted in the work by Ref. [16] that involves logistic regression, decision tree, and random forest. Publicly available dataset of German credit data served for evaluation among the algorithms. Results extracted from analysis reveals that random forest proved superior. Still on random forest algorithm, Ref. [17] focused on two variations of random forest namely; random-tree based and classification and regression tree (CART) random forest. Applying the forest-based models on dataset collected from China, the CART accounted for superlative percentages to tree-based algorithm. The inbuilt advantages provided by hyper-heuristic evolutionary algorithms opened the pathway for the development of an intelligent Bayesian network classifier for credit card fraud detection [18]. Empirical analysis when compared to other traditional Bayesian network algorithms and some learning algorithms rated the proposed method better to others in terms of economy efficiency. Data mining methods has always been in the fore front in tackling credit card fraud, which is further echoed in the work presented by Ref. [19]. The support vector machine, random forest, and logistic regression were used for data analysis. As recorded above, random forest once again shows its prowess by generating high performance. In handling big amount of data, the convolution neural network was drafted for use in the detection of behaviours deemed fraudulent in credit card data patterns [20]. State of the art algorithms such as SVM, NN and RF were compared with the proposed model. The RF proved it mettle but could not be stronger than CNN in overall performance. Ref. [4] investigated the credibility of three machine learning models namely; logistic regression, k-nearest neighbor, and naïve bayes, for finding suspicious behavioural patterns in fraud data. The principal component analysis act as feature reduction technique before the processed data is injected into the classifiers. A higher accuracy was accrued by k-nearest neighbor to other models. Ref. [21] proposed a strategy based on feature engineering for credit card fraud detection. A sequence classification task with reliance on Long Short-Term Memory (LSTM) was used for addressing the issue of fraud detection [22]. By deploying a deep learning technology of generative adversarial networks, a boost in the classification effective was achieved for credit card fraud detection [23].

3 Conceptual Characteristics of Fuzzy-Rough Nearest Neighbor, Sequential Minimal Optimization, and Logistic Regression

3.1 Fuzzy rough set

On the condition that there is crisp set $B \subseteq S$, the lower and upper approximation of Pawlak [24], [25], are defined in (1) and (2) with regards to equivalence E_r owning to *z* in *S*,

$$z \in E_r{\downarrow}B \text{ iff } [z]_{E_r} \subseteq B \tag{1}$$

$$z \in E_r{\uparrow}B \text{ iff } [z]_{E_r} \cap B \neq \varnothing \tag{2}$$

as seen identically in (3) and (4),

$$z \in E_r{\downarrow}B \text{ iff } (\forall s \in S)\big((s,z) \in E_r \Rightarrow s \in B\big) \tag{3}$$

$$z \in E_r{\uparrow}B \text{ iff } (\exists s \in S)\big((s,z) \in E_r \wedge s \in B\big) \tag{4}$$

With *B* and E_r denoting a set and relation in *S* that are fuzzy, it is possible to expand the equations in (3) and (4) with fuzzy implicator and *t*-norm depicted as *I* and *T* in (5) and (6) respectively.

$$(E_r{\downarrow}B)(z) = \inf_{s \in S}(I(E_r(s,z), B(s)) \tag{5}$$

$$(E_r{\uparrow}B)(z) = \sup_{s \in S}(T(E_r(s,z), B(s)) \tag{6}$$

3.2 Vaguely quantified rough set

The inf and sup operators in equation (5) and (6), processed from fuzzy rough sets are closely related to $\forall$ and $\exists$ quantifiers in (3) and (4). Such interconnections can have an immense influence on approximations when one entity changes. This makes fuzzy rough sets susceptible to meaningless and corrupted data. Hence, a decision to substitute $\forall$ and $\exists$ with abstract quantifiers like *most* and *some* was put forward to address this restriction [26], [27]. Vague quantifies are modelled mathematically via continuously growing fuzzy quantifier [28]: a growing [0,1] → [0,1] maps *Q* meeting the borderline specifications $Q(0) = 0$ and $Q(1) = 1$. In (7), the construction of instances defining fuzzy quantifiers is created using accompanying parameterized formula, for $0 \le \delta < \lambda \le 1$, and *s* in [0,1].

$$Q_{\delta,\lambda}(s)=\begin{cases} 0, & s\le\delta, \\ \dfrac{2(s-\delta)^2}{(\lambda-\delta)^2}, & \delta\le s\le\dfrac{\delta+\lambda}{2} \\ 1-\dfrac{2(s-\lambda)^2}{(\lambda-\delta)^2}, & \dfrac{\delta+\lambda}{2}\le s\le\lambda, \\ 1, & \lambda\le s. \end{cases} \tag{7}$$

The determination of a pair (Q_l, Q_u) leads to the description of approximations termed Q_l–lower and Q_u–upper of a fuzzy set B interpreted in (8) and (9) by relation E_r , for every element of z in S,

$$(E_r{\downarrow}_{Q_l}B)(z)=Q_l\left(\frac{\left|[z]_{E_r}\cap B\right|}{\left|[z]_{E_r}\right|}\right) \tag{8}$$

$$(E_r{\uparrow}_{Q_u}B)(z)=Q_u\left(\frac{\left|[z]_{E_r}\cap B\right|}{\left|[z]_{E_r}\right|}\right) \tag{9}$$

3.3 Fuzzy nearest neighbor

The process of classifying a test object owing to the similarity with respect to a specified *K*-nearest neighbor and their respective membership degrees is ascribed to the proposition of fuzzy *K*-nearest neighbor (FNN) algorithm [29], [30]. The FNN pseudocode is shown in Algorithm 1. Given that an object *z* resides within class *C*, the similarity is formulated as:

$$C'(z)=\sum_{s\in N}E_r(s,z)C(s) \tag{10}$$

where *N* connotes the set of object *z*'s *K*-nearest neighbors. $E_r(s,z)$ is similarity of *s* and *z* and is located inside [0,1]. It can also be defined traditionally as:

$$E_r(s,z)=\frac{\|z-s\|^{-2/(m-1)}}{\sum_{j\in N}\|z-j\|^{-2/(m-1)}} \tag{11}$$

where $\|\cdot\|$ depicts Euclidean norm, and *m* is used for handling the similarity's weight.

