Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2747~2757
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.11159

Toward optimal bankruptcy prediction: evaluating ensemble methods using Taiwan and U.S. financial bankruptcy data

Nur Farahaina Idris[1], Mohd Arfian Ismail[1,2]
[1]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia
[2]Centre of Excellence for Artificial Intelligence and Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Gambang, Malaysia

Article Info

Article history:

Received Aug 5, 2025
Revised Apr 3, 2026
Accepted Apr 18, 2026

Keywords:

Adaptive boosting
Bankruptcy
Ensemble
Random forest
Stacking

ABSTRACT

Bankruptcy filings have increased significantly in many countries, causing widespread concern across society and triggering various economic issues. One contributing factor to the global rise in corporate bankruptcies is the unstable nature of companies' growth. This issue often driven by unclear financial strategies and weak business direction. Thus, bankruptcy prediction plays a vital role, enabling earlier intervention and allowing business owners to improve their financial strategies proactively. This research investigates the effectiveness of ensemble machine learning (ML) methods using random forest (RF), stacking, and adaptive boosting (AdaBoost) for the prediction of corporate bankruptcy using datasets from Taiwan and the U.S. In the experimental phase, the performance is assessed using accuracy, precision, recall, F1 score, relative absolute error, and time. RF scored the highest accuracy in the classification of Taiwan's bankruptcy data with 97.067%, meanwhile, AdaBoost M1 obtained the highest accuracy in the classification of the U.S.'s bankruptcy data with 94.0075%. The research shows that these methods, particularly AdaBoost M1, can improve early-warning systems and provide actionable insights for financial risk management. The main contribution of this research is its cross-country comparison of ensemble methods for bankruptcy prediction.

Corresponding Author:

Mohd Arfian Ismail
Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Pahang, Malaysia
Email: arfian@umpsa.edu.my

1. INTRODUCTION

Bankruptcy has risen significantly worldwide, including in more developed countries like the U.S. and European regions such as Germany, Ireland, and Luxembourg. This issue is driven by various factors such as personal loans, credit card debt, housing loans, and, more critically, corporate bankruptcy. According to 2024 data from Eurostat, corporate bankruptcies increased by 2.7% in the third quarter, while new business registrations also rose by 2.2% [1]. Corporate bankruptcy is driven by severe financial losses stemming from unclear financial strategies and government policies. The macroeconomic impacts of government policies, such as tariffs or increases in government taxes, can also cause disruptions in supply and demand chains. These issues may lead to price inflation that reduces society's purchasing power [2] and decreases consumer willingness to spend, ultimately impacting corporate businesses. These trends highlight the complex and multifaceted causes behind corporate failures in today's business landscape. It can lead to widespread financial distress, including job losses and a significant financial burden on business owners and society.

While bankruptcy has emerged as a critical financial concern, financial analysts have taken greater notice of the situation, prompting many corporations to adopt advanced business strategies aimed at improving

financial planning and resilience. As technology evolves, artificial intelligence (AI) is being utilized to assist financial analysts in making strategic business plans. One key area of AI is machine learning (ML), which focuses on algorithms and analyzing patterns in data. Nowadays, ML is applied in all kinds of fields, including healthcare, engineering, and finance. The growing reliance on ML-based decision-making has opened new opportunities in the financial field, particularly in more critical areas like risk management and bankruptcy prediction. ML methods are helpful as they can assist in detecting early warning signs of financial distress. Hence, giving business owners a valuable head start in adjusting strategies and avoiding potential losses.

Various renowned single-model ML methods have been utilized for the classification of financial data, such as support vector machines (SVM), neural networks (NN), Naïve Bayes (NB), and decision tree (DT) [3]–[5]. However, these techniques have not consistently shown a great performance in classification tasks, and their results often vary significantly depending on the type of financial data [6], [7]. Hence, after reviewing various studies, this research identified that the ensemble method architecture that combines multiple models may enhance model performance. Ensemble methods have already shown promising results in other fields, including healthcare [8], [9]. Furthermore, existing research has proven that ensemble methods perform relatively well in financial classification tasks such as fraud detection and credit scoring [10], [11]. Thus, making them a potentially suitable approach for bankruptcy prediction.

However, most existing bankruptcy prediction research focuses on datasets from a single country, which limits the generalizability and accuracy of their findings across different economic environments. Differences in regulatory frameworks and market structures across countries suggest that models validated in one country may not perform consistently in another [12], [13]. Therefore, a clear research gap exists in the comparative evaluation of ensemble ML methods. This paper contributes a comparative evaluation of ensemble methods for corporate bankruptcy prediction using datasets from both Taiwan and the U.S. By conducting a cross-country analysis, this research enhances the practicality and effectiveness of bankruptcy prediction methods and supports the development of more reliable early warning systems for financial risk management. The aim of this research is to identify the most effective method for corporate bankruptcy prediction and determine whether the ensemble methods are indeed more effective for bankruptcy classifications than single classifiers. In order to achieve this aim, the following objectives are briefly outlined:

- To study the behaviors and processes of ML methods, including ensemble methods.
- To review existing literature about ensemble methods for classification in bankruptcy prediction.
- To compare the results of the studied ensemble methods using various evaluation metrics such as accuracy, precision, recall, F1 score, relative absolute error (RAE), and also the time taken to build the model.
- To verify whether the best-performing ensemble method is a state-of-the-art technique by comparing it with many other existing methods, including single-model approaches.

