CENTRE FOR RESPONSIBLE BANKING & FINANCE
CRBF

University of St Andrews
School of Management

WORKING PAPERS IN RESPONSIBLE BANKING & FINANCE

Modelling and Predicting Failure in US Credit Unions

By *Qiao Peng, Donal McKillop, Barry Quinn, Kailong Liu*

Abstract: This study presents a Random Forest (RF) - based machine learning model to predict the liquidation of US Credit Unions one year in advance, using biannual data from December 2001 to September 2020. The model demonstrates impressive accuracy with the training set (89.1%) and the test set (97.9%) when utilising all 44 factors. Simplifying the model to only the top five factors based on feature importance analysis results in a slightly lower, but still significant, accuracy for both sets (83.6% and 92.2%, respectively). Comparisons with seven other classification methods verify the superiority of the RF model. This study also uses the Cox proportional-hazards model and Shapley value-based approaches to interpret the significance and interactions of key features. The model offers valuable tools for regulators and Credit Unions, providing an early warning system for potential failures and an opportunity for corrective measures or strategic mergers, ultimately protecting the National Credit Union Share Insurance Fund.

WP Nº 24-002

1st Quarter 2024

University of St Andrews
Scotland's first university

600 YEARS
1413 – 2013

Modelling and Predicting Failure in US Credit Unions

Qiao Peng [a], Donal McKillop [a], Barry Quinn [a], Kailong Liu [b*]

[a] Queen's Management School, Queen's University Belfast, Belfast, BT9 5EE, UK
[b] School of Control Science and Engineering, Shandong University, China

*** Corresponding author:** Kailong Liu

E-mail addresses:
qiao.peng@qub.ac.uk (Q. Peng)
dg.mckillop@qub.ac.uk (D. McKillop)
b.quinn@qub.ac.uk (B. Quinn)
kliu02@qub.ac.uk (K. Liu)

CRediT Authorship Statement
Qiao Peng: Data curation, Methodology, Software, Validation, Writing-original draft.
Donal McKillop: Conceptualization, Methodology, Validation, Writing-original draft.
Barry Quinn: Conceptualization, Methodology, Validation, Writing-original draft.
Kailong Liu: Conceptualization, Methodology, Software.

Total words (exclude abstract, figures, tables, references and appendix): 5,488

Modelling and Predicting Failure in US Credit Unions

Abstract

This study presents a Random Forest (RF) - based machine learning model to predict the liquidation of US Credit Unions one year in advance, using biannual data from December 2001 to September 2020. The model demonstrates impressive accuracy with the training set (89.1%) and the test set (97.9%) when utilising all 44 factors. Simplifying the model to only the top five factors based on feature importance analysis results in a slightly lower, but still significant, accuracy for both sets (83.6% and 92.2%, respectively). Comparisons with seven other classification methods verify the superiority of the RF model. This study also uses the Cox proportional-hazards model and Shapley value-based approaches to interpret the significance and interactions of key features. The model offers valuable tools for regulators and Credit Unions, providing an early warning system for potential failures and an opportunity for corrective measures or strategic mergers, ultimately protecting the National Credit Union Share Insurance Fund.

Keywords: Random Forest; Interpretable machine learning; Explainable AI; Failure prediction; Feature selection; Credit unions.

1. Introduction

The Federal Credit Union Act of 1934 in the US initiated a surge in establishing Credit Unions, reaching a peak of 23,866 institutions in 1969. However, despite the continuous growth of assets and memberships, which increased to $1.85 billion and 124.3 million by 2020, there has been a consistent decline in the number of Credit Unions, falling to 5,099 by the end of 2020. This number reduction has been mainly due to mergers and many failures, especially during substantial economic downturns. Given this trend, stakeholders such as the National Credit Union Administration (NCUA), the Credit Union National Association (CUNA), the Credit Union board of directors, and associated organisations are keenly interested in reliable models that can accurately predict the likelihood of Credit Union failure.[1]

Despite the scarcity of information regarding Credit Union failures[2], several studies have unearthed key contributing factors. In the US, younger, smaller, and less capitalised Credit Unions are more prone to failure (Wilcox, 2005), especially during adverse macroeconomic conditions (Smith and Woodbury, 2010). Furthermore, Credit Unions with small asset sizes, high liquidity, low loans-to-asset ratios, and overly high capital ratios are at an increased risk (Goddard et al., 2014). International studies also offer insights; machine learning models have effectively predicted early financial distress in Australian Credit Unions (Tan and Dihardjo, 2001). In Canada, the Equity/Asset ratio is a robust

---

[1] Although such events occur stochastically (to a degree), information from capital markets can be utilised to develop predictive models (Barboza et al., 2017).

[2] An extensive literature exists regarding bank default prediction. A review of this literature is found in Demyanyk and Hasan (2009).

predictor of failure (Pille and Paradi, 2002). In the UK, a multi-period logit model based on macroeconomic and institutional level indicators obtained from CAMEL ratios effectively characterises troubled Credit Unions a year in advance (Coen et al., 2019). These findings underscore the importance of macroeconomic conditions and institutional-level indicators in predicting Credit Union failures.

Machine learning, a branch of artificial intelligence, leverage intelligent algorithms rooted in statistical methods that enable computers to learn from experience and make predictions or classifications. Such algorithms analyse known data, establishing laws applied to anonymous data (Russell and Norvig, 2021; Alpaydin, 2020). These data-intensive methods have permeated financial sectors, enhancing decision-making in financial markets forecasting, risk modelling, and failure prediction (Panos and Wilson, 2021; Jordan and Mitchell, 2015; Liu and Pun, 2022). Random Forest (RF) is recognised for its accuracy and power among various machine learning algorithms. It randomises variables and data to construct multiple decision trees, where the individual trees' outputs determine the final output. The advantages of RF include efficient computation for large datasets, resilience to missing and unbalanced data, unbiased error estimates, and revealing insights about variable importance and interaction (Emerson et al., 2019; Shi and Horvath, 2006; Van et al., 2012). However, machine learning models have limitations compared to classical survival analysis methods. Statistical models typically offer more substantial explanatory power and extensive diagnostic capabilities, whereas machine learning models rely on numerical optimisation and lack a clearly defined process for generating results (Dixon et al., 2020). Thus, selecting an analytical method should reflect the specific needs and constraints of a given problem.

This study presents an innovative application of machine learning, specifically the RF model, to predict the potential failure of US Credit Unions one year in advance. The data utilised spans from December 2001 to September 2020, a period which has seen a decrease in the number of Credit Unions but an increase in membership and asset size (Wilcox, 2005). This paradox underlines the importance of developing reliable prediction models for regulatory bodies, an objective of critical importance to bodies such as the NCUA and the CUNA. The RF model is selected due to its high accuracy, computational efficiency, and ability to handle large volumes of variables and missing data (Emerson et al., 2019). We undertake a comparative analysis of the RF model with seven alternative machine learning techniques including AdaBoost, Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression, Linear Discriminant Analysis, Naive Bayes and K-Nearest Neighbours (KNN) (Alpaydin, 2020; Dixon et al., 2020). Furthermore, we employ the Cox proportional-hazards (Cox PH) model to elucidate the econometric implications of critical variables.

We regard a Credit Union as failed if it has been liquidated or subjected to a purchase and assumption enforced by the NCUA. The data subset from January 2003 to June 2015 serves as model training data, and July 2015 to September 2020 as the testing period. Emphasising failure prediction and post hoc explanations, we employ interpretable machine learning techniques, including feature importance, accumulated local effects (ALE), and Shapley value-based approaches (SHapley Additive exPlanations (SHAP) and Shapley Additive Global importancE (SAGE)) (Apley and Zhu, 2016; Lundberg and Lee, 2017). Findings indicate that the interactions between various features are crucial in predicting Credit Union failures. RF models showcase remarkable predictive accuracy (89.1% for the training set, 97.9% for the test set), outperforming logistic regression. While focusing on the five most crucial features, accuracy remains high (83.6% for the training set, 92.2% for the test set), demonstrating that these factors encompass most of the predictive power. The comparison of RF with seven alternative methods underlines the superiority of RF across the predictive classifications.

Overall, this study primarily demonstrates the efficiency and flexibility of machine learning models in feature selection, circumventing the challenges of traditional econometric specification testing. These models effectively quantify and visualise the impact of features on predicted outcomes, including non-monotonic effects, which linear regression struggles to characterise. This study shows that a reliable future prediction of Credit Union failure can be achieved using as few as five factors, thus providing an early warning system that allows for corrective measures or merger partner identification, safeguarding the National Credit Union Share Insurance Fund (NCUSIF). Furthermore, the study employs a unique methodology combining the Cox PH model and an explainable machine learning model, underscoring the significance of feature interactions in predicting Credit Union failures.

The most significant contribution of this study is the efficacy of using a reduced set of critical features in predicting Credit Union failure. Remarkably, even when the model is constrained to the top five features, it retains a high degree of accuracy for both the training and test sets. This finding offers a streamlined approach to risk assessment and provides a manageable framework for decision-makers. Additionally, the study employs various interpretable machine learning techniques to provide a detailed understanding of decision-making processes (Apley and Zhu, 2016; Lundberg and Lee, 2017). These techniques offer differing insights into the model's function and a comprehensive understanding of the factors influencing Credit Union failure. Minor contributions include providing an early warning system for at-risk Credit Unions, allowing these institutions to implement corrective measures or find appropriate merger partners. This approach safeguards the NCUSIF. The findings also open avenues for future research. If expanded to encompass a broader spectrum of Credit Unions, the approach detailed in this study could provide invaluable insights for regulatory bodies.

The rest of this paper is structured into sections addressing literature review, data introduction, the three-stage research method, detailed explanation of the RF-based model framework and interpretable machine learning techniques, results discussion, and concluding remarks.

2. Confronting Financial Institution Research Using Machine Learning

Integrating machine learning into econometrics has sparked significant discussions and transformations in financial institution research. This literature review explores the intersection between machine learning and econometrics, building upon influential works such as Varian's (2014) and Mullainathan and Spiess's (2017) examination of the predictive power of machine learning in econometrics. The debate concerns balancing predictive performance and traditional econometric models' inferential capabilities. Furthermore, challenges arise regarding causality, as econometrics traditionally emphasises causal relationships, while machine learning focuses on prediction. Ethical considerations, interpretability issues and the potential bias of machine learning algorithms have been highlighted (Kleinberg et al., 2017). Addressing these concerns, Athey (2018) offers insights into the strengths of machine learning for semi-parametric estimation, systematic model selection and modifying algorithms to provide valid confidence intervals. Ultimately, the combination of machine learning and new datasets will profoundly impact financial institution research, influencing research questions, collaborative approaches and the role of economists in policy implementation. This review highlights the dynamic and transformative relationship between machine learning and econometrics, underscoring the benefits and challenges of pursuing advanced financial analysis.

Recent research has increasingly applied machine learning methods to issues relevant to financial institutions, outperforming traditional statistical techniques in several applications. Min and Lee (2005) use machine learning to predict bankruptcy with superior explanatory power and stability. Khandani et al. (2010) utilise these methods to build non-linear consumer credit risk prediction models, resulting in significant cost savings. Liu and Zhang (2010) develop an effective machine learning model to detect suspicious activities for regulators and financial institutions, aiding in combating money laundering. Wang et al. (2011) apply integrated learning using multiple classifiers for credit scoring and achieved optimal performance with the stacking and bagging decision tree. Barboza et al. (2017) test machine learning models for predicting corporate bankruptcy, outperforming traditional approaches and improving prediction accuracy, especially with bagging, boosting, and RF techniques. Bellotti et al. (2021) use a series of regression techniques and machine learning models to predict the recovery rate of non-performing loans. The results suggest that rule-based models such as RF, Boosted Trees and Cubism perform significantly better. Petropoulos et al. (2020) use a range of modelling approaches to predict bank failures for a sample of US financial institutions. The results show that RF is superior in out-of-sample and out-of-time predictions, and that neural networks perform almost as well as RF in out-of-time prediction. Conlon et al. (2021) suggest that machine learning techniques can improve the performance of factor-based portfolio optimisations. Lastly, Chang et al. (2018) argue the impact of data's heterogeneous nature on classification accuracy, successfully using the XGBoost classifier to build a superior credit risk evaluation model.

Machine learning has been applied to tackle issues pertinent to Credit Unions. Desai et al. (1997) find that neural networks and genetic algorithms surpass traditional methods like linear discriminant analysis and logistic regression in classifying Credit Union loans. Tan and Dihardjo (2001) demonstrate the effectiveness of an Artificial Neural Network (ANN) model over a Probit model for developing early warning signals of financial distress in Australian Credit Unions. Harris (2013) compares two SVM-based credit-scoring models using different default definitions, concluding that broader default definitions performed better, and quantitative models improved risk assessment. Gozer et al. (2014) find that statistical models based on ANN and SVM technologies excel at predicting bankruptcy in Brazilian Credit Unions, with the SVM-based LibSVM algorithm performing best. Finally, Paula et al. (2019) propose a method combining credit risk modelling and profit scoring for improved loan provision effectiveness in Brazilian Credit Unions, finding the RF model superior in estimating credit and profit scores compared to logistic and ordinary least squares regression models. These papers are summarised in detail in Appendix 1.

3. Data and Methodology

This study predicts Credit Union failures by analysing financial data and status information from December 2001 to September 2020. We exclude Credit Unions have abnormal data or have more than two features without data. We use data from at least one year before each observation time point.

The model training data uses a subset covering January 2003 to June 2015. This dataset contains information on 110 Credit Unions that failed during this period and additionally information on a further 110 Credit Unions that existed unaltered over this period. The failed Credit Unions in the training set contain all Credit Unions in the filtered database that are categorised as having ‘failed’ during this period. The same number of surviving Credit Unions are chosen for the training set to balance out the dataset. Specifically, there were 6,021 surviving Credit Unions in 2015. We employ the random under-sampling method (Kotsiantis and Pintelas, 2003) to select 110 of them to match the 110 failed Credit Unions in that period. Random under-sampling is a non-heuristic method that aims to overcome the idiosyncrasies of the machine learning algorithm by randomly eliminating examples from the majority class. A limitation of random under-sampling is that potentially useful data that may be important to the inductive process may be discarded (Kotsiantis et al., 2006). To mitigate this limitation, we match surviving Credit Unions with failed Credit Unions based on the basis of their location (state) as they adopt a community-based strategy (Deller and Sundaram-Stukel, 2012). The test set includes all Credit Unions that are filed as failed, 34 in total, from July 2015 to September 2020. The test set also includes 5,166 Credit Unions that survived over the course of the same period. For testing purposes, data on these surviving Credit Unions is chosen from a random year during the period under assessment.

We use the PEARLS[3] monitoring toolkit to identify 44 financial indicators to analyse a Credit Union's financial condition. We also consider 16 additional factors from the literature. The variables used in our predictive models are in Appendix 2, and their descriptive statistics are in Appendix 3. The data includes both failed and surviving Credit Unions.

This study aims to predict Credit Union failures using a three-part method. We use machine learning classifiers, followed by explainable AI algorithms to analyse the feature space of the best model, and classical survival models to clarify feature complexities. See Figure 1 for the design of the whole paper.

