Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 2, August 2026, pp. 640~650
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i2.pp640-650

Advanced personal bankruptcy prediction using tree-based deep learning models

Nhat Nguyen Minh, Duy Ngo Hoang Khanh
Faculty of Banking, Ho Chi Minh University of Banking (HUB), Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Sep 14, 2024
Revised Jun 29, 2026
Accepted Jul 23, 2026

Keywords:

Deep forest
Neural oblivious decision ensembles
Personal bankruptcy prediction
Tabular convolutional neural networks
Tree-based deep learning

ABSTRACT

Due to the unstable economic conditions, worsened by the post-COVID-19 environment and persistent foreign wars in 2024, financial institutions have growing difficulties in accurately predicting customer default probability. This study examines the use of sophisticated tree-based deep learning and deep neural network models for forecasting personal bankruptcy. This research utilises a dataset of roughly 9,800 individuals from Vietnamese financial institutions, spanning from 2012 to 2022, to evaluate the efficacy of models including neural decision tree, deep forest, tabular convolutional neural networks (TBCNN), and neural oblivious decision ensembles (NODE). The results demonstrate that the Deep Forest model far surpasses its competitors, providing nearly flawless predicted accuracy and enhanced interpretability. The findings highlight the efficacy of tree-based deep learning and deep neural network models as effective instruments for financial risk management, especially in volatile and unpredictable economic environments.

Corresponding Author:

Nhat Nguyen Minh
Faculty of Banking, Ho Chi Minh University of Banking (HUB)
36 Ton That Dam street, Nguyen Thai Binh ward, District 1
Ho Chi Minh City, Vietnam
Email: nhatnm@hub.edu.vn

1. INTRODUCTION

Precisely forecasting personal bankruptcy is crucial for financial organisations, since it profoundly impacts risk management techniques and lending policies [1]. Conventional models like logistic regression and simple decision trees, however common, frequently inadequately address the complexity and non-linear correlations present in contemporary financial data [2]. The growing complexity of consumer financial behaviour and the accessibility of high-dimensional datasets underscore the need for increasingly sophisticated prediction models [3]. In particular, the prediction of personal financial distress has become increasingly important in lending environments where borrower characteristics, repayment capacity, and credit behaviour interact in complex and dynamic ways. In such settings, models that can capture non-linear structures more effectively are expected to improve the quality of risk assessment and credit-related decision-making [4].

Prior studies have examined bankruptcy and financial distress prediction using a range of statistical, machine learning, and deep learning techniques. Traditional approaches have shown the practical value of structured prediction models in identifying high-risk borrowers, but they often face limitations when relationships among predictors are highly non-linear or when tabular financial data become more complex [2], [4]. Recent advancements have therefore shifted attention toward more sophisticated machine learning models capable of learning richer patterns from structured datasets. Among these, Deep Forest has been highlighted as an effective ensemble framework that integrates multiple decision trees and can achieve high

predictive accuracy in complex forecasting tasks [5]. This approach is particularly relevant in financial applications because it can model hierarchical relationships in the data while preserving a level of interpretability that is important for risk management practice [6], [7].

Alongside Deep Forest, tree-based deep learning models such as the Neural Decision Tree have attracted attention for combining the interpretability of conventional decision trees with the learning capacity of deep models [8]. By optimising decision boundaries across layers, the Neural Decision Tree can improve accuracy and generalization [9], while retaining a more transparent decision process that is easier to communicate to managers, stakeholders, and regulators [10]. At the same time, deep learning models designed for tabular data, such as TBCNN and NODE, have also been proposed as promising alternatives for structured financial datasets [11]. These models are effective in capturing complex non-linear dependencies in borrower-level data, but their practical use in financial decision-making remains constrained by concerns over opacity and limited interpretability [12].

Although prior studies have shown the value of both tree-based deep models and tabular deep neural models, several issues remain unresolved. Direct comparisons of these advanced models within the same personal bankruptcy prediction framework are still limited, especially in emerging markets such as Vietnam, where borrower behaviour, credit conditions, and data structures may differ from those in other settings. In addition, the literature has not yet established whether the greater complexity of neural models leads to a practically meaningful advantage over tree-based models when predictive performance and interpretability are considered together [12]. These gaps motivate the present study.

Based on the above discussions, this study aims to investigate the core question: Can tree-based deep learning models achieve prediction performance comparable to or better than deep neural networks in multi-class personal bankruptcy prediction while retaining stronger interpretability? Both techniques have proven highly useful in different settings; nonetheless, the optimal choice likely hinges on the application's unique requirements whether the emphasis is on maximising accuracy or assuring interpretability. Accordingly, this study compares four advanced models including Neural Decision Tree, Deep Forest, TBCNN and NODE to identify the model that provides the most suitable balance between predictive performance and interpretability in the context of personal bankruptcy prediction [13]. More specifically, the paper contributes to the literature in three ways. First, it provides a comparative assessment of tree-based deep learning and deep neural network models for multi-class personal bankruptcy prediction. Second, it offers empirical evidence based on a dataset supported by the research Institute of Ho Chi Minh University of Banking, consisting of 9,800 personal loan customers at commercial banks and credit institutions in Vietnam during the period from 2012 to 2022. Third, it evaluates model suitability not only in terms of accuracy and F1-score, but also from the perspective of interpretability and practical applicability in financial risk management [13]. The study was conducted on a dataset supported by the research Institute of Ho Chi Minh University of Banking. In the following sections of the study, the author will present the method, empirical results, discussion of the findings, and conclusion.