The objectives of this research are divided into four vital components, which start with an in-depth study of ML methods, including single classifiers and ensemble methods. Secondly, various existing studies on the bankruptcy classification task will be thoroughly reviewed to understand the nature and expected outcomes. Based on the review, prominent ensemble methods, such as random forest (RF), stacking, and adaptive boosting M1 (AdaBoost M1), have been selected for the classification-based study of corporate bankruptcy prediction [11], [14], [15]. Then, a comparative analysis of the studied ensemble methods' results classification using the Taiwan and U.S. financial datasets will be conducted using several evaluation metrics to identify the best-performing method. Finally, the best-performing method will be compared with other approaches, including single classifiers, to verify and validate whether it is a state-of-the-art technique. To enhance clarity, this paper is structured as follows: related works, method, results, discussion, and conclusion.

2. METHOD

Classification is widely recognized by computer scientists as a fundamental technique to achieve accurate prediction and forecasting, as it involves an extraction of information from the data [16], [17]. Various ML methods can be applied to solve classification problems. However, this research identified that ensemble methods display superior capability in financial classification tasks, which is supported by multiple findings from existing literature. For instance, the stacking ensemble incorporated with convolutional neural network-long short-term memory (CNN-LSTM) as base classifiers had achieved a classification accuracy of 52.13% using a real-world Bitcoin price dataset collected from January 1, 2018, to August 31, 2019 [18]. This result is considered reasonable given the volatile nature of cryptocurrency markets and the challenging task of Bitcoin time-series data classification. Then, AdaBoost managed to achieve an accuracy of 86.96%, while RF and extra trees classifier both scored 88.41% using the Australian Credit dataset acquired from the UCI ML repository [19]. These results greatly outperformed traditional single classifiers such as logistic regression (LR) (83.33%), DT (82.61%), and K-Nearest Neighbors (71.74) [19]. Additionally, both bootstrap aggregation (bagging) methods that utilized DT and multilayer perceptron (MLP) as base classifiers achieved a perfect classification

accuracy of 100% when classifying the dataset obtained from the New York Stock Exchange (NYSE) [20]. Based on this same research, MLP with Boosting attained a high-performing accuracy score of 96.32% for stock market prediction [20]. Given the strong performance of various ensemble methods in the existing studies related to finance prediction, this research decided to adopt ensemble methods for corporate bankruptcy classification, despite the limited comparison and evaluation of results across different bankruptcy datasets.

2.1. Ensemble method

Ensemble methods are regarded as more advanced ML techniques because they involve more intricate processes compared to single classifiers. Technically, their operations and processes can differ significantly depending on the ensemble type. Nonetheless, the fundamental idea behind every ensemble method is to aggregate predictions from multiple models and leverage diverse perspectives before making a final decision output [21]. This approach assists in increasing the generalization capability and stability while simultaneously reducing the risk of overfitting [22]. By combining multiple models, ensemble methods are better able to capture complex patterns in data, which leads to better performance [23]. Various types of ensemble methods are available to improve classification performance, including bagging, stacking, and boosting. Some ensemble methods, such as bagging, RF, and AdaBoost are composed of homogeneous base classifiers, while others are heterogeneous, combining different types of base classifiers. In this context, base classifiers serve as individual models, whereas the ensemble acts as the overarching architecture. Based on the existing theoretical explanations, the effectiveness of ensemble methods is partly due to the reduction of bias [24]. This research focuses on ensemble methods for corporate bankruptcy prediction to evaluate and verify their potential in financial risk classification. The analysis and review in this research concentrate on the three most prevalent ensemble methods, which are stacking, RF, and AdaBoost M1, consecutively.

2.1.1. Stacking

Stacking is a classic ensemble method that has seen a gain in interest in recent years, largely due to advances in modern computing power. It employs a two-stage ensemble architecture. In the first stage, referred to as level 0, base classifiers are trained on the training set while predictions are made for the validation set, as the training data is split into training and validation sets. Then, for the second stage, known as level 1, a meta-classifier would be used for another round of classification. However, rather than learning from the original training data, the meta-classifier conducts classification using prediction outputs of the base classifiers to determine the test data prediction. This approach is designed to improve generalization and discover more complex data patterns. The existing research by Watono *et al.* [25] that implemented stacking using six diverse base classifiers acquired the classification accuracy and F1 score of 92.29% and 91.87%, respectively, when using the Bank Term Deposit dataset. These results are based on SVM as a meta-classifier and classification of the original non-resampled dataset. Overall, this result outperforms extreme gradient boost (XGBoost), which obtained an accuracy of 91.62%, and RF with an accuracy of 91.36%. Then, Muslim proposed a sophisticated method that combines a genetic algorithm and SVM for feature selection, with stacking as the classification method. The research reported that this novel stacking approach achieved an outstanding accuracy of 99.58% when classifying the Taiwanese Bankruptcy prediction dataset [26]. This result shows the high potential of stacking in bankruptcy prediction. However, despite its good performance, stacking may impose a high computational burden, particularly when it involves a combination of high complexity classifiers. In this research, stacking is implemented as a heterogeneous ensemble, where the base classifiers are of different types to ensure diversity in level 0 outputs. The base classifiers used are NB, LR, random tree (RT), and DT with C4.5 (J48) as the learning algorithm, while the meta-classifier, LR, is selected following the rule of thumb.

2.1.2. Random forest

Bagging is an ensemble method that uses bootstrapped and resampled data to train multiple models [8]. Each model was trained independently using different bootstrapped samples of training data. However, there are slight differences between bagging and RF that significantly affect the outcomes, particularly the introduction of feature randomness, in which at each split in the tree, it selects only a random subset of features. The feature randomness helps to decorrelate the trees and reduces the likelihood of constructing similar trees. Hence, this further reduces variance and solves the overfitting issues. RF is also suitable for both classification and regression problems, where it predicts the output using the mode for classification and the mean for regression. RF builds multiple DTs during the training phase, which reduces the impact of variance from individual trees, thereby increasing model stability and making it more robust to outliers. However, the implementation of multiple DTs also reduces the overall interpretability compared to a single DT. Based on existing research by Zhang [27], RF had a lower standard deviation of mean square error (MSE), which ranged only from 0.01 to 0.05, compared to DT, ranging between 0.03 to 0.07, indicating RF is less overfitting and has better generalization than DT when classifying COMPUSTAT financial data. It also had a better anomaly

detection accuracy compared with DT, ranging between 0.94 to 0.98 versus DT, with only between 0.9 to 0.96. Then, an existing research by Patel *et al.* [28] demonstrated that RF with an accuracy of 81.45% outperformed several methods, including other ensemble approaches such as bagging (74.16%), LogitBoost (78.65%), and Voting (53.25%) when classifying the Bombay Stock Exchange dataset. Nevertheless, a notable weakness of RF is its large model size, resulting from the use of unpruned trees and the typical requirement of around 100 base classifiers [29]. In this research, RF is implemented using RT as the base classifier, following the WEKA default setting.