3.1 Stage 1: Machine Learning Horse Race

In Appendix 4, we provide a detailed presentation of the RF model, which is the best-performing model. To evaluate the effectiveness of the RF-based model, we compared it with seven other mainstream classification methods, namely AdaBoost, DT, SVM, Logistic Regression, Linear Discriminant Analysis, Naive Bayes, and KNN. Appendix 5 contains the classification results of these methods after 5-fold cross-validation, with all 44 variables used in the comparative tests. Our analysis shows that the RF procedure outperforms all other methods, with an overall accuracy of 97.9%. The false negatives rate for the RF model (the Credit Union continues to survive but is predicted to fail) is only 2%, while the false positive rate (the Credit Union fails but is predicted to survive) is 8.8%. Only the AdaBoost method comes close to the RF procedure in terms of performance, with an overall accuracy of 94.9%, a false negatives rate of 5.0%, and a false positive rate of 10.4%. Additionally, we also evaluated the performance of both training and test sets for the RF procedure when using only the five most important variables. Our findings indicate that even with only these five variables, the RF model provides a high level of predictive accuracy for the test sets, which is quite pleasing.

[3] PEARLS was developed by the World council of Credit Unions (WOCCU). It is a set of financial ratios or indicators that help to standardize terminology between institutions. In total, there are 44 quantitative financial indicators that facilitate an integral analysis of the financial condition of any financial institution. PEARLS is an acronym which stands for (Protection, Effective financial structure, Asset quality, Rates of return and cost, Liquidity and Signs of growth).

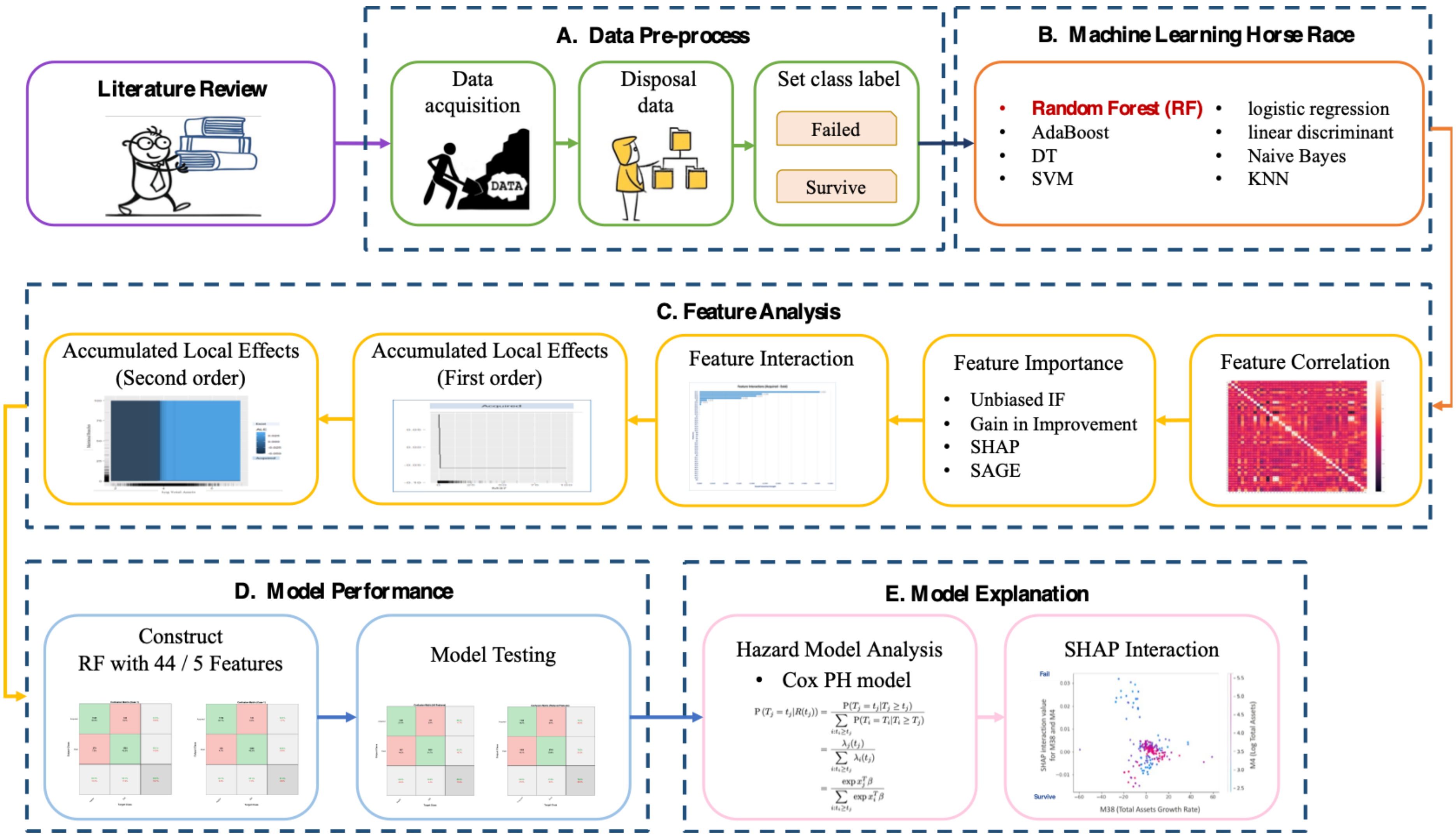


Figure 1. Research Design

3.2 Stage 2: Explain the Machines

In order to interpret the prediction process of the RF model, four techniques of interpretable machine learning are integrated. The structure of the interpretable RF classification model used in this research is illustrated in Figure 2. The 44 variables listed in Appendix 2 are used as the inputs. The RF classification model output is the labelled classes of future performance of the Credit Unions concerned (Failed or Survived).

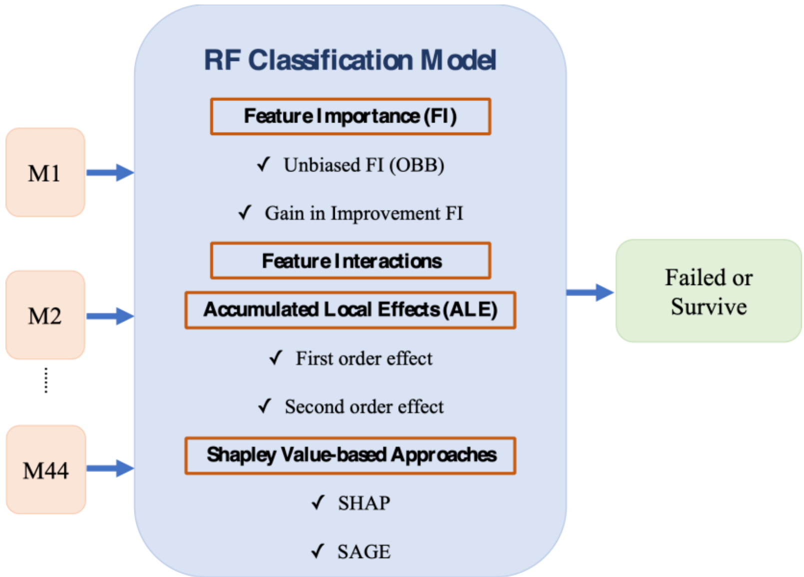


Figure 2. Structure of Designed RF Classifier

Traditional machine-learning models lack explanations for their results, making it difficult to prevent errors. This study shows that financial indicators are intertwined and nonlinear, requiring a method that provides classification results and details the decision-making process. Feature importance, interactions, and ALE are essential for understanding the mechanics behind predictions.

■ **Feature Importance (FI)**

In order to predict the future performance of the Credit Union, we need to determine the importance of various factors. We use an unbiased FI obtained through Out-of-Bag (OOB) prediction to do this. This method allows us to see how the classification error varies when the values of a particular feature in the OOB prediction test are randomly changed. Another way to assess feature importance is by calculating the gain in improving the Gini impurity changes resulting from the splits per feature. A larger Gini value indicates that the feature has more potential to improve the classification result. For a detailed explanation of determining unbiased FI and Gini impurity FI, please refer to Liu et al. (2021).

■ **Feature Interaction**

When developing a prediction model, the relationships between different factors can often be complex and interactive. This means that the effects of each factor on the predicted outcome are not simply cumulative but more intricate. To address this, more advanced algorithms, particularly tree-based ones, are better equipped to capture these interactions and therefore tend to perform better (Boehmke and Greenwell, 2019). To further estimate the strength of these interactions, Friedman et al. (2008) introduced the H-statistic, which measures the influence of feature interaction on the degree of change in prediction results. Mathematically, the H-statistic for the interaction between factors is described as follows:

$$H_{jk}^2 = \sum_{i=1}^{n}[PD_{jk}\left(x_j^{(i)}, x_k^{(i)}\right) - PD_j\left(x_j^{(i)}\right) - PD_k\left(x_k^{(i)}\right)]^2 / \sum_{i=1}^{n} PD_{jk}^2(x_j^{(i)}, x_k^{(i)})|$$

where PD_{jk} is the two-way partial dependence function of factors j and k, and $PD_{jk}\left(x_j, x_k\right) = PD_j\left(x_j\right) + PD_k(x_k)$, PD_j is the partial dependence function of factor j. Similarly, the H-statistic of the factor j interacting with any other factor is shown as follows:

$$H_j^2 = \sum_{i=1}^{n}[\hat{f}\left(x^{(i)}\right) - PD_j\left(x_j^{(i)}\right) - PD_{-j}\left(x_{-j}^{(i)}\right)]^2 / \sum_{i=1}^{n} \hat{f}^2\left(x^{(i)}\right)$$

where the prediction function $\hat{f}(x)$ is the sum of partial dependence functions, $PD_{-j}\left(x_{-j}\right)$ the partial dependence function depends on all factors apart from the j^{th} factor (Molnar, 2020).

- **Accumulated Local Effects (ALE)**

ALE explain how features affect the prediction of a machine learning model on average. Its essence is simplifying the function by averaging other factors' influence, thereby reducing the complicated prediction function f to a function that only relies on one/two factors. The ALE plots average the variations of the predictions and then accumulate them on the grid. The uncentered effect is estimated by:

$$\hat{\tilde{f}}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j{}^i \in N_j(k)} [f\left(z_{k,j}, x_{\setminus j}^{(i)}\right) - f(z_{k-1}, x_{\setminus j}^{(i)})]$$

Where $z_{k,j}$ is the boundary value of the k^{th} feature interval, $n_j(k)$ the number of samples in the interval, the sample point in the interval and the features other than a feature j. When the effect is centred, the mean effect is zero. So, the ALE main effect estimator is expressed by:

$$\hat{f}_{j,ALE}(x) = \hat{\tilde{f}}_{j,ALE}(x) - \frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{f}}_{j,ALE}\left(x_j^{(i)}\right)$$

The ALE plots can be estimated for a single numerical feature and present the interaction effect of two factors. In this instance, the two factors' overall mean and main effects are adjusted. The ALE plots for the two factors only estimate the second-order effect (the additional interaction effect of the two factors) but do not contain their main effects (Molnar, 2020). To estimate the ALE second-order effect of x_j and x_l, the sample range is divided into K^2rectangular cells rather than using intervals. (k, m) represents the indices into the grid with k corresponding to x_j and m corresponding to x_l. The uncentered effect is estimated by:

$$\hat{h}_{\{j,l\},ALE}\left(x_j, x_l\right) = \sum_{k=1}^{k_j(x_j)} \sum_{m=1}^{k_l(x_l)} \frac{1}{n\{j,l\}(k,m)} \sum_{i:x_{\{j,l\}}^{(i)} \in N_{\{j,l\}}(k,m)} \triangle_f^{\{j,l\}}\left(K, k, m; x_{\{j,l\}}^{(i)}\right)$$

where $\triangle_f^{\{j,l\}}\left(K,k,m;x_{\{j,l\}}^{(i)}\right)$ means the second-order finite difference of $f\left(x_i,x_l,x_{\{j,l\}}^{(i)}\right)$ for (x_i,x_l) across cells $\left(z_{k-1,j}^K,z_{k,j}^K\right]\times\left(z_{m-1,l}^K,z_{m,l}^K\right]$. Then, the ALE main effects of X_j and X_l for $\hat{h}_{\{j,l\},ALE}\left(x_j,x_l\right)$ are calculated and subtracted from $\hat{h}_{\{j,l\},ALE}\left(x_j,x_l\right)$:

$$\hat{\tilde{f}}_{\{j,l\},ALE}\left(x_j,x_l\right)$$
$$= \hat{h}_{\{j,l\},ALE}\left(x_j,x_l\right)$$
$$-\sum_{k=1}^{k_j(x_j)}\frac{1}{n_j(k)}\sum_{m=1}^{k}n_{\{j,l\}}(k,m)\left\{\hat{h}_{\{j,l\},ALE}\left(z_{k,j},z_{m,l}\right)-\hat{h}_{\{j,l\},ALE}\left(z_{k-1,j},z_{m,l}\right)\right\}$$
$$-\sum_{m=1}^{k_l(x_l)}\frac{1}{n_l(m)}\sum_{k=1}^{k}n_{\{j,l\}}(k,m)\left\{\hat{h}_{\{j,l\},ALE}\left(z_{k,j},z_{m,l}\right)-\hat{h}_{\{j,l\},ALE}\left(z_{k,j},z_{m-1,l}\right)\right\}$$

So, the centred ALE second-order effect is expressed as follows (Apley & Zhu, 2020):

$$\hat{f}_{\{j,l\},ALE}\left(x_j,x_l\right)=\hat{\tilde{f}}_{\{j,l\},ALE}\left(x_j,x_l\right)-\frac{1}{n}\sum_{k=1}^{K}\sum_{m=1}^{K}n_{\{j,l\}}(k,m)\,\hat{\tilde{f}}_{\{j,l\},ALE}\left(z_{k,j},z_{m,l}\right)$$

■ **Shapley Values, SHAP and SAGE**

The Shapley value is a solution derived from cooperative game theory that can be used to calculate the contributions of variables for individual predictions in various machine learning models. To calculate the exact Shapley value, estimating all possible alliances (sets) of factor values with and without the factor is necessary. During each iteration of the curious case, a random sample is chosen, and a random order of factors is created. Two new cases are generated by combining values, with the first being of particular interest. The values of the first case substitute all previous values and include the value of the factor in question. The second case is similar to the first, but the value of the factor is not included in the replacement process. This allows for the calculation of the difference in prediction.

$$\phi_j^m=\hat{f}\left(x_{+j}^m\right)-\hat{f}\left(x_{-j}^m\right)$$

The average of all these differences is:

$$\phi_j(x)=\frac{1}{M}\sum_{m=1}^{M}\phi_j^m$$

As more factors are involved, the number of probable alliances grows exponentially, making it, in this instance, more challenging to find the exact solution. In such a situation, the approximation with Monte-Carlo sampling developed by Štrumbelj and Kononenko (2014) solves this problem. The approximation is:

$$\hat{\phi}_j=\frac{1}{M}\sum_{m=1}^{M}\left(\hat{f}\left(x_{+j}^m\right)-\hat{f}\left(x_{-j}^m\right)\right)$$

All Shapley values can be obtained by repeating the above process.