2. METHOD

Figure 1 illustrates the experimental workflow adopted in this study. The procedure begins with data preprocessing, followed by feature preparation, dataset splitting, model development, hyperparameter optimisation, and out-of-sample evaluation. This sequential design was adopted to ensure that the comparison among TBCNN, Neural Decision Tree, NODE, and Deep Forest was conducted under the same data conditions and evaluation protocol, thereby allowing the research question stated in the Introduction to be examined in a transparent and reproducible manner [14]-[16].

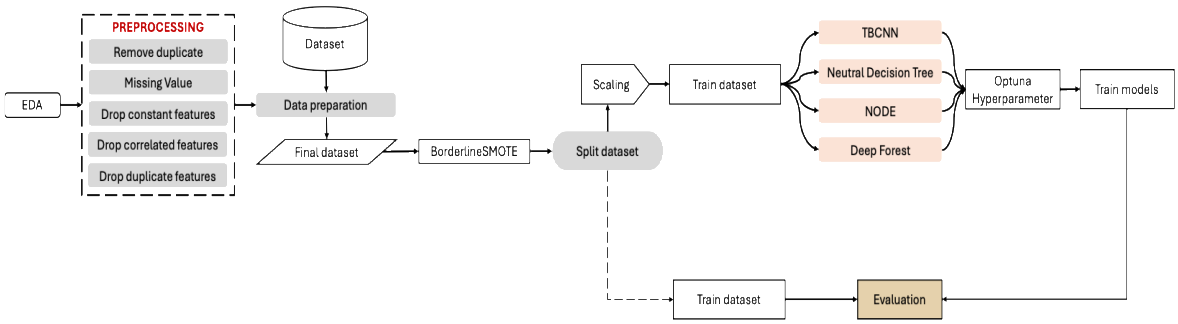

Figure 1. The process of the research method

2.1. Preprocessing

Preprocessing is the initial and crucial phase in machine learning, guaranteeing that raw data is effectively converted into a format suitable for model training. During this step, eliminating duplicates is crucial to avoid the over-representation of any one data point, which may distort the findings. Addressing absent data is an essential aspect of preprocessing. A prevalent method is single imputation, when absent values are substituted using statistical metrics such as the mean, median, or mode, contingent upon the data's characteristics [17]. This approach preserves the dataset's integrity while reducing the influence of absent data on model precision.

In this study, preprocessing was conducted in four consecutive steps. First, duplicate observations were removed to avoid over-representing repeated customer records. Second, missing values were handled using single imputation. Numerical variables were imputed using the median when the distribution was skewed and the mean when the distribution was approximately symmetric, whereas categorical variables were imputed using the mode [18]. Third, constant features with zero variance were removed because they do not contribute useful discriminatory information. Fourth, highly correlated predictors were screened using a pairwise Pearson correlation threshold of 0.90, and one variable from each highly correlated pair was removed to reduce redundancy and improve model stability [19].

Furthermore, the elimination of constant and strongly correlated features is executed to diminish noise in the dataset and prevent complications like multicollinearity, which can skew model predictions [20]. Multicollinearity arises when independent variables exhibit significant correlation, hence compromising the precision of feature significance and inflating model coefficients. This phase guarantees that the models concentrate on significant and non-redundant data items.

Multicollinearity can be quantitatively represented by the variance inflation factor (VIF):

$$VIF_j = \frac{1}{1-R_j^2}$$

where R_j^2 is the coefficient of determination of the regression model that predicts the j^{th} feature using all the other features. A higher VIF indicates a high correlation, suggesting the feature should be removed [21].

After the initial correlation screening, multicollinearity was further assessed using the Variance Inflation Factor. Variables with a VIF greater than 10 were excluded from the final feature set, as this threshold is commonly used to indicate serious multicollinearity. This two-stage procedure helped ensure that the retained variables were both informative and sufficiently independent for robust model training.

2.2. Data preparation

The data preparation phase entails scaling and normalising features to guarantee that all variables contribute uniformly to the model [20]. In the absence of scaling, characteristics with broader ranges (e.g., income) may overshadow those with narrower ranges (e.g., age), resulting in skewed model performance. Min-Max scaling is a prevalent method in which each feature is normalised to a range of [0, 1] or [-1, 1]. This is denoted by the equation:

$$x' = \frac{x-x_{min}}{x_{max}-x_{min}}$$

where x represents the original value, and x_{min} and x_{max} denote the minimum and maximum values of the feature, respectively [22].

Categorical data must be converted using one-hot encoding or label encoding, based on the characteristics of the data. One-hot encoding transforms categorical variables into binary vectors, with each category represented as a distinct column with values of 0 or 1, hence facilitating the effective utilisation of categorical data by machine learning models.

In this study, numerical features were scaled to the range [0,1] using Min-Max normalisation. Nominal categorical variables were transformed using one-hot encoding, while ordinal variables, when present, were encoded using label encoding. To avoid data leakage, the scaling and encoding transformers were fitted only on the training data and then applied to the validation and test data using the fitted parameters.

Mitigating class imbalance is essential at this phase, particularly in personal bankruptcy prediction, because the majority of persons are unlikely to fail, rendering the minority class (those who do default) more challenging to identify. Borderline synthetic minority over-sampling technique (SMOTE) generates synthetic data points for the minority class, concentrating on samples situated around the decision border as shown in Figure 2. The model is more prone to misclassifying these borderline situations; hence, oversampling them

enhances the model's capacity to differentiate between classes [23]. The safe-level variation of Borderline SMOTE enhances this method by creating synthetic points in areas with a greater probability of accurate classification, hence reducing the danger of noise introduction. This approach enhances dataset balance and improves the model's sensitivity to predictions of the minority class [24].