2.1.3 Adaptive boosting M1

AdaBoost M1, which is commonly referred to as AdaBoost, was developed by Freund and Schapire [30] and has lower complexity than the later AdaBoost M2 variant. This method applies the boosting algorithm and utilizes base classifiers that are widely known as weak learners. The process of AdaBoost begins by assigning equal weights to all instances in the dataset, which forms a uniform weights distribution. In the first round, a weak learner is trained using these initial weights. The process then proceeds iteratively and sequentially [31], where in each round, the instance weights are adjusted to emphasize the misclassified instances. Specifically, misclassified instances would receive higher weights so that they would become more prominent for the next weak learner classification. Simultaneously, each trained weak learner is also assigned a weight based on its accuracy. More accurate models would receive higher weights and have a greater impact on the final prediction. This iterative training of weak learners is repeated for a set number of iterations. Lastly, all the weak learners are combined, and the final prediction is made using a weighted majority vote [32]. Based on research by Chang *et al.* [33] shows that AdaBoost obtained an accuracy of 84.7% for test data prediction of the Chinese stock market dataset, even with a 3% tolerance for uncertainties. Aside from achieving a strong accuracy performance, the same research revealed that AdaBoost achieved a remarkable precision score of 89.2% and a recall of 94.3%. Moreover, research by Tsai and Hung [34] revealed that traditional AdaBoost managed to acquire outstanding accuracies of 86.12%, 87.64%, 86.02%, and 86.84% for the Enterprise Performance datasets of China, Japan, Korea, and Taiwan, respectively. Based on the analysis, AdaBoost got a favorable average accuracy of 86.66% for two-class classification, which is just slightly lower than modified AdaBoost with 88.04% but relatively better than back propagation neural network with 84.64%. Despite the superior performance in existing financial studies, AdaBoost has a drawback involving high sensitivity to noise and outliers, and can be detrimental with class noise [35]. C4.5 (J48) is selected as the base classifier for AdaBoost M1 in our research because it is a stronger classifier than the decision stump that is typically applied for AdaBoost studies.

3. RESULTS

The evaluation metrics applied to assess the methods' performance are accuracy, precision, recall, and F1 score, which are computed using the confusion matrix. Accuracy is defined as the count of correctly predicted instances divided by the total instances. Precision is basically the ratio of true positives to the sum of true positives and false positives. Meanwhile, recall is the ratio of true positives to the sum of true positives and false negatives. The F1 score is the harmonic mean of both precision and recall, providing a balanced evaluation of both metrics. Additionally, the time taken to build models (measured in seconds) was used as a time-based performance evaluation while RAE was used to measure how well a model performs relative to a simple mean-based predictor, which is lower values of RAE indicate higher predictive accuracy. For the experimentation, stratified k-fold cross-validation with k set to ten. This approach was more reliable as each split fold of data has approximately the same percentage of samples as the full dataset, thereby reducing the risk of bias in results. The experiment was carried out thoroughly using the WEKA tool on a MacBook M1. Preprocessing was limited to WEKA's standard handling of missing values. All studied ensemble methods used WEKA's default hyperparameters, as no manual tuning was performed.

3.1. Results of the U.S. Bankruptcy dataset and Taiwan Bankruptcy dataset classification

In this research, two corporate bankruptcy datasets were utilized to evaluate the classification performance and predictive capability of the methods in the context of financial risk prediction. The first dataset is the U.S. Bankruptcy dataset, which was obtained from Kaggle and is publicly available. This novel dataset consists of 8,262 American public companies listed on the NYSE and NASDAQ, spanning the years 1999 to 2018. It originally comprised 21 attributes (including the target variable), but the company name feature was removed during preprocessing as it did not contribute to classification, resulting in a total of 20 attributes. The dataset includes 78,682 instances, with all features represented as numerical continuous values. The target variable, referred to as "class," indicates the status of a company, whether the company is still alive (sustained) or failed (bankrupt). This dependent variable is predicted based on a set of independent variables, which

represent the key financial indicators of a company's operational performance and fiscal stability. These include current assets (X1), cost of goods sold (X2), depreciation and amortization (X3), EBITDA (X4), inventory (X5), net income (X6), and total receivables (X7). Additional variables such as market value (X8), net sales (X9), total assets (X10), and long-term debt (X11) provide insights into the company's valuation and capital structure. Other vital features include EBIT (X12), gross profit (X13), current liabilities (X14), retained earnings (X15), total revenue (X16), total liabilities (X17), year (X19), and total operating expenses (X20), which collectively span a comprehensive range of financial performance.