Based on the theory of Shapley values, SHAP are developed by Lundberg and Lee (2017) as an additive feature attribution approach with a local sensitivity focus. It defines the explanation as follows:

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j$$

where g represents the explanation model, M means the maximum coalition size, and $\phi_j \in \mathbb{R}$ stands for the feature attribution for the feature j. SHAP can be used to estimate the importance of features. Specifically, features with larger average absolute Shapley values are considered to be more critical:

$$I_j = \frac{1}{n} \sum_{i=1}^{n} \left| \phi_j^{(i)} \right|$$

The SHAP method for determining feature importance differs from the unbiased and Gini impurity methods in that it measures importance based on the magnitude of feature attributions. This approach allows for creating a SHAP summary plot, which shows the Shapley value for each feature in a given case. The y-axis of the plot represents the feature importance ranking, while the x-axis represents the Shapley value. Additionally, the SHAP dependence plot provides a more detailed view of the relationship between a feature's value and its impact on predicted outcomes. Unlike the ALE plot, which only shows average effects, the SHAP dependence plot also displays variance. When feature interactions are significant, the points on the y-axis are more dispersed. Finally, the SHAP interaction method estimates the combined effect of multiple features after accounting for their individual effects, using the Shapley interaction index to measure this effect.

$$\phi_{i,j} = \sum_{S \subseteq \backslash \{i,j\}} \frac{|S|!\,(M - |S| - 2)!}{2(M-1)!} \delta_{ij}(S)$$

where S represents the set of all possible feature coalitions, $\delta_{ij}(S) = \hat{f}_x(S \cup \{i,j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$ and $i \neq j$. When the values of SHAP interaction for all features are calculated, one matrix can be obtained per instance with dimensions, where are the feature numbers.

An additional approach to estimating the feature importance based on the theory of Shapley value is SAGE, which summarises the importance of each feature by the predictive power it contributes. Different from SHAP, SAGE focuses on global scopes and predictive power metrics. The detailed formula derivation for this approach can be found in Covert et al. (2020).

3.3 Stage 3: Classic Survival Model Explanations

In this stage, we use a Cox PH model to extract traditional econometric meaning from the features. This type of semiparametric regression model analyses the effects of multiple factors on survival outcomes over the study period, taking into account the survival outcome (Survived or Failed) and survival time as dependent variables. The model has been applied to bank failures and acquisitions in the past, with Wheelock and Wilson (2000) being the pioneers. We use hazard function estimations, as detailed in Minder and Bednarski (1996), to address failure prediction for Credit Unions.

To prevent the issue of multicollinearity in modelling, we retain the feature with higher importance estimated in Stage 2 if the correlation between two variables is more significant than 0.5, and discard the other. The correlations between pairs of the 44 variables are displayed in Figure 3, and the features used to build the Cox PH model are listed in Appendix 6. We also utilise a post-hoc visual explanation (SHAP dependence plot) to verify the reliability of the results of this conventional statistical approach for failure prediction.

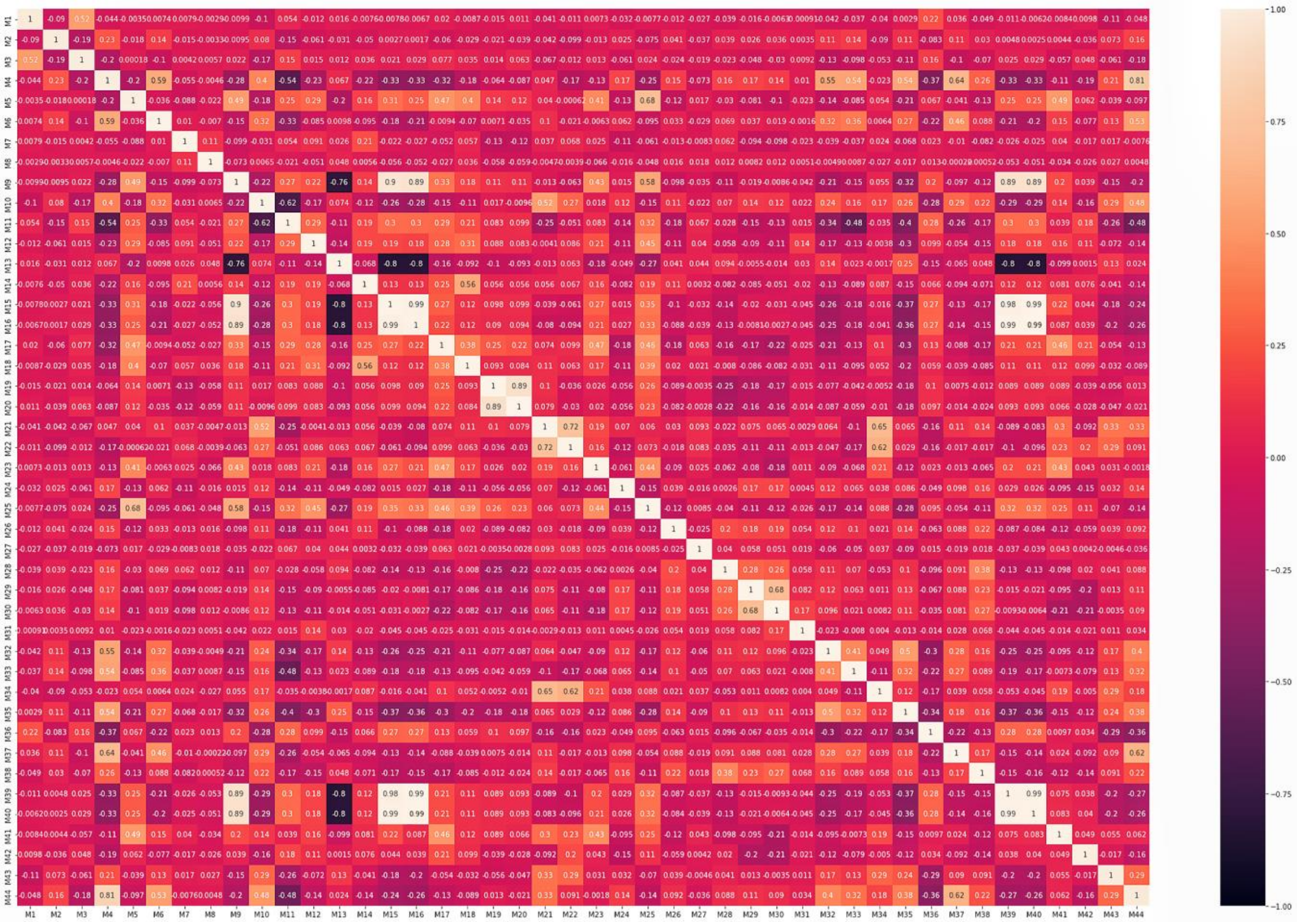


Figure 3. Feature Linear Correlation Matrix

4. Empirical Results

Our RF-based framework analyses the decision-making process to classify Credit Unions based on survival. We train with 110 Failed and 110 Survived Credit Unions and test with 34 Failed and 5,166 Survived. The RF classification model uses all 44 features as inputs to predict future performance. We conduct a sensitivity analysis of the features and explain the RF model's classification performance. Finally, we use the filtered 29 variables to build the Cox PH model and discuss insights gained.

4.1 Interpretable Machine Learning

We analyse 44 factors to predict Credit Union performance. Log total assets and total assets growth rate are the most important factors for survival. High nonperforming loans and low profitability and liquidity also increase the chance of failure. Other variables have a minor impact. SHAP and SAGE techniques confirm our findings (See Appendices 7 and 8 for detailed exposition).

The H-statistic in Figure 4 reveals that feature interaction impacts prediction results. M4 (Log total assets) and M25 (Cash & cash equivalents/Total shares and deposits) have a strong interaction with other features, explaining 15.82% and 14.91% of the variance, respectively. Figure 5 shows a complex non-linear relationship between M4 and M25, with a strong interaction of 35.16%. The top five variables from Table 1 also have significant impact and interaction with other features, contributing significantly to the prediction result.

In Figure 6 presents the one-dimensional ALE plots illustrating the impact of the five most significant variables (M4, M38, M17, M24, and M25) on the likelihood of a 'Failed' prediction. The plot shows that all five variables significantly influence the prediction of failure. For M4 (Log total assets), M38 (Total assets growth rate), and M24 (Net income/Assets), an increase in their values decreases the probability of Credit Union failure. On the other hand, M17 (Nonperforming loans/Total loans) and M25 (Cash & cash equivalents/Total shares and deposits) have a positive effect, indicating that the probability of Credit Union failure increases as these ratios increase. As mentioned earlier, the literature suggests that factors such as size, profitability (Return on Assets), loan write-offs, and excessive liquidity are also contributing factors to the failure of US Credit Unions (Wilcox, 2005; Smith and Woodbury, 2010; Goddard et al., 2014).

Figure 7 shows the two-dimensional ALE plots for the top 4 variables: M4, M38, M17, and M24, reporting second-order effects. The six plots demonstrate the effects of these variables on predicting Credit Union performance (Failed or Survived). Lighter shades indicate a higher probability of failure, while darker shades mean it's more likely to survive. The plots suggest that smaller Credit Unions with lower M38 are more likely to fail, while larger ones are more likely to survive. Other variables, such as Nonperforming loans/Total loans and Net income/Assets, also impact Credit Union failure rates.

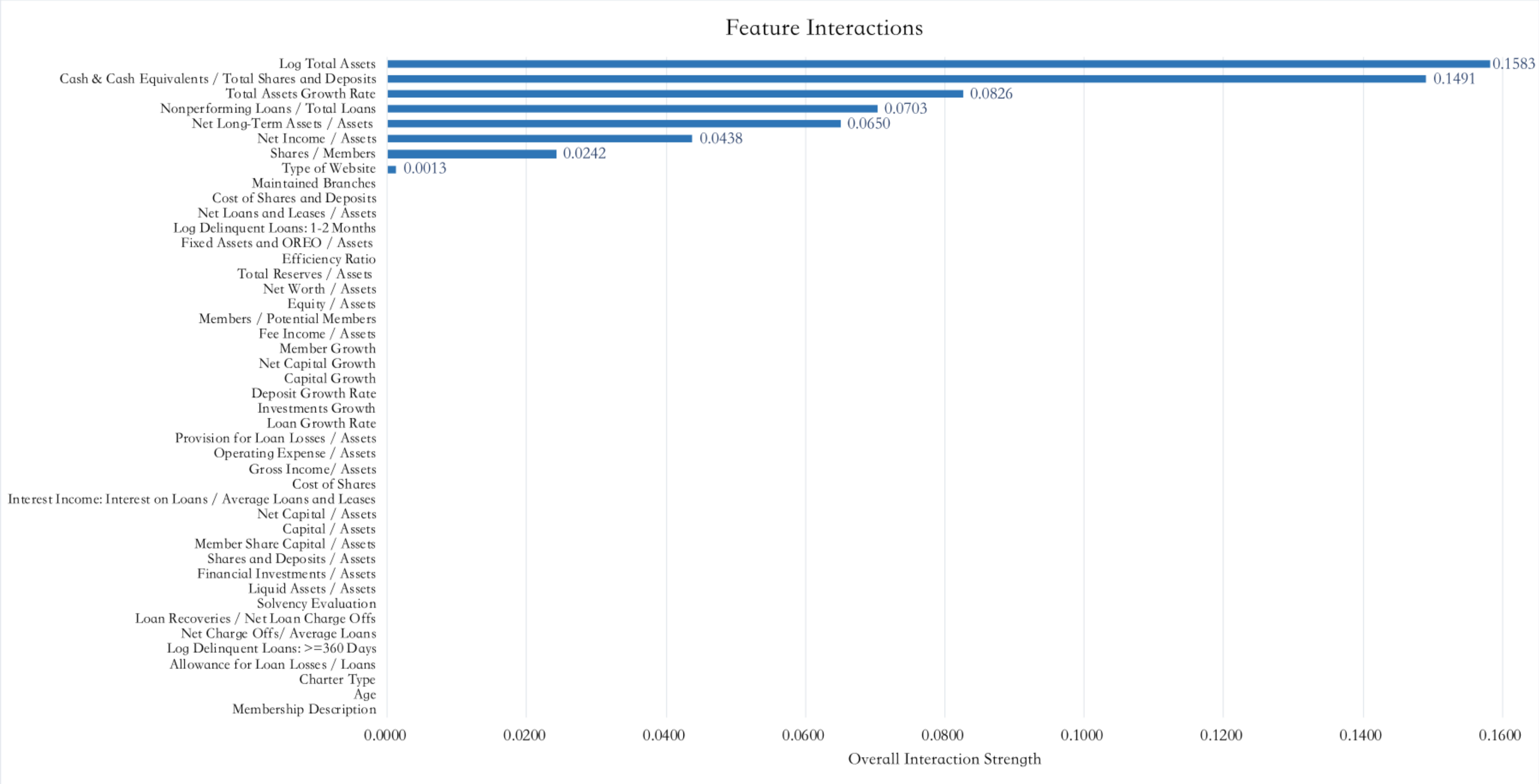
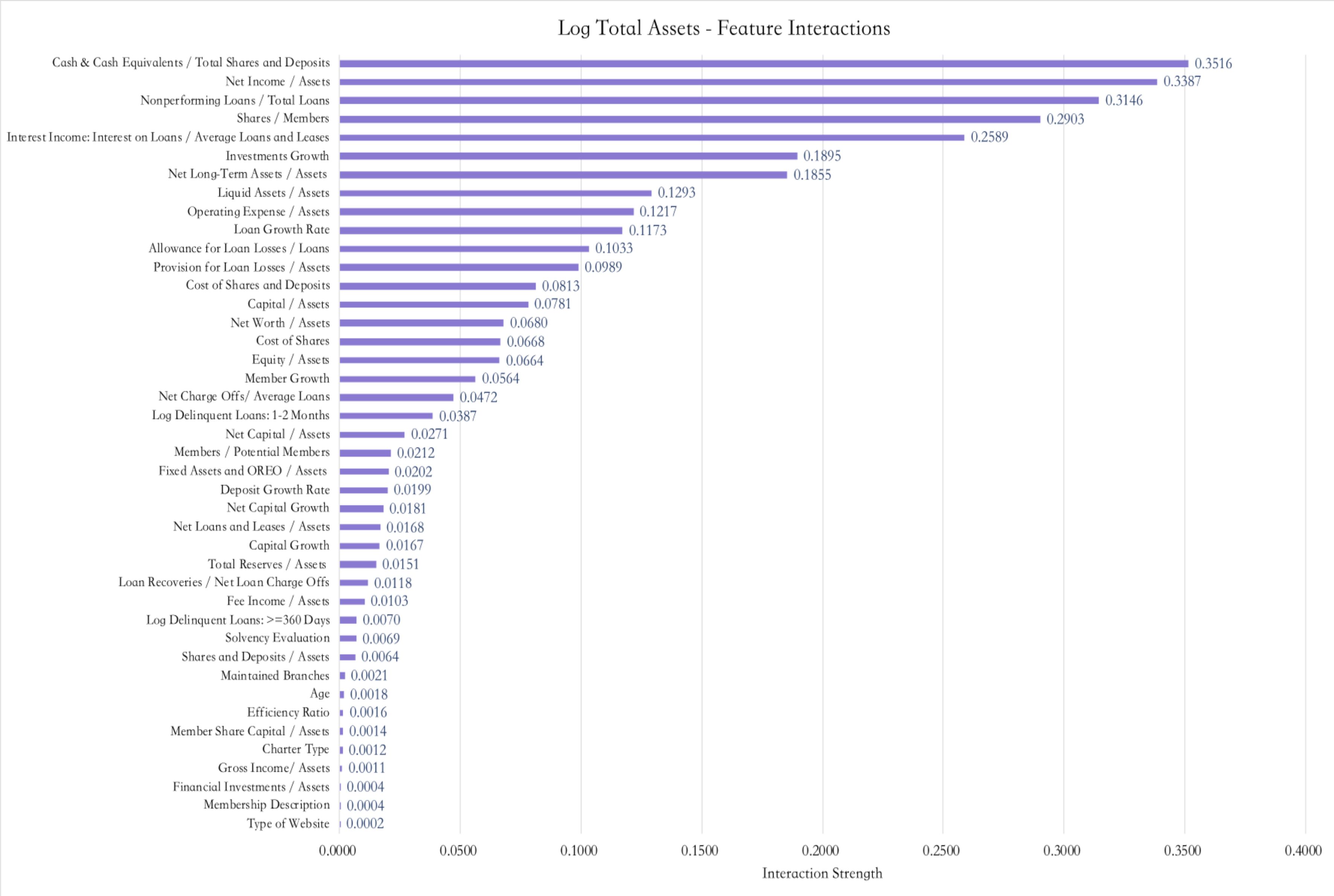


