Received: 6 October 2022 | Revised: 4 March 2023 | Accepted: 13 March 2023

DOI: 10.1002/for.2977

RESEARCH ARTICLE

WILEY

Forecasting nonperforming loans using machine learning

Mohammad Abdullah[1] | **Mohammad Ashraful Ferdous Chowdhury**[2] | **Ajim Uddin**[3] | **Syed Moudud-Ul-Huq**[4]

[1]Faculty of Business and Management, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

[2]Interdisciplinary Research Center (IRC) for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia

[3]Martin Tuchman School of Management, New Jersey Institute of Technology, Newark, New Jersey, USA

[4]Department of Business Administration, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh

Correspondence
Mohammad Ashraful Ferdous Chowdhury, Interdisciplinary Research Center (IRC) for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia.
Email: ashraful_ferdous@yahoo.com; mohammad.chowdhury@kfupm.edu.sa

Abstract

Nonperforming loans play a critical role in financial institutions' overall performance and can be controlled by forecasting the probable nonperforming loans. This paper employs a series of machine learning techniques to forecast bank nonperforming loans on emerging countries' financial institutions. Using quarterly cross-sectional data of 322 banks from 15 emerging countries, this study finds that advanced machine learning-based models outperform simple linear techniques in forecasting bank nonperforming loans. Among all 14 linear and nonlinear models, the random forest model outperforms other models. It achieves a 76.10% accuracy in forecasting nonperforming loans. The result is robust in different performance metrics. The variable importance analysis reveals that bank diversification is the most critical determinant for future nonperforming loans of a bank. Additionally, this study revealed that macroeconomic factors are less prominent in predicting nonperforming loans compared with bank-specific factors.

KEYWORDS

bagged CART, banking, forecasting, machine learning, nonperforming loans (NPLs)

1 | INTRODUCTION

Nonperforming loans (hereafter, NPLs) are a significant barrier to banks' financial stability for both developed and emerging nations (Lee et al., 2022). The 2007–2008 global financial crisis is a prime example of the adverse effect of NPLs on financial institutions. Numerous established financial institutions failed due to a high percentage of nonperforming mortgages and derivatives on their balance sheet (Manz, 2019). Since the global financial crisis, banks all over the world have experienced a rising share of NPLs in their loan portfolio. A high level of NPLs deteriorates banks' asset quality and signifies a greater risk to banks' liquidity. Studies also suggest a strong linkage among NPLs, macrofinancial shocks, and banking sector vulnerability (Karadima & Louri, 2020).

The problem of NPLs is more severe for emerging countries. Emerging countries characterized by lenient macroprudential regulation often prioritize growth over stability and overlook the lending behavior of their financial institution. In addition, due to poor governance and control, the financial institutions of these countries also prioritize profitability over stability by taking high-risk initiatives, for example, issuing a high number of questionable loans (Arham et al., 2020). As a result, in recent years, banks in emerging countries have experienced a significant surge in their NPLs. Although depending on the economic development and institutional quality, the level and recovery of NPL vary from country to country, it remains a significant source of risk for the financial stability of these countries. Hence, reducing NPLs is necessary to ensure a sound banking system and promote the overall financial stability of emerging countries. Designing effective policy alternatives to minimize bank NPLs require accurately forecasting future NPLs and properly identifying the factors that lead to NPLs (Cepni, Demirer, et al., 2022; Hajja, 2020; Plakandaras et al., 2015).

This study uses advanced machine learning models to forecast NPLs of emerging nations' banks. We analyzed a series of linear and nonlinear models to test whether state-of-the-art machine learning-based models accurately forecast the level of banks' future NPLs. In addition, we evaluated 15 bank-specific, macroeconomic, and global determinants of NPLs in the machine learning spectrum to identify their contribution to driving NPLs. In recent years, machine learning-based models have shown remarkable success in forecasting. Due to their capacity to uncover embedded trends in financial data, machine learning models have also been widely used in various financial applications, such as asset pricing, bankruptcy prediction, high-frequency trading, and inflation rate forecasting (Abdullah, 2021; Akyildirim et al., 2021; Cepni, Gupta, & Onay, 2022; Gu et al., 2021; Liu et al., 2022; Sulong et al., 2022; Tang et al., 2020; Uddin et al., 2021; Zhang et al., 2022). Following their success, in this paper, we exploit the powerful nonlinear modeling capacity of machine learning to improve the forecasting accuracy of NPLs. Alongside the conventional econometric models, we tested 14 linear and nonlinear algorithms to identify the best machine learning model for forecasting NPLs.

Our study uncovers several interesting findings by analyzing 322 banks of 15 emerging countries from 2001Q1 to 2022Q3. First, the machine learning-based model is superior in forecasting bank NPLs in emerging countries. Compared with linear models' machine learning-based models reduce prediction error by 16% to 50%. The results are robust across multiple performance metrics. Our findings of machine learning-based models' superiority are consistent with the findings of related studies that applied machine learning-based predictive models in other areas of finance. For example, Gu et al. (2020), Gu et al. (2021), Uddin et al. (2022), Cepni, Gupta, Pienaar, and Pierdzioch (2022), and Bonato et al. (2023) show that machine learning-based models can outperform traditional models in financial time series forecasting.

Second, among all studied models, we find that the random forest is the most appropriate model for forecasting NPLs. It achieves the highest prediction accuracy, has low bias–variance tradeoff, and indicates little to no overfitting problem. The result is consistent across multiple robustness parameters. The finding also sheds light on the overfitting issue of machine learning algorithms and signifies the importance of robustness tests in identifying the appropriate machine learning model for real-world applications.

Third, the variable importance test shows that compared with macroeconomic factors, bank-specific factors are more influential in determining bank NPLs. Bank diversification, credit quality, inefficiency, and solvency ratio are some of the most prominent factors that affect banks' future NPLs. In addition, based on the result of this study, we developed a web application, "Non-Performing Loans Amount Prediction by Machine Learning (NPLML)."[1] The application is freely available for public use. Policymakers and bank authorities can use the application to predict the future NPLs of banks and understand how each factor affects the level of bank NPLs.

Our study and the associated findings contribute to the existing literature in the following ways. First, our study provides a detailed analysis of a machine learning-based model on banks' NPL forecasting. Although the literature related to the determinant of banks NPL is abundant (Ghosh, 2015), the application of the machine learning model in NPL is novel. Prior studies mainly focused on finding the determinates of NPLs, such as bank-specific factors (Karadima & Louri, 2020), macroeconomic factors (Manz, 2019; Vouldis & Louzis, 2018), and loan-specific factors (Kuzucu & Kuzucu, 2019). Therefore, this study fills this gap. This study adds the first evidence of the successful application of machine learning for NPL forecasting in the growing body of big data analytics in finance literature. Second, the findings related to variable importance provide a deeper understanding of the underlying determinants of NPLs. As a result, it can help policymakers design appropriate policies to reduce bank NPLs. Finally, the machine learning web application NPLML, designed based on the result of this study, will allow emerging nations' bank authorities to access timely information for policy formulation without developing new machine learning models.

The reminder this paper is organized as follows: Section 2 reviews the related literature on NPLs. Section 3 describes the methodology, data, and machine learning algorithms. Section 4 presents the empirical findings of the study. Section 5 provides a detailed discussion related to the findings and associated policy recommendations. Finally, Section 6 concludes the paper with future research directions.

2 | REVIEW OF LITERATURE

This study is related to two groups of literature: determinants of bank NPLs and the application of machine learning models in forecasting financial time series. The first stream, bank NPLs, and their determinants help us identify the input and output of the machine learning models we applied based on the second stream of the literature.

Several studies have been carried out to identify the influential factors of NPLs. This includes macroeconomic factors (Manz, 2019; Vouldis & Louzis, 2018), bank-specific factors (Dimitrios et al., 2016; Karadima & Louri, 2020; Partovi & Matousek, 2019), and loan-specific factors (Bashir et al., 2017; Kılıç Depren & Kartal, 2020; Kuzucu & Kuzucu, 2019; Vouldis & Louzis, 2018). These studies cover a wide range of theories and vary in their findings and contributions (Manz, 2019).

To determine the impact of macroeconomic factors on NPL, Radivojević et al. (2019) applied the dynamic generalized method of moments (GMM) model to all Latin American Banks from 2000 to 2015. He finds that GDP has a significant impact on NPLs in emerging countries. They also argued that the inflation rate does not have any statistically significant relationship with NPL. Ghosh (2015) analyzed the effects of regional economic and banking industry-specific factors on NPLs using fixed effect and dynamic GMM models on commercial banks from 1984 to 2013. They find that GDP and unemployment rates significantly impact NPLs and suggest that improving the economic condition would reduce the NPLs. Bougatef (2015) assessed the impact of corruption on NPLs in 22 emerging countries and found a robust positive association between corruption and NPLs. He suggests that imposing strong collateral and bankruptcy law would reduce NPLs.