Then, the second dataset used in this research is the Taiwan Bankruptcy dataset, which was collected from the publicly accessible UCI ML Repository. This dataset originates from the Taiwan Economic Journal and covers corporate records between 1999 and 2009, while the companies' bankruptcy status was determined following the criteria set by the Taiwan Stock Exchange. It consists of 6819 instances with 96 attributes related to a company's operational performance and solvency status. The solvency status acts as a target variable with the class label of 0 representing bankruptcy, and a label of 1 signifies non-bankruptcy. The 95 independent variables used as input features for the classification task are represented with various financial ratios and accounting figures that provide insights into the companies' financial health. These features consists of debt performance indicators like cost of interest-bearing debt (X1), interest expenses to total revenue (X5), and total liability to equity ratio (X6), as well as liquidity and leverage indicators such as the current ratio (X3), quick assets to current liability (X16), and degree of financial leverage (X36). Profitability and efficiency ratios are also included, such as operating income to capital (X10), gross profit to net sales (X54), and return on total assets (X52). Moreover, the dataset contains detailed cash flow and per-share metrics, such as cash flow per share (X44), earnings per share (EPS-Net Income) (X61), and cash flow to total assets (X84). Features consist of growth measures like net income growth (X90) and total asset growth (X93), incorporated in this dataset to capture dynamic financial trends. All the features are in numerical format with no missing data. Then, due to class imbalance in both datasets, the weighted average technique is used to compute evaluation metrics, including precision, recall, and F1 score. This technique is advantageous as it assigns greater importance to certain classes based on their significance and frequency. Tables 1 and 2 display the performance comparison of the ensemble methods for the U.S. and Taiwan Bankruptcy datasets, respectively.

Table 1. Performance comparison of ensemble methods on the U.S. bankruptcy dataset using accuracy, precision, recall, F1 score, RAE, and computation time

Method	Accuracy	Precision	Recall	F1-score	RAE	Time (s)
RF	93.9338	93.9	93.9	91.5	81.6658	43.72
AdaBoost M1	94.0075	92.7	94.0	92.9	48.2407	222.96
Stacking	93.5907	92.2	93.6	91.0	91.7633	102.01

Table 2. Performance comparison of ensemble methods on the Taiwan bankruptcy dataset using accuracy, precision, recall, F1 score, RAE, and computation time

Method	Accuracy	Precision	Recall	F1 score	RAE	Time (s)
RF	97.067	96.4	97.1	96.1	75.276	2.12
AdaBoost M1	96.7737	96.0	96.8	96.2	50.3425	12.59
Stacking	96.8177	95.7	96.8	95.8	92.6746	18.11

The classification results of the three ensemble methods differed marginally, only within 1% in terms of accuracy, precision, and recall for both datasets. However, a stark difference was observed in the time taken to build the classification models for the U.S. Bankruptcy dataset. Specifically, RF required only 43.72 seconds, while AdaBoost M1 and Stacking took 222.96 seconds and 102.01 seconds, respectively. A higher time taken to build the AdaBoost M1 and Stacking models reflects a greater computational cost and complexity. Despite the substantial difference in runtime, AdaBoost M1 achieved the best classification performance for the U.S. Bankruptcy dataset, with an accuracy of 94.0075%, a recall of 94%, and an F1 score of 92.9%, outperforming both RF and stacking. Furthermore, it recorded the lowest RAE at 48.2407%, compared to RF's 81.6658% and stacking's 91.7633%, demonstrating stronger generalization performance, as both RF and stacking exceeded 81%. Figure 1 illustrates the classification performance comparison of ensemble methods on the U.S. Bankruptcy dataset. For the Taiwan Bankruptcy dataset, RF achieved the highest accuracy at 97.067%, the highest precision at 96.4%, and the highest recall at 97.1%. It also had the shortest training time with only 2.12 seconds, compared to AdaBoost M1's 12.59 seconds and stacking's 18.11 seconds. The differences in accuracy performance on both datasets indicate that model performance is influenced by dataset characteristics such as feature structure, heterogeneity, and class imbalance. A detailed discussion of these effects is presented in the Discussion section. However, in terms of the F1 score, which is the harmonic mean of precision and recall,

AdaBoost M1 yielded a better result at 96.2%, outperforming RF. Additionally, AdaBoost M1 also achieved the lowest RAE with this dataset, with only 50.3425%, highlighting its stronger predictive capability. Figure 2 shows the classification performance comparison of ensemble methods on the Taiwan Bankruptcy dataset. These findings indicate that even though AdaBoost M1 did not always achieve the highest accuracy, it demonstrated the greatest potential for financial forecasting.

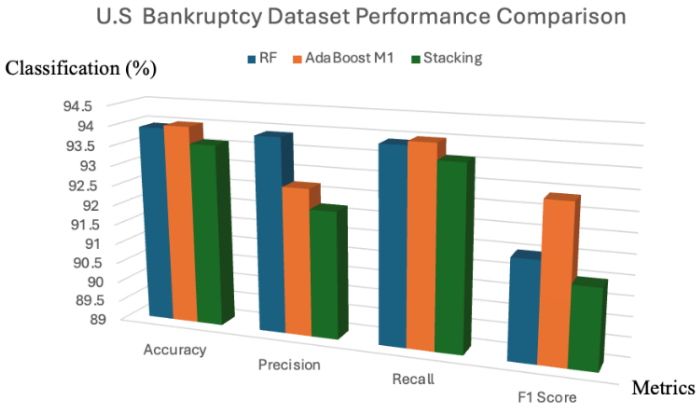

Figure 1. Classification performance comparison of ensemble methods (RF, AdaBoost M1, and stacking) on the U.S. bankruptcy dataset using accuracy, precision, recall, and F1 score

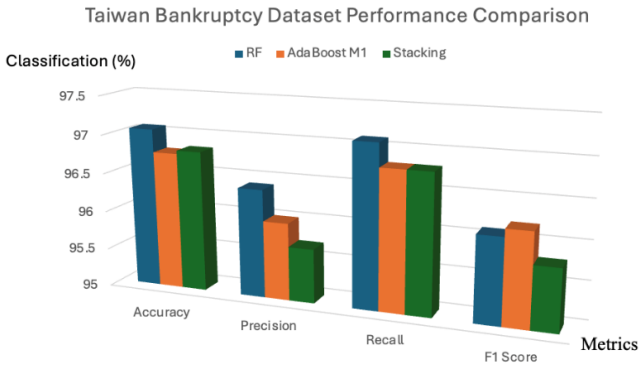

Figure 2. Performance comparison of ensemble methods (RF, AdaBoost M1, and stacking) on the Taiwan bankruptcy dataset using accuracy, precision, recall, and F1 score