Algorithm 1: The fuzzy nearest neighbor (FNN) algorithm

```
Require: $S$: the training data, $Ϛ$: the class set of decision, $z$: the object to be classified,
$K$: the number of nearest neighbors
$N$ ← get Nearest Neighbors($z$,$K$)
$\forall C \in Ϛ$ do
    $C'(z) = \sum_{s\in N} E_r(s,z)C(s)$
end
return $\underset{C\hat{I}\,V}{\arg\max}(C¢(z))$ as the output
end
```

3.4 Fuzzy rough nearest neighbors

The concatenation of approximations of fuzzy rough set with that of traditional FNN schematics gave birth to the proposition of fuzzy-rough nearest neighbours (FRNN) algorithm [31]. The algorithm, as revealed in Algorithm 2, relies solely on building fuzzy lower and upper decision class approximations using the nearest neighbours. Classification procedure of instances is based on linkage of membership to approximations.

The FRNN algorithm is thickened by selecting fuzzy tolerance relation E_r . Suppose there is an existence a set of conditional attributes E_r is constructed and outlined in (12):

$$E_r(s,z) = \min_{q\in\square} E_{r(q)}(s,z) \tag{12}$$

where $E_{r(q)}(s,z)$ is degree of correlation between object *s* and *z* for attribute *q*. Equation (13) establishes $E_{r(q)}(s,z)$ as shown.

$$E_{r(q)}(s,z) = 1 - \frac{|q(s)-q(z)|}{|q_{\max} - q_{\min}|} \tag{13}$$

where the maximum and minimum value of attribute *q* denoted as $q_{\max}$ and $q_{\min}$ respectively. A high $(E_r{\downarrow}C)(z)$ signifies the inclusion of all of *z's* neighbor to class *C*, whereas as $(E_r{\uparrow}C)(z)$ goes high, it indicates that at least one neighbor belongs to *C*.

Algorithm 2: The fuzzy-rough nearest neighbor (FRNN) algorithm

```
  Require: S: the training data, Ϛ: the class set of de-
cision, z: the object to be classi-fied,
  1: N ← getNearestNeighbors(z, K)
  2:
  3: C Ϛ do
  4:
  5:
  6:
  7: end
  8: return Class as the output
  6:  end
```

3.5 Sequential minimal optimization

The goal of sequential minimal optimization (SMO) is to train the support vector machines (SVMs). Basically to dissolve associated SVM deficiencies in handling large sized problems [32]. The concept of SVM goes thus; With reference to [33], if there exist collection of data points $\{(H_\varepsilon, w_\varepsilon)\}_\varepsilon^p$; H_ε and p are input vector and all training data. The process involved in training SVM for the purpose of classification is analogous to finding solution to the following:

$$\text{maximize:}\quad Q(\delta_\varepsilon) = \sum_{\varepsilon=1}^{p} \delta_\varepsilon - \frac{1}{2} \sum_{\varepsilon=1}^{p} \sum_{\mu=1}^{p} \delta_\varepsilon \delta_\mu w_\varepsilon w_\mu \kappa(H_\varepsilon, H_\mu) \tag{14}$$

$$\text{subject to:}\quad \sum_{\varepsilon=1}^{p} \delta_\varepsilon w_\varepsilon = 0, \qquad 0 \le \delta_\varepsilon \le c, \qquad \varepsilon = 1, \ldots, p \tag{15}$$

where $\kappa(H_\varepsilon, H_\mu)$, δ_ε, and c connotes kernel function, Lagrange multiplier, user-determined regularization constant respectively. The widespread kernel function is the Gaussian function. If the problem in (14) becomes resolved, a unique data sequence is identified by decision function in (16) for the class label.

$$\text{function}(H) = \sum_{\varepsilon=1}^{p} \delta_\varepsilon w_\varepsilon \kappa(H_\varepsilon, H) + b \tag{16}$$

with *b* acquired from Equation (14).

The SVM fails to deal with QP problems of large sizes. In resolving this, the SMO disintegrate enormous QP task into sub-problems. Optimization of a training data sequence subset in each phase, which is called a working set. Two working sets are used to mitigate the QP sub-problems with a simple systematic technique [34]. A set of rules are vital in specifying two δ_{ε} . SMO adjusts quadratically the total data sequence.

3.6 Logistic regression

Logistic regression (LR) is a statistical technique for assessing the likelihood of a binary result determined by a number of reasonable factors. This explains the effect of the considered variables on the dependent variable examined. Contrary, if the explanatory factors include a minimum of three unsorted subgroups, then multinomial logistic regression (MLR) is deployed. Compliance with the notion of binomial logistic regression, the MLR approach was conceived on the same fundamental arrangement. It can therefore be stated that the logistic regression is being extended [35]–[37].

