A Hybrid Meta-Learner Technique for Credit Scoring of Banks' Customers

Ali Ghasemy Armaki
Department of Management
Qazvin Branch
Islamic Azad University
Qazvin, Iran
alghasemy@yahoo.com

Mir Feiz Fallah
Tehran Central Branch
Islamic Azad University
Tehran, Iran
fallahshams@gmail.com

Mahmoud Alborzi
Information Technology
Management Department
Science and Research Branch
Islamic Azad University
Tehran, Iran
Mahmood_alborzi@yahoo.com

Amir Mohammadzadeh
Department of Management
Qazvin Branch
Islamic Azad University
Qazvin, Iran
amn_1378@yahoo.com

Abstract—**Financial institutions are exposed to credit risk due to issuance of consumer loans. Thus, developing reliable credit scoring systems is very crucial for them. Since, machine learning techniques have demonstrated their applicability and merit, they have been extensively used in credit scoring literature. Recent studies concentrating on hybrid models through merging various machine learning algorithms have revealed compelling results. There are two types of hybridization methods namely traditional and ensemble methods. This study combines both of them and comes up with a hybrid meta-learner model. The structure of the model is based on the traditional hybrid model of 'classification + clustering' in which the stacking ensemble method is employed in the classification part. Moreover, this paper compares several versions of the proposed hybrid model by using various combinations of classification and clustering algorithms. Hence, it helps us to identify which hybrid model can achieve the best performance for credit scoring purposes. Using four real-life credit datasets, the experimental results show that the model of (KNN-NN-SVMPSO)-(DL)-(DBSCAN) delivers the highest prediction accuracy and the lowest error rates.**

Keywords-credit scoring; hybrid machine learning models; stacking; deep learning

I. INTRODUCTION

Owing to the recent global financial crisis and European sovereign debt crisis, credit risk assessment has turn out to be an increasingly vital issue for banks and credit institutions throughout the world. Also, the sharp competition in financial sector has caused a large decline in banking profit. This leads banks toward more consumer loans to make higher interest profits. However, the expected profitability depends on the quality of consumer loans issued by the banks, which requires a vigilant credit scoring process. It is worthwhile to mention that even 1% enhancement on the accuracy of credit scoring system would significantly increase the profit of banks and other financial institutions [1]. Traditionally, credit decisions were made by human experts based on past experiences, historical performances, and some guidelines specially the classic five C's of credit: character, capacity, capital, collateral and conditions [2]. But this approach suffers from some drawbacks including inconsistent decisions, repeated incorrect decisions, and high training costs. Therefore, with the quick development

in credit industry, various credit scoring techniques are being used for the credit evaluation. The credit scoring models have been developing at a fast pace to distinguish bad credit applicants from good ones through their associated features such as gender, age, education, income, job and marital status or based on their historical credit performance. The advantages of credit scoring models can be enumerated as cost reduction of credit analysis, faster credit decisions, higher rate of credit collections, efficient performance monitoring of the model, mitigating possible risks and changes in economic conditions or policies can simply be integrated into the model [3-5]. Even a minor betterment in the accuracy of credit scoring models may diminish a significant amount of credit risks and generate noteworthy future savings. Due to both the impacts of financial crisis and soaring risk appetite, the number of non-performing loans has sharply intensified along with banks giving more credits to applicants without sufficient assessments. Thus, the use of efficient credit scoring models seems to be inevitable for the banks and other credit institutions. There are several approaches employed by financial institutions over the past decades to model the credit risk which are mainly classified into two groups of statistical and Artificial Intelligence (AI) techniques. Generally, the statistical methods include Logistic Regression (LR) and Linear Discriminant Analysis (LDA). On the other hand, AI approaches mainly comprise of machine learning techniques such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT) and so many other machine learning (classification and clustering) algorithms. There are some pros and cons associated with these methods. For instance, LDA assumes a normal distribution of the variables and a linear relationship between explanatory variables but it is unable to verify fulfillment of these assumptions [4, 5]. LR is used for forecasting on a dataset with binary outcomes. Although, the normality assumption is not required by LR, but linear relationship among variables is a basic assumption for both models. Therefore, some researchers [4-7] are having doubts about predictive performance of these models for credit scoring. In contrast, artificial intelligence techniques recently draw attention from many scholars for coping with credit scoring problems. These techniques are best known for their higher predictive accuracy compared to statistical models and usually do not require abovementioned

assumptions. For example, ANN which simulates the human brain's mechanism on the computer environment does not need any assumptions and in the field of credit scoring, it performs much better than its classical rivals including LR and LDA [8-13]. In general, it can be said that AI methods are superior to traditional ones [14-16]. In recent years, many researchers have focused on the development of machine learning techniques for credit scoring applications. One of the methods that they are using to improve the performance machine learning algorithms is the hybridization. These researchers believe that the credit scoring models which are built by combing classification (supervised learning) and clustering (unsupervised learning) techniques have the ability to outperform sole machine learning methods [12, 17-20].

In this study, a new hybrid method is introduced for credit scoring which is based on a combination of traditional hybrid and stacking ensemble methods. The idea comes from the traditional hybrid model of classification plus clustering. This is because clustering is considered as an unsupervised learning method and it cannot differentiate data precisely like supervised methods. Accordingly, a classifier or set of classifiers can be trained first, and then its output is used as the input for the clustering method to enhance the clustering outcomes [21]. In this model, instead of using a single classification algorithm in the first part of the hybrid model, we adopt a stacking ensemble method and in the second part several clustering algorithms will be interchangeably used. Also, this model benefits from a deep learning algorithm as the meta-learner classifier. It is believed that the superior learning capacity of deep learning can improve the predictive accuracy of the new hybrid credit scoring model. This study has chosen various types of classifiers and clusterers to be used in this hybrid model. In the relevant literature, many studies have developed hybrid credit rating models only by choosing single learning algorithms as the baselines (traditional hybrid models) but this study adopts the stacking ensemble method as the baseline of the hybrid model. Moreover, this paper tries to compare several versions of the proposed hybrid model by using various combinations of classification and clustering algorithms. Thus, it helps us to identify which hybrid model can achieve the best prediction accuracy for credit scoring purposes. The structure of the paper is as follows: Section II reviews the literature in terms of different hybrid and ensemble credit scoring models. Section III explains the data and methodology of the study and Section IV presents the experimental results and analysis. Finally, Section V concludes the study and discusses future work opportunities.