In relation to bank-specific determinants of NPLs, Berger and DeYoung (1997) first draw the links between bank-specific variables and bank nonperformance loans. Analyzing the US commercial banks from 1985 to 1994, they find that a decrease in cost inefficiency increases banks' NPL. A similar conclusion is reached by Podpiera and Weill (2008) by studying the Czech banking industry from 1994 to 2005. Breuer (2006) examines a wide range of intuitional variables and their influence on NPLs. The author finds that corruption and deposit insurance raise NPLs, whereas disclosure related to off-balance sheet items, risk management practices, and imposition of sanctions on bank management and directors help reduce NPLs. By analyzing banks from five European counties, Karadima and Louri (2020) examined the impact of bank market power on NPLs by adopting a quantile regression approach on 646 banks of 19 European banks from 2005 to 2017. They find that profit margin and bank competition exert a significant positive effect on NPLs. Hajja (2020) uses a dynamic panel and Vector autoregressive model on 19 Malaysian sample banks from 2002 to 2011 to analyze the effects of bank capital on NPLs. They find strong evidence of the threshold effect of capital. They suggest that NPLs increase with bank capital until it reaches the threshold for moral hazard effects. However, after reaching the threshold, NPLs decreased due to the regulatory effect. For a detailed review of the literature on bank NPLs and their determinants, interested readers can refer to Manz (2019).

The second stream of literature related to this study is the application of machine learning approaches for forecasting financial time series. The most prominent use of machine learning methods in finance is forecasting asset prices. Gu et al. (2020) compare multiple linear and nonlinear forecasting models to predict asset return on the US equity and show that the machine learning-based deep learning model performs best. Uddin et al. (2021) and Gu et al. (2021) use machine learning-based autoencoder to build asset pricing models. Their studies suggest that machine learning-based models significantly outperform traditional asset pricing models. Cepni, Gupta, Pienaar, and Pierdzioch (2022) examined the ability of investor sentiment to forecast housing returns in China. He documented that machine learning algorithms are adaptable enough to capture systemic reform and time-varying information in a set of predictors to forecast housing returns. Akyildirim et al. (2021) used high-frequency intraday data to forecast Bitcoin price movement and showed machine learning algorithms outperform benchmark models such as ARIMA and random walk. Several other studies, that is, Bonato et al. (2023), Cepni, Gupta, Pienaar, and Pierdzioch (2022), and Cepni et al. (2019), also documented the implication of machine learning models in financial and macroeconomic time series forecasting. Furthermore, studies also show that portfolios developed based on machine learning models, including random forests, gradient-boosted trees, and deep neural networks, could earn positive alpha (Pan et al., 2017; Uddin et al., 2021, 2022).

In recent years, machine learning methods have also been used to study the banking sector. Credit scoring and identifying customer default probability are among the most popular uses of machine learning techniques.

Studies suggest that machine learning-based models benefit in bank risk management, including liquidity risk and operational risk (Tavana et al., 2018), fraud detection in financial system (Song et al., 2014), and loan delinquencies (Giannopoulos & Aggelopoulos, 2019).

Overall, based on the determinant identified in the existing literature, we can develop forecasting models for NPLs. Existing literature tried to forecast NPLs using the traditional econometrics approach, for example, VAR (Hajja, 2020). However, machine learning application in other areas of finance shows that advanced machine learning techniques can significantly improve prediction accuracy in financial time series. Therefore, this paper fills the research gap between these two areas of the literature and develops a forecasting methodology for bank NPLs using advanced machine learning approaches.

3 | METHODOLOGY

The forecasting framework of this study is demonstrated in Figure 1. Following standard practice in machine learning, we first start with data acquisition and feature selection, which is followed by data cleansing and dividing the cleaned data into train-test subsamples. In the analysis stage, we train all our machine learning models in the training sample and test their performance in the test sample. Finally, we select the best model based on the performance metrics.

3.1 | Data collection and feature engineering

The first step of machine learning model development is to select features. Based on the previous literature discussed in Section 2, we select 15 features for NPL prediction. The feature list, along with the data sources, is presented in Table 1. The next step of machine learning model development is data collection. As this study aims to study emerging market samples, we select all available banks from 16 emerging counties. This includes five BRICS countries and NEXT 11 (N11) countries. However, as the Iranian bank's data are not available at Datastream, we drop Iran from our sample. The final sample includes 322 banks from 15 emerging countries. These banks are selected based on the data availability of bank-specific variables. The bank-specific variables of these banks and country-specific macroeconomic variables for all the selected emerging countries are collected from Datastream. Additionally, we select global factors such as implied volatility, dollar index return, US Treasury yield, S&P 500 return, and economic policy uncertainty to proxy the shock in global financial conditions. The rationale for considering global variables is that emerging countries often use external funding to supply credit in local markets due to their low saving rates. As a result, the lending rate of emerging countries is significantly influenced by global factors (Ashraf & Shen, 2019). We considered quarterly data from the first quarter of 2001 to

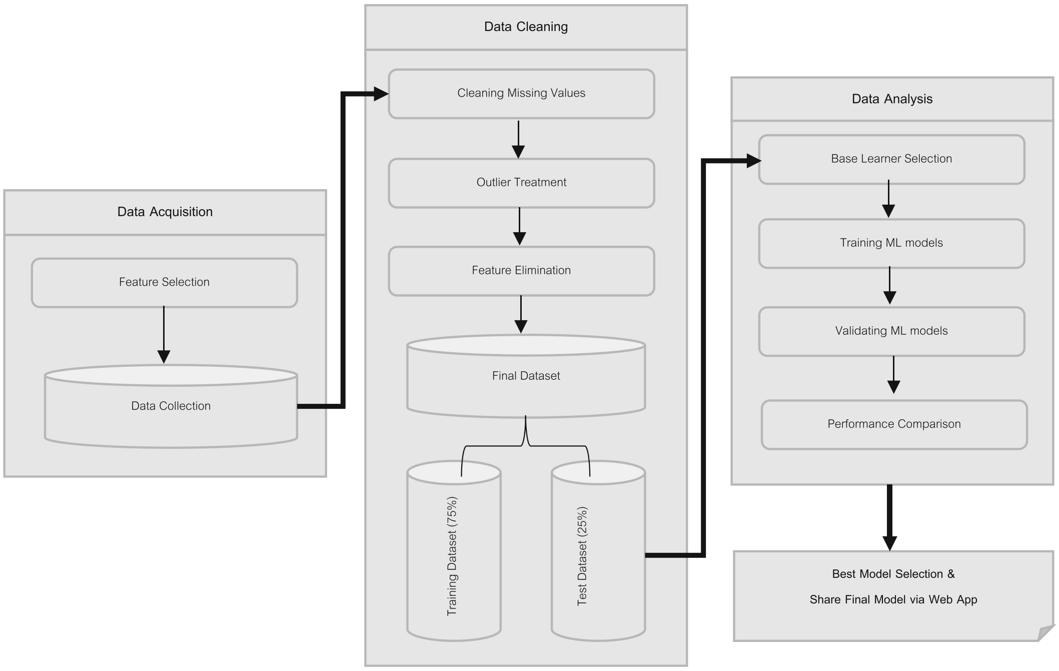

FIGURE 1 Forecasting framework of the study.

TABLE 1 Feature selection.

Variables	Description	VIF	VarImp	Data Source
Outcome variable				
NPL	NPL ratio = nonperforming loans/total loans	NA	NA	Datastream
Bank-specific variable				
LDR	Loans to deposit ratio = total loans/total deposits	1.192	56.604	Datastream
ROA	Return on assets	2.242	35.623	Datastream
LNTA	Bank size = log total asset	1.371	166.837	Datastream
SR	Solvency ratio = owned capital/total asset	1.525	47.686	Datastream
INEFF	Inefficiency = operating expenses/operating income	2.046	45.734	Datastream
DIVER	Diversification = noninterest income/total income	1.083	167.42	Datastream
CQ	Credit quality: loans loss provision to total loans	1.121	61.282	Datastream
Age	Age of the bank	1.065	98.583	Datastream
Macroeconomic variables				
INF	Quarterly inflation rate	1.220	138.341	Datastream
GDP	Quarterly GDP growth rate	1.263	76.228	Datastream
Global factors				
VIX	Quarterly changes in implied volatility	2.105	38.778	Datastream
DXY	Quarterly dollar index return	1.209	23.011	Datastream
USTY	Quarterly changes in US Treasury yield	1.175	70.895	Datastream
SP500	S&P 500 quarterly return	2.076	36.32	Datastream
EPU	Quarterly changes in economic policy uncertainty	1.245	30.525	Baker et al. (2016)

Note: This table reports the features we use in this study as the inputs for machine learning models. VIF represents the variance inflation factor to measure multicollinearity using multiple regressor model. Variables are dropped with VIF value higher than 5. VarImp represents variable importance, calculated using the random forest algorithm.

the third quarter of 2022. This sample period also gives us enough data points on both sides of the 2007–2008 financial crisis, COVID-19, and the recent Russia–Ukraine war. Following Vithessonthi (2016), we winsorize data at one percentile level to avoid extreme values and outliers. The final sample summary is reported in Table A.1.

Table 1 also reports the variables' importance of all selected factors. We calculate the variable importance using a random forest algorithm (details of the random forest algorithm are presented in Section 3). The results of variable importance[2] are listed in Table 1. Because all variables had important factors, all variables were selected for model development. Afterward, multicollinearity is detected using the variance inflation factor (VIF).[3] The results of VIF testing are presented in Table 1. We have used all listed features as all of their VIF value is lower than the tolerable range.

The full data are split into train and test subsamples using a 75:25 split. That is, 75% of the data is used to train the model, and 25% of the data is used to test the model's performance. The final data includes a total of 14,394 bank-quarter observations. Among them, 10,795 observations are used to train the machine learning model parameters, and 3599 observations are used to test the model performance using the learned parameters from training data. Table 2 presents the descriptive statistics of all studied variables from the full dataset. The descriptive statistics for train and test data are reported in Table A.2. The descriptive statistics indicate that our main studied variable, the mean NPL, is 0.052. The standard deviation (SD) of all variables in each dataset is not too high, indicating much less volatility across the sample. Skewness results suggest all variables are positively skewed except INEFF, CQ, AGE, GDP, and SP500. Kurtosis values of all variables indicate the variable's distribution is not fat-tailed.