In the work done by Le Cessie and Van Houwelingen [38], a ridge values of 1×10^{8} was recommended for the log probability computation. There exist modifications to for the classification purpose [39]. If *n* cases with *m* features have *k* classes, the $m*(k-1)$ matrix points towards component *B* being computed. The probability for class *j* with the exception of the class is as in (17).

$$P_j(X_i)=\frac{\exp(X_iB_j)}{\sum_{j=1}^{k-1}\exp(X_iB_j)+1} \tag{17}$$

The last class has probability as shown in (18).

$$1-\sum_{j=1}^{k-1}P_j(X_i)=\frac{1}{\sum_{j=1}^{k-1}\exp(X_iB_j)+1} \tag{18}$$

Therefore, the negative multinomial log-likelihood is represented as follows:

$$L=-\sum_{i=1}^{n}\left\{\sum_{j=1}^{k-1}Y_{ij}*\mathrm{In}(P_j(X_i))+\left(1-\sum_{j=1}^{k-1}Y_{ij}\right)*\mathrm{In}\left(1-\sum_{j=1}^{k-1}P_j(X_i)\right)\right\} \\ +ridge*(B^2) \tag{19}$$

A Quasi-Newton process is employed for discovering enhanced values of $m*(k-1)$ elements to locate matrix *B* where *L* is reduced. The matrix *B* is compressed to a $m*(k-1)$ vector prior to the optimization approach [38], [39].

4 Proposed Methodology

The step-by-step procedure of the proposed ensemble algorithm consisting of fuzzy rough nearest neighbor (FRNN), sequential minimal optimization (SMO), and logistic regression (LR) is described in this section.

To begin execution of the ensemble algorithm, the original training data is loaded into base classifiers which are FRNN and SMO algorithms. They are trained to form a combined prediction of the FRNN and SMO. The resulting predictive outcome ultimately serves as input for the meta-classifier to give a final prediction. Figure 1 illustrates the proposed ensemble model.

The figure below can be simplified for better understanding. It encompasses the following steps:

a) The original training data ***D***, having m instances and n attributes is prepared for the base classifiers.
b) The two algorithms of FRNN and SMO that represents the base classifiers train on ***D***.
c) Individual predictions from FRNN and SMO are combined into a single data D^{level2} (that is second level data) with m instances and M attributes.
d) A meta-classifier (LR) is thus trained on the second level data to generate a final predictive outcome for proper classification.

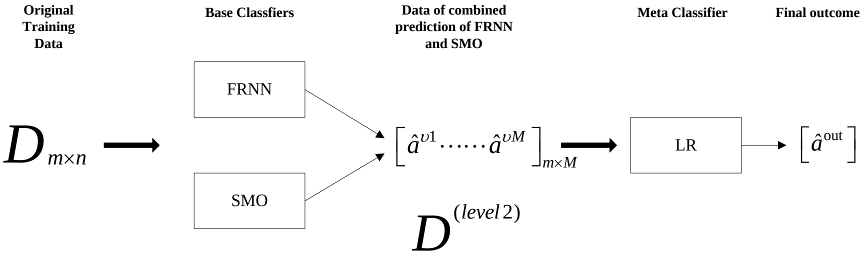

Original Training Data
Base Classfiers
Data of combined prediction of FRNN and SMO
Meta Classifier
Final outcome
FRNN
SMO
LR
$D_{m\times n}$
$\left[\hat{a}^{\upsilon 1}\cdots\cdots\hat{a}^{\upsilon M}\right]_{m\times M}$
$D^{(level2)}$
$\left[\hat{a}^{\text{out}}\right]$

Fig. 1. The proposed ensemble model consisting of FRNN, SMO and LR classifier algorithms

5 Experimental Setup and Results

The credit card fraud datasets used for experimentations are provided and retrieved from UCI Machine Learning Repository through http://archive.ics.uci.edu/ml [40]. The datasets are Australian credit approval data and German credit data. The Australian Credit Approval is composed of credit application data and has 14 attributes with one class label, + or -, as well as 690 instances. 307 instances are categorized as positive (credit approved) and 383 instances as negative (credit denied). The dataset is a good mixture of attributes, including nominal and numerical values. Usage of all numerical attributes version for Australian Credit Approval is employed for use. With respect to the German Credit data, the numeric version is adopted for use. It consists of 700 instances of creditworthy applicants and 300 instances of non-creditworthy applicants. It describes the credit details for each applicant with 24 input variables. Both datasets are trained with the ensemble model of FRNN, SMO and LR algorithms. Popular algorithms within the domain of credit card fraud detection are selected for comparison namely; multi-layer perceptron (MLP), IBk or K-nearest neighbour algorithm, Naïve Bayes, and random forest (RF). The Waikato environment for knowledge analysis (WEKA) takes the centre stage for running all the experiments. Training and assessment is done with a 10-fold cross-validation. This involves the dataset divide into ten subsets of the same size by allocating nine subsets for the training data. An average mean of each results are collated.

5.1 Assessment measures

The performance metrics to evaluating algorithms' effectiveness are the detection rate (*DR*) (true positive rate), false alarm rate (*FAR*) (false positive rate), specificity (*SP*), positive predictive value (*PPV*), and F-measure. The terms are described in (20) to (24):

$$DR = \frac{TP}{TP + FN} \tag{20}$$

$$FAR = \frac{FP}{FP + TN} \tag{21}$$

$$SP = \frac{TN}{TN + FP} \tag{22}$$

$$PPV = \frac{TP}{TP + FP} \tag{23}$$

$$F - measure = 2 \times \frac{Positive\,Predictive Value \times Sensitivity}{Positive\,Predictive Value + Sensitivity} \tag{24}$$

where TP and FP are the true positives and false positives, while FN and TN are the false negatives and true negatives.