II.   LITERATURE REVIEW

In this section, the literature of machine learning in the field of credit scoring will be reviewed. When banks want to grant credit to their customers, they evaluate their credit. Through adopting a good credit scoring system, banks can classify their customers in terms of risk (probability of default), so offer them risk adjusted loans with different interest rates and collateral conditions. Therefore, optimal credit decisions can be made based on the outputs of the credit scoring models. Since the emergence of AI systems like neural networks, genetic algorithm and expert systems, these methods have been

increasingly used in financial researches and also implemented by many financial institutions specially banks. Authors in [22] adopt the four different types of traditional hybrid machine learning techniques to identify which method can achieve the best predictive results. They combine different classification and clustering algorithms such as Naïve Bayesian, Decision Tress, Logistic regression, Neural Network, K-means and Expectation Maximization. Then, they apply these hybrid models on a real credit dataset from Taiwan. Comparative results show that the "classification + classification" hybrid model outperforms the other hybrid models. This model utilizes Logistic regression and Neural Network as the first and second classifiers (LR + NN), respectively. They state that these hybrid credit scoring models can help financial institutions make more correct decisions for issuing consumer loans with high confidence in the future. Authors in [23] study the behavior of imbalanced credit scoring datasets by different machine learning methods. Data imbalances take place when the number of defaulting customers in a dataset is typically much lower than the number of non-defaulting ones. They tested various models on five real-world credit datasets. Finally, they showed that when datasets are imbalanced, machine learning methods like decision trees, KNN, linear discriminant analysis (LDA) do no perform well. On the other hand, models such as gradient boosting and random forests have much better predictive performance. Although, decision trees (DT) is one of the most popular algorithms used in machine learning and credit scoring, it suffers from two drawbacks: 1) it's very sensitive to noise and 2) redundant features may falsify the learning process. Hence, authors in [24] suggest two ensemble methods namely Bagging-RS DT and RS-Bagging DT to deal with these problems. In these models they adopt Random Subspace (RS) and Bootstrap Aggregating (Bagging) strategies. They test these two models on Australian and German credit datasets and results show that these two models perform better than other base models. Author in [25] introduces a new solution for credit scoring problems which is based on a modified version of SVM. He mentions that since most of real credit datasets are pretty big, the use of conventional nonlinear SVMs even with high levels accuracy are computationally suboptimal. Consequently, he proposes a clustered support vector machine (CSVM) to cope with this problem. He concludes that the CSVM, despite gaining similar prediction performance, can stay relatively cheap from computational point of view. In another attempt to create an optimal credit scoring model, authors in [26] have proposed the Ensemble Classification based Supervised Clustering (ECSC) method. The main idea behind this model is that data samples from the identical class might have dissimilar characteristics or patterns. By means of supervised clustering, samples with similar characteristics or patterns are categorized into the same cluster. Hence, the training subsets, formed by mixture of clusters from diverse classes, could well express various patterns of samples, which is beneficial to enhance the variety and accuracy of base classifiers. In this paper they use base classifiers such as logistic regression, decision trees, SVM and also K-means for supervised clustering. They have applied this model along with random subspace bagging (RS-Bagging), Bagging-RS, and dynamic classifier ensemble using classification confidence (DCE-CC) on German and Australian

credit datasets. Results show the ECSC is relatively more accurate than other models.

One of the heuristic methods used for credit scoring is fuzzy SVM based on Support Vector Data Description (SVDD) which is introduced in [27]. SVDD is based on the SVM classifier, which looks for a spherical-shaped border around a dataset to identify outliers or unique data. This approach uses SVDD to mitigate the impact of outliers and noisy data in order to improve the Fuzzy SVM learning rate. Authors adopted this model to test against the ordinary linear and nonlinear fuzzy SVM on the Australian and German credit datasets. What can be drawn as a conclusion is that although the best result is obtained by the SVDD-FSVM its superiority is negligible. In a recent study, authors in [28] introduced a model based on the combination of hybrid and ensemble methods. They believe that merging filtering and feature selection methods can perform as an effective pre-processor for machine learning models. For this reason, they have combined Multivariate Adaptive Regression Splines (MARS) and Gabriel Neighborhood Graph editing (GNG) in the hybrid modeling stage. As base classifiers, they have selected decision trees, ANN, random forests, Bayesian network and SVM. They have applied these models on seven real world credit datasets. Results illustrate that the authors' proposed model relatively improves the predictive performance relative compared with base learners.

III. METHODOLOGY

This section describes the procedure of developing the credit scoring system introduced by this study. Generally, there are two ways to establish a hybrid machine learning model which are traditional and ensemble methods. The traditional hybridization method offers four different ways to combine machine learning algorithms. These options are (1) merging two classification algorithms, (2) merging one classification algorithm with one clustering algorithm, (3) merging one clustering algorithm with one classification algorithm, and (4) merging two clustering algorithms [22, 29]. On the other hand, ensemble methods offer sophisticated ways of hybridizing machine learning techniques. Employing ensembles is beneficial as they can overcome the three problems of base learning algorithms namely statistical, computational, and representational problems [30]. When the size of a dataset is too small compared with the potential space of hypotheses, a learning algorithm may select to yield a hypothesis from a group having the equal predictive accuracy on the training data. Thus, the statistical problem emerges in such cases if the selected hypothesis is unable to forecast new data. When a learning algorithm is trapped in an incorrect local minimum rather than finding the best hypothesis within the hypotheses space, the computational problem will arise. Lastly, the representational problem occurs when no hypothesis inside the hypotheses space is a good estimate to the correct function [31]. There are several forms of ensembles including bagging, boosting, and stacking. These techniques are frequently used in the literature of machine learning and credit scoring. Findings suggest that ensemble methods usually achieve superior predictive performance compared to other single algorithms or traditional hybrid models [26, 28, 32-36]. Unlike the bagging

and boosting which are used in many papers, few researchers have employed the Stacking method. Stacking (stacked generalization) is designed to enhance predictive performance through combining the predictions of several machine learning algorithms [37]. It consists of training a combiner algorithm to amalgamate the predictions of various learning algorithms. First, an ensemble of classifiers (base classifiers) is trained using the available data via bootstrapped sampling (Tier 1 classifiers). Then the output of base classifiers are used as an input to train a meta-classifier (Tier 2 classifier) [36]. In other words, stacking trains a set of classifiers parallelly and then learning is done by a meta-learner. Author in [38] emphasizes that classifiers which are functioning in a collaborative way can significantly outpace those working separately, showing the importance of using such a model.

The meta-learner (meta-classifier) in the stacking algorithm generates a vector of weight distribution by assigning a weight to each base classifier that is proportional to their performances [31]. Stacking can be considered as a fully customizable hybrid machine learning system as it hosts various types of base- and meta-classifiers. Also, it has been successfully employed on both supervised and unsupervised learning tasks [39-41]. In a recent study [32] is showed that a hybrid ensemble machine learning system with stacking is superior to other types of ensemble methods. The proposed hybrid meta-learner model is built based on the combination of traditional hybrid and ensemble modeling of credit scoring systems. The foundation of the model is based on the traditional hybrid model of "classification+clustering" which uses a classification technique as a pre-processor for the clustering algorithm. The only difference is that this paper adopts the stacking ensemble method in the first part instead of using a single classification method. Also, several classification and clustering techniques are used in this study which are briefly described in following sub sections.