Because the target variable in this study consists of four progressive classes of financial distress, class imbalance was addressed using Borderline-SMOTE applied only to the training data. The oversampling procedure was configured with 5 nearest neighbours for synthetic instance generation and 10 nearest neighbours for identifying borderline samples. Each minority class in the training fold was oversampled until it matched the size of the majority class in that fold. A fixed random seed of 42 was used to ensure reproducibility. This sequencing was important because resampling the full dataset before splitting could introduce information leakage and overstate model performance.

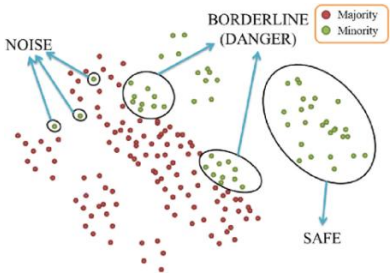

Figure 2. Specific groups of imbalanced data in the Borderline-SMOTE

2.3. Splitting the dataset

During this step, the dataset is divided into training and testing subsets, commonly according to an 80/20 ratio. The training set is utilised to calibrate the model, whilst the testing set assesses the model's capacity to generalise to novel data. This method guarantees that the model is not excessively tailored to the training data and can excel with fresh, real-world data.

In this study, the cleaned dataset was divided into training and testing subsets using a stratified 80/20 split, with the same class proportions preserved across both subsets. The random seed was set to 42 to ensure that the partition could be reproduced exactly [24]. The training subset was used for model fitting and hyperparameter tuning, whereas the test subset was reserved strictly for final out-of-sample evaluation.

Cross-validation is implemented to augment model resilience. In k-fold cross-validation, the dataset is partitioned into k equally sized segments. The model is trained using k-1 folds and verified on the remaining fold. This procedure is executed k times, and the outcomes are averaged to get a more dependable assessment of the model's efficacy. This strategy is crucial for alleviating overfitting and guarantees consistent model performance across various data subsets.

Within the training subset, stratified 5-fold cross-validation was used during model selection and hyperparameter optimisation. In each iteration, four folds were used for training and one fold was used for validation, and the average performance across the five folds was used to assess a candidate hyperparameter configuration. Stratification was maintained throughout the cross-validation process in order to preserve the class distribution of the multi-class bankruptcy variable.

2.4. Model training and evaluation

Four advanced models were selected for comparison: TBCNN, Neural Decision Tree, NODE, and Deep Forest. These models were chosen because they represent two relevant methodological families discussed in the Introduction: deep neural models designed for tabular learning and tree-based deep architectures that may provide a better balance between predictive performance and interpretability. All models were trained on the same preprocessed data and evaluated under the same protocol so that their relative performance could be compared fairly.

2.4.1. Tabular convolutional neural network (TBCNN)

This model employs convolutional layers on tabular data, utilising convolution processes normally applied to picture data. The convolution function in TBCNN is denoted as:

$$h^{ij} = f\left(\sum_k W_k X_{i+k,j} + b_k\right)$$

where h^{ij} is the ouput feature map, W_k represents the convolutional weights, and f is the activation function [25].

In the present study, TBCNN was implemented as a neural benchmark for modelling non-linear relationships in structured borrower-level data. The TBCNN hyperparameters tuned by Optuna included the learning rate (10^{-4} to 10^{-2}), batch size {32, 64,128}, number of hidden units {64,128, 256}, dropout rate (0.10 to 0.50), and number of epochs up to 50. The Adam optimiser and ReLU activation function were used, and early stopping with a patience of 10 epochs was applied to reduce overfitting.

2.4.2. Neural decision tree

This approach integrates the architecture of decision trees with the adaptability of neural networks. It substitutes rigid divides with flexible, probabilistic divisions. The soft decision function is represented as:

$$P(x) = \frac{1}{1+e^{-\theta^{T}x}}$$

where $P(x)$ is the probability of selecting a branch, and $-\theta^{T_x}$ is the dot product of the weight vector and the input features.

Neural decision tree was included because it combines hierarchical decision logic with differentiable learning. The tuned hyperparameters for this model included tree depth {3, 4, 5, 6, 7}, hidden dimension {32, 64,128}, learning rate (10^{-4} to 10^{-2}), batch size {64,128, 256}, and dropout rate (0.10 to 0.50). The model was trained for a maximum of 50 epochs, with early stopping after 10 epochs without validation improvement.

2.4.3. Neural oblivious decision ensembles (NODE)

This model employs an ensemble of decision trees in which all nodes at the same level utilise the identical characteristic for partitioning. The ultimate forecast is calculated as follows:

$$y = \sum_{t=1}^{T} \alpha_t f_t(x)$$

where y is the final output, α_t is the weight for each tree, and $f_t(x)$ is the prediction of the t-th tree [26].

NODE was selected because it is specifically designed for tabular data and has been reported as an effective deep learning approach for structured prediction tasks. The tuned hyperparameters included the number of ensemble layers {2, 3, 4}, number of trees {128, 256, 512}, tree depth {4, 6, 8}, learning rate (10^{-4} to 10^{-2}), batch size {32, 64, 128}, and dropout rate (0.10 to 0.40). As with the other neural models, training was limited to 50 epochs with early stopping patience set to 10 epochs.