5.2 Simulation results

The execution of the simulations relies on WEKA having a 3.40GHz Intel® Core i7 Processor with 4GB of RAM. The findings are tabled and diagrammatically visualized following series of experiments for each dataset. The performance results in Table 1 accommodate the Australian credit approval datasets. With respect to detection rate, FRNN, SMO and LR generated rates of 81.00%, 84.60%, and 85.40% respectively.

Other algorithms such as the MLP, IBk, naïve bayes and random forest accounted for detection rates at 83.80%, 82.00%, 77.50%, and 84.90% accordingly. The proposed ensemble model is rated second at 84.90% alongside random forest. Assigned with the lowest detection rate is naïve bayes algorithm, and LR shows to produce highest rate at 85.40%. Regarding false alarm rate, the lower the rate, the algorithm shows to be better. The proposed ensemble model of FRNN, SMO, and LR, gave the lowest and best rate at 13.80%. Naïve bayes has the poorest false rate of 26.10%.

Table 1. Results for Australian Credit Approval dataset

Algorithms	DR (%)	FAR (%)	SP (%)	PPV (%)	F-measure (%)
MLP	83.80	16.40	83.60	83.80	83.80
IBk	82.00	18.70	81.30	82.00	82.00
Naïve Bayes	77.50	26.10	73.90	79.20	76.70
Random Forest	84.90	15.20	84.80	85.00	84.90
FRNN	81.00	19.90	80.10	81.00	81.00
SMO	84.60	14.10	85.90	85.70	84.60
LR	85.40	14.40	85.60	85.60	85.40
Proposed model	84.90	13.80	86.20	85.90	85.00

It can be revealed that in terms of specificity, the proposed model supersedes all other algorithms with a rate of 86.20%. Also, the proposed ensemble model certifies its superiority over the compared algorithms when positive predictive value is concerned. An 85.90% PPV is accredited to the proposed model. With f-measure, in second place is the proposed model at 85.00%. LR proved better overall with a rate of 85.40%. Scanning through the results of Australian credit approval, naïve bayes performed poorly to others overall, while the proposed model proved the best on the overall comparison.

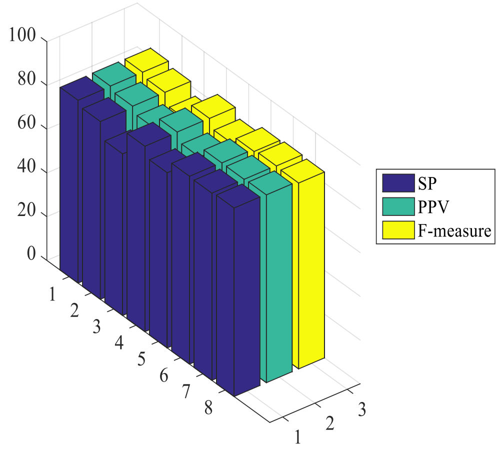


Fig. 2. The graph plots of SP, PPV, and F-measure pertaining to Australian Credit Approval

The performance results in Table 2 accommodate the German credit datasets. With respect to detection rate, FRNN, SMO and LR generated rates of 68.50%, 76.40%, and 76.30% respectively. Other algorithms such as the MLP, IBk, naïve bayes and random forest accounted for detection rates at 70.20%, 66.00%, 75.40%, and 73.80% accordingly. The proposed ensemble model is rated second at 76.30% alongside logistic regression. Assigned with the lowest detection rate is IBk algorithm, and SMO shows to produce highest rate at 76.40%. The proposed ensemble model gave a fasle alarm rate at 40.40%, and is ranked fourth. Random forest has the poorest false rate of 49.70% with naïve bayes having the best at 38.70%.

Table 2. Results for German Credit dataset

Algorithms	DR (%)	FAR (%)	SP (%)	PPV (%)	F-measure (%)
MLP	70.20	43.40	56.60	69.50	69.80
IBk	66.00	47.50	52.50	65.80	65.90
Naïve Bayes	75.40	38.70	61.30	74.40	74.70
Random Forest	73.80	49.70	50.30	71.80	70.90
FRNN	68.50	45.70	54.30	67.80	68.10
SMO	76.40	40.20	59.80	75.20	75.20
LR	76.30	40.10	59.90	75.10	75.20
Proposed model	76.30	40.40	59.60	75.10	75.10

It can be revealed that in terms of specificity, the proposed model was able to supersede some other algorithms with a rate of 56.60%. Also, the proposed ensemble model certifies its superiority over the compared algorithms when positive predictive value is concerned. A 75.10% PPV is accredited to the proposed model. With f-measure, in second place is the proposed model at 75.10%, tied with LR. SMO proved better overall with a rate of 75.20%. Observations acquired with results of German credit dataset is that the proposed model performed significantly well in par with rest of the algorithms.