A. Classification Techniques

Classification (or supervised learning) methods are capable of mapping input vectors into one of various preferred output classes through learning by examples. A classifier can be learned by computing the rough distance between input–output instances and correctly labeling outputs out of training set. This procedure is named as the model generation stage. After generating the model, the resulting classifier is able to classify an unidentified example based on the learned classes in the training set. Various classification techniques are employed in this paper which are Artificial Neural Network (NN), Automated Multilayer Perceptron (AMLP), Decision Tree (DT), K-nearest Neighbors (KNN), Logistic Regression (LR), Naïve Bayesian (NB), Support Vector Machines (SVM), and Support Vector Machines optimized by Particle Swarm Optimization (SVM-PSO).

B. Clustering Techniques

Clustering (or unsupervised learning) methods can be viewed as the way toward combining similar examples into a cluster. Unlike the classification, labeled examples are not available in clustering. The main aim of clustering approach is

to increase the resemblance between the group members. Moreover, the data between different clusters should have the highest dissimilarity. Conversely, the highest data similarity should exist within each cluster [42]. There are two categories for clustering algorithms, which are partitional and hierarchical clustering algorithms but the former is much more popular [43]. Partitional clustering has been widely implemented in many credit scoring problems. K-means and expectation maximization are two renowned partitional clustering algorithms. In contrast, hierarchical clustering generates clusters according to a hierarchy by means of the accumulation algorithm. Then, a different single cluster will be merged individually until fulfilling some rules. The outcome will create a series of branching partitions. This study uses five clustering algorithms namely Expectation Maximization (EM), K-means (KM), Fuzzy C-means (FCM), Density-based spatial clustering of applications with noise (DBSCAN), and Self-organizing Maps (SOM).

C. The Hybrid Meta-Learner Model

This study introduces a new hybrid method for credit scoring which is a mixture of traditional hybrid and stacking ensemble methods. The idea comes from the traditional hybrid model of "classification plus clustering" due to the fact that clustering is an unsupervised learning method and it is unable to distinguish data accurately like supervised approaches. Therefore, a classifier or set of classifiers can be trained first, and then the output can be used as the input for the clustering technique to improve the clustering results. In this process, instead of using a single classification algorithm in the first part of the hybrid model, a stacking ensemble method will be used. This stacking model utilizes three different base classifiers (level 0 generalizers) to train the meta-classifier (level 1 generalizer). In the second part, several clustering techniques will be interchangeably used in order to find which combination of algorithms yield the best results. One of the advantages on this hybrid system is placing a deep learning algorithm (DL) in the heart of the proposed model as the meta-learner. Owing to the great learning ability of DL, the predictive performance of the model is expected to improve significantly. For the first time, the concept of DL is proposed in 2006. This algorithm is defined in the framework of deep belief networks (DBN). Later, DL has caused considerable amount of scientific researches in several fields [45-47]. This algorithm as a feature selection technique, tries to get feature abstractions at the high-levels through learning various feature

structures in the training process. Every DL iteration is an unsupervised learning process for feature extraction, and the mix of different layers has the ability to create a deep supervised predictor [48]. DL has various theoretical frameworks, but this study utilizes the H2O version, which is based on the feedforward architecture. As shown in the Figure 1, the main part of the DL model is the neuron which is inspired by the human neural system. In this model, the weighted mix of input signals (α) is combined, and then an output signal f(α) conveyed by the connected neuron. The nonlinear activation function and neuron's activation threshold (bias) are denoted by f and b, respectively [49].
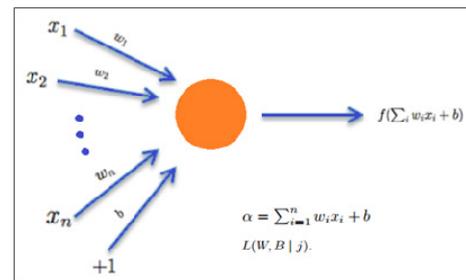

Fig. 1.   The neuron architecture in the deep learning model

The weights which are connecting neurons and biases with other neurons define the output of the whole network. The error on the labeled training data should be minimized through weight adjustment procedure in order to make learning possible. Specifically, the aim is to minimize the loss function of $L(W; B \mid j)$ for every training example j [49]. DL as the meta-learner in the stacking algorithm creates a vector of weight distribution by giving a weight to each base classifier that is proportional to their performances. Stacking can be seen as a completely customizable hybrid machine learning technique since it embraces various types of base- and meta-classifiers. This study employs different types of classifiers and clusterers to be used in this hybrid model. In the literature, many works have been done by developing hybrid credit rating models only by choosing single learning algorithms as the baselines but this study adopts the stacking ensemble method as the baseline of the hybrid model. Furthermore, the aim is to compare several versions of the proposed hybrid model by selecting various mix of classification and clustering algorithms in order to find the best model. Figure 2 illustrates the framework for the proposed hybrid meta-learner model.
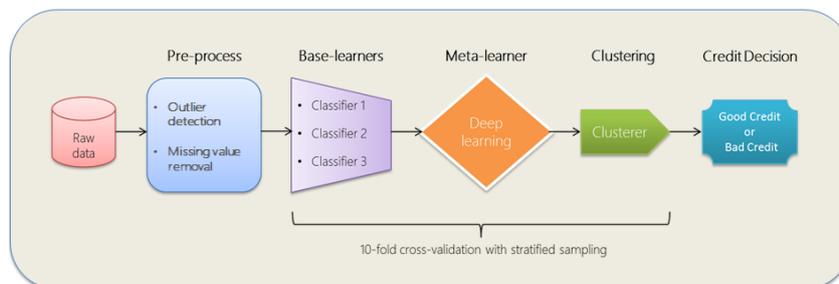

Fig. 2.   The hybrid meta-learner model

After collecting the desired datasets, the pre-process treatment should be applied on the data. In this stage, those data points which are considered as outliers or anomalies should be removed from datasets. For this reason, a distance-based outlier detection algorithm is used to identify n outliers in the given dataset based on the distance to their k nearest neighbors [50]. Also, those examples with missing feature values are removed from datasets. Then, a combination of three classifier algorithms is interchangeably selected out of various classifiers as the base learners. As mentioned before, the stacking model is equipped with DL algorithm as the meta-learner. The stacking model uses the results of base learners (level 0 generalizers) to train the meta-learner (level 1 generalizers). In the next step, the results of the DL as the processed data are fed into the clustering unit. Hence, it is expected to see significant improvement in the prediction accuracy of the model. The performances of models are compared after applying 10-fold cross-validation with stratified sampling during the training and testing stages. Therefore, each dataset is divided into 10 unique subsets (strata), in which any

9 of the 10 subsets are used for training and the leftover for testing. In other words, each model will be trained and tested 10 times.