3.2. Comparative analysis of the ensemble method

As this research identified that ensemble methods achieved superior classification performance, all exceeding 90% accuracy, a comparative analysis was conducted on the results of AdaBoost M1, which was found to have the highest potential for financial forecasting and prediction. The purpose of this comparative analysis is to validate whether AdaBoost M1 is comparable to other existing methods, including single-model approaches. Therefore, this research compares AdaBoost M1's results with other methods, such as C4.5 or J48, in the context of WEKA, Bayes network, MLP, and NB, all of which are renowned single classification techniques. Additionally, LogitBoost, another boosting method, was also selected for the comparative analysis for further validation. The classification results were derived from a preliminary study conducted using the WEKA tool, given that neither dataset is commonly utilized in benchmark research. For LogitBoost, REPTree was selected instead of the C4.5 base classifier, which had been used for AdaBoost M1, as it is not supported due to the production of multi-branch trees, which may destabilize boosting in the U.S. Bankruptcy dataset. Accuracy is selected as the evaluation metric in this analysis, as it reflects the overall correctness of classification. The performance analysis of the methods was conducted using both the U.S. and Taiwan bankruptcy datasets. Tables 3 and 4 present the comparative performance analyses of classification methods for the U.S. and Taiwan bankruptcy datasets, respectively.

Table 3. Comparative performance analysis of classification methods on the U.S. bankruptcy dataset

Method	Accuracy (%)
C4.5	93.1369
Bayes network	88.6861
MLP	93.3657
NB	14.1722
LogitBoost (REPTree base classifier)	93.2157
AdaBoost M1	94.0075

Table 4. Comparative performance analysis of classification methods on the Taiwan bankruptcy dataset

Method	Accuracy (%)
C4.5	95.9965
Bayes network	86.2736
MLP	96.1725
NB	69.805
LogitBoost (REPTree base classifier)	96.6124
AdaBoost M1	96.7737

Among the methods of C4.5, Bayes network, MLP, NB, LogitBoost, and AdaBoost M1, the latter managed to score the best classification accuracy with 94.0075%, while MLP ranked as the second-best with 93.3657% when classifying the U.S. Bankruptcy dataset. NB showed the lowest classification performance, with only 14.1722%, indicating its lack of predictive power, as real-world financial datasets often contain numerical features that are skewed and correlated, causing conflict with the method's assumption of a normal (Gaussian) distribution for continuous features. Moreover, the Bayes network also underperformed, achieving an accuracy of only 88.6861%. Despite having greater flexibility than NB, Bayes network still heavily relies on probabilistic assumptions and often requires discretization of features for better classification performance. Additionally, it requires extensive parameter tuning to effectively manage complex, continuous, and correlated data. Meanwhile, for the Taiwan Bankruptcy dataset, AdaBoost M1 also achieved the highest classification accuracy at 96.7737%, followed by the LogitBoost ensemble with 96.6124%. In contrast, Naive Bayes (69.805%) and Bayes network (86.2736%) showed comparatively weaker performance on this dataset, suggesting lower potential for financial prediction and a possible need for further parameter tuning. Figures 3 and 4 show the accuracy comparison of the classification methods for the U.S. and Taiwan Bankruptcy datasets, respectively. Based on the in-depth analysis, AdaBoost M1 outperforms those methods on both datasets.

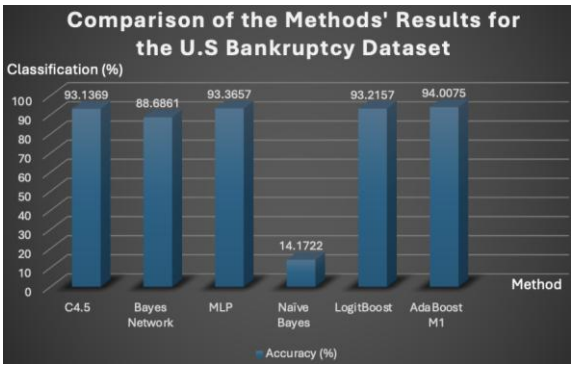

Figure 3. Accuracy comparison of classification methods on the U.S. bankruptcy dataset

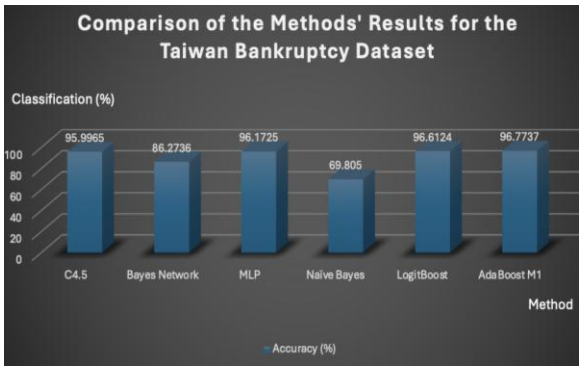

Figure 4. Accuracy comparison of classification methods on the Taiwan bankruptcy dataset