Figure 4. Feature Interactions


Log Total Assets - Feature Interactions
Cash & Cash Equivalents / Total Shares and Deposits 0.3516
Net Income / Assets 0.3387
Nonperforming Loans / Total Loans 0.3146
Shares / Members 0.2903
Interest Income: Interest on Loans / Average Loans and Leases 0.2589
Investments Growth 0.1895
Net Long-Term Assets / Assets 0.1855
Liquid Assets / Assets 0.1293
Operating Expense / Assets 0.1217
Loan Growth Rate 0.1173
Allowance for Loan Losses / Loans 0.1033
Provision for Loan Losses / Assets 0.0989
Cost of Shares and Deposits 0.0813
Capital / Assets 0.0781
Net Worth / Assets 0.0680
Cost of Shares 0.0668
Equity / Assets 0.0664
Member Growth 0.0564
Net Charge Offs/ Average Loans 0.0472
Log Delinquent Loans: 1-2 Months 0.0387
Net Capital / Assets 0.0271
Members / Potential Members 0.0212
Fixed Assets and OREO / Assets 0.0202
Deposit Growth Rate 0.0199
Net Capital Growth 0.0181
Net Loans and Leases / Assets 0.0168
Capital Growth 0.0167
Total Reserves / Assets 0.0151
Loan Recoveries / Net Loan Charge Offs 0.0118
Fee Income / Assets 0.0103
Log Delinquent Loans: >=360 Days 0.0070
Solvency Evaluation 0.0069
Shares and Deposits / Assets 0.0064
Maintained Branches 0.0021
Age 0.0018
Efficiency Ratio 0.0016
Member Share Capital / Assets 0.0014
Charter Type 0.0012
Gross Income/ Assets 0.0011
Financial Investments / Assets 0.0004
Membership Description 0.0004
Type of Website 0.0002
0.0000
0.0500
0.1000
0.1500
0.2000
0.2500
0.3000
0.3500
0.4000
Interaction Strength

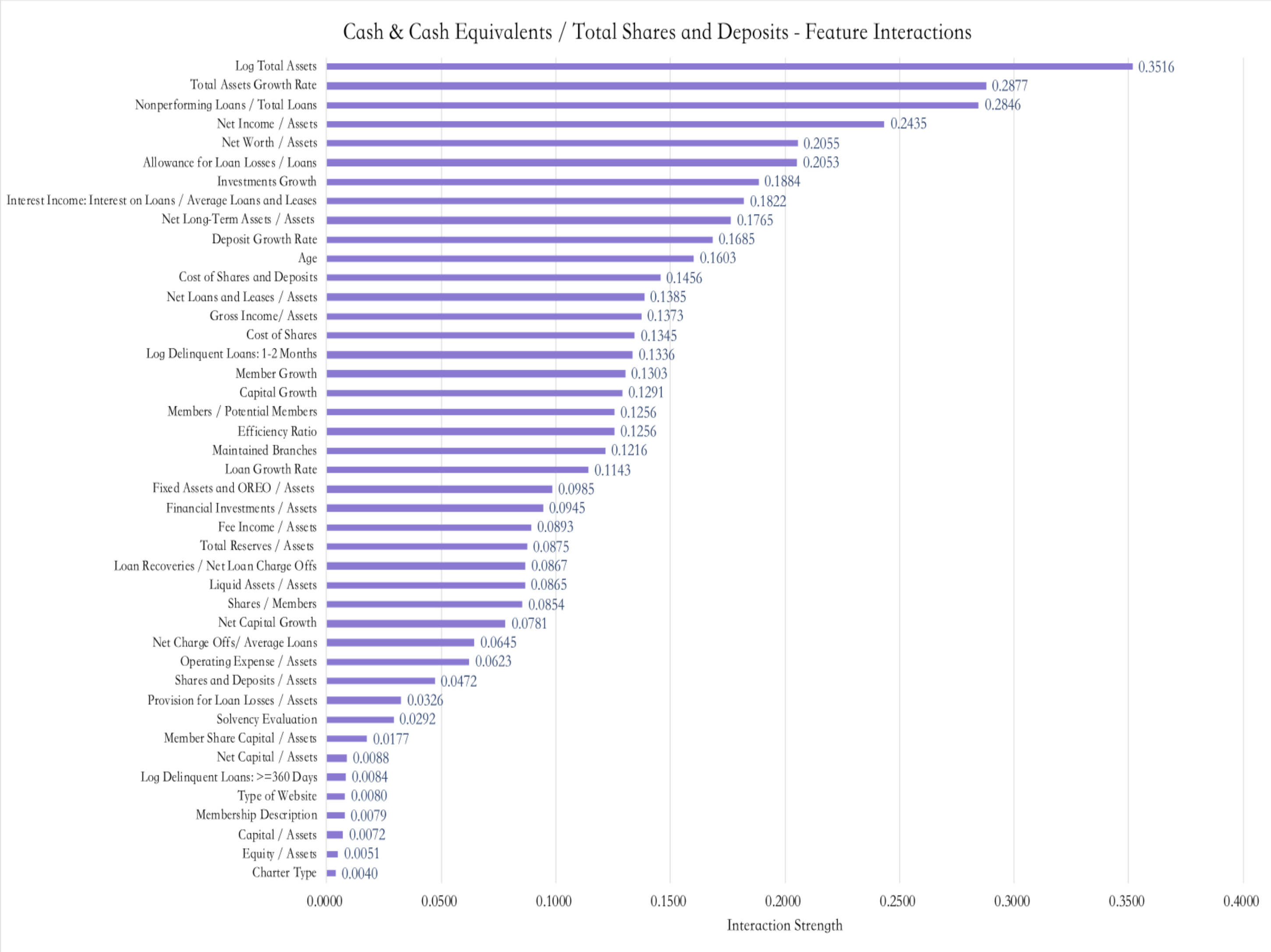


Figure 5. Interaction Strength

(a) Log total assets (M4); (b) Cash & cash equivalents/Total shares and deposits (M25)

Table 1. Feature Importance (Value and Ranking)

Number	Name	Value (Unbiased FI)	Ranking (Unbiased FI)	Value (Gain in Improvement FI)	Ranking (Gain in Improvement FI)	Comprehensive Ranking
M4	Log Total Assets	17.29	1	11.56	1	1
M38	Total Assets Growth Rate	15.60	2	6.01	2	2
M17	Nonperforming Loans/Total Loans	10.94	5	5.86	3	3
M24	Net Income/Assets	11.56	4	4.76	4	4
M25	Cash & Cash Equivalents/Total Shares and Deposits	13.59	3	4.40	6	5
M32	Shares/Members	7.97	8	4.59	5	6
M18	Interest Income: Interest on Loans/Average Loans and Leases	8.96	6	3.11	7	7
M33	Net Long-Term Assets/Assets	8.90	7	2.87	8	8
M5	Allowance for Loan Losses/Loans	7.41	9	2.70	9	9
M11	Liquid Assets/Assets	6.72	13	2.40	10	10
M27	Investments Growth	6.90	12	1.99	11	11
M26	Loan Growth Rate	6.31	14	1.41	12	12
M40	Net Worth/Assets	7.32	10	1.15	17	13
M16	Net Capital/Assets	7.08	11	1.11	18	14
M19	Cost of Shares and Deposits	6.02	16	1.22	15	15
M23	Provision for Loan Losses/Average Assets	5.11	20	1.24	14	16
M44	Log Delinquent Loans: 1-2 Months	4.82	22	1.35	13	17
M41	Total Reserves/Assets	4.79	23	1.20	16	18
M7	Net Charge Offs/Average Loans	5.49	18	1.06	21	19
M10	Net Loans and Leases/Assets	4.97	21	1.11	19	20
M9	Solvency Evaluation	6.09	15	0.85	26	21
M39	Equity/Assets	5.94	17	0.87	25	22
M28	Deposit Growth Rate	4.40	24	0.97	22	23
M2	Age	3.37	27	1.09	20	24
M15	Capital/Assets	5.12	19	0.62	30	25
M31	Member Growth	3.96	26	0.93	24	26
M35	Type of Website	2.57	30	0.95	23	27
M22	Operating Expense/Average Assets	3.10	28	0.70	29	28
M30	Net Capital Growth	2.47	32	0.84	27	29
M37	CU Maintained Branches	4.16	25	0.47	35	30
M29	Capital Growth	2.16	33	0.80	28	31
M8	Loan Recoveries/Net Loan Charge Offs	2.56	31	0.62	31	32
M43	Fixed Assets and OREO/Assets	2.87	29	0.53	33	33
M42	Efficiency Ratio	1.32	35	0.60	32	34
M20	Cost of Shares	1.42	34	0.44	37	35
M13	Shares and Deposits/Assets	1.11	37	0.46	36	36
M34	Fee Income/Average Assets	0.49	41	0.50	34	37
M21	Gross Income/Average Assets	1.11	38	0.40	38	38
M12	Investments: Securities and Other Investments/Total Assets	1.16	36	0.32	42	39
M14	Membership Capital Corporate Cus/Assets	0.55	40	0.40	39	40
M36	Members/Potential Members	1.06	39	0.39	40	41
M6	Log Delinquent Loans: >=360 Days	0.05	42	0.36	41	42
M3	Charter Type	-0.90	43	0.07	43	43
M1	Membership Description	-1.00	44	0.03	44	44