D. Data

Financial institutions need to have a system for evaluating the credit risk of their customers when granting loans. But before that, they need to make sure that the accuracy of their system is at the acceptable level. Since, it is very difficult to assess the performance of a credit scoring system against private datasets, it is necessary to measure its performance against some benchmark datasets. As a result, we study four real-life datasets to evaluate the predictive power of our proposed hybrid meta-learner model. Table I displays the characteristics of the datasets. Particularly, the first three datasets are considered as benchmark datasets in the literature. They are related to consumer credit card loans from Australia, Germany, and Japan; and collected from the UCSD data repository. The last dataset is related to consumer loans in Iran and collected from Mellat Bank.

TABLE I.   CHARACTERISTICS OF THE DATA SETS

Data set	#Attributes	#Good	#Bad	Total	Source
Australian (AUS)	14	307	383	690	UCI Machine Learning Database Repository
German (GER)	24	700	300	1000	UCI Machine Learning Database Repository
Japanese (JPN)	15	296	357	653	UCI Machine Learning Database Repository
Iranian (IRA)	12	9351	649	10000	Mellat Bank

E. Evaluation Strategies

In order to assess the predictive power of the developed models, prediction accuracy rate and F-measure which is the harmonic mean of precision and recall are taken into account. Precision shows the accuracy degree of classification results and recall is the success rate of identifying classification results. Moreover, besides from these evaluation methods, type I and II errors are also shown for the best model in each dataset (Table II). These evaluation methods can be calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad F - \text{measure} = \frac{2TP}{2TP + FP + FN}$$

TABLE II.   CONFUSION MATRIX FOR A CREDIT SCORING PROBLEM

	Good applicant	Bad applicant
Predicted as Good	TN	FN
Predicted as Bad	FP	TP

Note: Positive class is Bad.

IV. EMPIRICAL RESULTS

A. The Single Baseline Classifiers

Table III shows the prediction accuracy, F-measure and average rank of single baseline classifiers. In order to rank the performance of the models, we choose the average rank method as presented in [23].

TABLE III.   ACCURACY, F-MEASURE AND AVERAGE RANK OF THE SINGLE BASELINE MODELS

Model	Accuracy				F-measure				Average Rank	
	AUS	*GER*	*JPN*	*IRA*	*AUS*	*GER*	*JPN*	*IRA*	*Accuracy*	*F-measure*
SVMPSO	97.54	93.90	98.47	98.92	97.15	88.68	98.62	91.54	1.00	1.00
NN	87.54	81.20	87.29	98.66	86.65	66.19	87.70	89.25	3.00	3.00
AMLP	87.10	77.40	87.90	98.57	85.53	65.65	88.47	88.62	3.75	3.75
LR	87.68	73.40	88.82	96.21	86.82	64.53	89.62	76.65	4.00	3.25
DT	86.81	68.00	86.68	98.89	86.32	46.67	86.80	90.65	5.00	4.50
SVM	85.51	77.90	86.52	94.75	85.03	55.71	86.67	48.58	5.50	6.50
KNN	73.33	74.80	74.89	95.48	63.35	37.31	77.66	54.34	6.75	7.50
NB	80.72	74.40	80.40	93.29	75.77	60.25	83.55	52.11	7.00	6.50

Note: All the numbers are in percentage form. The analyses are carried out in RapidMiner 7.2 program.

This method helps us compare various model performances and identify the top performers. As shown, the SVMPSO has shown the best predictive accuracy and F-measure and it is ranked first among the other baseline classifiers. SVMs are first

introduced by authors in [51] as a form of linear classifiers. SVMs can be utilized for twofold classification with the aim of generating a best hyperplane (line) that sorts the input information into two classes (bad or good credit) [52]. As a

population-based stochastic optimization approach, particle swarm optimization (PSO) works via simulating the birds' behavior inside a flock. This algorithm is introduced by authors in [53] and [54]. PSO can be used to improve the accuracy of SVMs through identifying the best hyperplane which separates two classes. As can be seen from the results, the SVM-PSO is one of the highly accurate heuristic classifiers but only few studies have applied this method. According to the results, the second and third best performers are NN and AMLP. Conversely, the worst predictive performance on average belongs to NB. Here, F-measure shows the overall performance of the model via combining precision and recall values.

B. The Hybrid Meta-Learner Models

Table IV represents the prediction accuracy, F-measure and average rank of the best hybrid meta-learner models. As it is shown in this table, the (KNN-NN-SVMPSO)-(DL)-(DBSCAN) model has achieved the best accuracy rate and F-measure compared with other hybrid models on all credit

datasets. It is interesting that other high performing models also have the same stacking combination as the best model. It is obvious that employing an optimized version of support vector machine, SVMPSO, has a significant role in improving the accuracy of the results. Moreover, among the clustering algorithms, DBSCAN has highest performance and followed by SOM and FCM. DBSCAN is a data clustering algorithm introduced in [55] and it is a density-based clustering algorithm that groups data points which are closely located together. DBSCAN is known as one of the top performing clustering algorithms in the literature. The advantages of DBSCAN algorithm can be enumerated as follows:

- Unlike the K-means, DBSCAN does not need the number of clusters in the dataset to be specified a priori.

- It has the ability to detect clusters which are arbitrarily shaped. It can even discover a cluster entirely bounded by (but not linked to) a different cluster.