The correlation coefficients and distribution plots of all studied variables are illustrated in Figure 2. The correlation coefficients indicate that the variable of interest NPL is positively correlated with SR, AGE, and INF and

TABLE 2 Descriptive Statistics.

	N	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
NPL	14,394	0.052	0.052	0.006	0.203	1.630	1.844
LDR	14,394	0.752	0.229	0.372	1.249	0.359	−0.442
ROA	14,394	0.004	0.003	−0.002	0.012	0.434	−0.165
LNTA	14,394	22.814	1.960	19.422	26.770	0.363	−0.517
SR	14,394	0.105	0.047	0.050	0.230	1.203	0.859
INEFF	14,394	0.455	0.359	−0.360	1.207	−0.133	0.237
DIVER	14,394	1.397	1.654	−0.626	6.526	1.820	2.970
CQ	14,394	−0.044	0.034	−0.127	−0.005	−1.062	0.216
Age	14,394	3.372	0.633	2.110	4.425	−0.129	−0.747
INF	14,394	1.509	0.920	0.337	3.700	0.826	−0.108
GDP	14,394	1.175	0.691	−0.473	2.272	−0.696	0.048
VIX	14,394	0.034	0.324	−0.331	0.833	1.144	0.563
DXY	14,394	0.004	0.035	−0.052	0.071	0.215	−0.796
USTY	14,394	0.090	0.375	−0.381	1.292	1.768	3.559
SP500	14,394	0.024	0.062	−0.111	0.112	−0.577	−0.593
EPU	14,394	0.029	0.167	−0.266	0.376	0.260	−0.353

Note: SD represents standard deviation; *N* represents number of observations.

negatively correlated with LDR, ROA, LNTA, INEFF, CQ, GDP, and USTY. However, in the studied period, we find that NPL in not significantly correlated with DIVER, VIX, DXY, SP500, and EPU. The correlation coefficients and distribution plots from the train and test data separately are available in the appendix (Figures A.1 and A.2).

3.2 | Model selection and hyperparameter tuning

In order to forecast the NPL of cross-sections of banks, we apply 14 machine learning models and perform a comparative analysis. In this section, we discuss the specification of each machine learning algorithm. Table 3 reports the list of selected machine learning models, their hyperparameters, and base learners. Hyperparameters can influence the performance of machine learning models. In order to select the optimal parameters for each model, we perform hyperparameter tuning using 10-fold cross-validation on the training data. For each model, the best parameters are selected using a grid search on the hyperparameters reported in Table 3. In addition, to avoid overfitting, we also perform three repetitions on the training set with multiple random seeds (Gu et al., 2020; Uddin et al., 2022). We implement all the machine learning models using the Caret package from R statistical programing language.

3.2.1 | Linear models

This study applied five linear models: linear regression, linear regression with stepwise selection, partial least squares (PLSs), lasso regression, and elastic net. These linear models are used as a benchmark to evaluate the performance of machine learning models. First, linear regression models the relationship between explanatory variables and the dependent variable by fitting a linear equation to observed data (Gallagher, 2007). The second linear model we use for comparison is stepwise regression. Stepwise linear regression is a technique for regressing many variables while concurrently deleting those that are not significant. Stepwise regression essentially does numerous regressions, each time deleting the least associated variable (Yamashita et al., 2007). The third applied model in our work is PLS regression. PLS reduces the predictors to a smaller set of uncorrelated components and conducts least squares regression on these components rather than the original data. For cases when predictors are very collinear or where there are more predictors than observations, the regular least squares regression coefficients can have high standard errors (Geladi & Kowalski, 1986). In such cases, PLS regression

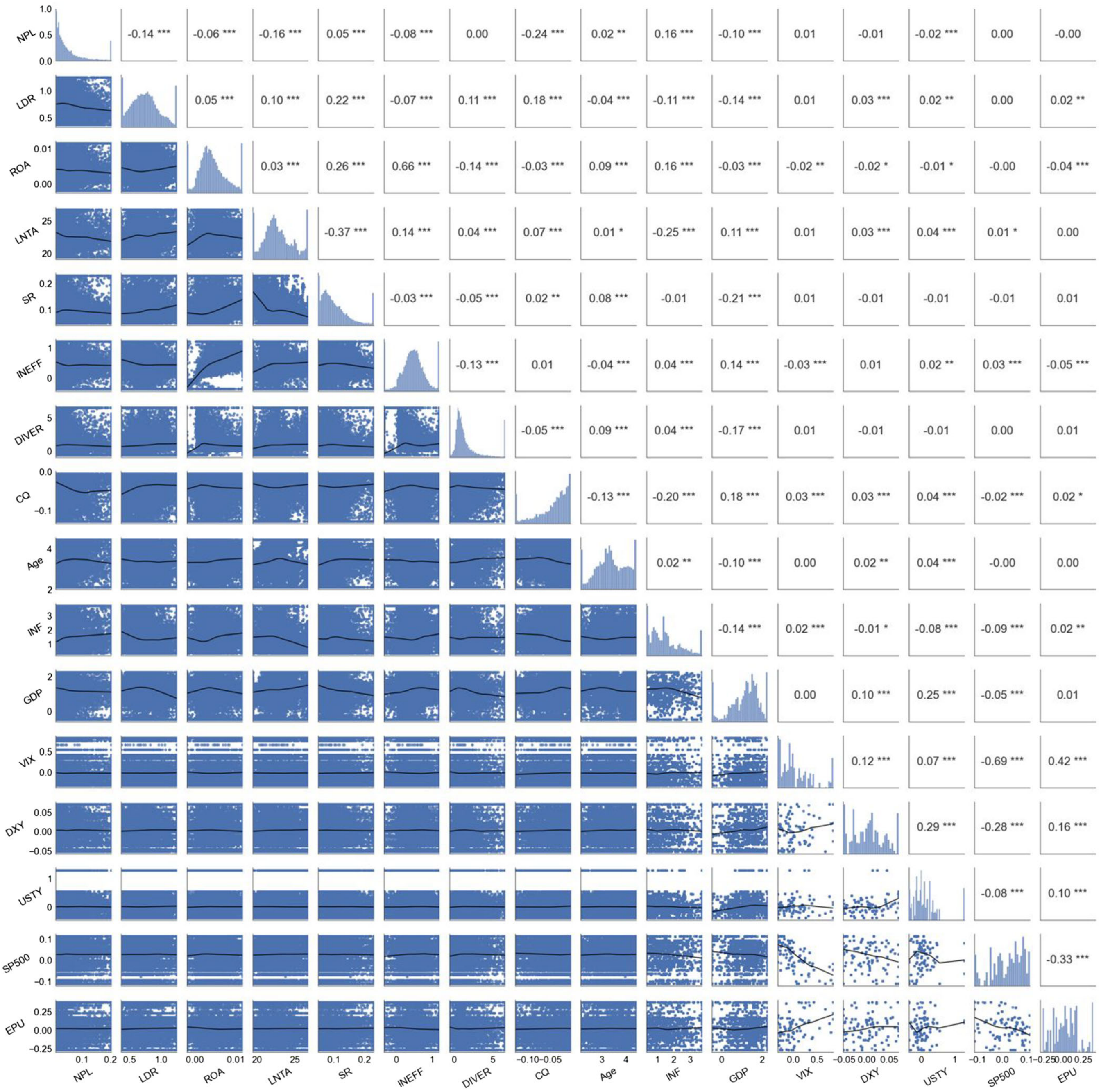

FIGURE 2 Correlation matrix and distribution plot of full sample. ***, **, and * indicate correlation is significant at 1%, 5%, and 10% significance level.

is especially effective. Unlike multiple regression, PLS does not presume that the predictors are fixed. As a result, the predictors can be assessed with greater accuracy, making PLS more resistant to measurement uncertainty (Geladi & Kowalski, 1986).

Lasso is a linear regression technique that employs shrinkage. In shrinkage operation, data values are shrunk towards a central point, such as the mean. Lasso favors models that are simple and sparse. It is ideal for models with high degrees of multicollinearity or for automating variable selection (Santosa & Symes, 1986). The final linear model we used for our analysis is elastic net. Elastic net employs penalties from both the lasso and ridge approaches to regularize regression models (Zou & Hastie, 2005). If the variables are in highly connected groups, lasso often selects one variable from each group and disregards the others (Zou & Hastie, 2005). The elastic net overcomes the lasso's shortcomings by adding a quadric part and allows incorporating *n* variables until saturation.

TABLE 3 Selected models and their base learners.