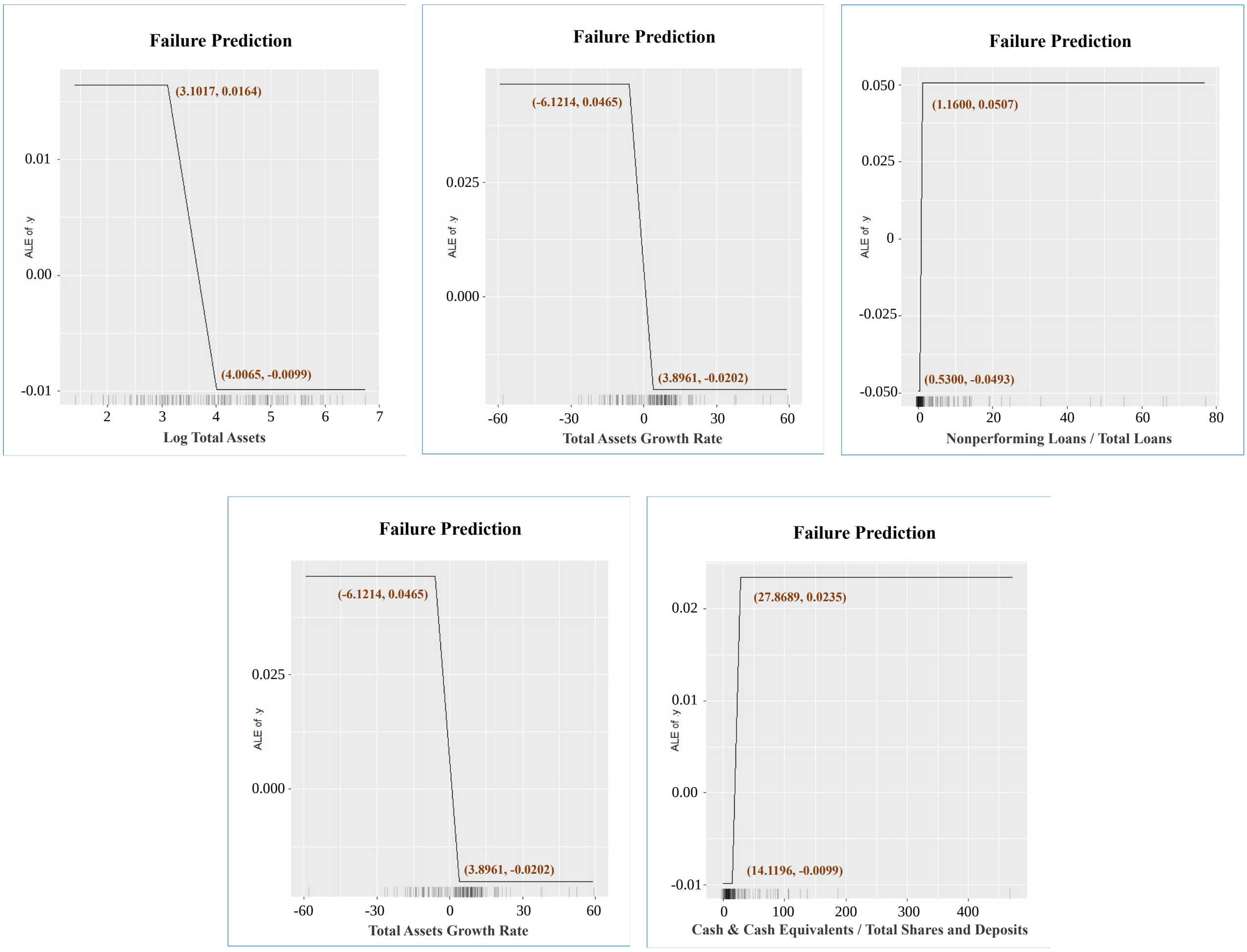


Figure 6. One-Dimensional ALE Plots

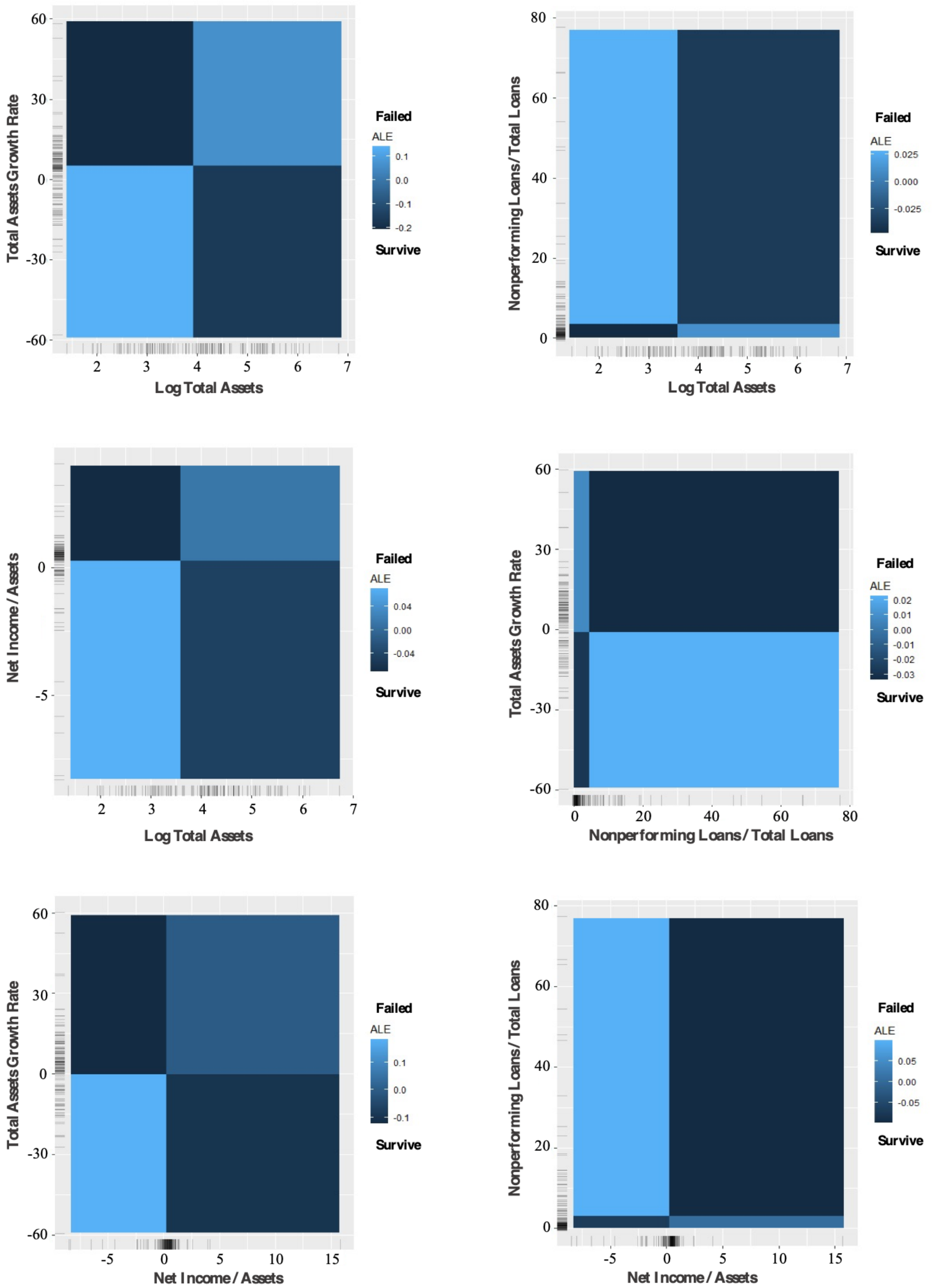


Figure 7. Two-Dimensional ALE Plots

4.2 Performance of the RF-based Model

A test was done to assess the effectiveness of the proposed RF framework in predicting Credit Union performance as Failed or Survived. All 44 variables are used in an OOB prediction test, resulting in a confusion matrix (CM) shown in Figure 8 (a). The overall accuracy achieved is 89.1%, with only a small percentage of misclassifications.

An evaluation is conducted to determine the effectiveness of a Credit Union's future performance (Failed or Survived) classification using the proposed RF framework. The evaluation included an OOB prediction test that used the five most important features (M4, M38, M17, M24 and M25) identified earlier. The resulting CM for the Failed/Survived classification is shown in Figure 8 (b), with a slight decrease in classification accuracy compared to using all 44 variables. The overall accuracy rate dropped by 5.5%, with the accuracy rate for the Survived class decreasing by 3.6% and the accuracy rate of the Failed class falling by 7.3%. However, this analysis shows that using only the five factors (M4, M38, M17, M24 and M25) has a relatively small negative impact on prediction accuracy compared to using all 44 factors.

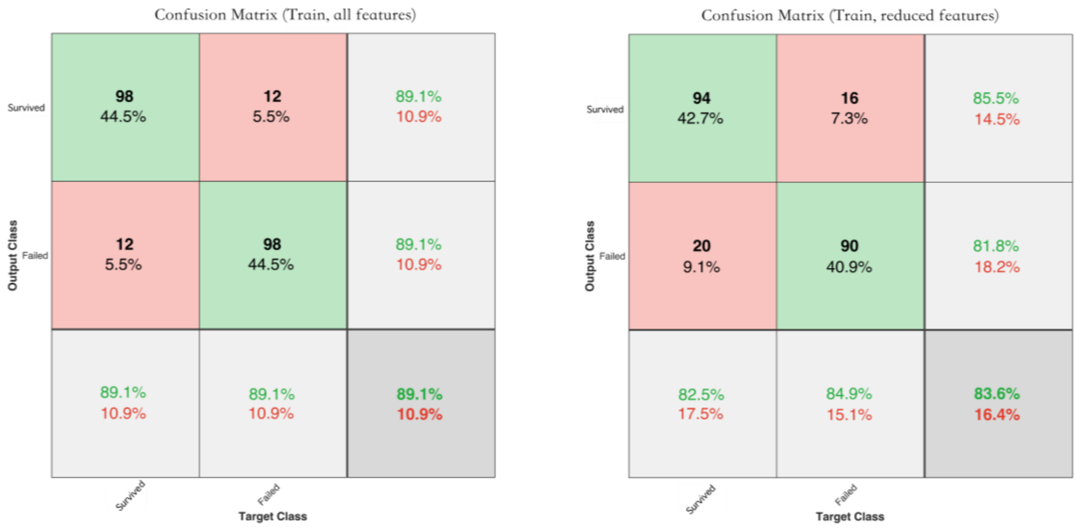


Figure 8. Confusion Matrix (Training)
(a) 44 Features; (b) 5 Features

The established RF models are tested with test datasets and have an accuracy of 97.9% and 92.2% for the 44-variable and 5-variable models, respectively. The RF model with 44 variables had a 2.0% false negatives rate and an 8.8% false positive rate, while the RF model with 5 variables had a 7.8% false negatives rate and a 17.6% false positive rate. Focusing on the five most important factors does not significantly affect classification accuracy.

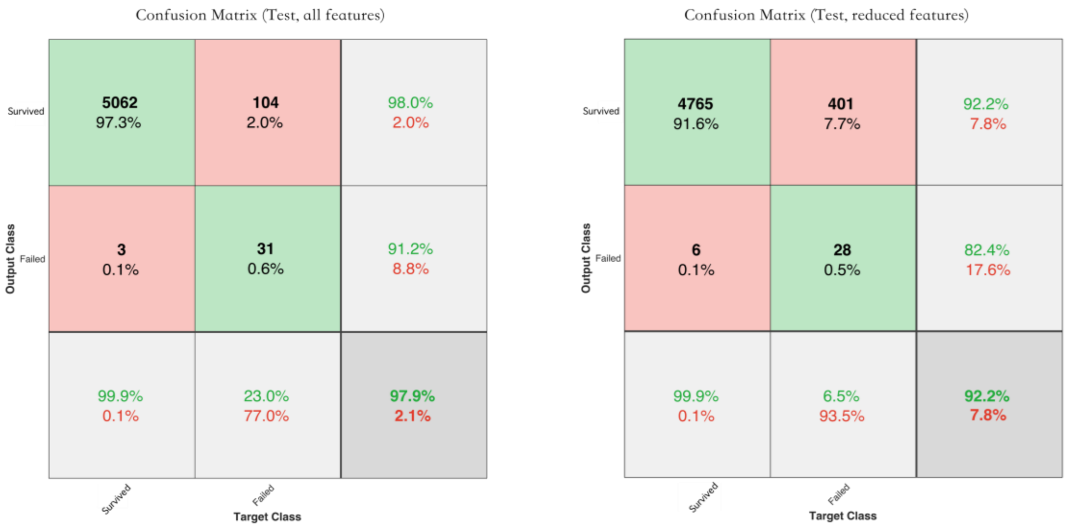


Figure 9. Confusion Matrix (Test)
(a) 44 Features; (b) 5 Features

4.3 Hazard Model Analysis

In this section, we utilised the 29 variables listed in Appendix 6 to fit the Cox PH model, and the results are presented in Table 2. During this modelling section, all variables are standardized, and the survived Credit Unions are matched with the failed Credit Unions based on location (State) with the same number (144 survived Credit Unions and 144 failed Credit Unions are included).

Higher liquidity (M25: Cash & Cash Equivalents/Total Shares and Deposits) increases the probability of failure, whereas the total assets (M4) is negatively significant, highlighting smaller Credit Unions' increased probability of failure. These findings align with the ALE plots' results in Figure 7. M18 (Interest Income), representing the cost of borrowing from Credit Unions, is positively significant, which may reflect a 'last ditch' attempt to generate an increase in surpluses, or it may simply reflect the uncompetitive nature of the failing Credit Union's business model.

The positive coefficient of M12 (Investments) and the negative coefficient of M10 (Net Loans and Leases/Assets) indicate that Credit Unions with a higher percentage of deposits (lower percentage of loans) have a greater probability of failure, which may be due to their limited income from loans. M34 (Fee Income/Average Assets) and M19 (Cost of Shares and Deposits) are positively significant, indicating that failed Credit Unions charge their members higher fees, but their savings members achieve a higher return. Combining these results with M18's discussed findings may reflect a bias towards savings members, the inappropriate nature of the business model, or perhaps the need to offer high rates to mitigate against members withdrawing their funds.

Both M2 (Age) and M26 (Loan Growth Rate) are negatively significant, implying that younger Credit Unions are more likely to fail. Additionally, when the loans of a Credit Union stop growing (or even decline), it is also more likely to fail. Finally, M31 (Member Growth) and M36 (Members/Potential

Members) are both positively significant, indicating that the more a Credit Union has exhausted its potential membership, the more likely it is to fail.

Table 2. Cox PH Model Results

Variables	Coef.	Robust Std. Err.	95% Conf. Interval	
			Lower band	Upper band
M25 (Cash & Cash Equivalents/Total Shares and Deposits)	0.3035***	0.1061	0.0957	0.5114
M4 (Log Total Assets)	-1.6430***	0.1945	-2.0242	-1.2619
M24 (Net Income/Assets)	0.0795	0.1234	-0.1623	0.3213
M18 (Interest Income: Interest on Loans/Average Loans and Leases)	0.2360**	0.1085	0.0233	0.4487
M38 (Total Assets Growth Rate)	-0.0182	0.1362	-0.2851	0.2486
M17 (Nonperforming Loans/Total Loans)	-0.0343	0.1472	-0.3227	0.2542
M41 (Total Reserves/Assets)	-0.1424	0.1249	-0.3872	0.1024
M12 (Investments: Securities and Other Investments/Total Assets)	0.1637**	0.0767	0.0134	0.3140
M7 (Net Charge Offs/Average Loans)	0.0257	0.1069	-0.1838	0.2352
M23 (Provision for Loan Losses/Average Assets)	0.0132	0.1145	-0.2113	0.2376
M10 (Net Loans and Leases/Assets)	-0.4153***	0.0995	-0.6103	-0.2204
M34 (Fee Income/Average Assets)	0.1708***	0.0661	0.0412	0.3003
M27 (Investments Growth)	-0.0255	0.1213	-0.2613	0.2103
M30 (Net Capital Growth)	-0.2379	0.2147	-0.6587	0.1830
M39 (Equity/Assets)	-0.1795	0.1264	-0.4273	0.0683
M2 (Age)	-0.3824***	0.1207	-0.6191	-0.1458
M26 (Loan Growth Rate)	-0.1755*	0.1064	-0.3841	0.0331
M42 (Efficiency Ratio)	0.0189	0.0848	-0.1473	0.1851
M19 (Cost of Shares and Deposits)	0.4605***	0.1550	-0.1567	0.7643
M28 (Deposit Growth Rate)	0.1130	0.1050	-0.0928	0.3188
M8 (Loan Recoveries/Net Loan Charge Offs)	-0.0455	0.0691	-0.1809	0.0898
M31 (Member Growth)	0.2350***	0.0747	0.0885	0.3815
M3 (Charter Type)	0.0462	0.1382	-0.2247	0.3171
M36 (Members/Potential Members)	0.2198*	0.1155	-0.0065	0.4460
M43 (Fixed Assets and OREO/Assets)	0.0540	0.1029	-0.1476	0.2556
Intercept	-10.4413***	0.9424	-12.2883	-8.5942
ln_p	1.1322***	0.0913	0.9533	1.3111
*****: 1% significance level; **:5% significance level; *:10% significance level**				
p	3.1025	0.2832	2.5942	3.7104
1/p	0.3223	0.0294	0.2695	0.3855

The Cox PH model only considers the main effects of features on predicted outcomes, ignoring interactions between pairs of features. To address this, the SHAP interaction technique is used to examine the RF-based model's second-order effect. The result is shown in Figure 10, highlighting the significant role of feature interactions, with the largest pure interaction effect found between M4 and M38. Adding an interaction term to reconstruct the Cox PH model emphasizes the importance of considering feature interactions (see Table 3). The SHAP interaction plot for M4 and M38 shows that the impact of the interaction on the predicted outcome is complex rather than linear (see Figure 11). Smaller Credit Unions with lower Total Assets Growth Rate are more likely to fail.