2.4.4. Deep forest (GCForest)

This model constructs an ensemble of decision trees and implements them in a cascading architecture. Every layer enhances the predictions of the preceding layer, augmenting accuracy with each iteration. The result of the cascade layer is:

$$\hat{y}^{(l)} = \frac{1}{M_l} \sum_{m=1}^{M_l} h_m^{(l)}(x)$$

where $\hat{y}^{(l)}$ is the output from layer ll, and $h_m^{(l)}(x)$ is the prediction from tree m in layer l [26].

Deep Forest was included as the principal tree-based ensemble benchmark because it can model complex non-linear structures through layered tree ensembles while retaining a comparatively interpretable decision process. The tuned hyperparameters for Deep Forest included the number of trees per forest {100, 200, 300}, maximum tree depth {5, 10, 15}, minimum samples per leaf {1, 2, 5}, and the number of cascade layers {2, 3, 4}. The final prediction was obtained from the best-performing cascade configuration selected on the training folds.

Optuna is utilised to optimise the hyperparameters of these models, employing the tree-structured parzen estimator (TPE) to effectively navigate the hyperparameter space. The objective function for hyperparameter optimisation may be represented as:

$$Minimize\ f(x) = 1 - Accuracy(x)$$

where x denotes the hyperparameter configuration. Optuna's dynamic sampling and pruning features enable the framework to rapidly converge on optimal hyperparameters, enhancing model performance without necessitating laborious searches.

In this study, Optuna was run for 100 trials per model using the TPE sampler with a fixed random seed of 42. The optimisation objective was the mean cross-validated accuracy across the five training folds, which is equivalent to minimising $1 - \text{Accuracy}(x)$. A median pruning strategy was used to terminate unpromising trials early and reduce computational cost. This optimisation protocol was adopted to ensure that each model was evaluated under a comparable and systematic tuning procedure.

2.5. Performance evaluation and model comparison

After hyperparameter optimisation, the best version of each model was retrained on the full training set and evaluated on the hold-out test set. Because the study aims to compare the predictive ability of four alternative models, the final analysis focused on model-by-model out-of-sample performance rather than a weighted ensemble. Performance was assessed using Accuracy, Precision, Sensitivity, Specificity, F1-score, and AUC; for the multi-class setting, these metrics were computed using macro-averaging, while AUC was estimated through a one-vs-rest approach across the four bankruptcy classes. Confusion matrices and class-wise ROC curves were also examined to provide a more detailed evaluation of how effectively each model distinguished among the four levels of financial distress. This evaluation design ensures a consistent comparison between tree-based deep models and deep neural models under the same preprocessing, resampling, and tuning conditions.

3. RESULTS AND DISCUSSION

3.1. Data research

The dataset for this study on forecasting personal defaults comprises information from 9,800 clients of several commercial banks and credit institutions in Vietnam, covering the period from 2012 to 2022. All client data was encrypted during collection in compliance with rigorous data protection regulations. The target variable, Bankruptcy, is classified into four progressive tiers of financial distress:

- Class 0: Denotes debts that are either in good standing or overdue by fewer than 10 days, however deemed recoverable.
- Class 1: Debts that are overdue for a period ranging from 10 to 180 days or have been restructured for the initial instance, signifying preliminary indicators of financial distress. This group encompasses borrowers who may encounter challenges yet are still expected to recuperate, necessitating increased oversight by lenders.
- Class 2: Pertains to loans that are overdue for a duration of 181 to 360 days or those that have undergone restructuring for the second time. This Class indicates significant financial difficulty, characterised by uncertain recovery and borrowers' difficulties in fulfilling their financial commitments. A more stringent risk management strategy is required at this juncture.
- Class 3: Denotes debts that are overdue for over 360 days, categorised as probably irrecoverable. Borrowers in this category have considerable financial instability, rendering recovery exceedingly unlikely, therefore requiring substantial assistance or write-offs by lenders.

This revised categorisation provides a more detailed representation of borrowers' financial difficulties, which is important for developing accurate prediction models and stronger risk assessment frameworks. Alongside the target variable, the dataset includes explanatory factors such as loan status, outstanding amounts, loan duration, income, job history, homeownership, loan purpose, monthly debt commitments, and credit history. Using these financial characteristics, the models aim to predict default risk more accurately and support proactive risk mitigation by financial institutions. Importantly, the four-class structure captures not only whether a borrower is distressed, but also the severity of that distress, which is particularly useful for decisions related to monitoring, restructuring, and recovery strategies.

3.2. Comparison results on the predictive ability of the models

The out-of-sample test results of the tree-based deep-learning and deep neural networks models including TBCNN, Neural Decision Tree, NODE, and Deep Forest are presented in detail in Table 1. The performance measures in the table demonstrate notable discrepancies across the models in forecasting personal bankruptcy. TBCNN demonstrates a relatively low accuracy of 0.6575, accompanied by subpar precision and recall scores, reflecting its inadequate capacity to accurately categorize both positive and negative instances. The Neural Decision Tree model exhibits subpar performance, evidenced by an accuracy of 0.6233, a precision of 0.8116, and an F1 score of 0.6429, indicating an imbalance in its capacity to detect genuine positives while minimizing false negatives. Conversely, NODE exhibits superior performance compared to the Neural Decision Tree and TBCNN, attaining an accuracy of 0.7967 and an AUC of 0.7302, indicating a moderate capacity for class differentiation. Nonetheless, its F1 score of 0.8531 signifies a reasonable balance between accuracy and recall, although there remains room for improvement in prediction

tasks. The preeminent model is definitely Deep Forest, which surpasses the other models with an accuracy of 0.9929 and nearly flawless precision and recall. The F1 score of 0.9953 and AUC of 0.9905 indicate an effective balance between false positives and false negatives, while preserving excellent discriminatory capability. This outcome highlights the superiority of tree-based ensemble models such as Deep Forest, especially in managing the intricate, non-linear interactions inherent in financial information.