Model	Method of Caret	Grid range of tuning parameters	Final tuning parameters
Linear regression	lm	intercept = True	intercept = True
Partial least squares	pls	ncomp = 1–10	ncomp = 3
Linear regression with stepwise selection	leapSeq	nvmax = 1–10	nvmax = 4
Lasso regression	lasso	fraction = 1–5	fraction = 0.9
Elastic net	enet	fraction = 1–5, lambda = 1–100	fraction = 1, lambda = 1e-04
Bayesian ridge regression	bridge	None	None
Random forest	rf	mtry = 4–15, ntree = 250–750	mtry = 9, ntree = 650
Bagged CART	treebag	nbagg = 1–200	nbagg = 85
Bagged MARS	bagEarth	nprune = 1–50, degree = 1–3	nprune = 17, degree = 1
Quantile random forest	qrf	mtry = 1–10	mtry = 8
Boosted tree	blackboost	mstop = 1–200, maxdepth = 3–5	mstop = 150, maxdepth = 3
k-nearest neighbors	kknn	kmax = 5–15, distance = 1–4, kernel = optimal	kmax = 9, distance = 2, kernel = optimal
Support vector regression	svmLinear	C = 1–5	C = 1
Bayesian regularized neural networks	brnn	nurons = 1–5	nurons = 3

Note: C = cost, degree = polynomial degree, distance = parameter of Minkowski distance, fraction = fraction of full solution, kernel = kernel parameter of *k*-NN, kmax = maximum number of neighbors, lambda = quadratic penalty parameter, mstop = number of boosting iterations, mtry = number of randomly selected predictors, ncomp = number of components, neurons = number of neurons, nvmax = maximum number of predictors, prune = number of iterations determined by optimal Akaike information criterion (AIC) value across all iterations, and sigma = width of Gaussian distribution.

Each of these above discussed linear models can be expressed as per the following equation:

$$\frac{\arg}{\beta_0}, \frac{\min}{\beta} \| Y - X\beta \|_2^2 + \gamma\left((1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|1\right), \quad (1)$$

where $Y = (Y_1, \ldots, Y_n)$ represents the vector of the outcome variable in the training dataset, X symbolizes the $(n \times p)$ matrix of predictors, and $\left(\beta_1, \beta_2, \ldots, \beta_p\right)$ is the vector of regression beta coefficients. Moreover, different models' error terms are defined as $\lambda \geq 0$ and mixing term as $0 \leq \alpha \leq 1$, the mixing term varies with the model. In specific, $\lambda = 0$ for normal linear regression, $\lambda > 0$ and $\alpha = 1$ for lasso regression, and $\lambda \geq 0$ and $0 < \alpha < 1$ for elastic net. For lasso regression, λ is selected based on the value of the highest lambda and optimum RMSE.

3.2.2 | Bagging

Bagging algorithms select a slice of data for creating independent variables and afterward combine them by mean for outputting an aggregated independent variable (Breiman, 1996). It is also known as bootstrap aggregation. Bagging is an ensemble learning method for reducing variance in a noisy dataset. It selects a random sample of data from a training set with replacement, which means that individual data points can be chosen more than once. Bagged CART (Bauer & Kohavi, 1999) and bagged MARS (Friedman & Roosen, 1995) models are applied in this study for analysis. The following equation constructs the bagging models:

$$\widehat{f_T} = \widehat{f_1}(X) + \ldots + \widehat{f_m}(X), \quad (2)$$

where T denotes the function of the number of decision trees, m denotes the number of bootstrapping, and X number of independent variables for model training. Particularly, the bagging algorithm uses the training dataset to randomize observation by including and excluding the trees to output an average prediction.

3.2.3 | Boosting

Freund and Schapire (1997) initially developed boosting technique as a classification algorithm. However, over time boosting proved successful for regression analysis too. Boosting is a method based on the decision tree for improving prediction accuracy. In short, it is an ensemble modeling strategy that tries to create a strong classifier from a large number of weak learners. First, a model is constructed using the training data. The second model is then constructed in an attempt to address the faults in the previous model. This approach is repeated until the entire training data set is properly predicted or the maximum number of models is added. The Boosting algorithm can be defined as

$$\begin{gathered}\widehat{g}(\chi)=\widehat{\beta}^{\widehat{\delta}}\chi^{\widehat{\delta}},\\ \widehat{\beta}(j)=\sum_{i=1}^{n}X_i^{(j)}U_i/\sum_{i=1}^{n}\left(X_i^{(j)}\right)^2,\\ \widehat{\delta}=\underset{1\leq j\leq p}{\arg\min}\sum_{i=1}^{n}\left(U_i-\widehat{\beta}^{(j)}X_i^{(j)}\right)^2,\\ \widehat{\beta}^{[m]}=\widehat{\beta}^{[m-1]}+\upsilon.\widehat{\beta}^{\left[\widehat{\delta m}\right]}.\end{gathered} \tag{3}$$

Boosting method generates m different training samples and allocates weight to each sample, where $\widehat{\delta}$ indicates the index of each predictor in each m sample and $\widehat{\beta}^{[m]}$ is the update coefficients estimates. This method fits small tress to gradually increase performance when required (Bühlmann & Hothorn, 2007). In this study, we use boosted trees as a base learner.

3.2.4 | Artificial neural networks (ANNs)

ANNs are one of the most popular machine learning algorithms. ANN has multiple advantages over conventional regression-based techniques. Given adequate training data, ANN can easily model complex interactions and nonlinearity in the data. However, ANN is also often criticized for its low interpretability, high computational requirement, and overdependency on data. Researchers over the years have proposed multiple regularization techniques to overcome these challenges. In this paper, we used Bayesian regularized ANN proposed by Burden and Winkler (2008) for ANN model training. Bayesian regularized ANN incorporates Bayes' theorem to regularize the neural network by changing a nonlinear regression into a statistical problem. As a result, it is more robust and reduces the requirement for extensive cross-validation. The cost function for regularized ANN can be defined as:

$$S(w)=\beta\sum_{i=1}^{N_D}[y_i-f(X_i)]^2+\alpha\sum_{j=1}^{N_w}w_j^2, \tag{4}$$

where N_D is the number of data points and N_w is the number of weights. Given the initial hyperparameters α and β, the cost function $S(w)$ is minimized with respect to w. With Bayes' theorem, the inference for the hyperparameters α and β can be defined as

$$P(\alpha,\beta|D)=\frac{P(D|\alpha,\beta)P(\alpha,\beta)}{P(D)}, \tag{5}$$

where D is the Hessian of the data, only maximizing $P(D|\alpha,\beta)$ will be enough and the priors $P(D)$ and $P(\alpha,\beta)$ can be ignored without any significance estimation error. Therefore, in the training phase of the Bayesian regularized ANN, we need to perform two optimizations: minimize Equation (4) with respect to weights and maximize $P(D|\alpha,\beta)$ in Equation (5) until it converges. For a more detailed discussion regarding Bayesian regularized ANN, see Burden and Winkler (2008).

3.2.5 | Naive Bayes regression

Naive Bayes algorithms are referred to as probabilistic regression by adopting Bayes' hypothesis with robust prediction between the inputs. The structure of the Naive Bayes algorithm can be represented as a direct acyclic graph comprised of nodes and edges, where the nodes represent inputs, and the edges express the link between nodes and output (Bishop, 1995). In this study, we use Bayesian ridge regression as the base learner for the naïve Bayes model. Bayesian ridge is one of the most popular Bayesian regression models. It overcomes the challenge of limited and poorly distributed data by expressing linear regression using probability distributions rather than point estimates (Bishop & Tipping, 2003). Bayesian ridge regression can be defined as follows:

$$\underline{\widehat{\beta}}_{Ridge}=\left(X^TX+I\lambda\right)^{-1}X^T\underline{Y}, \tag{6}$$

here X^T is the transpose of predictors vector, I is the identity matrix, λ is the regularization penalty parameter, and $\lambda>0$, a higher value of λ, indicates a larger penalty on the coefficient.

3.2.6 | k-nearest neighbors (k-NN)

k-NN regression is a nonparametric method that approximates the relationship between independent variables and continuous outcomes by averaging data in the same neighborhood. The algorithm divides the observations space into k closest training dataset for prediction. Cross-validation can be used to pick the size of the neighborhood that minimizes the mean-squared error

(Altman, 1992). The prediction algorithm based on k-NN can be defined as

$$\widehat{f}=\frac{1}{k}\sum_{x_i \in N_k(x)} y_i, \tag{7}$$

where k is the nearest observations in the training dataset and $N_{k(x)}$ implies the nearest neighborhood of predictors x.

3.2.7 | Support vector regression (SVR)

SVR is a kernel-based machine learning algorithm. SVR develops its regression algorithm using the classification concepts from the support vector machines (SVM) (Drucker et al., 1996). In the case of regression, a tolerance margin is set to approximate the SVM. The algorithm is applicable for restraining the effects of outliers on the model fitness (Cortes & Vapnik, 1995). SVR algorithms can be constructed as follows:

$$\frac{\arg}{\beta_0},\frac{\min}{\beta}\, c\sum_{i=1}^{N} L_{\in}(y_i - f(x_i)) + \sum_{j=1}^{p} \beta_j^2, \tag{8}$$

$$f(x)=\omega_0\sum_{i=1}^{N} \omega_i K(x,x_i), \tag{9}$$

where c is the error assigned to residuals equal or larger than ϵ and L_ϵ is the cost function. Equation (8) can be used to estimate the set of unknown weights ω. In Equation (9), the kernel function is symbolized by k. Model fitness is determined by training dataset weights.

3.2.8 | Random forest

First introduced by Ho (1995), the random forest algorithm is a tree-based algorithm for regression. This algorithm selects predictors randomly using bootstrapped observations from the training dataset and selects response variables for each tree. As a result, it can provide information about the response variable's entire conditional distribution. Random forest algorithm can be constructed using Equation (2). Quantile regression forests, an extension of random forests, can be used to infer conditional quantiles. Quantile regression forests estimate conditional quantiles for high-dimensional predictor variables in a nonparametric manner. In this study, we use both random forest and quantile random forest models as the base learner.