TABLE IV.   ACCURACY, F-MEASURE AND AVERAGE RANK OF THE BEST HYBRID MODELS

Model	Accuracy				F-measure				Average Rank	
	AUS	GER	JPN	IRA	AUS	GER	JPN	IRA	Accuracy	F-measure
(KNN-NN-SVMPSO)-(DL)-(DBSCAN)	99.71	99.80	99.85	99.90	99.67	99.67	99.86	99.23	2.00	2.00
(KNN-NN-SVMPSO)-(DL)-(SOM)	99.86	99.70	99.85	99.80	99.84	99.50	99.86	98.45	2.13	2.13
(KNN-NN-SVMPSO)-(DL)-(FCM)	99.71	99.70	99.85	99.74	99.67	99.50	99.86	97.99	3.25	3.25
(KNN-NN-SVMPSO)-(DL)-(EM)	99.86	99.60	99.69	99.78	99.84	99.33	99.72	98.29	3.50	3.50
(KNN-NN-SVMPSO)-(DL)-(KM)	99.71	99.70	99.69	99.72	99.67	99.50	99.72	97.82	4.13	4.13
(DT-KNN-NN)-(DL)-(FCM)	95.65	97.70	94.33	99.48	95.07	96.19	94.71	95.95	11.88	11.38
(DT-KNN-NN)-(DL)-(KM)	95.94	97.90	93.88	99.48	95.39	96.45	94.30	95.91	12.13	13.50
(KNN-NN-SVM)-(DL)-(FCM)	95.22	98.10	94.33	99.46	94.60	96.82	94.78	95.78	15.00	14.63
(NB-KNN-NN)-(DL)-(FCM)	95.65	97.80	94.03	99.43	95.03	96.33	94.42	95.61	16.00	16.75
(KNN-NN-SVM)-(DL)-(DBSCAN)	95.80	97.70	94.03	99.42	95.25	96.10	94.48	95.53	16.50	15.38
(KNN-NN-LR)-(DL)-(SOM)	95.65	97.60	94.18	99.42	95.07	95.99	94.54	95.44	17.25	17.63
(DT-KNN-NN)-(DL)-(EM)	95.80	97.10	93.87	99.55	95.27	95.09	94.40	96.45	17.38	16.75
(DT-KNN-NN)-(DL)-(DBSCAN)	95.36	97.80	93.72	99.55	94.61	96.28	94.12	96.51	17.75	18.50
(NB-KNN-NN)-(DL)-(EM)	95.36	97.40	94.49	99.45	94.70	95.65	94.87	95.74	17.88	18.00
(KNN-NN-AMLP)-(DL)-(EM)	96.52	95.90	94.03	99.46	96.03	92.79	94.45	95.73	17.88	18.25
(KNN-NN-AMLP)-(DL)-(FCM)	95.51	95.70	94.49	99.47	94.99	92.77	94.99	95.80	17.88	17.75
(KNN-NN-AMLP)-(DL)-(DBSCAN)	93.91	98.20	94.49	99.41	93.02	97.01	94.89	95.28	18.87	19.50
(KNN-NN-AMLP)-(DL)-(KM)	95.51	96.00	94.18	99.47	94.82	93.22	94.49	95.87	19.00	20.13
(KNN-NN-AMLP)-(DL)-(SOM)	94.93	98.30	94.33	99.39	94.14	97.12	94.78	95.27	19.63	19.13
(KNN-NN-LR)-(DL)-(DBSCAN)	95.51	98.10	94.18	99.37	94.89	96.85	94.69	95.17	19.63	18.00
(DT-KNN-NN)-(DL)-(SOM)	95.07	97.30	94.03	99.52	94.48	95.43	94.42	96.23	20.13	20.88
(KNN-NN-LR)-(DL)-(KM)	95.36	97.50	94.18	99.44	94.81	95.83	94.53	95.71	20.50	18.88
(NB-KNN-NN)-(DL)-(KM)	95.65	97.50	93.72	99.43	95.15	95.81	94.23	95.48	21.13	21.00
(KNN-NN-LR)-(DL)-(FCM)	95.80	97.20	93.87	99.42	95.19	95.29	94.32	95.47	21.25	21.88
(NB-KNN-NN)-(DL)-(DBSCAN)	95.51	96.90	94.18	99.41	95.06	94.70	94.60	95.52	22.75	19.88
(KNN-NN-LR)-(DL)-(EM)	95.51	97.00	93.87	99.43	95.01	94.92	94.43	95.52	23.63	22.38
(LR-NN-AMLP)-(DL)-(KM)	95.65	95.40	93.11	99.46	95.03	91.87	93.58	95.76	23.88	24.38
(KNN-NN-SVM)-(DL)-(KM)	95.36	97.70	93.42	99.41	94.82	96.10	94.07	95.39	24.38	24.13
(KNN-NN-SVM)-(DL)-(EM)	95.65	97.50	93.87	99.37	95.02	95.77	94.32	95.00	24.38	25.88
(NB-KNN-NN)-(DL)-(SOM)	95.36	97.20	93.87	99.41	94.74	95.38	94.33	95.39	25.88	25.63
(LR-NN-AMLP)-(DL)-(EM)	93.04	97.70	92.50	99.45	92.23	96.17	93.17	95.62	26.50	26.75
(KNN-NN-SVM)-(DL)-(SOM)	95.22	97.50	93.72	99.40	94.53	95.83	94.22	95.41	27.00	26.13
(LR-NN-AMLP)-(DL)-(DBSCAN)	93.33	97.60	92.8	99.43	92.51	96.00	93.41	95.47	27.75	28.62
(LR-NN-AMLP)-(DL)-(SOM)	93.91	97.20	94.03	99.39	92.91	95.29	94.44	95.15	28.00	28.63
(LR-NN-AMLP)-(DL)-(FCM)	93.91	97.60	93.42	99.36	93.16	95.86	93.99	94.94	30.75	30.75
(DT-NB-LR)-(DL)-(KM)	94.49	89.30	90.66	99.38	93.75	83.56	91.30	95.07	35.88	35.63
(DT-NB-LR)-(DL)-(FCM)	93.62	88.80	91.27	99.39	92.59	82.22	91.80	95.07	36.25	37.38
(DT-NB-LR)-(DL)-(DBSCAN)	94.20	88.50	91.27	99.38	93.38	80.21	91.82	95.02	36.25	36.50
(DT-NB-LR)-(DL)-(EM)	93.33	89.20	90.81	99.39	92.65	80.99	91.71	95.12	37.00	37.25
(DT-NB-LR)-(DL)-(SOM)	93.48	89.60	91.73	99.36	92.66	82.61	92.22	94.83	37.13	37.25

Note: See Table III

- It is robust to outliers so it can successfully deal with noises in datasets.

Table V compares the best baseline and hybrid models in terms of accuracy, F-measure, type I and II errors. The best hybrid model has improved the accuracy rate and F-measure results of the best baseline model by 2.68% and 5.97% on average, respectively. In contrast, the best hybrid model has reduced the type I and II errors by 79.84% and 95.16% on average, respectively. Furthermore, as we have included three public real-world credit datasets in this study, it is easy to compare the predictive performance of our hybrid models with other studies in the literature.

Table VI presents a comprehensive summary of credit scoring performances of various models which are used in the literature by several researchers. As the Australian and German credit data sets are the most frequently used datasets in the relevant literature, we also collected the results of the papers which have used these two datasets as their benchmarks. As shown, different authors have applied different models in their studies and they found various results in terms of predictive performance. It is interesting that the best performing hybrid model of this study is placed in the first rank based on the prediction accuracy measure. This hybrid credit scoring model has reached to an astonishing 99.71% and 99.80% accuracy on Australian and German datasets, respectively.