4. DISCUSSION

This research explores various ML methods, with a particular focus on ensemble methods for corporate bankruptcy classification. The collected datasets utilized for this research are the U.S. and Taiwan Bankruptcy datasets. The main limitation of these datasets is that they are heavily imbalanced. In the Taiwan Bankruptcy dataset about 6599 negative classes with a value of 0 (non-bankrupt) and only 220 positive classes with a value of 1 (bankrupt). Similarly, with the U.S Bankruptcy dataset, 73462 with alive as the class, while only 5220 with the failed class. The number of instances in both datasets is substantial for classification

purposes. The performance of RF, AdaBoost M1, and stacking was thoroughly analyzed using those datasets. The overall findings of the experiment indicate that the AdaBoost M1 achieved the highest classification accuracy for the U.S. bankruptcy classification, while RF obtained the highest accuracy result for the Taiwan dataset. This contrast in performance among ensemble methods across the two datasets can be largely attributed to differences in dataset characteristics, such as feature distributions and data size, as well as how each model responds to these variations. The Taiwanese dataset consists contains well-structured patterns and relatively low noise, despite a high degree of class imbalance. In this context, RF performs particularly well in terms of overall accuracy by aggregating multiple decision trees trained on bootstrapped samples, which effectively reduces variance and captures dominant patterns in the majority class. This leads to RF having better accuracy performance on the Taiwanese dataset with 97.067%. The result outperforms other existing studies like LR-Kendal Tau with 86.36% [36], SVM-Domain adaptation learning with 82%, and XGBoost with 94.5% [37]. The Taiwan Bankruptcy dataset consists of 6819 instances, which is slightly smaller and potentially cleaner.

In contrast, the U.S. bankruptcy dataset is larger, noisier and more heterogeneous, reflecting greater variability in the dataset. This increased complexity, combined with class imbalance, makes AdaBoost M1 more effective for the U.S. bankruptcy dataset. In terms of accuracy, it achieved 94.0075%, outperforming methods like artificial neural network, which acquired 78% and LR with 57% [38]. This is due to its iteratively emphasizes misclassified instances, which normally belong to the minority bankruptcy class. Overall, these distinctions in dataset size, quality, and complexity contribute to the observed differences in model performance, reinforcing the importance of aligning model choice with data characteristics. To determine which method offers overall greater predictive performance, this research also utilized the F1 score as the primary evaluation metric for ensemble methods comparison, as it balances both precision and recall. The selection of both accuracy and F1 score as the main evaluation metrics is common in ML studies, as they allow for a more accurate and balanced comparison of results across methods. In terms of F1 score, AdaBoost M1 recorded the highest values for both datasets, making it the best-performing ensemble method in this study. AdaBoost's superior F1 score performances on both datasets can be attributed to its ability to focus on misclassified instances, which improves recall for minority classes. While RF yields higher overall classification accuracy for the U.S Bankruptcy dataset by correctly classifying the majority class, it can sometimes underperform on the minority class, which leads to lower precision or recall. In contrast, AdaBoost iteratively reweights misclassified samples, giving it an edge in balancing both precision and recall, which is particularly reflected in a higher F1 score.

Then, the primary advantage of utilizing AdaBoost M1 is its robustness to overfitting. Overfitting happens when a model performs well on training data but fails to generalize to unseen test data. In contrast, AdaBoost M1 recorded the lowest RAE among the studied ensemble methods. For instance, in the classification of the U.S. Bankruptcy dataset, it had only 48.2407% RAE, and for the Taiwan Bankruptcy dataset classification, it had 50.3425% RAE, indicating that overfitting was not a major issue. The overfitting issue normally emerges when a model becomes too complex, leading the model to capture noise instead of meaningful data patterns. The excellent performance of AdaBoost in corporate bankruptcy prediction is heavily attributed to its boosting mechanism, which iteratively adjusts the weights of misclassified instances. This process allows the model to focus more on the minority class samples that are harder to classify. The reweighting strategy improves prediction for minority classes, making AdaBoost a strong candidate for imbalanced classification problems like bankruptcy prediction.

In this research, a comparative analysis is being carried out between AdaBoost M1 with several renowned single-model approaches, including MLP, NB, Bayes network, and C4.5, and another boosting method known as LogitBoost, using corporate bankruptcy datasets of two countries, was conducted to evaluate AdaBoost M1's effectiveness. Based on the analysis, the comparison based on classification accuracy revealed that AdaBoost M1 consistently delivered superior performance across both datasets. Aside from identifying AdaBoost M1 as the top-performing ensemble method, the findings also emphasized the overall strength of ensemble methods that outperformed single classifiers. An exception was observed with the U.S. Bankruptcy dataset, where the MLP slightly surpassed LogitBoost by obtaining a higher accuracy score of 93.3657%. However, ensemble methods clearly manage to outperform all the single classifiers for the Taiwan Bankruptcy dataset. These results act as evidence that further reinforces the advantage of ensemble methods, especially AdaBoost M1, in bankruptcy prediction tasks.

Despite its remarkable performance, AdaBoost has a critical limitation involving the training time. For instance, AdaBoost M1 took 12.59 seconds for the classification of the Taiwan Bankruptcy dataset. This duration is significantly longer compared to RF, which completed the training time in just 2.12 seconds. Aside from that, AdaBoost required 222.96 seconds to build the model for the U.S. Bankruptcy dataset. This duration is noticeably greater than both RF and stacking. All these results indicate that AdaBoost tends to be more time-consuming, reflecting its higher computational cost and complexity. The trade-off between performance and computation time is particularly relevant for large-scale financial datasets, where training time could increase

extensively. Hence, it impacts the practical deployment despite the method's superior F1 score and ability to handle class imbalance effectively. Worst case in scenarios, where fast training time is essential, such as real-time applications, this delay may cause a challenge. The long training time of AdaBoost stems from its iterative reweighting process, which proceeds until it fine-tunes the decision boundary for optimal results. Nevertheless, the need for split-second decisions is less critical in corporate bankruptcy prediction compared to fields like emergency healthcare or cybersecurity of malware intrusion. Thus, longer training time is not necessarily a major drawback for AdaBoost. It can be practical for financial monitoring and risk assessment. Banks can use it to identify firms at risk of bankruptcy. Thus, enabling early intervention and informed lending decisions. Not only that, but corporations can also leverage the cross-country predictions to manage international portfolio risk, while regulators may integrate the method into monitoring systems to detect emerging bankruptcy risks.