3.3 | Measuring variable importance

Usually, machine learning models are considered black boxes due to their lack of interpretability. Unlike conventional econometric models, machine learning models typically do not provide simplified coefficients to interpret the results. However, machine learning-based models can be explained by variable importance. Variable importance measures the significance of a particular variable in the outcome of a machine learning model. It can be used to determine which features are essential for making predictions and which can be removed or replaced with a more effective feature. We use local interpretable model-agnostic explanations (LIME) to estimate the variable importance of the model (Ribeiro et al., 2016). LIME can explain machine learning models and understand the impact of each feature on the expected output by repeatedly feeding perturbed data from the original data points. Observing the corresponding outputs of each perturb data, it generates a local approximation of the underlying model (Ribeiro et al., 2016). It uses the following approximation to calculate the feature weights:

$$\xi(x)=\underset{g\in G}{\operatorname{argmin}}\, \mathcal{L}(f,g,\pi_x)+\Omega(g), \tag{10}$$

where $f(x)$ is the probability of outcome variable value. π_x denotes the proximity between z and x. Moreover, $\mathcal{L}(f,g,\pi_x)$ measures the error of g in forecasting. Here, linear models are used to measure the weights based on a search using permutations. In this study, we use each features' LIME average weights to measure their importance.

4 | EMPIRICAL RESULTS

4.1 | Model performance

In order to evaluate the performance of the models discussed in Section 3.2, we employed three performance metrics, root mean square error (RMSE), mean absolute error (MAE), and R-squared (Rsq). In machine learning literature, these three metrics are the most common and are generally used to compare model predictability (Chai & Draxler, 2014; Uddin et al., 2022). The three metrics are defined as follows:

$$RMSE=\sqrt{\frac{1}{N_\Omega}\sum_{i\in\Omega}(y-\widehat{y})^2}, \tag{11}$$

$$MAE = \frac{1}{N_\Omega}\sum_{i\in\Omega}|y-\hat{y}|, \tag{12}$$

$$Rsq = 1 - \frac{\sum_{i\in\Omega}|y-\hat{y}|}{\sum_{i\in\Omega}|y-\bar{y}|}, \tag{13}$$

where y is the observed value, $\hat{y}$ is the predicted value, $\bar{y}$ is the mean of observed value, Ω is the set of test data, and N_Ω is the total number of test observations. As these equations suggest, we evaluate each model's performance on out-of-sample, that is, test data.

Table 4 reports the result from all models in forecasting NPL. The first three columns report result from training data, and the last three columns report result from test data. In all performance metrics of test data, nonlinear machine learning-based models perform better than linear models. For example, the RMSE of linear regression, stepwise regression, PLSs, ridge regression, lasso, and elastic net is above 0.017, whereas the RMSE of random forest is 0.16. A similar trend is also visible in Rsq and MAE. Among the 14 algorithms we consider, random forest outperforms all models with the lowest RMSE and MAE and highest Rsq on the test data. The test sample Rsq of the random forest model is 76.10%, indicating a better explanatory power than any other model. According to RMSE and MAE of test data, the second-best model is quantile random forest, and the third-best model is SVR.

The prediction result from training and test data also provides important insights into machine learning-based models' overfitting problem. Most algorithms have little variation between the training and testing results except PLSs, lasso regression, elastic net, and bagged MARS. These three models have high variation between training and testing results, signifying the potential of overfitting problems. In our top three best-performing models, that is, random forest, quantile random forest, and SVR, train and test data performance are almost identical in all three performance metrics. The slight variation and outperformance in training data are expected as machine learning models' parameters are learned on training data. However, the best model cannot be selected only relying on the performance metrics (RMSE, Rsq, and MAE). Overfitting and underfitting should be checked to choose the best model. The following subsection elaborates more on the overfitting and underfitting diagnostics of the models.

4.2 | Bias–variance tradeoff

For a machine learning-based model, it is highly important to check the model's fitness. Overfitting and underfitting problems can significantly influence the real-life application of a machine learning algorithm. When the

TABLE 4 Training and testing performance of all models.

	Training Performance			Testing Performance		
Algorithms	**RMSE**	**Rsq**	**MAE**	**RMSE**	**Rsq**	**MAE**
Linear regression	0.017	0.751	0.013	0.018	0.746	0.013
Partial least squares	0.023	0.564	0.018	0.019	0.712	0.014
Linear regression with stepwise selection	0.024	0.532	0.019	0.020	0.659	0.015
Lasso regression	0.031	0.531	0.024	0.018	0.745	0.013
Elastic net	0.033	0.530	0.026	0.018	0.746	0.013
Bayesian ridge regression	0.017	0.751	0.012	0.018	0.746	0.013
Random forest	**0.016**	**0.761**	**0.013**	**0.016**	**0.776**	**0.012**
Bagged CART	0.022	0.588	0.017	0.023	0.590	0.017
Bagged MARS	0.028	0.456	0.022	0.018	0.747	0.013
Quantile random forest	0.018	0.757	0.012	0.017	0.771	0.011
Boosted tree	0.023	0.614	0.018	0.017	0.758	0.012
k-nearest neighbors	0.021	0.648	0.015	0.020	0.683	0.015
Support vector regression	0.018	0.750	0.012	0.018	0.744	0.012
Bayesian regularized neural networks	0.017	0.751	0.013	0.017	0.759	0.012

Note: This table presents the result of training and testing performance of all models. Best model is in boldface considering all diagnostics tests.

learned parameters of a model perfectly learn the training data, it can result in a lack of generalization and an overfitting problem. In contrast, when the parameters of a model cannot represent training data properly, it can induce an underfitting problem (van der Aalst et al., 2010). These problems arise from bias–variance tradeoff errors. In machine learning, the bias–variance tradeoff is the key metric of a model. It indicates how the variance of the estimated parameters can be reduced by increasing the bias (Kohavi & Wolpert, 1996).

We perform three nonparametric tests to find out the overfitting or underfitting properties of each model. These are (i) Kolmogorov–Smirnov (KS) test, which tests the null hypothesis that the two samples have the same variance (Marsaglia et al., 2003); (ii) Cucconi (CC) test, which tests the null hypothesis that the locations and scales of the two population distributions are equal (Marozzi, 2009); and (iii) Lepage (LP) test, which tests the null hypothesis that the locations or scales of the two population distributions are equal (Lepage, 1971). The results of these three nonparametric tests are presented in Table 5. Results indicate that all algorithms' null hypotheses are accepted as their *p*-values are higher than 5% except for linear regression, quantile random forest, and *k*-NN models. This indicates that linear regression, quantile random forest, and *k*-NN models are not a good fit because their training and testing sample residuals are highly volatile and differ from the distribution.

For further diagnostics tests, learning curves for all models are presented in Figure 3. Here, RMSE is used as the cost function metric. The learning curve demonstrates that linear regression, quantile random forest, and *k*-NN have high bias–variance, indicating an overfitting problem. However, the other 11 algorithms have low bias and low variance, indicating those models are a good fit. As an implication of this finding, the quantile random forest model cannot be considered the second-best model within the same hyperparameter tuning and training sample size. These results of nonparametric tests and Figure 3 indicate that random forest is the best model for predicting NPL. It can predict 76.10% of NPL, according to Rsq.

Figure 4 illustrates the actual versus forecasted values of NPL using the random forest model. In this figure, NPL values are sorted, and the first 20 banks out of 322 with higher NPL are plotted. Each subplot's title represents the Reuters instrument code for that bank. The black horizontal line in 2016Q separates the training and test period of the data, where the left side of the line is the training dataset, and the right side of the line is the test dataset. The figure indicates a similar variation in actual and forecasted values. For most banks, the predicted values of NPL almost mirror the actual values. Even for banks with high variance in NPL, for example, BBAS3, UBP, the random forest model predicts the future NPL with a very high level of accuracy. The figures also demonstrate little to no variation between training and test data accuracy. These results signify that the suggested best model random forest has high prediction accuracy with a low bias and variance.

TABLE 5 Overfitting or underfitting test results.

Algorithms	KS test	CC test	LP test
Linear regression	0.000	0.000	0.000
Partial least squares	0.613	0.280	0.362
Linear regression with stepwise selection	0.711	0.273	0.485
Lasso regression	0.599	0.920	0.810
Elastic net	0.752	0.918	0.808
Bayesian ridge regression	0.517	0.878	0.683
Random forest	0.752	0.925	0.808
Bagged CART	0.416	0.025	0.085
Bagged MARS	0.875	0.735	0.762
Quantile random forest	0.000	0.000	0.000
Boosted tree	0.033	0.000	0.000
k-nearest neighbors	0.000	0.000	0.000
Support vector regression	0.885	0.807	0.701
Bayesian regularized neural networks	0.992	0.745	0.717

Note: This table presents overfitting or underfitting test *p*-values. KS test represents Kolmogorov–Smirnov tests; CC test represents Cucconi test; and LP test represents Lepage test.