TABLE V.   AVERAGE PERFORMANCE COMPARISON OF THE BEST BASELINE AND HYBRID MODELS

	Best baseline SVMPSO				Best hybrid (KNN-NN-SVMPSO)-(DL)-(DBSCAN)			
	Accuracy	F-measure	Error I	Error II	Accuracy	F-measure	Error I	Error II
AUS	97.54	97.15	0.00	5.54	99.71	99.67	0.26	0.33
GER	93.90	88.68	0.00	20.33	99.80	99.67	0.14	0.33
JPN	98.47	98.62	3.38	0.00	99.85	99.86	0.34	0.00
IRA	98.92	91.54	0.46	10.02	99.90	99.23	0.03	1.08
Average	97.21	94.00	0.96	8.97	99.82	99.61	0.19	0.43
Improvement	-	-	-	-	2.68	5.97	-79.84	-95.16

Note: "Improvement" shows the percentage growth rate for the accuracy and F-measure. It also show the type I and II error reduction rates in percentage.

TABLE VI.   PERFORMANCE COMPARISON OF VARIOUS CREDIT SCORING MODELS IN THE LITERATURE

Model	Australian	German	Rank	Author(s)	Year
(KNN-NN-SVMPSO)-(DL)-(DBSCAN)	99.71	99.80	1	(This study)	2017
MC-LR (Intersection)	99.11	99.18	2	Tsai,Hsu [56]	2013
Hybrid SOM-KM-NN	97.98	98.46	3	Hsieh [17]	2005
MLP + FS	-	97.20	4	Analide [57]	2011
LibSVM	86.38	94.00	5	Peng, Kou, Shi,Chen [58]	2008
Hybrid NN	91.61	87.45	6	Tsai,Hung [59]	2014
AMMLP	92.75	84.67	7	Marcano-Cedeno, Marin-De-La-Barcena, Jiménez-Trillo, Pinuela,Andina [60]	2011
Gaussian classifier	92.6	83.80	8	Somol, Baesens, Pudil,Vanthienen [61]	2005
ANN	97.32	78.97	9	Tsai,Wu [62]	2008
VBDTM	91.97	81.64	10	Zhang, Zhou, Leung,Zheng [63]	2010
DeepSVM	88.98	83.70	11	Qi, Wang, Tian,Zhang [64]	2016
PSO-SVM	91.03	81.62	12	Lin, Ying, Chen,Lee [65]	2008
SVM	85.70	-	13	Martens, Baesens, Van Gestel,Vanthienen [66]	2007
CLC	86.52	84.80	14	Luo, Cheng,Hsieh [67]	2009
MLP	90.20	79.11	15	Tsai [68]	2008
2SGP	89.17	79.49	16	Huang, Tzeng,Ong [69]	2006
GNG + MARS	88.10	79.00	17	Ala'raj,Abbod [28]	2016
RS-Bagging DT	88.17	78.36	18	Wang, Ma, Huang,Xu [24]	2012
Genetic programming	88.27	77.34	19	Ong, Huang,Tzeng [16]	2005
Parallel Random Forest	89.40	76.20	20	Van Sang, Nam,Nhan [70]	2016
LS-SVM	90.40	74.60	21	Baesens, Van Gestel, Viaene, Stepanova, Suykens,Vanthienen [71]	2003
FA-MLP	86.08	78.76	22	Tsai [72]	2009
SVM + GA	86.90	77.92	23	Huang, Chen,Wang [73]	2007
SVDD-FSVM	87.25	77.30	24	Shi,Xu [27]	2016
RBF-SVM	87.52	76.60	25	Ping,Yongheng [74]	2011
Genetic Fuzzy classifier	88.60	75.00	26	Lahsasna, Ainon,Wah [11]	2010
Mixture-of-experts network	87.25	76.30	27	West [5]	2000
LDA + SVM	86.52	76.70	28	Chen,Li [75]	2010
Bayes	86.70	76.00	29	Hoffmann, Baesens, Mues, Van Gestel,Vanthienen [76]	2007
GR-GA-SVM	86.84	75.75	30	Liu, Fu,Lin [77]	2010
RS-LMNC	87.05	74.67	31	Nanni,Lumini [78]	2009
Adopted CBA	86.96	74.40	32	Lan, Janssens, Chen,Wets [79]	2006
HGA-NN	-	78.90	33	Oreski, Oreski,Oreski [80]	2012
ECSC	86.86	70.60	34	Xiao, Xiao,Wang [26]	2016

Note: The above models are ranked based on their average performances on Australian and German credit datasets.

V. CONCLUSION

As financial institutions are exposed to credit risk when issuing consumer loans, developing reliable credit scoring systems is crucial for them. Since, machine learning methods have demonstrated their applicability and merit, this study develops and compares several hybrid machine learning approaches for the credit scoring problem. In this paper, a novel framework is proposed for hybrid meta-learning to improve the predictive performance of credit scoring models. Based on the selected datasets, the results show that the hybrid meta-learner model of (KNN-NN-SVMPSO)-(DL)-(DBSCAN) outpaces all the literature's baseline classifiers in terms of accuracy rate and type I/II errors. This model also outperforms the best models used in the relevant literature in terms of accuracy rate with a significant margin. The findings of this study give us the insight to realize which type of hybrid machine learning techniques is capable of achieving higher accuracy and lower error rates in the case of credit scoring. Also, it is clear that the optimized version of support vector machine, SVMPSO, and deep learning algorithms have significant roles in enhancing predictive power of the proposed models. As a result, it is believed that using the best credit scoring model identified by this study can help financial institutions to make more accurate and confident credit decisions in the future. For further studies, several issues can be studied in the future. One of them can be the reduction of feature dimensionality. To be precise, the pre-process treatment of selected datasets for dimensionality reduction or feature selection can be beneficial to achieve better prediction precision [72]. Although, this paper is quite rich in terms of employing various machine learning algorithms, there are other techniques which can be applied for further comparisons especially those algorithms which are heuristically optimized. Lastly, since this paper specifically concentrates on the credit scoring problem, future studies can examine other problem areas such as corporate loans, house and car loans to identify which hybrid method has the best performance or if the empirical outcomes differ from the results of this paper.