5. CONCLUSION

The main contribution of this research lies in its cross-country comparison of ensemble ML methods for corporate bankruptcy prediction. Specifically, this study evaluated the performance of well-known ensemble ML methods, including RF, AdaBoost M1, and stacking, for corporate bankruptcy prediction using datasets from Taiwan and the U.S. The ensemble methods managed to consistently outperform single classifiers, as they combine multiple models to reduce prediction errors and improve generalization. The results show that RF achieved the highest accuracy for the Taiwan dataset, while AdaBoost M1 had the highest classification accuracy for the U.S. dataset. In order to further evaluate the robustness of the method, other metrics such as precision, recall, F1 score, RAE, and time taken were used during the experiment. This led to the identification that AdaBoost M1 scored best in both datasets through the highest F1 score among the ensemble methods. Hence, highlighting the robustness of AdaBoost M1 in cross-national bankruptcy prediction. However, to further validate the generalization capability of AdaBoost M1 in financial risk classification tasks beyond bankruptcy prediction, the method should be tested using a broader range of financial datasets. Future research may explore more complex and larger datasets, such as cryptocurrency markets, real-time bankruptcy prediction and GDP growth indicators. In addition, future research should also focus on enhancing AdaBoost M1 by curating the selection of base classifiers, exploring deep learning–based ensembles and conducting more extensive data preprocessing, including techniques such as feature selection and fuzzification of data.

ACKNOWLEDGMENTS

All authors are delighted to express deep gratitude to our colleagues in the Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, for their support during the research implementation.

FUNDING INFORMATION

This research was funded by the Ministry of Higher Education, Malaysia, under the Konsortium Kecemerlangan Penyelidikan (KKP) grant JPT(BPKI)1000/016/018/25(96) through the Artificial Intelligence in Financial Investment Research Consortium (AIFIC). The grant was managed by Universiti Malaysia Pahang Al-Sultan Abdullah (Grant No.: RDU263903).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nur Farahaina Idris	✓	✓	✓			✓	✓	✓	✓	✓	✓			
Mohd Arfian Ismail				✓	✓				✓	✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY

Both datasets used in this study are publicly available: The U.S. Bankruptcy dataset was obtained from Kaggle. The Taiwan Bankruptcy dataset was retrieved from the UCI Machine Learning Repository.

REFERENCES

[1] "Business registration and bankruptcy index by NACE Rev.2 activity - quarterly datatle," Eurostat, 2024.

[2] D. M. F. Ikhsan, N. A. Aziz, and E. Mahyudin, "Case study on the implementation of goods and services tax (GST) in Malaysia and Singapore," *Journal of International Studies*, vol. 18, pp. 159–189, 2022, doi: 10.32890/jis2022.18.6.

[3] M. Brygała and T. Korol, "Personal bankruptcy prediction using machine learning techniques," *Economics and Business Review*, vol. 10, no. 2, pp. 118–142, Jun. 2024, doi: 10.18559/ebr.2024.2.1149.

[4] H. Xu, "Stock price prediction using decision tree classifier and LSTM network," *Applied and Computational Engineering*, vol. 37, no. 1, pp. 222–229, Feb. 2024, doi: 10.54254/2755-2721/37/20230512.

[5] I. Setiani, M. N. Tentua, and S. Oyama, "Prediction of banking stock prices using Naive Bayes method," in *Second UPY International Conference on Applied Science and Education (2nd UPINCASE)*, vol. 1823, no. 1, Mar. 2021, doi: 10.1088/1742-6596/1823/1/012059.

[6] Y. Ling and P. P. Wang, "Ensemble machine learning models in financial distress prediction: Evidence from China," *Journal of Mathematical Finance*, vol. 14, no. 02, pp. 226–242, 2024, doi: 10.4236/jmf.2024.142013.

[7] T. Yu and Y. Huo, "Classification of imbalanced data set in financial field based on combined algorithm," *Mobile Information Systems*, vol. 2022, pp. 1–7, Sep. 2022, doi: 10.1155/2022/1839204.

[8] A. Y. Mahmoud, "Novel efficient feature selection: Classification of medical and immunotherapy treatments utilising random forest and decision trees," *Intelligence-Based Medicine*, vol. 10, 2024, doi: 10.1016/j.ibmed.2024.100151.

[9] M. Ali, M. N. Haider, S. A. Lashari, W. Sharif, A. Khan, and D. A. Ramli, "Stacking classifier with random forest functioning as a meta classifier for diabetes diseases classification," *Procedia Computer Science*, vol. 207, pp. 3459–3468, 2022, doi: 10.1016/j.procs.2022.09.404.

[10] C. Liu, Y. Chan, S. H. A. Kazmi, and H. Fu, "Financial fraud detection model: Based on random forest," *International Journal of Economics and Finance*, vol. 7, no. 7, Jun. 2015, doi: 10.5539/ijef.v7n7p178.

[11] Z. Zhang, Y. Ma, and Y. Hua, "Financial fraud identification based on stacking ensemble learning algorithm: Introducing MD&A text information," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, Sep. 2022, doi: 10.1155/2022/1780834.

[12] C. Dewi *et al.*, "Feature selection for financial data classification using random forest, boruta, and recursive feature elimination," *Ingénierie des systèmes d information*, vol. 30, no. 08, pp. 2165–2173, Aug. 2025, doi: 10.18280/isi.300822.

[13] N. Elsayed, S. A. Elaleem, and M. Marie, "Improving prediction accuracy using random forest algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, pp. 436–441, 2024, doi: 10.14569/IJACSA.2024.0150445.