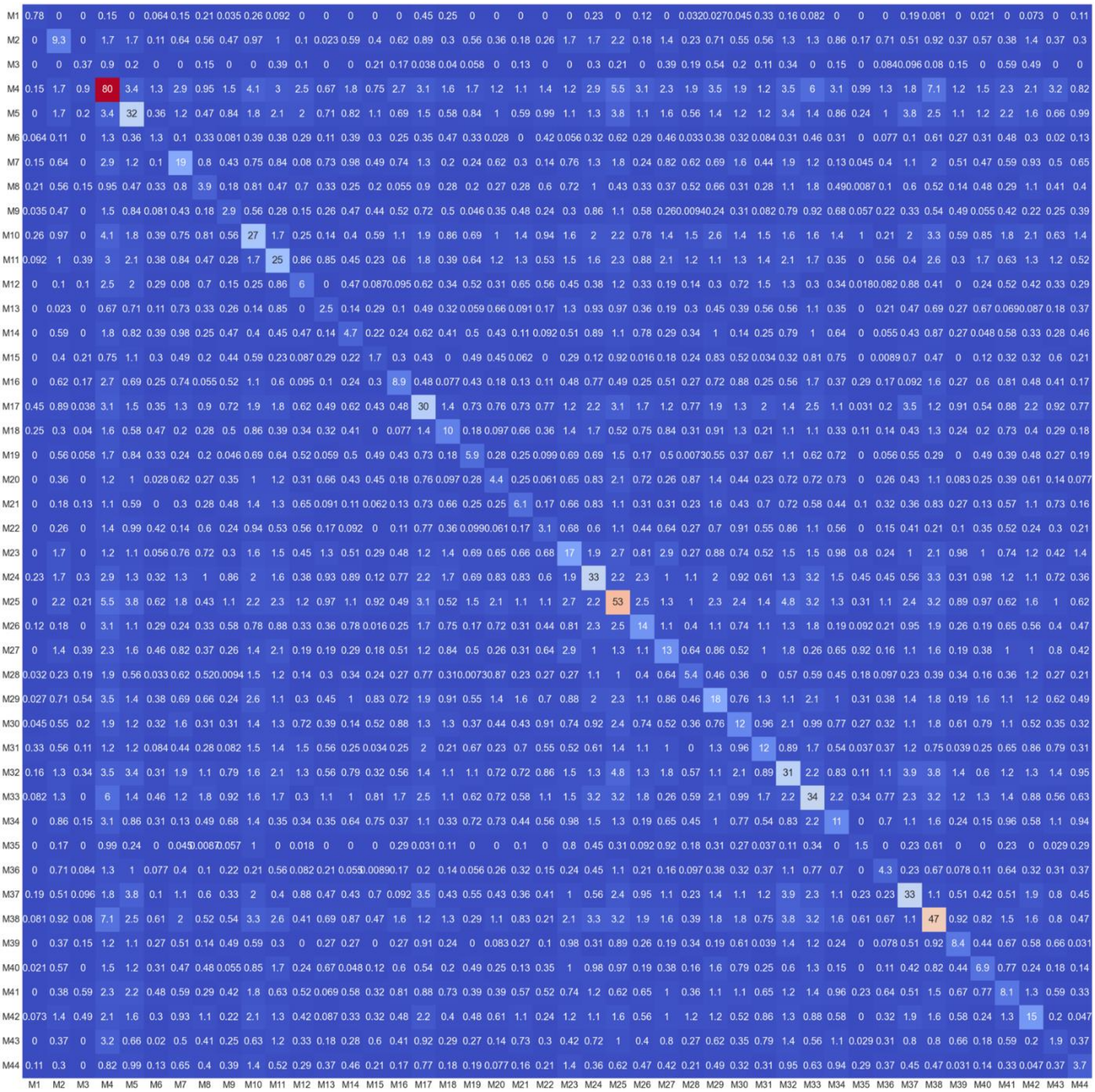


Figure 10. SHAP Interaction

Table 3. Cox PH Model with the Interaction of M4 and M38

Variables	**Coef.**	**Robust Std. Err.**	**95% Conf. Interval**	
			Lower band	**Upper band**
M25 (Cash & Cash Equivalents/Total Shares and Deposits)	0.2898***	0.1061	0.0817	0.4978
M4 (Log Total Assets)	-1.6303***	0.1968	-2.0118	-1.2489
M24 (Net Income/Assets)	0.0725	0.1231	-0.1688	0.3138
M18 (Interest Income: Interest on Loans/Average Loans and Leases)	0.2341**	0.1098	0.0190	0.4492
M38 (Total Assets Growth Rate)	-0.1165	0.2148	-0.5375	0.3046
M17 (Nonperforming Loans/Total Loans)	-0.0247	0.1444	-0.3077	0.2584
M41 (Total Reserves/Assets)	-0.1688	0.1187	-0.4015	0.0639
M12 (Investments: Securities and Other Investments/Total Assets)	0.1547**	0.0748	0.0081	0.3014
M7 (Net Charge Offs/Average Loans)	0.0470	0.1179	-0.1841	0.2781
M23 (Provision for Loan Losses/Average Assets)	0.0431	0.1180	-0.1882	0.2744
M10 (Net Loans and Leases/Assets)	-0.4062***	0.0982	-0.5988	-0.2136
M34 (Fee Income/Average Assets)	0.1559**	0.0689	0.0209	0.2908
M27 (Investments Growth)	-0.0279	0.1198	-0.2628	0.2069
M30 (Net Capital Growth)	-0.2117	0.2261	-0.6548	0.2315
M39 (Equity/Assets)	-0.1731	0.1229	-0.4140	0.0678
M2 (Age)	-0.3756***	0.1226	-0.6159	-0.1354
M26 (Loan Growth Rate)	-0.2033*	0.1179	-0.4344	0.0278
M42 (Efficiency Ratio)	0.0280	0.0832	-0.1351	0.1911
M19 (Cost of Shares and Deposits)	0.4766***	0.1461	0.1902	0.7629
M28 (Deposit Growth Rate)	0.1151	0.0975	-0.0760	0.3062
M8 (Loan Recoveries/Net Loan Charge Offs)	-0.0476	0.0669	-0.1786	0.0835
M31 (Member Growth)	0.2264***	0.0778	0.0740	0.3788
M3 (Charter Type)	0.0523	0.1384	-0.2189	0.3236
M36 (Members/Potential Members)	0.2214*	0.1153	-0.0045	0.4474
M43 (Fixed Assets and OREO/Assets)	0.0446	0.1006	-0.1525	0.2418
Interaction of M4 and M38	-0.1496	0.1736	-1.4898	0.1906
Intercept	-10.4151***	0.9350	-12.2476	-8.5826
ln_p	1.1319***	0.0906	0.9543	1.3096
*****: 1% significance level; **:5% significance level; *:10% significance level**				
p	3.1016	0.2811	2.5968	3.7046
1/p	0.3224	0.0292	0.2699	0.3851

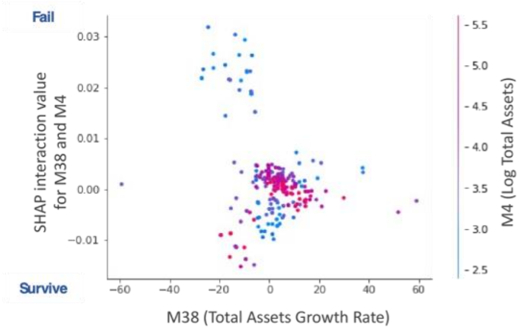


Figure 11. SHAP Interaction Plot

Finally, this study uses an explainable machine learning technique and tests its robustness through SHAP dependence plots. The results show that SHAP captures more comprehensive relationships between variables (see Figure 12). For example, the effect of M10 (Net Loans and Leases/Assets) on the prediction outcome is not monotonic. While the Cox PH model only captures the negative correlation between the percentage of loans and Credit Union failure, the interpretable machine learning technique through SHAP can more accurately and comprehensively describe the prediction decision process.

5. Conclusion

Credit Unions are often assessed using monitoring systems like PEARLS and CAMEL, which are also used for banks. These systems help manage risk and regulation by providing a framework to compare the performance of individual Credit Unions against sectoral standards. To expand the PEARLS framework, our study develops a RF model based on the PEARLS monitoring system and findings from a literature review of financial institution failure. Our model predicts Credit Union failure one year in advance with high levels of accuracy and identifies the decision-making process involved in generating the prediction.

Unlike other machine learning research, we use interpretable machine learning to determine the decision-making process of the prediction. This allows us to analyse feature importance, feature interactions, ALE, and Shapley value-based approaches, which all provide valuable insights. Our study proves the predictive superiority of the RF model when benchmarked against seven alternative classification methods. We also use an extensive dataset covering economic growth and stagnation periods to ensure the results are robust to different points in the economic cycle.

Our analysis identifies the five most important features for the RF model, namely Log total assets, Total assets growth rate, Nonperforming loans/Total loans, Net income/Assets, and Cash & cash equivalents/Total shares and deposits. These features measure size, size growth, return on assets, bad debt and liquidity. Using only these essential features, the RF model still achieved a high-test set accuracy of 92.2%.

Our study offers a means for organisations like the US Credit Union Regulator and the national representative body to establish an accurate assessment of future Credit Union failure one year in advance. This early warning system allows at-risk Credit Unions time to undertake corrective action or identify a merger partner in order to protect the NCUSIF.

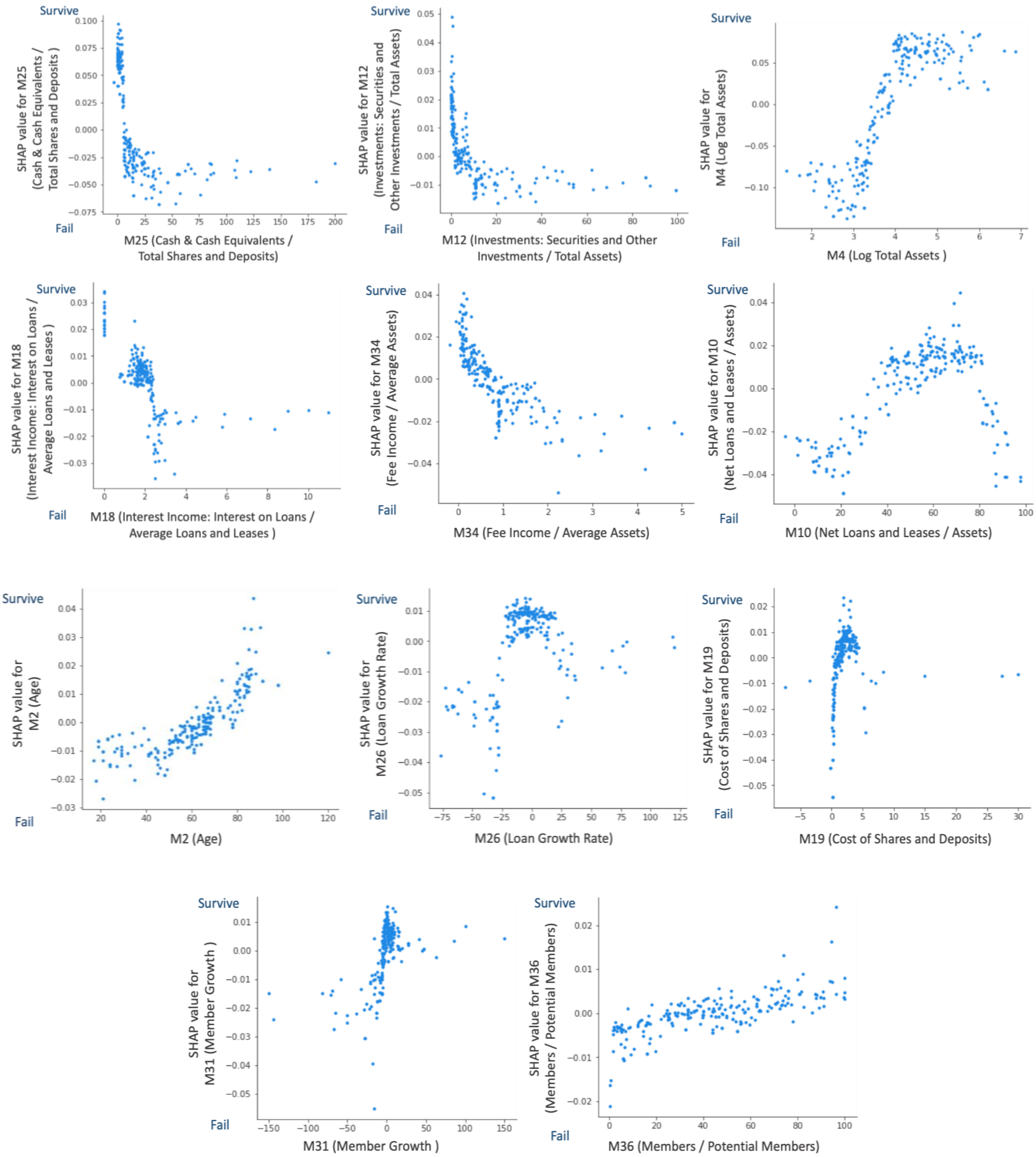


Figure 12. SHAP Dependence Plots

Reference

[1] Alpaydin, E., 2020. Introduction to machine learning. MIT Press.

[2] Apley, D.W. and Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), pp.1059-1086. https://doi.org/10.1111/rssb.12377

[3] Athey, S., 2018. The impact of machine learning on economics. The Economics of Artificial Intelligence: An Agenda (pp. 507-547). University of Chicago Press.

[4] Barboza, F., Kimura, H. and Altman, E., 2017. Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, pp.405-417. https://doi.org/10.1016/j.eswa.2017.04.006

[5] Bellotti, A., Brigo, D., Gambetti, P. and Vrins, F., 2021. Forecasting recovery rates on non-performing loans with machine learning. International Journal of Forecasting, 37(1), pp.428-444. https://doi.org/10.1016/j.ijforecast.2020.06.009

[6] Boehmke, B. and Greenwell, B.M., 2019. Hands-on machine learning with R. CRC Press.

[7] Chang, Y.C., Chang, K.H. and Wu, G.J., 2018. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Applied Soft Computing, 73, pp.914-920. https://doi.org/10.1016/j.asoc.2018.09.029

[8] Coen, J., Francis, W.B. and Rostom, M., 2019. The determinants of credit union failure: insights from the United Kingdom. International Journal of Central Banking, 15(4), pp.207-240.

[9] Conlon, T., Cotter, J. and Kynigakis, I., 2021. Machine learning and factor-based portfolio optimization. arXiv preprint arXiv:2107.13866. https://doi.org/10.48550/arXiv.2107.13866

[10] Covert, I., Lundberg, S.M. and Lee, S.I., 2020. Understanding global feature contributions with additive importance measures. Advances in Neural Information Processing Systems, 33, pp.17212-17223.

[11] Demyanyk, Y. and Hasan, I., 2010. Financial crises and bank failures: a review of prediction methods. Omega, 38(5), pp.315-324. https://doi.org/10.1016/j.omega.2009.09.007

[12] Deller, S. and Sundaram-Stukel, R., 2012. Spatial patterns in the location decisions of US credit unions. The Annals of Regional Science, 49, pp.417-445. https://doi.org/10.1007/s00168-011-0457-1

[13] Desai, V.S., Conway, D.G., Crook, J.N. and Overstreet Jr, G.A., 1997. Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. IMA Journal of Management Mathematics, 8(4), pp.323-346. https://doi.org/10.1093/imaman/8.4.323

[14] Dixon, M.F., Halperin, I. and Bilokon, P., 2020. Machine learning in Finance (Vol. 1406). New York, NY, USA: Springer International Publishing.

[15] Emerson, S., Kennedy, R., O'Shea, L., O'Brien, J., 2019. Trends and applications of machine learning in quantitative finance (May 30, 2019). 8th International Conference on Economics and Finance Research (ICEFR 2019). [online] Available at: https://ssrn.com/abstract=3397005.

[16] Friedman, J.H. and Popescu, B.E., 2008. Predictive learning via rule ensembles. The Annals of Applied Statistics, 2(3), pp.916-954. https://www.jstor.org/stable/30245114

[17] Goddard, J., Liu, H., Mckillop, D. and Wilson, J.O., 2014. The size distribution of US banks and credit unions. International Journal of the Economics of Business, 21(1), pp.139-156. https://doi.org/10.1080/13571516.2013.835970

[18] Gozer, I.C., leite de Albuquerque, A.R.P., Isotani, S., Gimenes, R.E.M.A.T., Moreira, W.H., Alberton, O. and Menezes, E.A.U., 2014. Evaluation of insolvency in mutual credit unions by the models of

artificial neural networks and support vector machines. African Journal of Agricultural Research, 9(16), pp.1227-1237. https://doi.org/10.5897/AJAR2014.8654

[19] Harris, T., 2013. Quantitative credit risk assessment using support vector machines: broad versus narrow default definitions. Expert Systems with Applications, 40(11), pp.4404-4413. https://doi.org/10.1016/j.eswa.2013.01.044

[20] Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science, 349(6245), pp.255-260. https://doi.org/10.1126/science.aaa8415

[21] Khandani, A.E., Kim, A.J. and Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), pp.2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

[22] Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2018. Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), pp.237-293. https://doi.org/10.1093/qje/qjx032

[23] Kotsiantis, S.B. and Pintelas, P.E., 2003. Mixture of expert agents for handling imbalanced data sets. Annals of Mathematics, Computing & Teleinformatics, 1(1), pp.46-55.

[24] Kotsiantis, S., Kanellopoulos, D. and Pintelas, P., 2006. Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering, 30(1), pp.25-36.

[25] Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R News, 2(3), pp.18-22.