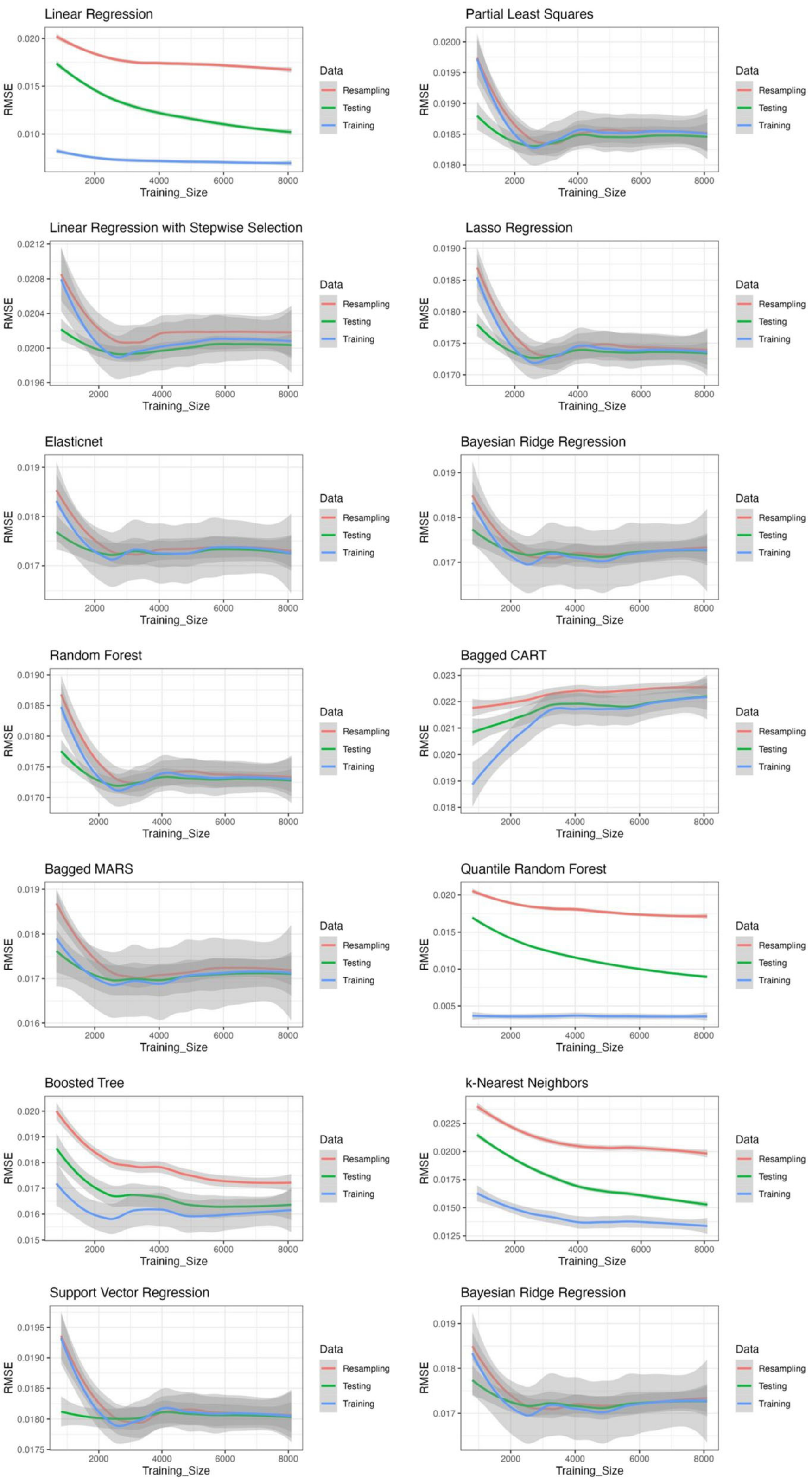

FIGURE 3 Learning curves for overfitting or underfitting specification.

4.3 | Variable importance

The results reported in the above two subsections indicate that NPL can be forecasted using machine learning techniques. However, policymakers are more concerned about the contribution of each variable in the model. Therefore, to provide interpretability of the model, we calculate variable importance using LIME and presented in Figure 5. The variables ranking among models are very consistent. As illustrated in Figure 5, size, income

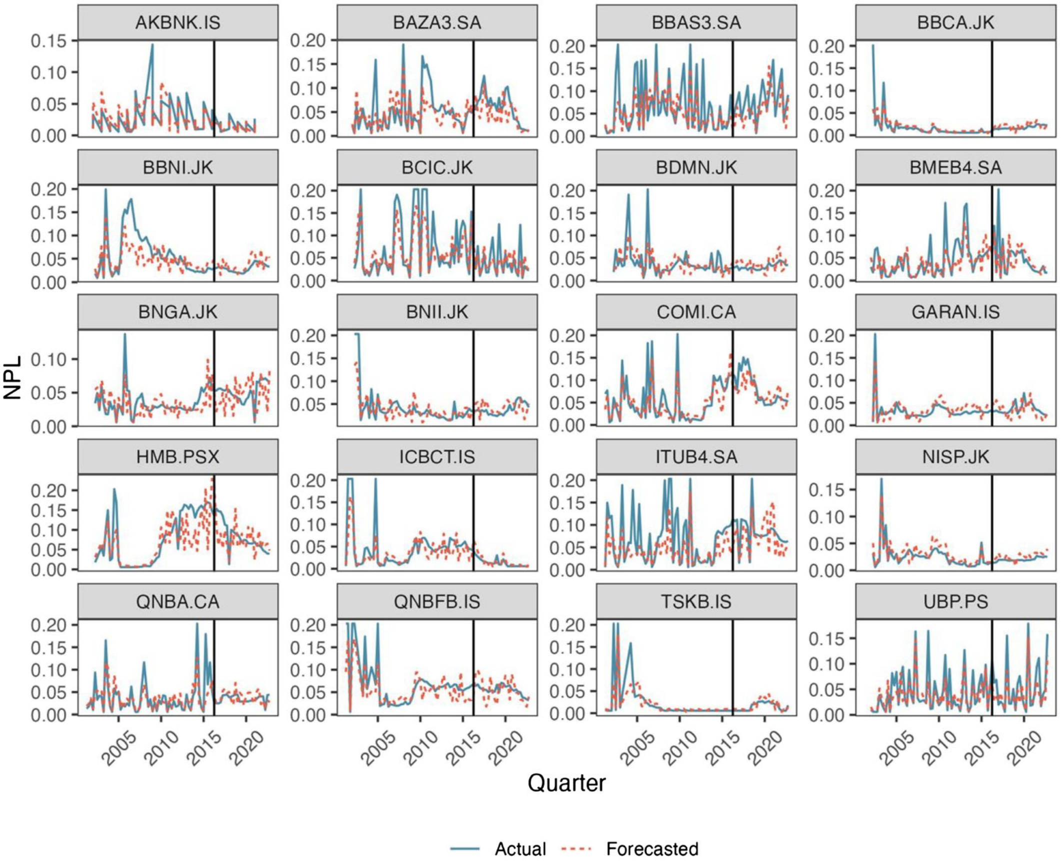

FIGURE 4 Actual versus forecasted NPL of 20 banks.

diversification, inefficiency, credit quality, and inflation are the most important determinant of a bank's NPL.

To comprehensively evaluate the best-performing model, we present the variable importance according to the random forest model in Table 6. We ranked the variables according to their importance and reported the value in different intervals trees (ntree = 550, 600, 650, 700, and 750). The final selected parameter ntree = 650 is presented in boldface. We find that DIVER has the highest variable importance weight in forecasting NPL among all variables. This indicates that the value of NPL increases along with bank diversification. Several previous studies also support this finding (Louzis et al., 2012; Riahi, 2019). INF, CQ, and INEFF are other important factors in NPL forecasting. In addition, we find that LNTA has the lowest importance among bank-specific characteristics, and USTY has the lowest importance weight among macroeconomic factors. These findings explain that DIVER contributes positively, and LNTA contributes negatively to NPL. For example, on average, one SD increase in DIVER caused an increase in our forecasting of NPL of about 0.202, whereas LNTA reduced the forecasted NPL by 0.328.

4.4 | Robustness test

To check the robustness of the results, we employ three additional tests. First, we train and test our models using two additional train-test splits, 50:50 and 90:10. Table 7 presents the result of the robustness test, where Panel A shows 50:50 train-test split results and Panel B shows 90:10 train-test split results. The finding in Panel A shows random forest is the best model considering RMSE, MAE, and Rsq. It achieved an Rsq of 76.90%. Panel B also shows similar results. The random forest model outperforms other models with 76.50% Rsq. This finding further corroborates our baseline results with a 75:25 train-test split, suggesting random forest is an appropriate model for NPL forecasting. Our conclusion is also in line with earlier studies on machine learning-based models. For example, Ozgur et al. (2021), Bonato et al. (2023), and Akyildirim et al. (2021) report better output with the random forest model in different forecasting scenarios.