REFERENCES

[1] D. J. Hand, W. E. Henley, "Statistical classification methods in consumer credit scoring: a review", Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 160, No. 3, pp. 523-541, 1997

[2] C. R. Abrahams, M. Zhang, Fair lending compliance: Intelligence and implications for credit risk management, John Wiley & Sons, 2008

[3] E. Rosenberg, A. Gleit, "Quantitative methods in credit management: a survey", Operations Research, Vol. 42, No. 4, pp. 589-613, 1994

[4] L. C. Thomas, D. B. Edelman, J. N. Crook, Credit scoring and its applications, SIAM, 2002

[5] D. West, "Neural network credit scoring models", Computers & Operations Research, Vol. 27, No. 11-12, pp. 1131-1152, 2000

[6] R. C. Lacher, P. K. Coats, S. C. Sharma, L. F. Fant, "A neural network for classifying the financial health of a firm", European Journal of Operational Research, Vol. 85, No. 1, pp. 53-65, 1995

[7] T.-S. Lee, I.-F. Chen, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines", Expert Systems with Applications, Vol. 28, No. 4, pp. 743-752, 2005

[8] H. Abdou, J. Pointon, A. El-Masry, "Neural nets versus conventional techniques in credit scoring in Egyptian banking", Expert Systems with Applications, Vol. 35, No. 3, pp. 1275-1292, 2008

[9] M.-C. Chen, S.-H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques", Expert Systems with Applications, Vol. 24, No. 4, pp. 433-441, 2003

[10] V. S. Desai, J. N. Crook, G. A. Overstreet, "A comparison of neural networks and linear scoring models in the credit union environment", European Journal of Operational Research, Vol. 95, No. 1, pp. 24-37, 1996

[11] A. Lahsasna, R. N. Ainon, T. Y. Wah, "Enhancement of transparency and accuracy of credit scoring models through genetic fuzzy classifier", Maejo International Journal of Science and Technology, Vol. 4, No. 1, pp. 136-158, 2010

[12] T.-S. Lee, C.-C. Chiu, C.-J. Lu, I.-F. Chen, "Credit scoring using the hybrid neural discriminant technique", Expert Systems with applications, Vol. 23, No. 3, pp. 245-254, 2002

[13] M.-C. Tsai, S.-P. Lin, C.-C. Cheng, Y.-P. Lin, "The consumer loan default predicting model–An application of DEA–DA and neural network", Expert Systems with applications, Vol. 36, No. 9, pp. 11682-11690, 2009

[14] J. N. Crook, D. B. Edelman, L. C. Thomas, "Recent developments in consumer credit risk assessment", European Journal of Operational Research, Vol. 183, No. 3, pp. 1447-1465, 2007

[15] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study", Decision Support Systems, Vol. 37, No. 4, pp. 543-558, 2004

[16] C.-S. Ong, J.-J. Huang, G.-H. Tzeng, "Building credit scoring models using genetic programming", Expert Systems with Applications, Vol. 29, No. 1, pp. 41-47, 2005

[17] N.-C. Hsieh, "Hybrid mining approach in the design of credit scoring models", Expert Systems with Applications, Vol. 28, No. 4, pp. 655-665, 2005

[18] A. Jain, A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting", Applied Soft Computing, Vol. 7, No. 2, pp. 585-592, 2007

[19] H. Kim, K. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets", Applied Soft Computing, Vol. 7, No. 2, pp. 569-576, 2007

[20] R. Malhotra, D. K. Malhotra, "Differentiating between good credits and bad credits using neuro-fuzzy systems", European Journal of Operational research, Vol. 136, No. 1, pp. 190-211, 2002

[21] J. Huysmans, B. Baesens, J. Vanthienen, T. Van Gestel, "Failure prediction with self organizing maps", Expert Systems with Applications, Vol. 30, No. 3, pp. 479-487, 2006

[22] C. Tsai, M. Chen, "Credit rating by hybrid machine learning techniques", Applied Soft Computing, Vol. 10, No. 2, pp. 374-380, 2010

[23] I. Brown, C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets", Expert Systems with Applications, Vol. 39, No. 3, pp. 3446-3453, 2012

[24] G. Wang, J. Ma, L. Huang, K. Xu, "Two credit scoring models based on dual strategy ensemble trees", Knowledge-Based Systems, Vol. 26, pp. 61-68, 2012

[25] T. Harris, "Credit scoring using the clustered support vector machine", Expert Systems with Applications, Vol. 42, No. 2, pp. 741-750, 2015

[26] H. Xiao, Z. Xiao, Y. Wang, "Ensemble classification based on supervised clustering for credit scoring", Applied Soft Computing, Vol. 43, pp. 73-86, 2016

[27] J. Shi, B. Xu, "Credit Scoring by Fuzzy Support Vector Machines with a Novel Membership Function", Journal of Risk and Financial Management, Vol. 9, No. 4, pp. 13, 2016

[28] M. Ala'raj, M. F. Abbod, "A new hybrid ensemble credit scoring model based on classifiers consensus system approach", Expert Systems with Applications, Vol. 64, pp. 36-55, 2016

[29] M. J. Lenard, G. R. Madey, P. Alam, "The design and validation of a hybrid information system for the auditor's going concern decision", Journal of Management Information Systems, Vol. 14, No. 4, pp. 219-237, 1998

[30] T. G. Dietterich, "Ensemble learning", The handbook of brain theory and neural networks, Vol. 2, pp. 110-125, 2002

[31] M. Tavana, K. Puranam, Handbook of Research on Organizational Transformations through Big Data Analytics, IGI Global, 2014

[32] F. Anifowose, J. Labadin, A. Abdulraheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines", Applied Soft Computing, Vol. 26, pp. 483-496, 2015

[33] J. Kittler, M. Hatef, R. P. Duin, J. Matas, "On combining classifiers", IEEE transactions on pattern analysis and machine intelligence, Vol. 20, No. 3, pp. 226-239, 1998

[34] D. Opitz, R. Maclin, "Popular ensemble methods: An empirical study", Journal of Artificial Intelligence Research, Vol. 11, pp. 169-198, 1999

[35] M. Pal, "Ensemble learning with decision tree for remote sensing classification", World Academy of Science, Engineering and Technology, Vol. 1, No. 12, pp. 3839-3841, 2007

[36] D. H. Wolpert, "Stacked generalization", Neural networks, Vol. 5, No. 2, pp. 241-259, 1992

[37] J. Sill, G. Takacs, L. Mackey, D. Lin, "Feature-weighted linear stacking", arXiv:0911.0460, 2009

[38] M. Tan, "Multi-agent reinforcement learning: Independent vs. cooperative agents", Tenth International Conference on Machine Learning pp. 330-337, 1993

[39] . Breiman, "Stacked regressions", Machine learning, Vol. 24, No. 1, pp. 49-64, 1996

[40] M. Ozay, F. T. Y. Vural, "A new fuzzy stacked generalization technique and analysis of its performance", arXiv:1204.0171, 2012

[41] P. Smyth, D. Wolpert, "Linearly combining density estimators via stacking", Machine Learning, Vol. 36, No. 1-2, pp. 59-83, 1999

[42] T. M. Mitchell, Machine learning. 1997, Burr Ridge, IL: McGraw Hill, 1997

[43] A. K. Jain, M. N. Murty, P. J. Flynn, "Data clustering: a review", ACM Computing Surveys (CSUR), Vol. 31, No. 3, pp. 264-323, 1999