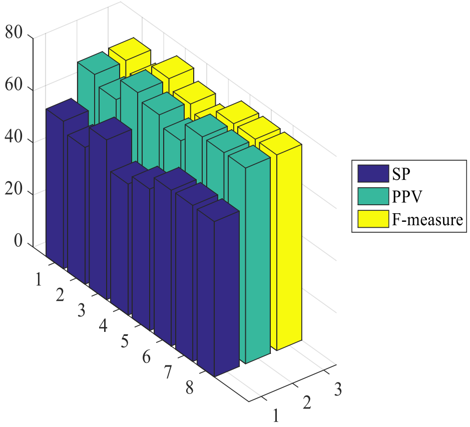


Fig. 3. The graph plots of SP, PPV, and F-measure pertaining to German Credit

Illustrated in Figure 4 through to Figure 7 are the receiver operating characteristic (ROC) curves for all the algorithms. It is analogous to its corresponding area under the curves values that are generated from the ROC curves tabulated in Table 3. For the Australian Credit Approval, the proposed model reveals a better AUC than other algorithms at 0.8555. The LR came close with an AUC of 0.8550. Naïve Bayes recorded the lowest AUC value of 0.7570. Focusing on the AUC results for German credit data, eclipsing four of the algorithm is the proposed system. Only three algorithms of naïve bayes, SMO, and LR with AUC values at 0.6835, 0.6810, and 0.6810 were superior to the AUC of proposed model of 0.6795.

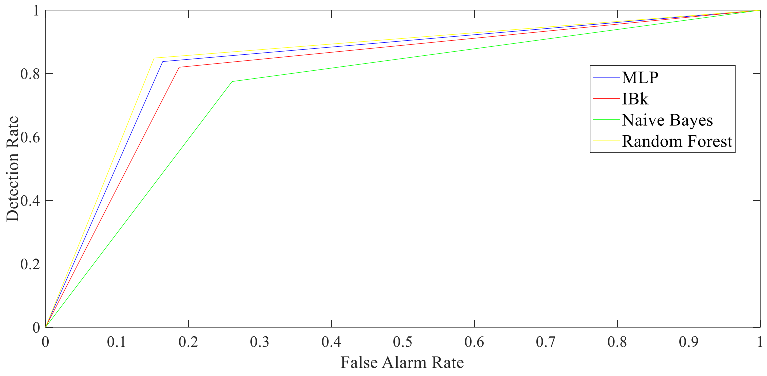


Fig. 4. ROC curves for Australian credit approval for MLP, IBk, naïve bayes, and RF

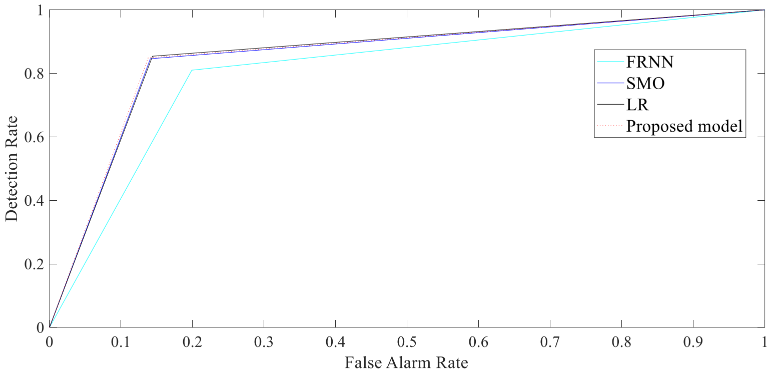


Fig. 5. ROC curves for Australian credit approval for FRNN, SMO, LR, and proposed model

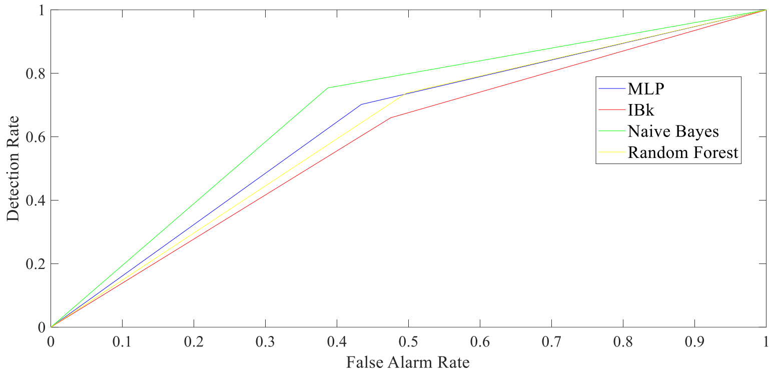


Fig. 6. ROC curves for Australian credit approval for MLP, IBk, naïve bayes, and RF

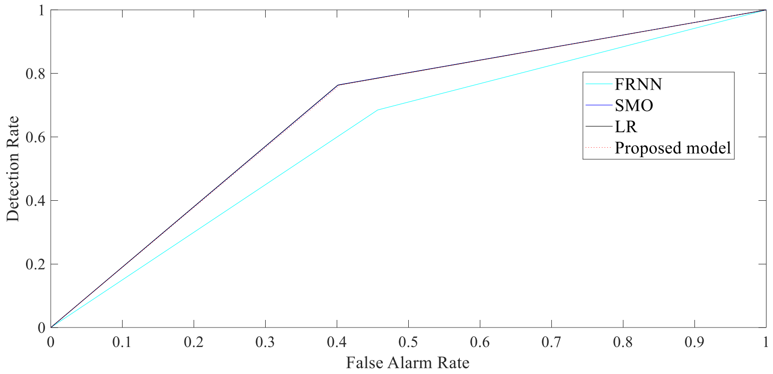


Fig. 7. ROC curves for Australian credit approval for FRNN, SMO, LR, and proposed model