Second, we estimate the cumulative sum of squares for error difference (CSSED) to examine the performance improvement of the model against the benchmark. CSSED is the sum of the squares of error estimates of the

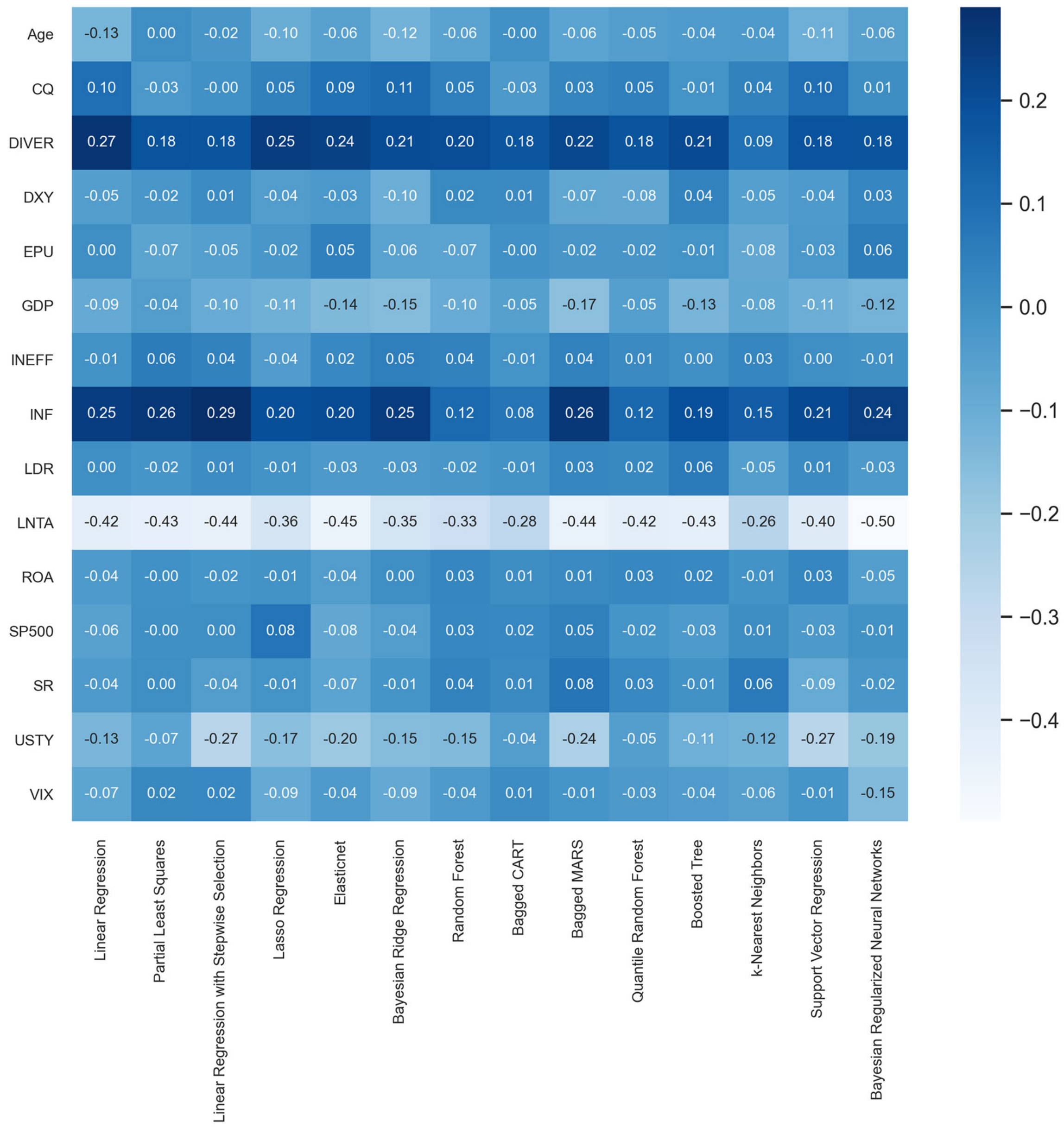

FIGURE 5 Variable importance of all models.

difference between the actual and forecasted values. Following Welch and Goyal (2007), CSSED is estimated as follows:

$$CSSED_{m,t}^{h} = \sum_{i=R}^{t} \left((y-\bar{y})^2 - \left(y-\widehat{Y}\right)^2 \right), \tag{14}$$

where y is the observed value, $\widehat{y}$ is the predicted value, and $\bar{y}$ is the mean of observed value. R and t are the starting and ending point of the rolling window. A positive $CSSED_{m,t}^{h}$ for each t indicates that model m is outperforming the rolling mean as of time t. Model m is improving in comparison with the mean rolling benchmark at that particular point in time t, and vice versa for declines in the $CSSED_{m,t}^{h}$ line. We use 10% interval for the rolling window.

Figure 6 illustrates the result of CSSED of all models using Equation (14). The findings show a linear trend of CSSED in all models. The random forest model has the highest CSSED, and the bagged CART model has the lowest CSSED. The higher value of CSSED indicates random forest model is improving over the sample. This finding further corroborates our baseline results and indicates that the random forest model is the best model for NPL forecasting.

Finally, the random forest model hyperparameter is parallelized to check the increase or decrease in RMSE. For this test, we use mtry = 4 to 15 and ntree = 250 to 750 (intervals of 50). Figure 7 illustrates the error curves

TABLE 6 Random forest model variable importance.

Rank	Variables	ntree = 550	ntree = 600	ntree = 650*	ntree = 700	ntree = 750
1	DIVER	0.103	0.151	**0.202**	0.101	0.171
2	INF	0.206	0.11	**0.115**	0.111	0.173
3	CQ	0.123	0.03	**0.048**	−0.021	−0.02
4	INEFF	0.017	−0.001	**0.042**	0.009	−0.001
5	SR	−0.119	−0.038	**0.038**	−0.010	0.018
6	ROA	−0.005	0.059	**0.029**	−0.014	0.000
7	SP500	0.122	0.015	**0.027**	0.007	0.008
8	DXY	0.029	−0.035	**0.017**	−0.006	−0.008
9	LDR	0.045	0.009	**−0.021**	−0.007	−0.006
10	VIX	−0.059	−0.051	**−0.036**	−0.001	−0.031
11	Age	0.003	−0.063	**−0.063**	−0.065	0.001
12	EPU	−0.029	0.006	**−0.065**	0.009	−0.007
13	GDP	−0.135	−0.023	**−0.095**	−0.034	0.002
14	USTY	−0.111	−0.055	**−0.146**	−0.042	−0.014
15	LNTA	−0.457	−0.498	**−0.328**	−0.459	−0.497

Note: This table presents the findings of random forest model variable importance over different tree size using mtry = 9 using average LIME feature weights. *Final model variable importance.

TABLE 7 Robustness test using different train and test split.

	Panel A: 50:50 train-test split						Panel B: 90:10 train-test split					
	Training (50%)			Testing (50%)			Training (90%)			Testing (10%)		
Algorithms	RMSE	Rsq	MAE	RMSE	Rsq	MAE	RMSE	Rsq	MAE	RMSE	Rsq	MAE
Linear regression	0.017	0.746	0.013	0.018	0.752	0.013	0.017	0.750	0.013	0.018	0.748	0.013
Partial least squares	0.023	0.563	0.018	0.019	0.713	0.014	0.023	0.564	0.018	0.019	0.715	0.013
Linear regression with stepwise selection	0.024	0.530	0.018	0.021	0.661	0.015	0.024	0.531	0.019	0.020	0.655	0.015
Lasso regression	0.030	0.529	0.024	0.018	0.750	0.013	0.031	0.530	0.024	0.018	0.747	0.013
Elastic net	0.032	0.526	0.025	0.018	0.752	0.013	0.033	0.530	0.026	0.018	0.748	0.013
Bayesian ridge regression	0.017	0.746	0.012	0.018	0.752	0.013	0.017	0.750	0.013	0.018	0.748	0.012
Random forest	**0.018**	**0.750**	**0.014**	**0.017**	**0.769**	**0.012**	**0.018**	**0.765**	**0.013**	**0.016**	**0.779**	**0.012**
Bagged CART	0.022	0.594	0.017	0.023	0.593	0.017	0.022	0.590	0.017	0.023	0.577	0.017
Bagged MARS	0.028	0.446	0.022	0.017	0.755	0.013	0.028	0.478	0.022	0.017	0.749	0.013
Quantile random forest	0.018	0.745	0.013	0.017	0.761	0.012	0.018	0.761	0.012	0.017	0.771	0.011
Boosted tree	0.023	0.608	0.018	0.017	0.757	0.013	0.023	0.615	0.018	0.017	0.760	0.012
k-nearest neighbors	0.021	0.624	0.015	0.020	0.680	0.015	0.020	0.657	0.015	0.019	0.694	0.014
Support vector regression	0.018	0.745	0.012	0.018	0.751	0.012	0.018	0.749	0.012	0.018	0.746	0.012
Bayesian regularized neural networks	0.017	0.746	0.013	0.017	0.766	0.012	0.017	0.750	0.013	0.017	0.761	0.012

Note: This table presents the robustness test result of training and testing performance of all models based on 50:50 and 90:10 train-test split. Best model is in boldface.

of random forest model hyperparameter tuning. We can see that RMSE decreases with the number of trees, whereas the highest RMSE is recorded with 250 trees. Furthermore, the result shows that an increase in mtry lead to a decrease in RMSE, and the lowest RMSE is achieved with mtry = 9. Therefore, these findings suggest that mtry = 9 and ntree = 650 is the most optimum hyperparameter for the random forest model. These results

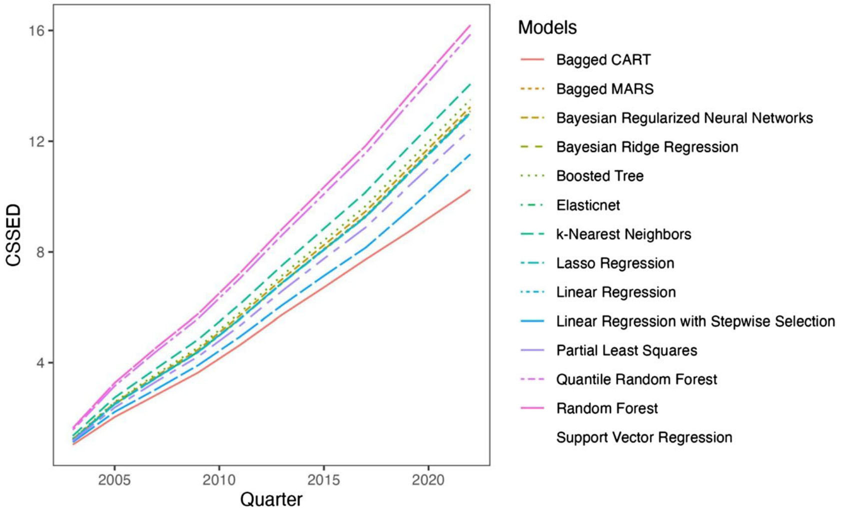

FIGURE 6 Cumulative sum of squares for error difference of all models.