[14] I. D. Mienye and Y. Sun, "A survey of ensemble learning: concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[15] E. M. Ferrouhi and I. Bouabdallaoui, "A comparative study of ensemble learning algorithms for high-frequency trading," *Scientific African*, vol. 24, Jun. 2024, doi: 10.1016/j.sciaf.2024.e02161.

[16] N. Idris, "A review of homogenous ensemble methods on the classification of breast cancer data," *Przegląd Elektrotechniczn*, vol. 1, no. 1, pp. 103–106, Jan. 2024, doi: 10.15199/48.2024.01.21.

[17] N. S. Ruzgar, "Comparison of classification algorithms on financial data," *WSEAS Transactions on Computers*, vol. 18, pp. 256–263, 2019.

[18] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble deep learning models for forecasting cryptocurrency time-series," *Algorithms*, vol. 13, no. 5, May 2020, doi: 10.3390/a13050121.

[19] A. S. Parvin and B. Saleena, "An ensemble classifier model to predict credit scoring-comparative analysis," in *2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*, Dec. 2020, pp. 27–30, doi: 10.1109/iSES50453.2020.00017.

[20] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00299-5.

[21] M. S. Ali, M. M. Hossain, M. A. Kona, K. R. Nowrin, and M. K. Islam, "An ensemble classification approach for cervical cancer prediction using behavioral risk factors," *Healthcare Analytics*, vol. 5, Jun. 2024, doi: 10.1016/j.health.2024.100324.

[22] N. D. Ponnaganti and R. Anitha, "A novel ensemble bagging classification method for breast cancer classification using machine learning techniques," *Traitement du Signal*, vol. 39, no. 1, pp. 229–237, Feb. 2022, doi: 10.18280/ts.390123.

[23] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," *Heliyon*, vol. 10, no. 2, Jan. 2024, doi: 10.1016/j.heliyon.2024.e24536.

[24] L. Breiman, "Bias, variance, and arcing classifiers," Technical Report 460, Statistics Department, University of California, 1996.

[25] B. Watono, E. Utami, and D. Ariatmanto, "Implementation of stacking ensemble learning for bank term deposit acceptance classification," in *2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, Jun. 2024, pp. 98–104, doi: 10.1109/SIML61815.2024.10578260.

[26] M. A. Muslim *et al.*, "An ensemble stacking algorithm to improve model accuracy in bankruptcy prediction," *Journal of Data Science and Intelligent Systems*, vol. 2, no. 2, pp. 79–86, Mar. 2023, doi: 10.47852/bonviewJDSIS3202655.

[27] S. Zhang, "Application of random forest algorithm in accounting data analysis and prediction," in *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Dec. 2024, pp. 1–5, doi: 10.1109/ICMNWC63764.2024.10871989.

[28] H. Patel, S. Parikh, A. Patel, and A. Parikh, "An application of ensemble random forest classifier for detecting financial statement manipulation of Indian listed companies," in *Advances in Intelligent Systems and Computing*, 2019, pp. 349–360, doi: 10.1007/978-981-13-1280-9_33.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[30] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth*

InternationalConference, 1996, pp. 148–156.
[31] L. T. Afolabi, F. Saeed, H. Hashim, and O. O. Petinrin, "Ensemble learning method for the prediction of new bioactive molecules," *PLOS ONE*, vol. 13, no. 1, Jan. 2018, doi: 10.1371/journal.pone.0189538.
[32] Y. Ding, H. Zhu, R. Chen, and R. Li, "An efficient adaboost algorithm with the multiple thresholds classification," *Applied Sciences*, vol. 12, no. 12, Jun. 2022, doi: 10.3390/app12125872.
[33] V. Chang, T. Li, and Z. Zeng, "Towards an improved Adaboost algorithmic method for computational financial analysis," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 219–232, Dec. 2019, doi: 10.1016/j.jpdc.2019.07.014.
[34] J.-K. Tsai and C.-H. Hung, "Improving adaboost classifier to predict enterprise performance after Covid-19," *Mathematics*, vol. 9, no. 18, Sep. 2021, doi: 10.3390/math9182215.
[35] T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Comparing boosting and bagging techniques with noisy and imbalanced data," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 3, pp. 552–568, May 2011, doi: 10.1109/TSMCA.2010.2084081.
[36] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, "Optimizing bankruptcy prediction through filter- based feature selection," in *2024 International Conference on Informatics Engineering, Science & Technology (INCITEST)*, Oct. 2024, pp. 1–8, doi: 10.1109/INCITEST64888.2024.11121476.
[37] T. Ansah-Narh, E. N. N. Nortey, E. Proven-Adzri, and R. Opoku-Sarkodie, "Enhancing corporate bankruptcy prediction via a hybrid genetic algorithm and domain adaptation learning architecture," *Expert Systems with Applications*, vol. 258, Dec. 2024, doi: 10.1016/j.eswa.2024.125133.
[38] K. Samara and A. Shinde, "Bankruptcy prediction using machine learning and data preprocessing techniques," *Analytics*, vol. 4, no. 3, Sep. 2025, doi: 10.3390/analytics4030022.

BIOGRAPHIES OF AUTHORS

Nur Farahaina Idris received the Bachelor of Computer Science and Master's degree in Computer Science from Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). She is currently pursuing Ph.D. while also teaching at UMPSA. She can be contacted at email: farahaina@umpsa.edu.my.

Mohd Arfian Ismail is an Associate Professor at the Faculty of Computing in Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Malaysia. He received B.Sc., M.Sc., and Ph.D. degree in Computer Science from Universiti Teknologi Malaysia (UTM) in 2008, 2011, and 2016, respectively. His current research interests are in the areas of machine learning methods and optimization method. He can be contacted at email: arfian@umpsa.edu.my.