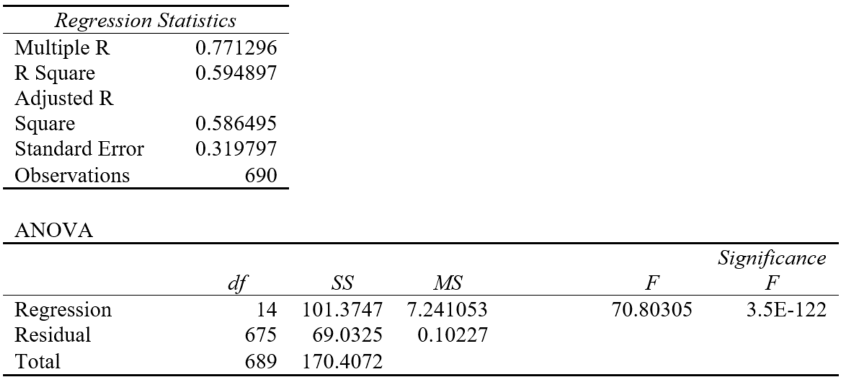
Table 3. Area Under the Curve (AUC) for all algorithms for the datasets

Algorithms	AUC for Australian Credit Approval	AUC for German Credit
MLP	0.8370	0.6340
IBk	0.8165	0.5925
Naïve Bayes	0.7570	0.6835
Random Forest	0.8485	0.6205
FRNN	0.8055	0.6140
SMO	0.8525	0.6810
LR	0.8550	0.6810
Proposed model	0.8555	0.6795

5.3 Statistical analysis of logistic regression using pseudo-R^2

The quality of regression model is assessed statistically by analyzing with the pseudo-R^2. Relating to Australian credit approval, the pseudo-R^2 value is 0.594897. P-value is 3.5E-122 which is less than (<) 0.05. So it is statistically significant. As with

German credit, the value of 0.236271 is accounted for by pseudo-R^2. It has a p-value of 1.83E-47, that is statistically significant.

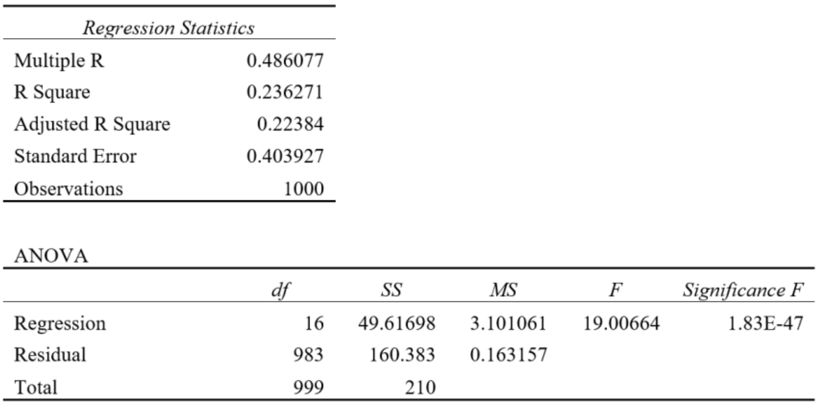


Regression Statistics	
Multiple R	0.771296
R Square	0.594897
Adjusted R Square	0.586495
Standard Error	0.319797
Observations	690

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	14	101.3747	7.241053	70.80305	3.5E-122
Residual	675	69.0325	0.10227		
Total	689	170.4072			

Fig. 8. Analysis for Australian credit approval



Regression Statistics	
Multiple R	0.486077
R Square	0.236271
Adjusted R Square	0.22384
Standard Error	0.403927
Observations	1000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	49.61698	3.101061	19.00664	1.83E-47
Residual	983	160.383	0.163157		
Total	999	210			

Fig. 9. Analysis for German credit

6 Conclusion

This paper presents a stacking ensemble classification model based on fuzzy-rough nearest neighbor algorithm, sequential minimal optimization, and logistic regression for credit card fraud detection. The ensemble method takes advantage of the prediction results of base classifiers by combining them. Afterwards, the meta-classifier accommodates the results accrued from base classifier to generate a final classification result. It also improves the efficiency of classification model. The experimental results on Australian credit approval and German credit datasets indicates that the proposed classification model is able to produce significant and promising classification results in terms of detection rate, false alarm rate, specificity, positive predictive value, f-measure, ROC curves and AUC area. A detection rate of 84.90% and AUC of 0.8555 is generated for Australian credit approval dataset and a 76.30% detection rate with 0.6795 AUC for German credit dataset using 10-fold cross validation procedure. The difference in results between the dataset could be attributed to the dataset features. Australian credit approval with 14 features and German credit having 24 features. A higher data feature may result in lower performance. Therefore, the proposed model through experimentation and analysis confirms that it is very suitable and proficient for the detection of credit card. Future works can be directed towards expanding the algorithms for ensemble in getting better classification results. Also, other techniques that are used in developing ensemble models aside from stacking should be considered for credit card fraud detection.