[26] Liu, K., Hu, X., Zhou, H., Tong, L., Widanalage, D. and Marco, J., 2021. Feature analyses and modelling of lithium-ion batteries manufacturing based on random forest classification. IEEE/ASME Transactions on Mechatronics, 26(6), pp.2944-2955. https://doi.org/10.1109/TMECH.2020.3049046

[27] Liu, R. and Pun, C.S., 2022. Machine-Learning-enhanced systemic risk measure: A Two-Step supervised learning approach. Journal of Banking & Finance, 136, p.106416. https://doi.org/10.1016/j.jbankfin.2022.106416

[28] Liu, X. and Zhang, P., 2010, August. A scan statistics based suspicious transactions detection model for anti-money laundering (AML) in financial institutions. In 2010 International Conference on Multimedia Communications, IEEE, pp. 210-213. https://doi.org/10.1109/MEDIACOM.2010.37

[29] Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[30] Min, J.H. and Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications, 28(4), pp.603-614. https://doi.org/10.1016/j.eswa.2004.12.008

[31] Minder, C.E. and Bednarski, T., 1996. A robust method for proportional hazards regression. Statistics in Medicine, 15(10), pp.1033-1047. https://doi.org/10.1002/(SICI)1097-0258(19960530)15:10<1033::AID-SIM215>3.0.CO;2-Y

[32] Molnar, C., 2020. Interpretable machine learning. [online] Available at: https://christophm.github.io/interpretable-ml-book/

[33] Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), pp.87-106. https://doi.org/10.1257/jep.31.2.87

[34] Panos, G.A. and Wilson, J.O., 2021. Introduction: Financial literacy and responsible finance in the FinTech era: capabilities and challenges. In Financial Literacy and Responsible Finance in the FinTech Era (pp. 1-5). Routledge.

[35] Paula, D.A.V.D., Artes, R., Ayres, F. and Minardi, A.M.A.F., 2019. Estimating credit and profit scoring of a Brazilian credit union with logistic regression and machine-learning techniques. RAUSP Management Journal, 54, pp.321-336.
[36] Petropoulos, A., Siakoulis, V., Stavroulakis, E. and Vlachogiannakis, N.E., 2020. Predicting bank insolvencies using machine learning techniques. International Journal of Forecasting, 36(3), pp.1092-1113. https://doi.org/10.1016/j.ijforecast.2019.11.005
[37] Pille, P. and Paradi, J.C., 2002. Financial performance analysis of Ontario (Canada) credit unions: an application of DEA in the regulatory environment. European Journal of Operational Research, 139(2), pp.339-350. https://doi.org/10.1016/S0377-2217(01)00359-9
[38] Russell, S., and Norvig, P., 2009. Artificial intelligence a modern approach, third edition. Prentice Hall, 2009.
[39] Shi, T. and Horvath, S., 2006. Unsupervised learning with random forest predictors. Journal of Computational and Graphical Statistics, 15(1), pp.118-138. https://doi.org/10.1198/106186006X94072
[40] Smith, D.M. and Woodbury, S.A., 2010. Withstanding a financial firestorm: credit unions vs. banks. Filene Research Institute, Madison, Wis.
[41] Štrumbelj, E. and Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems, 41(3), pp.647-665. https://doi.org/10.1007/s10115-013-0679-x
[42] Tan, C.N. and Dihardjo, H., 2001. A study of using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model. Managerial Finance, 27(4), pp.56-77. https://doi.org/10.1108/03074350110767141
[43] Van Essen, B., Macaraeg, C., Gokhale, M. and Prenger, R., 2012, April. Accelerating a random forest classifier: multi-core, GP-GPU, or FPGA. IEEE 20th International Symposium on Field-Programmable Custom Computing Machines, IEEE, pp. 232-239. https://doi.org/10.1109/FCCM.2012.47.
[44] Varian, H.R., 2014. Big data: new tricks for econometrics. Journal of Economic Perspectives, 28(2), pp.3-28. https://doi.org/10.1257/jep.28.2.3
[45] Wang, G., Hao, J., Ma, J. and Jiang, H., 2011. A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications, 38(1), pp.223-230. https://doi.org/10.1016/j.eswa.2010.06.048
[46] Wheelock, D.C. and Wilson, P.W., 2000. Why do banks disappear? The determinants of US bank failures and acquisitions. Review of Economics and Statistics, 82(1), pp.127-138. https://doi.org/10.1162/003465300558560
[47] Wilcox, J.A., 2005. Economies of scale and continuing consolidation of credit unions. FRBSF Economic Letter.

Appendix

Appendix 1. Application of Machine Learning in Credit Unions

Author	Journal	Methodology	Countries	Time period	Main results
Desai et al. (1997)	IMA Journal of Management Mathematics	Feedforward neural networks, genetic algorithms, linear discriminant analysis and logistic regression	US	1988 to 1991	The neural network only performs significantly better in the classification of poor loans since there are no important uniform nonlinear variables in other datasets.
Tan and Dihardjo (2001)	Managerial Finance	ANN, Probit model	Australia	1989 to 1991	The use of ANN models can provide early warning signals to Credit Unions that are in the early stages of financial stress, thereby helping them avoid future financial difficulties (as much as possible).
Harris (2013)	Expert Systems with Applications	SVM	Barbados	1997 to 2012	The performance of the model built using Broad default definitions is superior to the model developed using Narrow default definitions. The introduction of quantitative credit risk models can improve the credit risk assessment of Barbados-based financial institutions.
Gozer et al. (2014)	African Journal of Agricultural Research	ANN, SVM	Brazil	2013	In most cases, the LibSVM algorithm of SVM achieves the best results in all performance evaluations. The SVM is superior as a binary classifier for predicting bankruptcy.
Paula et al. (2019)	RAUSP Management Journal	RF, logit and ordinary least squares regression	Brazil	2015 to 2016	Compared with logit and ordinary least squares regression models, the RF model is superior in estimating credit and profit scores for Credit Unions.

Appendix 2. Variable List

Protection	**Effective Financial Structure**	**Rates of Return and Cost**
M5: Allowance for Loan Losses/Loans	M10: Net Loans and Leases/Assets	M18: Net Loan Income/Average Net Loan Portfolio
M6: Log Delinquent Loans: >= 360 Days	M11: Liquid Assets/Assets	M19: Total Interest Cost on Savings Deposits/Avg. Savings Dep.
M7: Net Charge-Offs/Average Loans	M12: Financial Investments/Assets	M20: Total Int. (Dividend) Cost on Shares/Avg. Member Shares
M8: Loan Recoveries/Net Loan Charge Offs	M13: Shares and Deposits/Assets	M21: Gross Income/Assets
M9: Solvency Evaluation[4]	M14: Member Share Capital/Assets	M22: Operating Expense/Assets
	M15: Institutional Capital/Assets	M23: Provision for Loan Losses/Assets
	M16: Net Institutional Capital/Assets	M24: Net Income/Assets
Signs of Growth	**Other (Identified from Literature Review)**	**Asset Quality**
M26: Loan Growth Rate[5]	M1: Membership Description[6]	M17: Nonperforming Loans/Total Loans
M27: Investments Growth	M2: Age[7]	**Liquidity**
M28: Deposit Growth Rate	M3: NCUA Charter Type[8]	M25: Cash & Cash Equivalents/Total Shares and Deposits
M29: Capital Growth	M4: Log Total Assets	
M30: Net Capital Growth	M32: Shares/Members[9]	
M31: Member Growth	M33: Net Long-Term Assets/Assets	
M38: Total Assets Growth Rate	M34: Fee Income/Assets	
	M35: Type of Website[10]	
	M36: Members/Potential Members	
	M37: Maintained Branches	
	M39: Equity/Assets	
	M40: Net Worth/Assets	
	M41: Total Reserves/Assets	
	M42: Efficiency Ratio[11]	
	M43: Fixed Assets and OREO/Assets[12]	
	M44: Log Delinquent Loans: 1-2 Months	

[4] Total assets less distributions of borrowings plus subordinated debt included in net worth less accounts payable and other liabilities less appropriation for non-conforming investments less accrued dividends payable on shares as a percent of total shares.
[5] The annualized change in loans and leases calculated as current period loans and leases less prior period loans and leases as a percent of prior period loans and leases.
[6] Dummy variable, multiple common bonds designated as 1, others common bond types designated as 0.
[7] The length of establishment of the Credit Unions as of October 2020 (in years).
[8] Dummy variable, federally owned designated as 1, state-owned designated as 0.
[9] Number of share accounts as a multiple of members.
[10] The type that best describes your website: informational designated as 1, interactive designated as 2, transactional designated as 3, no website designated as 0.
[11] Noninterest expense as a percent of the sum of net interest income after provisions, noninterest income, and provisions for loan and lease losses.
[12] Fixed assets plus other real estate owned assets as a percent of assets.

Appendix 3. Descriptive Statistics (Minimum, Median, Mean, Maximum, and Standard Deviation (SD)) of A. Failed; B. Survived)

	Variables	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**	**M11**	**M12**	**M13**	**M14**
	Min	1: 22.8%	17.0	1: 69.3%	1.4	0.1	0.0	-123.1	-300.0	100.3	-3.9	0.7	0.0	21.6	0.0
	Median		60.0		3.3	2.5	0.9	1.4	5.4	114.6	43.8	44.1	8.0	85.9	0.8
A	Mean	0: 77.2%	57.19	0: 30.7%	3.4	6.7	1.0	1.9	-0.1	124.3	43.5	43.9	18.3	81.6	1.5
	Max		98.0		6.2	166.7	5.5	94.1	350.0	357.1	97.8	97.8	99.5	99.2	41.0
	SD		17.63		0.9	17.1	0.9	23.3	97.2	31.4	27.3	26.9	26.6	13.3	4.7
	Min	1: 24.5%	17.0	1: 53.4%	2.1	0.0	0.0	-36.0	-390.9	102.9	1.5	0.3	0.0	22.8	0.0
	Median		67.0		4.8	0.7	1.4	0.3	0.2	112.6	60.9	16.6	0.1	87.5	0.2
B	Mean	0: 75.5%	68.2	0: 46.6%	4.9	1.0	1.4	0.5	1.4	114.1	59.1	19.1	1.0	86.4	0.4
	Max		120.0		7.4	44.4	4.7	61.2	350.0	195.0	100.0	81.4	160.5	96.7	3.4
	SD		14.8		0.7	1.2	0.8	2.2	23.7	6.2	17.5	11.3	5.4	4.8	0.3

	Variables	**M15**	**M16**	**M17**	**M18**	**M19**	**M20**	**M21**	**M22**	**M23**	**M24**	**M25**	**M26**	**M27**	**M28**	**M29**
	Min	2.9	1.3	0.0	-0.1	-7.4	-7.4	-1.2	-7.8	-9.4	-10.2	-3.4	-98.3	-400.0	-394.2	-183.8
	Median	13.0	11.9	2.3	1.9	1.2	1.1	5.8	4.2	0.6	-0.1	16.4	-3.3	-5.8	-3.1	-0.4
A	Mean	15.4	14.2	5.9	2.3	1.8	1.7	6.0	4.6	1.8	-0.3	24.2	-1.1	-38.9	-3.0	-6.7
	Max	90.2	80.4	76.9	33.3	78.8	78.8	24.7	21.4	59.3	4.0	182.1	263.2	400.0	293.7	42.1
	SD	9.9	9.5	9.9	2.4	5.1	5.1	2.7	2.9	5.7	1.3	24.6	34.9	149.3	38.7	28.7
	Min	3.0	1.3	0.1	0.8	-7.4	-7.4	-1.2	-7.8	-9.4	-8.3	-3.4	-98.3	-396.2	-394.2	-183.8
	Median	13.8	12.8	4.0	2.1	1.0	1.0	5.5	4.1	1.0	-0.1	17.3	-8.6	-19.2	-6.0	-3.3
B	Mean	18.7	17.1	9.7	3.7	2.3	2.3	5.8	4.9	3.1	-0.5	30.9	-6.1	-64.8	-9.3	-12.2
	Max	90.2	80.4	76.9	100.0	78.8	78.8	24.7	21.4	59.3	4.0	182.1	215.4	311.3	293.7	40.6
	SD	14.1	13.6	14.6	9.5	7.6	7.6	3.6	4.0	9.0	1.8	33.5	43.5	144.7	59.7	37.7

	Variables	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40	M41	M42	M43	M44
	Min	-319.4	-217.1	0.3	-0.3	-1.4	0: 58.3%	0.3	0.0	-59.2	-0.3	1.3	0.0	-200.0	0.0	0.0
	Median	1.7	-2.1	1.2	2.8	0.7	1: 18.1%	41.0	1.0	-0.6	11.9	11.9	0.7	88.0	0.8	1.0
A	Mean	-7.9	-5.1	1.3	9.1	1.1	2: 3.9%	43.3	1.5	-0.3	14.1	14.2	1.2	90.2	1.8	1.0
	Max	66.8	164.3	2.2	108.8	42.3	3: 19.7%	100.0	32.0	51.8	78.4	78.4	13.0	400.0	16.8	2.9
	SD	37.7	30.1	0.3	13.1	2.6		27.9	2.1	12.1	9.5	9.4	1.5	53.0	2.4	0.7
	Min	-319.4	-217.1	0.3	-0.3	-0.9	0: 9.3%	0.1	0.0	-38.5	0.3	2.8	0.0	-289.8	0.0	0.0
	Median	-2.0	-3.2	1.0	1.5	0.8	1: 3.8%	18.0	2.0	4.4	11.0	11.1	0.4	80.1	1.9	1.5
B	Mean	-18.5	-8.0	1.2	8.8	1.1	2: 2.6%	27.7	4.6	5.4	11.9	12.0	0.5	80.9	2.1	1.6
	Max	66.7	100.5	2.2	108.8	16.7	3: 84.3%	100.0	131.0	348.4	49.1	48.7	8.7	370.0	17.4	4.1
	SD	54.4	36.0	0.3	16.8	1.8		26.3	7.2	9.4	4.2	4.1	0.5	18.7	1.7	0.7

Appendix 4. Random Forest

Derived from the ensemble-learning framework, RF combines multiple individual decision trees (DTs). A random bootstrap dataset should well train each DT. Due to the superiority of nonparametric and simplification, the Classification and Regression Trees (CART) are usually adopted as the DTs of RF (Liaw and Wiener, 2002). Figure 13 provides a simple diagram describing the RF classifier's structure.