Table 1. Personal bankruptcy prediction results of models on out-of-sample datasets

Model	Accuracy	Precision	Sensitivity	Specificity	F1 score	AUC
TBCNN	0.6575	0.8239	0.7696	0.3088	0.6989	0.5392
Neural decision tree	0.6233	0.8116	0.7503	0.2509	0.6429	0.5006
NODE	0.7967	0.8903	0.8656	0.5948	0.8531	0.7302
Deep forest	0.9929	0.9953	0.9953	0.9858	0.9953	0.9905

These findings directly address the main research objective of the study, which was to determine whether tree-based deep learning models could achieve predictive performance comparable to, or better than, deep neural alternatives in personal bankruptcy prediction. The results show that this is indeed the case, with Deep Forest consistently outperforming TBCNN, Neural Decision Tree, and NODE across all evaluation criteria. This suggests that, for structured borrower-level financial data, a layered tree ensemble is more effective in capturing complex interactions and non-linear decision boundaries than the tested neural architectures. From a practical perspective, the results reported in Table 1 are also important because effective credit-risk prediction requires not only high sensitivity, so that distressed borrowers can be identified early, but also sufficient specificity to avoid excessive false alarms and inefficient intervention. In this respect, TBCNN and Neural Decision Tree perform weakly due to their low specificity, while NODE offers a clear improvement but still remains inferior to Deep Forest. Therefore, the main implication of Table 1 is not simply that Deep Forest achieves the highest accuracy, but that it provides the most balanced and operationally credible performance for personal bankruptcy prediction.

An analysis of the confusion matrices for Deep Forest, NODE, Neural Decision Tree, and TBCNN in Figure 3 indicates notable differences in model performance, with Deep Forest identified as the most precise and equitable model. It accurately identifies 843 non-bankrupt persons and 800 bankrupt individuals, with negligible misclassifications, merely 20 erroneous positives and 26 false negatives. This degree of prediction accuracy demonstrates that Deep Forest is exceptionally proficient at detecting potential defaulters while minimising the misclassification of non-bankrupt persons. The model's capacity to reduce both mistake types underscores its resilience, rendering it especially appropriate for financial applications where precision in risk differentiation is essential. Conversely, NODE has considerable difficulties, particularly in categorising non-bankrupt persons. NODE's shortcomings in distinguishing between default and non-default instances are clear, with 63 misclassified non-bankrupt persons and only 5 accurately recognised. NODE's strength resides in its ability to reliably identify insolvent persons, evidenced by its 873 right classifications. The elevated incidence of misclassifications in non-bankrupt instances diminishes its overall trustworthiness, especially in practical situations when false positives may result in detrimental loan choices. Neural Decision Tree demonstrate a significant bias, classifying the majority of instances as bankrupt and correctly recognising only three non-bankrupt persons. This biassed classification significantly prioritises the identification of bankrupt cases, as demonstrated by 860 accurate classifications. This pronounced bias towards a single category undermines the model's capacity to deliver balanced predictions, rendering it less appropriate for practical applications where differentiating between the two categories is equally critical. Likewise, TBCNN encounters difficulties in misclassifying non-bankrupt persons, erroneously categorising 43 non-bankrupt instances as bankrupt. Despite attaining a modest success rate in detecting bankrupt persons, evidenced by 828 accurate classifications, the model's failure to distinctly differentiate between the two groups constrains its utility in contexts necessitating precise risk evaluation. Deep Forest demonstrates the highest reliability, effectively distinguishing between bankrupt and non-bankrupt people while sustaining a minimal misclassification rate. The alternative models-NODE, Neural Decision Tree, and TBCNN-exhibit considerable deficiencies, especially in differentiating non-bankrupt persons, hence reducing their use in financial contexts. This comparison highlights Deep Forest's advantage in delivering a more thorough and precise risk assessment, essential for predicting personal bankruptcy.

Figure 3 shows more than a simple ranking of models; it also reveals the types of errors each model tends to make. This distinction is important because false positives and false negatives have different managerial consequences. Excessive false positives may lead institutions to classify relatively healthy borrowers as risky, resulting in unnecessary restrictions or monitoring costs, whereas excessive false negatives may delay intervention for borrowers who are genuinely deteriorating. Deep Forest is particularly

attractive because it reduces both types of error simultaneously, making it not only statistically superior but also more useful in practical lending and collection settings.