7 References

[1] Y. Sahin, S. Bulkan, and E. Duman, “A cost-sensitive decision tree approach for fraud detection,” *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916–5923, Nov. 2013. https://doi.org/10.1016/j.eswa.2013.05.021

[2] A. O. Adewumi and A. A. Akinyelu, “A survey of machine-learning and nature-inspired based credit card fraud detection techniques,” *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, no. 2, pp. 937–953, Nov. 2017. https://doi.org/10.1007/s13198-016-0551-y

[3] M. Zareapoor, K. R. Seeja, and M. A. Alam, “Analysis on credit card fraud detection techniques: based on certain design criteria,” *Int. J. Comput. Appl.*, vol. 52, no. 3, Jan. 2012. https://doi.org/10.5120/8184-1538

[4] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, “Credit card fraud detection using machine learning techniques: A comparative analysis,” in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 2017, pp. 1–9. https://doi.org/10.1109/iccni.2017.8123782

[5] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, “Credit card fraud detection using AdaBoost and majority voting,” *IEEE access*, vol. 6, pp. 14277–14284, Feb. 2018. https://doi.org/10.1109/access.2018.2806420

[6] H. John, “Payments companies are trying to fix the massive credit-card fraud problem with these 5 new security protocols,” Jul. 2015.[Online]. Available: http://www.businessinsider.com/how-payment-companies-are-trying-to-close-the-massive-hole-in-credit-card-security-2015-3 https://doi.org/10.1002/9781118386750.ch2

[7] (2016) “The Nelson Report,” [Online]. Available: https://nilson.report.com/upload/content_promo/The_Nilson_Report_10-17-2016.pdf

[8] N. S. Halvaiee and M. K. Akbari, “A novel model for credit card fraud detection using Artificial Immune Systems,” *Appl. Soft Comput.*, vol. 24, pp. 40–49, Nov. 2014. https://doi.org/10.1016/j.asoc.2014.06.042

[9] V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens, “APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions,” *Decis. Support Syst.*, vol. 75, pp. 38–48, Jul. 2015. https://doi.org/10.1016/j.dss.2015.04.013

[10] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, “BankSealer: A decision support system for online banking fraud analysis and investigation,” *Comput. Secur.*, vol. 53, pp. 175–186, Sep. 2015. https://doi.org/10.1016/j.cose.2015.04.002

[11] N. Mahmoudi and E. Duman, “Detecting credit card fraud by modified Fisher discriminant analysis,” *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2510–2516, Apr. 2015. https://doi.org/10.1016/j.eswa.2014.10.037

[12] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, “Learned lessons in credit card fraud detection from a practitioner perspective,” *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, Aug. 2014. https://doi.org/10.1016/j.eswa.2014.02.026

[13] A. Zakaryazad and E. Duman, “A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing,” *Neurocomputing*, vol. 175, pp. 121–131, Jan. 2016. https://doi.org/10.1016/j.neucom.2015.10.042

[14] M. Zareapoor, P. Shamsolmoali, and others, “Application of credit card fraud detection: Based on bagging ensemble classifier,” *Procedia Comput. Sci.*, vol. 48, no. 2015, pp. 679–685, Dec. 2015. https://doi.org/10.1016/j.procs.2015.04.201

[15] F. H. Chen, D.-J. Chi, and J.-Y. Zhu, “Application of Random Forest, Rough Set Theory, Decision Tree and Neural Network to Detect Financial Statement Fraud--Taking Corporate Governance into Consideration,” in *International Conference on Intelligent Computing*, 2014, pp. 221–234. https://doi.org/10.1007/978-3-319-09333-8_24

[16] S. Patil, V. Nemade, and P. K. Soni, “Predictive Modelling For Credit Card Fraud Detection Using Data Analytics,” *Procedia Comput. Sci.*, vol. 132, pp. 385–395, Jan. 2018. https://doi.org/10.1016/j.procs.2018.05.199

[17] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, “Random forest for credit card fraud detection,” in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 2018, pp. 1–6. https://doi.org/10.1109/icnsc.2018.8361343

[18] A. G. C. de Sá, A. C. M. Pereira, and G. L. Pappa, “A customized classification algorithm for credit card fraud detection,” *Eng. Appl. Artif. Intell.*, vol. 72, pp. 21–29, Jun. 2018. https://doi.org/10.1016/j.engappai.2018.03.011

[19] N. Carneiro, G. Figueira, and M. Costa, “A data mining based system for credit-card fraud detection in e-tail,” *Decis. Support Syst.*, vol. 95, pp. 91–101, Mar. 2017. https://doi.org/10.1016/j.dss.2017.01.002

[20] K. Fu, D. Cheng, Y. Tu, and L. Zhang, “Credit card fraud detection using convolutional neural networks,” in *International Conference on Neural Information Processing*, 2016, pp. 483–490. https://doi.org/10.1007/978-3-319-46675-0_53

[21] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, “Feature engineering strategies for credit card fraud detection,” *Expert Syst. Appl.*, vol. 51, pp. 134–142, Jun. 2016. https://doi.org/10.1016/j.eswa.2015.12.030

[22] J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen, “Sequence classification for credit-card fraud detection,” *Expert Syst. Appl.*, vol. 100, pp. 234–245, Jun. 2018. https://doi.org/10.1016/j.eswa.2018.01.037

[23] O. H. Yahya, H. Alrikabi, I. A. J. I. J. o. O. Aljazaery, and B. Engineering, "Reducing the Data Rate in Internet of Things Applications by Using Wireless Sensor Network,"international journal of interactive mobile technologies, vol. 16, no. 03, pp. 107-116, 2020. https://doi.org/10.3991/ijoe.v16i03.13021

[24] Z. Pawlak, “Rough sets,” *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.

[25] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*, vol. 9. Springer Science & Business Media, 2012.