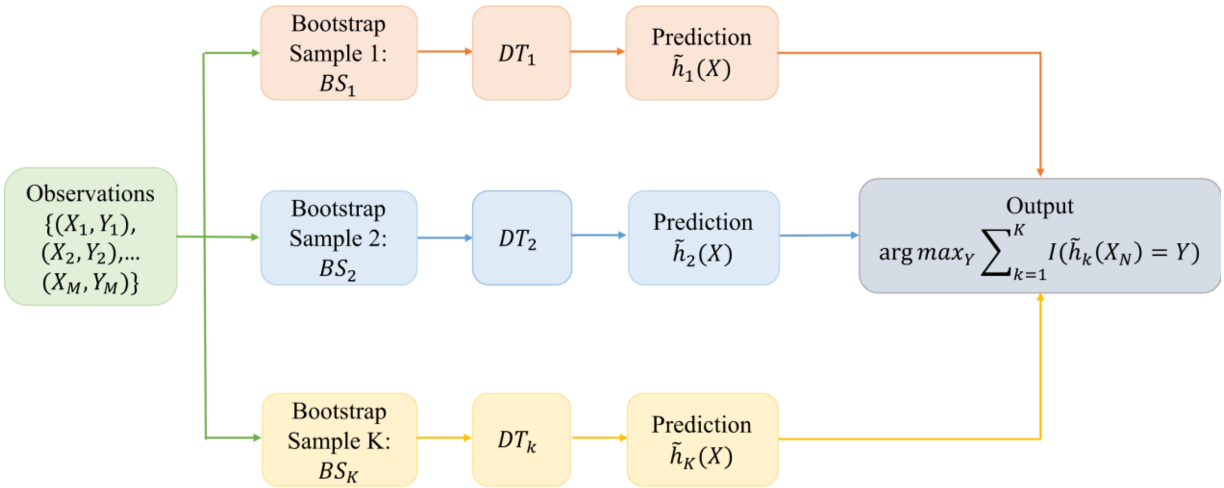


Figure 13. Structure of RF Classification Model

In terms of classification, let the training set $TS = \{(X_1, Y_1), (X_2, Y_2), \cdots, (X_m, Y_m)\}$ include m observations. $X_i = (X_{i1}, X_{i2}, \cdots, X_{i44})$ represents the input vector that has 44 features in the case of our study. Y_i refers to the output vector. The procedure used to conduct an RF classifier is described as follows (Liu et al., 2021):

1. **Procedure** RF model training
2. For $k = 1$ to K (K is the number of DTs):
3. Formulate a bootstrap sample BS_k with M size from TS (Training Set);
4. Use the formulated BS_k to fit a tree DT_k:
   A. Split a node with all observations from BS_k.
   B. Repeat the process below on each unsplit node recursively:
      a). Select n features randomly from 44 available features ($n \leq 44$).
      b). Discover the split way with the best impurity from all possible splits of n features.
      c). Split the node into two sub-nodes based on the obtained split way from step b).
5. Formulate the RF by combining all trained DT learners$h_k(\cdot)$.
6. **end procedure**
7. RF classification
8. For a new observation X_N, the outcome of classification $RF(X_N)$ is obtained by:

$RF(X_N) = arg\, max_Y \sum_{k=1}^{K} I\left(\tilde{h}_k(X_N) = Y\right),$

The prediction result, which adopts X_N as the input of the kth DT., $I(\cdot)$ is a zero-one judgement with the rule of $I\left(\tilde{h}_k(X_N) = Y\right) = 1$. $arg\,max_Y$ Outputs the class with the highest count number from all DTs.

9. **end procedure**

The primary target of the RF training phase is to build numerous de-correlated DTs. Each DT is trained from its specific bootstrap samples, thereby increasing the diversity of DT. Additionally, this method effectively reduces the correlations among DTs. In this context, the DTs within RF can be well-trained without pruning, resulting in a relatively small computational burden. Meanwhile, the noise immunity of RF is also improved by averaging various de-correlated DTs. Moreover, for each DT, some training data is reused in the bootstrap sample due to the bagging strategy, thus causing some other observations not to be chosen to fit this DT. These observations are called out-of-bag (OOB) samples and generally account for nearly 30% of all training datasets. Therefore, these OOB samples not used in the RF training process can be used to assess the classification effect of each DT after training. In this way, the RF can realise unbiased estimations without needing an external dataset. (Further explanation: Some machine-learning methods may need another set of data to realise validation in the training process, but sometimes there will be a significant difference between the two sets of data, which may lead to biased estimations. However, the RF with OOB can realise validation in the training process without needing another dataset as it uses the same dataset for training and validation. So, the RF can realise unbiased estimations without needing an external dataset). Furthermore, according to this built-in cross-validation behaviour, OOB samples can relieve the overfitting issue, thereby improving RF generalisation.

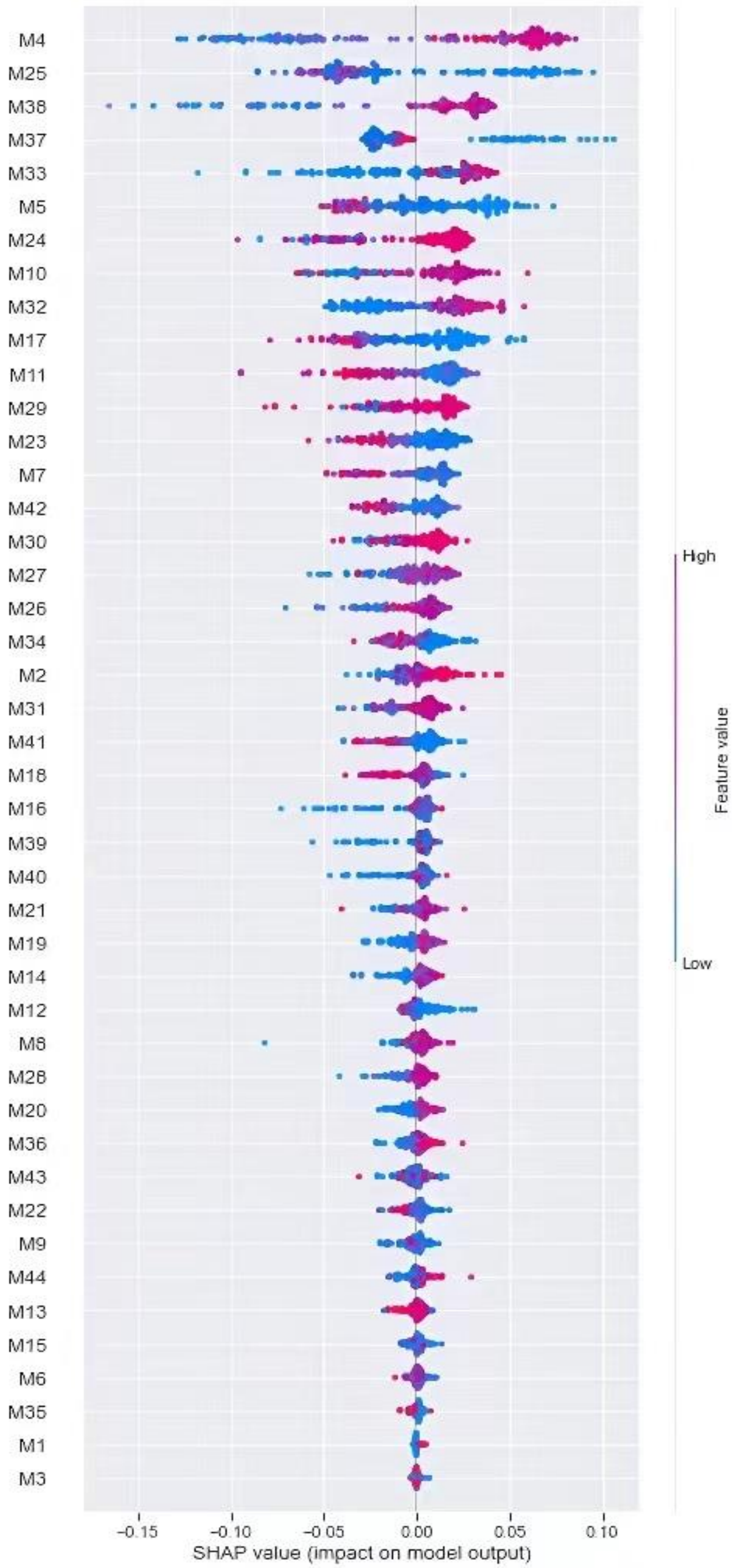
Appendix 5. Comparisons with Other Approaches

Test	**Name**	**Training Set**	**False Negatives Rate**	**False Positive Rate**	**Test Set**	**False Negatives Rate**	**False Positive Rate**
Failed/Survived with 44 variables	Random Forest (OOB)	89.1%	10.9%	10.9%	97.9%	2.0%	8.8%
	Random Forest (Cross - validation)	91.5%	6.8%	10.2%			
	Boosted tree	89.1%	4.5%	17.3%	94.9%	5.0%	10.4%
	Decision tree (Fine tree)	78.6%	19.1%	23.6%	81.8%	19.0%	25.0%
	SVM (Medium Gaussian)	75.9%	13.6%	34.5%	84.6%	15.0%	31.9%
	Logistic regression	75.7%	31.6%	17.1%	84.6%	15.5%	9.8%
	Linear discriminant	78.9%	28.9%	13.2%	83.8%	16.2%	17.8%
	Naïve Bayes (Kernel)	83.6%	14.5%	18.4%	84.1%	15.9%	17.8%
	KNN (Cosine)	85.5%	13.2%	15.8%	81.3%	18.7%	19.6%
Failed/Survived with 5 variables	Random Forest (OOB)	83.6%	14.5%	18.2%	92.2%	7.8%	17.6%

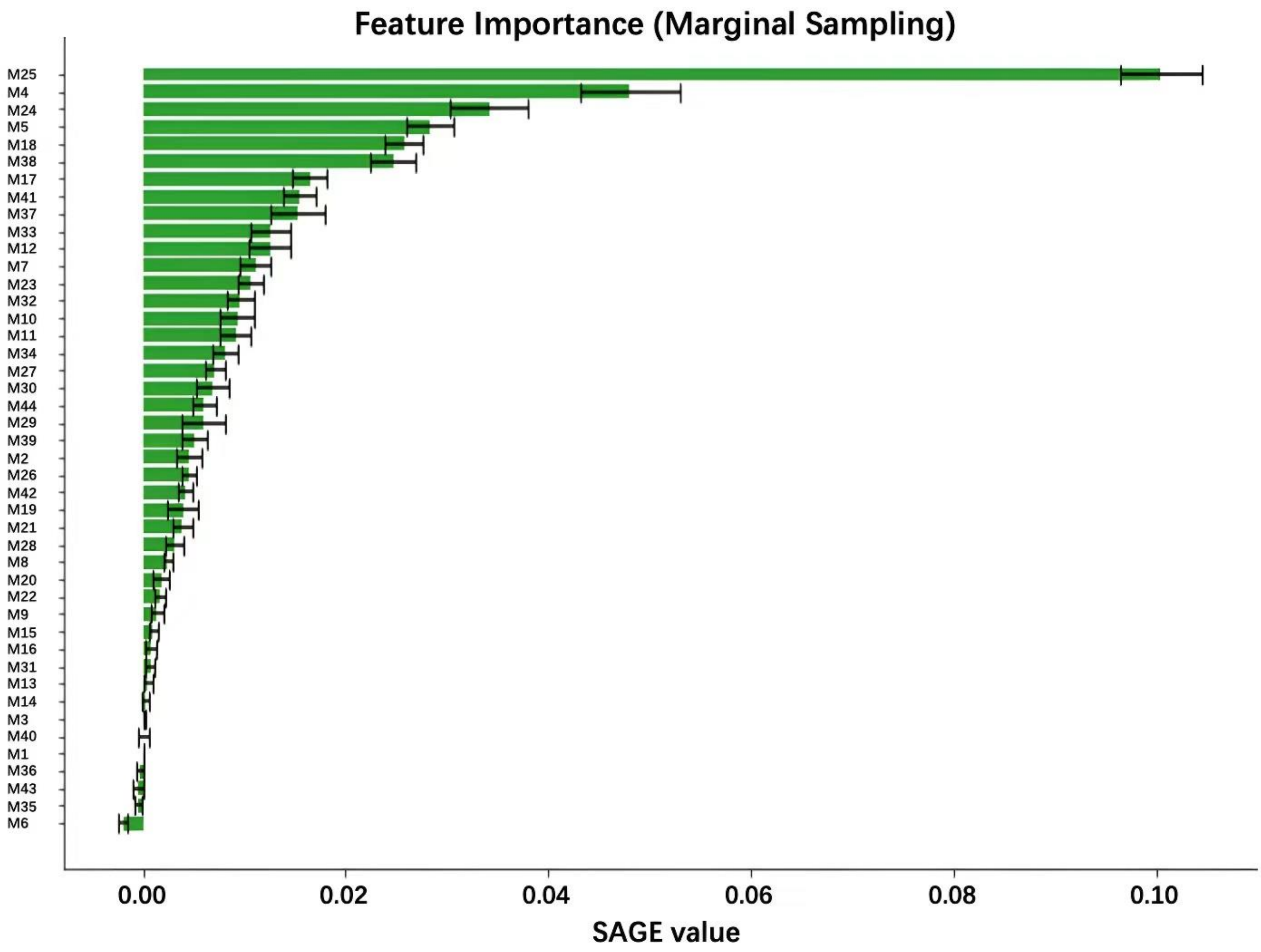
Appendix 6. Features Used to Build Cox PH Model

1	M25	Cash & Cash Equivalents/Total Shares and Deposits
2	M4	Log Total Assets
3	M24	Net Income/Assets
4	M18	Interest Income: Interest on Loans/Average Loans and Leases
5	M38	Total Assets Growth Rate
6	M17	Nonperforming Loans/Total Loans
7	M41	Total Reserves/Assets
9	M12	Investments: Securities and Other Investments/Total Assets
10	M7	Net Charge Offs/Average Loans
11	M23	Provision for Loan Losses/Average Assets
13	M10	Net Loans and Leases/Assets
14	M34	Fee Income/Average Assets
15	M27	Investments Growth
16	M30	Net Capital Growth
17	M39	Equity/Assets
18	M2	Age
19	M26	Loan Growth Rate
20	M42	Efficiency Ratio
21	M19	Cost of Shares and Deposits
22	M28	Deposit Growth Rate
23	M8	Loan Recoveries/Net Loan Charge Offs
24	M31	Member Growth
26	M3	Charter Type
28	M36	Members/Potential Members
29	M43	Fixed Assets and OREO/Assets

Appendix 7. SHAP Summary Plot



Appendix 8. SAGE Feature Importance Plot



The Centre for Responsible Banking & Finance
CRBF Working Paper Series
School of Management, University of St Andrews
The Gateway, North Haugh,
St Andrews, Fife,
KY16 9RJ.
Scotland, United Kingdom
https://crbf.wp.st-andrews.ac.uk/

Recent CRBF Working papers published in this Series

First Quarter | 2024

24-001 **José M. Liñares-Zegarra, John O.S. Wilson:** Access to Finance for UK Social Enterprises.

Fourth Quarter | 2023

23-020 **José M. Liñares-Zegarra, John O.S. Wilson:** The Impact of Brexit on the Internationalisation, Innovation and Turnover of UK SMEs: Implications for the UK's Industrial Strategy and the 'Levelling Up' Agenda.

23-019 **Oğuzhan Çepni, Ahmet Şensoy, Muhammed Hasan Yılmaz:** Climate Change Exposure and Cost of Equity.

23-018 **Ahmed Aboud, Niccolò Nirino, Giota Papadimitri:** Tax Avoidance and Sustainable Development Goal Disclosure: A Cross Country Analysis.

23-017 **Alfred Duncan, Charles Nolan:** Adam Smith and the Bankers: Retrospect and Prospect.

23-016 **Ivan Lim, Duc Duy Nguyen, Linh Nguyen, John O.S. Wilson:** Proximity to Bank Headquarters and Branch Efficiency: Evidence from Mortgage Lending.

Third Quarter | 2023

23-015 **Jeremy Burke, J. Michael Collins, Carly Urban:** Does State-mandated Financial Education Affect Financial Well-being?

23-014 **Isaac Swensen and Carly Urban:** Literature Review on the Effect of Physical and Mental Health on Financial Well-Being.

University of St Andrews
Scotland's first university
600 YEARS
1413 – 2013