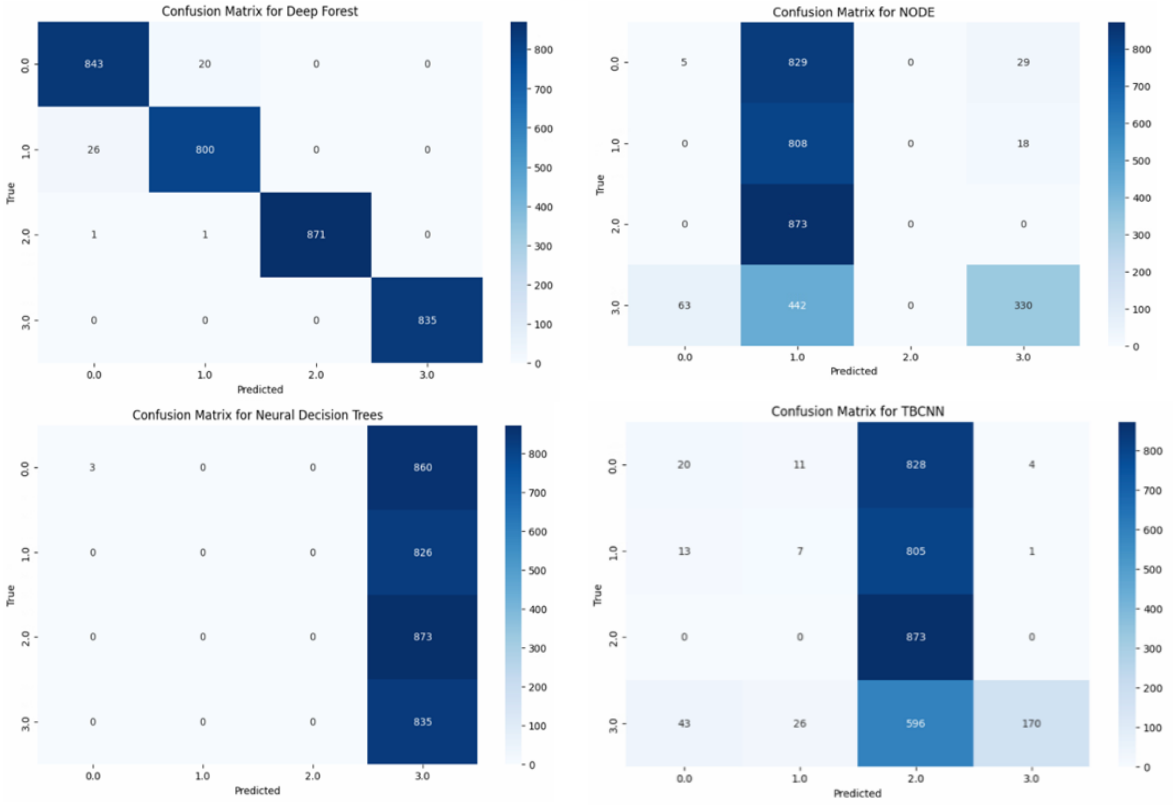

Figure 3. Prediction results of models on the confusion matrix

Figure 4 depicts a distinct performance differential across TBCNN, Neural Decision Tree, NODE, and Deep Forest across several classes, as seen by the ROC curves. Deep Forest distinguishes itself with exceptional categorisation, with an AUC of 0.9980 or above across all categories, including a flawless AUC of 1.000 in Class 3. This performance illustrates its strong capacity to precisely differentiate between default and non-default scenarios. Conversely, TBCNN continuously exhibits worse performance, with AUC values spanning from 0.3081 to 0.5146, signifying limited discriminatory capability, especially in Class 0, where it fails to adequately differentiate between true and false positives. Neural Decision Tree provide marginal enhancement compared to TBCNN, although they continue to display subpar performance with AUCs ranging from 0.4000 to 0.4757, reflecting a constrained capacity to elucidate the underlying data structure, particularly in Class 3. NODE has decent performance, achieving AUC values of 0.7726 in Class 3, however it is continuously surpassed by Deep Forest across all classes. The significant disparity in AUC values between Deep Forest and the other models underscores its exceptional capacity to generalise across all classes, establishing it as the most dependable model for personal bankruptcy prediction. The markedly reduced AUCs of TBCNN and Neural Decision Tree indicate their inadequacy in managing the dataset's complexity, highlighting the need for more sophisticated models such as Deep Forest to attain elevated prediction accuracy.

The ROC evidence is particularly informative because the target variable is defined in four levels of distress rather than as a simple binary outcome. In this context, a strong average metric is not enough; the model must also preserve discriminative ability across all bankruptcy classes. Figure 4 shows that Deep Forest remains consistently strong for every class, including the most severe class, which is particularly important for institutions that need reliable identification of high-risk borrowers. This result strengthens the interpretation from Table 1 and Figure 3: the advantage of Deep Forest is not confined to one category or one metric, but is visible throughout the entire classification structure.