[26] C. Cornelis, M. De Cock, and A. M. Radzikowska, “Vaguely quantified rough sets,” in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, 2007, pp. 87–94. https://doi.org/10.1007/978-3-540-72530-5_10

[27] L. D'eer, N. Verbiest, C. Cornelis, and L. Godo, "A comprehensive study of implicator--conjunctor-based and noise-tolerant fuzzy rough sets: definitions, properties and robustness analysis," *Fuzzy Sets Syst.*, vol. 275, pp. 1–38, Sep. 2015. https://doi.org/10.1016/j.fss.2014.11.018

[28] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," in *Computational linguistics*, Elsevier, 1983, pp. 149–184. https://doi.org/10.1016/0898-1221(83)90013-5

[29] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. Syst. Man. Cybern.*, no. 4, pp. 580–585, Jul. 1985. https://doi.org/10.1109/tsmc.1985.6313426

[30] D. T. Bui, Q. P. Nguyen, N.-D. Hoang, and H. Klempe, "A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS," *Landslides*, vol. 14, no. 1, pp. 1–17, Feb. 2017. https://doi.org/10.1007/s10346-016-0708-4

[31] N. S. Alseelawi, E. K. Adnan, H. T. Hazim, H. Alrikabi, and K. Nasser, "Design and Implementation of an E-learning Platform Using N-Tier Architecture," international journal of interactive mobile technologies, vol.14, issue.6, pp.171-185, 2020. https://doi.org/10.3991/ijim.v14i06.14005

[32] J. Platt and others, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[33] L. J. Cao, S. S. Keerthi, C. J. Ong, J. Q. Zhang, U. Periyathamby, X. J. Fu, and H. P. Lee, "Parallel sequential minimal optimization for the training of support vector machines," *IEEE Trans. Neural Networks*, vol. 17, no. 4, pp. 1039–1049, Jul. 2006. https://doi.org/10.1109/tnn.2006.875989

[34] A. Barbero and J. R. Dorronsoro, "Momentum sequential minimal optimization: An accelerated method for support vector machine training," in *The 2011 International Joint Conference on Neural Networks*, 2011, pp. 370–377. https://doi.org/10.1109/ijcnn.2011.6033245

[35] I. A. Aljazaery, H. T. S. Alrikabi, and M. R. J. I. J. o. I. M. T. Aziz, "Combination of Hiding and Encryption for Data Security,"international journal of interactive mobile technologies, vol. 14, no. 09, pp. 34-47, 2020. https://doi.org/10.3991/ijim.v14i09.14173

[36] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013. https://doi.org/10.1002/9781118548387

[37] A. J. Milewska, D. Jankowska, T. Wi\kesak, B. Acacio, and R. Milewski, "The application of multinomial logistic regression models for the assessment of parameters of oocytes and embryos quality in predicting pregnancy and miscarriage," *Stud. Logic, Gramm. Rhetor.*, vol. 51, no. 1, pp. 7–18, Sep. 2017. https://doi.org/10.1515/slgr-2017-0030

[38] S. Le Cessie and J. C. Van Houwelingen, "Ridge estimators in logistic regression," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 41, no. 1, pp. 191–201, Mar. 1992. https://doi.org/10.2307/2347628

[39] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform," *Soft Comput.*, vol. 18, no. 2, pp. 255–266, Feb. 2014. https://doi.org/10.1007/s00500-013-1055-1

[40] K. Bache and M. Lichman, "UCI machine learning repository," *URL* http//archive. ics. uci. edu/ml, vol. 901, 2013.

8 Authors

Ameer Saleh Hussein is presently working with the Directorate General of Education in Babylon. He was born in April, 3 1982 in Babylon Iraq. He went to school (Al-Entesar) in Babylon-Iraq to continue his study in secondary school. He furthered his studies at Babylon University in Babylon-Iraq and managed to earn bachelors (B.Sc) in Computer Science Department in 2004. He worked in several NGOs and got many training in Management, Design & Human Rights. Then he worked as teacher in secondary schools in (2005). Later he enrolled at University UTHM (Universiti Tun Hussein Onn Malaysia) in 2013, its place in Batu Pahat Johor- Malaysia to earn the Master Degree in Computer Science (Soft Computing).

Rihab Salah Khairy is currently attached to the Directorate General of Education in Babylon. She was born on November, 1 1982 in Babylon Iraq. She went to school (Al-Entesar) in Babylon-Iraq to continue her study in secondary school. She furthered her studies at Babylon University in Babylon-Iraq and managed to earn bachelors (B.Sc) in Computer Science Department in 2005. Then she worked as a teacher in secondary schools in (2005). Later she enrolled at University UTHM (Universiti Tun Hussein Onn Malaysia) in 2013, its place in Batu Pahat Johor- Malaysia to earn the Master Degree in Computer Science (Soft Computing).

Haider Th. Salim ALRikabi He is presently Asst. Prof and one of the faculty college of engineering, electrical engineering department, Wasit University in Al Kut, Wasit, Iraq. He received his B.Sc. degree in Electrical Engineering in 2006 from the Al Mustansiriya University in Baghdad, Iraq. his M.Sc. degree in Electrical Engineering focusing on Communications Systems from California state university / Fullerton, USA in 2014. His current research interests include Communications systems with mobile generation, Control systems, intelligent technologies, smart cities, and Internet of Things (IOT).

Al Kut city - Hay ALRabee, Wasit, Iraq.
Contact: - +9647732212637.
E-mail: - hdhiyab@uowasit.edu.iq.
The number of articles in national databases - 10.
The number of articles in international databases – 20.

Article submitted 2020-07-20. Resubmitted 2021-01-04. Final acceptance 2021-01-08. Final version published as submitted by the authors.