[44] G. E. Hinton, S. Osindero, Y.-W. Teh, "A fast learning algorithm for deep belief nets", Neural Computation, Vol. 18, No. 7, pp. 1527-1554, 2006

[45] I. Goodfellow, Q. V. Le, A. M. Saxe, H. Lee, A. Y. Ng, "Measuring invariances in deep networks", 23rd Annual Conference on Neural Information Processing Systems, pp. 646-654, 2009

[46] D. C. Ciresan, U. Meier, L. M. Gambardella, J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition", Neural Computation, Vol. 22, No. 12, pp. 3207-3220, 2010

[47] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures", arXiv:1206.5533, 2012

[48] Y. Bengio, A. Courville, P. Vincent, "Representation learning: A review and new perspectives", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, pp. 1798-1828, 2013

[49] A. Candel, V. Parmar, E. LeDell, A. Arora, Deep Learning with H2O, H2O Inc., 2016

[50] S. Ramaswamy, R. Rastogi, K. Shim, "Efficient algorithms for mining outliers from large data sets", ACM SIGMOD International Conference On Management Of Data, pp. 427-438, 2000

[51] C. Cortes, V. Vapnik, "Support-vector networks", Machine Learning, Vol. 20, No. 3, pp. 273-297, 1995

[52] S. Li, W. Shiue, M. Huang, "The evaluation of consumer loans using support vector machines", Expert Systems with Applications, Vol. 30, No. 4, pp. 772-782, 2006

[53] R. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory", Sixth International Symposium on Micro Machine and Human Science pp. 39-43, 1995

[54] J. F. Kennedy, J. Kennedy, R. C. Eberhart, Y. Shi, Swarm intelligence, Morgan Kaufmann, 2001

[55] M. Ester, H. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", Second International Conference on Knowledge Discovery and Data Mining, pp. 226-231, 1996

[56] C. F. Tsai, Y. F. Hsu, "A Meta‑learning Framework for Bankruptcy Prediction", Journal of Forecasting, Vol. 32, No. 2, pp. 167-179, 2013

[57] F. S. C. Analide, "Information asset analysis: credit scoring and credit suggestion", International Journal of Electronic Business, Vol. 9, No. 3, pp. 203-218, 2011

[58] Y. Peng, G. Kou, Y. Shi, Z. Chen, "A multi-criteria convex quadratic programming model for credit data analysis", Decision Support Systems, Vol. 44, No. 4, pp. 1016-1030, 2008

[59] C. Tsai, C. Hung, "Modeling credit scoring using neural network ensembles", Kybernetes, Vol. 43, No. 7, pp. 1114-1123, 2014

[60] A. Marcano-Cedeno, A. Marin-De-La-Barcena, J. Jimenez-Trillo, J. Pinuela, D. Andina, "Artificial metaplasticity neural network applied to credit scoring", International Journal of Neural Systems, Vol. 21, No. 04, pp. 311-317, 2011

[61] P. Somol, B. Baesens, P. Pudil, J. Vanthienen, "Filter‑ versus wrapper‑ based feature selection for credit scoring", International Journal of Intelligent Systems, Vol. 20, No. 10, pp. 985-999, 2005

[62] C. Tsai, J. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring", Expert Systems with Applications, Vol. 34, No. 4, pp. 2639-2649, 2008

[63] D. Zhang, X. Zhou, S. C. Leung, J. Zheng, "Vertical bagging decision trees model for credit scoring", Expert Systems with Applications, Vol. 37, No. 12, pp. 7838-7843, 2010

[64] Z. Qi, B. Wang, Y. Tian, P. Zhang, "When Ensemble Learning Meets Deep Learning: a New Deep Support Vector Machine for Classification", Knowledge-Based Systems, Vol. 107, pp. 54-60, 2016

[65] S. Lin, K. Ying, S. Chen, Z. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines", Expert Systems with Applications, Vol. 35, No. 4, pp. 1817-1824, 2008

[66] D. Martens, B. Baesens, T. Van Gestel, J. Vanthienen, "Comprehensible credit scoring models using rule extraction from support vector machines", European Journal of Operational Research, Vol. 183, No. 3, pp. 1466-1476, 2007

[67] S. Luo, B. Cheng, C. Hsieh, "Prediction model building with clustering-launched classification and support vector machines in credit scoring", Expert Systems with Applications, Vol. 36, No. 4, pp. 7562-7566, 2009

[68] C. F. Tsai, "Financial decision support using neural networks and support vector machines", Expert Systems, Vol. 25, No. 4, pp. 380-393, 2008

[69] J. Huang, G. Tzeng, C. Ong, "Two-stage genetic programming (2SGP) for the credit scoring model", Applied Mathematics and Computation, Vol. 174, No. 2, pp. 1039-1053, 2006

[70] H. Van Sang, N. H. Nam, N. D. Nhan, "A novel credit scoring prediction model based on Feature Selection approach and parallel random forest", Indian Journal of Science and Technology, Vol. 9, No. 20, 2016

[71] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring", Journal of the Operational Research Society, Vol. 54, No. 6, pp. 627-635, 2003

[72] C. Tsai, "Feature selection in bankruptcy prediction", Knowledge-Based Systems, Vol. 22, No. 2, pp. 120-127, 2009

[73] C. Huang, M. Chen, C. Wang, "Credit scoring with a data mining approach based on support vector machines", Expert Systems with Applications, Vol. 33, No. 4, pp. 847-856, 2007

[74] Y. Ping, L. Yongheng, "Neighborhood rough set and SVM based hybrid credit scoring classifier", Expert Systems with Applications, Vol. 38, No. 9, pp. 11300-11304, 2011

[75] F. Chen, F. Li, "Combination of feature selection approaches with SVM in credit scoring", Expert Systems with Applications, Vol. 37, No. 7, pp. 4902-4909, 2010

[76] F. Hoffmann, B. Baesens, C. Mues, T. Van Gestel, J. Vanthienen, "Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms", European Journal of Operational Research, Vol. 177, No. 1, pp. 540-555, 2007

[77] X. Liu, H. Fu, W. Lin, "A modified support vector machine model for credit scoring", International Journal of Computational Intelligence Systems, Vol. 3, No. 6, pp. 797-804, 2010

[78] L. Nanni, A. Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring", Expert Systems with Applications, Vol. 36, No. 2, pp. 3028-3033, 2009

[79] Y. Lan, D. Janssens, G. Chen, G. Wets, "Improving associative classification by incorporating novel interestingness measures", Expert Systems with Applications, Vol. 31, No. 1, pp. 184-192, 2006

[80] S. Oreski, D. Oreski, G. Oreski, "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment", Expert Systems with Applications, Vol. 39, No. 16, pp. 12605-12617, 2012