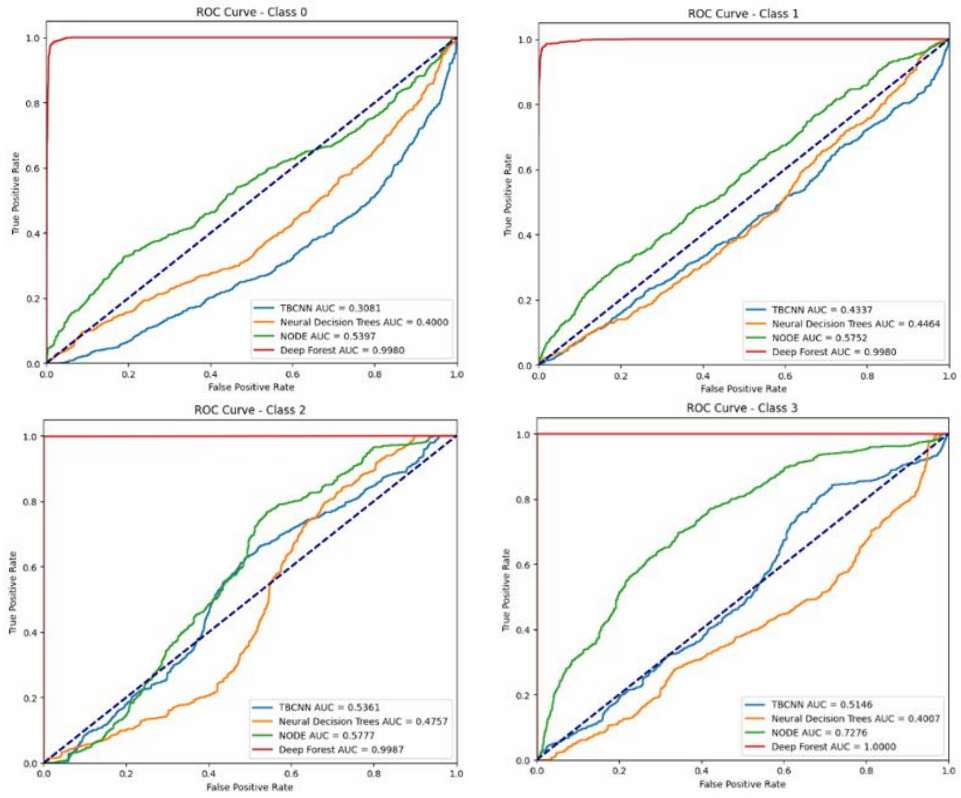

Figure 4. Prediction results of models on ROC Curves

3.3. Discussion and implications

The findings indicate that Deep Forest is the most effective model for multi-class personal bankruptcy prediction among the four evaluated approaches. This suggests that, for borrower-level financial data, a cascade ensemble of trees is better able to capture non-linear relationships, threshold effects, and variable interactions than the tested neural architectures. The result is important not only because Deep Forest achieves the best predictive performance, but also because its structure appears to be more compatible with the characteristics of financial data and the practical requirements of risk assessment. Although the initial expectation was that tree-based deep models would be competitive with deep neural alternatives, the evidence reveals a stronger pattern, with Deep Forest clearly outperforming the other models and NODE showing only intermediate performance. Nevertheless, the magnitude of this gap should be interpreted with caution, as it may partly reflect features of the present dataset that favour hierarchical tree-based partitioning.

This study contributes empirical evidence from a relatively large borrower-level dataset collected from commercial banks and credit institutions in Vietnam over the period 2012 to 2022 and evaluates multiple advanced models under a common experimental design. However, several limitations remain. The evidence is drawn from a single national context, the analysis focuses on predictive comparison rather than external validation, and further work is needed to examine the stability and drivers of the predictions over time. Overall, the results suggest that Deep Forest is the most suitable model in this setting because it combines strong discrimination, low misclassification, and practical relevance for financial risk assessment. Future research should validate these findings on external datasets, compare Deep Forest with other advanced tabular-learning models, and incorporate explainability techniques to identify the key determinants of borrower distress.

4. CONCLUSION

The Deep Forest model has greater efficacy in multi-class personal bankruptcy prediction, surpassing existing tree-based deep-learning and deep neural networks models including TBCNN, Neural Decision Tree, and NODE. Deep Forest has proficiency in capturing intricate, non-linear patterns across many classes, as seen by its nearly perfect AUC ratings, especially its impeccable performance in Class 3. The ensemble-based design of this model, which enhances predictions incrementally, enables superior

generalisation to varied and complex financial facts. Furthermore, the use of Optuna for hyperparameter optimisation significantly improved the model's performance by enabling the precise adjustment of critical parameters to get optimal predicted accuracy. Optuna's efficacy in navigating the hyperparameter space guaranteed that Deep Forest was optimised for superior multi-class differentiation compared to other models. The integration of Deep Forest's ensemble efficacy and Optuna's optimisation prowess underscores the model's versatility and precision, establishing it as the most dependable instrument for intricate, multi-class predictions in financial risk evaluations.

ACKNOWLEDGEMENTS

The authors thank all lecturers and managers of Ho Chi Minh University of Banking (HUB). This article is supported and funded by HUB, Vietnam.

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
Nhat Nguyen Minh		✓			✓	✓		✓	✓	✓	✓	✓		
Duy Ngo Hoang Khanh	✓		✓	✓		✓			✓		✓		✓	✓

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 STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author Nhat M. Nguyen on request.

REFERENCES

[1] J.-P. Lai, Y.-L. Lin, H.-C. Lin, C.-Y. Shih, Y.-P. Wang, and P.-F. Pai, "Tree-based machine learning models with optuna in predicting impedance values for circuit analysis," *Micromachines*, vol. 14, no. 2, p. 265, Jan. 2023, doi: 10.3390/mi14020265.
[2] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405–417, Oct. 2017, doi: 10.1016/j.eswa.2017.04.006.
[3] S. Smiti and M. Soui, "Bankruptcy prediction using deep learning approach based on borderline SMOTE," *Information Systems Frontiers*, vol. 22, no. 5, pp. 1067–1083, Aug. 2020, doi: 10.1007/s10796-020-10031-6.
[4] S. H. Syed Nor, S. Ismail, and B. W. Yap, "Personal bankruptcy prediction using decision tree model," *Journal of Economics, Finance and Administrative Science*, vol. 24, no. 47, pp. 157–170, Mar. 2019, doi: 10.1108/jefas-08-2018-0076.
[5] Q. Fu, K. Li, J. Chen, J. Wang, Y. Lu, and Y. Wang, "Building energy consumption prediction using a deep-forest-based DQN method," *Buildings*, vol. 12, no. 2, p. 131, Jan. 2022, doi: 10.3390/buildings12020131.
[6] T. Zhou, X. Sun, X. Xia, B. Li, and X. Chen, "Improving defect prediction with deep forest," *Information and Software Technology*, vol. 114, pp. 204–216, Oct. 2019, doi: 10.1016/j.infsof.2019.07.003.
[7] Y. Zhong and H. Wang, "Internet financial credit scoring models based on deep forest and resampling methods," *IEEE Access*, vol. 11, pp. 8689–8700, 2023, doi: 10.1109/access.2023.3239889.
[8] S. Popov, S. Morozov, and A. Babenko, "Neural oblivious decision ensembles for deep learning on tabular data," *arXiv preprint arXiv:1909.06312*, 2019.
[9] K. D. Humbird, J. L. Peterson, and R. G. Mcclarren, "Deep neural network initialization with decision trees," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1286–1295, May 2019, doi: 10.1109/tnnls.2018.2869694.
[10] F. M. Talaat, A. Aljadani, M. Badawy, and M. Elhosseini, "Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction," *Neural Computing and Applications*, vol. 36, no. 9, pp. 4847–4865, Dec. 2023, doi: 10.1007/s00521-023-09232-2.
[11] Q. Zhang, L. Sun, G. Yang, B. Lu, X. Ning, and W. Li, "TBNN: totally-binary neural network for image classification," *Optoelectronics Letters*, vol. 19, no. 2, pp. 117–122, Feb. 2023, doi: 10.1007/s11801-023-2113-2.
[12] M. Stevenson, C. Mues, and C. Bravo, "The value of text for small business default prediction: A Deep Learning approach," *European Journal of Operational Research*, vol. 295, no. 2, pp. 758–771, Dec. 2021, doi: 10.1016/j.ejor.2021.03.008.

[13] G. Li, H.-D. Ma, R.-Y. Liu, M.-D. Shen, and K.-X. Zhang, "A two-stage hybrid default discriminant model based on deep forest," *Entropy*, vol. 23, no. 5, p. 582, May 2021, doi: 10.3390/e23050582.
[14] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.
[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
[16] V. Pugliese, C. Hirata, and R. Costa, "Comparing supervised classification methods for financial domain problems," in *Proceedings of the 22nd International Conference on Enterprise Information Systems*, 2020, pp. 440–451, doi: 10.5220/0009426204400451.
[17] S. García, J. Luengo, and F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," *Knowledge-Based Systems*, vol. 98, pp. 1–29, Apr. 2016, doi: 10.1016/j.knosys.2015.12.006.
[18] Z. Zhang, "Missing data imputation: focusing on single imputation," *Annals of Translational Medicine*, vol. 4, no. 1, p. 9, 2016.
[19] S. García, J. Luengo, and F. Herrera, "Data preparation basic models," in *Data Preprocessing in Data Mining*, Springer International Publishing, 2014, pp. 39–57.
[20] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5–6, pp. 375–381, May 2003, doi: 10.1080/713827180.
[21] J. Cheng, J. Sun, K. Yao, M. Xu, and Y. Cao, "A variable selection method based on mutual information and variance inflation factor," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 268, p. 120652, Mar. 2022, doi: 10.1016/j.saa.2021.120652.
[22] B. Deepa and K. Ramesh, "Epileptic seizure detection using deep learning through min max scaler normalization," *International journal of health sciences*, pp. 10981–10996, May 2022, doi: 10.53730/ijhs.v6ns1.7801.
[23] Y. Sun *et al.*, "Borderline SMOTE algorithm and feature selection-based network anomalies detection strategy," *Energies*, vol. 15, no. 13, p. 4751, 2022, doi: 10.3390/en15134751.
[24] C. Chen, W. Shen, C. Yang, W. Fan, X. Liu, and Y. Li, "A new safe-level enabled borderline-SMOTE for condition recognition of imbalanced dataset," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023, doi: 10.1109/tim.2023.3289545.
[25] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, 2004, doi: 10.1145/1007730.1007735.
[26] D. Li, Z. Liu, D. J. Armaghani, P. Xiao, and J. Zhou, "Novel ensemble tree solution for rockburst prediction using deep forest," *Mathematics*, vol. 10, no. 5, p. 787, Mar. 2022, doi: 10.3390/math10050787.

BIOGRAPHIES OF AUTHORS

Nhat Nguyen Minh is a lecturer in the Department of Banking and Finance at Ho Chi Minh University of Banking, Vietnam. He earned his doctorate in Banking and Finance from Ho Chi Minh University of Banking. He also holds two master's degrees: a Master of Science in Finance from Johnson & Wales University in Rhode Island, USA, and a master's degree in quantitative and computational finance from Vietnam National University, Ho Chi Minh City. In 2025, he was recognized as an Associate Professor in Economics. Currently, he is responsible for managing the Fintech Department at Ho Chi Minh University of Banking. His research interests include financial market forecasting, machine learning, fintech, quantitative trading, corporate credit rating, and investment portfolio management. He can be contacted via email at nhatnm@hub.edu.vn.

Duy Ngo Hoang Khanh is a researcher at Ho Chi Minh City University of Banking, focussing on financial forecasting and the integration of new technologies in finance and banking. His principal research is on financial forecasting and the utilisation of technology within the finance and banking industries. His work encompasses several subjects, notably artificial intelligence (AI) and blockchain, especially within the realm of fintech enterprises. Possessing a robust foundation in data science, he has participated in several initiatives focused on enhancing finance solutions using cutting-edge AI and blockchain technologies. His proficiency in data science and fintech has rendered him a significant contributor to the domain, propelling innovations in financial technology for enterprises and institutions. He can be contacted at email: ngohoangkhanhduy.work@gmail.com.