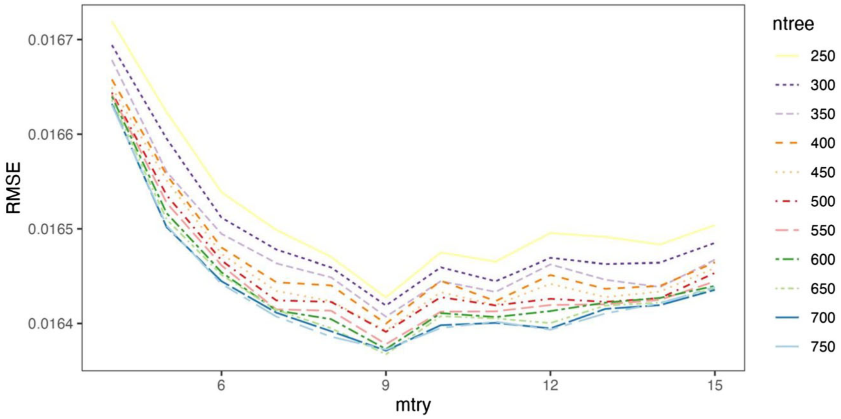

FIGURE 7 Random forest model hyperparameter tuning error curve.

signify our random forest model stability and further validate our findings. Overall, our robustness test indicates that our findings are robust, and the random forest model is an appropriate algorithm for NPL forecasting.

5 | DISCUSSION AND IMPLICATIONS

NPLs pose a significant risk in a bank lending channel. When the number of NPLs in a bank's loan portfolio grows, it erodes its assets and capital, posing a greater danger to its liquidity and profitability. High levels of NPLs can significantly hinder a country's banking sector's progress, which is one of the key factors in triggering banking and financial crisis (Konstantakis et al., 2016; Kuzucu & Kuzucu, 2019). In this study, we applied 14 machine learning algorithms to forecast NPLs. We find that random forest is the best model for forecasting NPLs among other models in terms of RMSE, MAE, Rsq, and bias–variance tradeoff. The Rsq values suggest that the NPL forecasting accuracy of the random forest model is 76.10%. The result is consistent with several previous studies, such as Ozgur et al. (2021), Bonato et al. (2023), and Akyildirim et al. (2021), who also documented the suitability of the random forest model in different financial series forecasting.

The most significant finding of this study is that bank diversification is the most influential variable for higher NPL. We find that it has the highest variable importance in predicting NPLs. Based on the "bad management" hypothesis, this finding suggests that NPLs increase alongside bank diversification, which indicates that due to bad management, banks are facing a higher level of inefficiency and leading to higher NPLs. Previous studies also report similar findings (Louzis et al., 2012; Riahi, 2019). Moreover, credit quality, inefficiency, and solvency ratio are other influential factors in NPL forecasting and imply that large banks take excessive risks by increasing leverage under the "too big to fail" hypothesis, resulting in more NPLs. Furthermore, inflation has the highest contribution in NPL forecasting amid the macroeconomic variables. This indicates NPLs are more sensitive to the inflation rate. Finally, we find that the GDP growth rate has the lowest importance in NPL prediction. Though a rise in GDP increases the borrower repayment capacity, other factors may influence bad repayment and lower the predictive power of GDP towards NPLs, that is, political instability and corruption. Arham et al. (2020) suggest that in emerging countries, corruption control, government effectiveness, and regulatory quality reduce the effect of GDP on NPLs.

The results of this study have four important implications. First, this study adds to the evidence of the application of machine learning models for NPL prediction in the literature. This study advocates that banks should implement machine learning models in their ERP system to track their NPLs. Second, the findings of this study suggest that banks suffering from a high level of NPLs should improve their diversification and efficiency. Banks should develop a strategy for NPL collection within the minimum requirement to reduce the loan loss provision. Third, regulators and policymakers of banking industries can consider the findings of this study for NPL-related policymaking, especially for risk diversification and macroprudential decisions. Finally, this study developed a web-based machine learning app for NPL forecasting, considering the outputs of this study. The emerging nations' bank authorities can use this app for policy decision-making without developing new machine learning models.

6 | CONCLUSION

In the backdrop of the recent financial crisis, NPLs have become a major issue for the banks of emerging markets. A higher level of NPL reduces banks' lending capability and liquidity, resulting in increased financial instability in these emerging countries. This study aims to develop a predictive machine learning model to forecast NPLs. After applying 14 machine learning algorithms on a large cross-section of banks in emerging markets, this study suggests that machine learning models work exceptionally well in predicting the future level of NPL. Among all the considered models, the random forest model outperforms other models in RMSE, MAE, and bias–variance tradeoff. Moreover, this study finds that diversification, credit quality, inefficiency, and solvency ratio among different bank-specific characteristics are the most significant variables affecting NPL.

This study provides several important implications in line with emerging nations' fiscal and macroprudential policy design. Having a clear understanding of factors that affect the probability of banks' future loan loss will allow policymakers to design effective policy alternatives to curve those factors. It also helps banks to make better credit decisions. Finally, our study sheds light on the future research direction on the application of machine learning techniques on financial institutions' performance evaluation. This study considers several competitive machine learning models to evaluate the effect of different bank-specific and macroeconomic factors on banks' NPL. One direction is to incorporate additional variables, for example, governance structure, loan approval policy, and other state-of-the-art machine learning techniques, for example, graph neural network, to increase the accuracy of the forecasting model. Another research direction involves applying explainable machine learning techniques to evaluate the source of predictability. Finally, future studies may focus on forecasting NPLs of consumer and commercial loans separately.

ACKNOWLEDGMENTS

None.

CONFLICT OF INTEREST STATEMENT

On behalf of all authors, the corresponding author states no conflict of interest.

DATA AVAILABILITY STATEMENT

Data are available on request from the authors.

ENDNOTES

[1] The web application “Non-Performing Loans Amount Prediction by Machine Learning (NPLML)” is publicly available at https://mabdullah.shinyapps.io/NPLML/.

[2] Variable importance is the measurement of variable contribution to a model which is calculated by summing up the decrease in error by the addition of a variable in a model (Archer & Kimes, 2008).

[3] VIF is the measurement of multicollinearity in a multiple regressor model.

REFERENCES

Abdullah, M. (2021). The implication of machine learning for financial solvency prediction: An empirical analysis on public listed companies of Bangladesh. *Journal of Asian Business and Economic Studies*, *28*(4), 303–320. https://doi.org/10.1108/JABES-11-2020-0128

Akyildirim, E., Cepni, O., Corbet, S., & Uddin, G. S. (2021). Forecasting mid-price movement of Bitcoin futures using machine learning. *Annals of Operations Research*, 1–32. https://doi.org/10.1007/s10479-021-04205-x

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician*, *46*(3), 175–185. https://doi.org/10.1080/00031305.1992.10475879

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics and Data Analysis*, *52*(4), 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Arham, N., Salisi, M. S., Mohammed, R. U., & Tuyon, J. (2020). Impact of macroeconomic cyclical indicators and country governance on bank non-performing loans in Emerging Asia. *Eurasian Economic Review*, *10*(4), 707–726. https://doi.org/10.1007/s40822-020-00156-z

Ashraf, B. N., & Shen, Y. (2019). Economic policy uncertainty and banks' loan pricing. *Journal of Financial Stability*, *44*, 100695. https://doi.org/10.1016/j.jfs.2019.100695

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bashir, U., Yu, Y., Hussain, M., Wang, X., & Ali, A. (2017). Do banking system transparency and competition affect nonperforming loans in the Chinese banking sector? *Applied Economics Letters*, *24*(21), 1519–1525. https://doi.org/10.1080/13504851.2017.1305082

Bauer, E., & Kohavi, R. (1999). Empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, *36*(1), 105–139. https://doi.org/10.1023/a:1007515423169

Berger, A. N., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, *21*(6), 849–870. https://doi.org/10.1016/S0378-4266(97)00003-4

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural Computation*, *7*(1), 108–116. https://doi.org/10.1162/neco.1995.7.1.108

Bishop, C. M., & Tipping, M. E. (2003). Bayesian regression and classification. In J. A. K. Suykens, I. Horvath, S. Basu, C. Micchelli, & J. V. Null (Eds.), *BT-advances in learning theory: Method* (pp. 267–285). IOS Press. http://www.iospress.nl/book/advances-in-learning-theory-methods-models-and-applications/

Bonato, M., Cepni, O., Gupta, R., & Pierdzioch, C. (2023). Climate risks and realized volatility of major commodity currency exchange rates. *Journal of Financial Markets*, *62*, 100760. https://doi.org/10.1016/j.finmar.2022.100760

Bougatef, K. (2015). The impact of corruption on the soundness of Islamic banks. *Borsa Istanbul Review*, *15*(4), 283–295. https://doi.org/10.1016/j.bir.2015.08.001

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/bf00058655

Breuer, J. B. (2006). Problem bank loans, conflicts of interest, and institutions. *Journal of Financial Stability*, *2*(3), 266–285. https://doi.org/10.1016/j.jfs.2006.07.001

Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, *22*(4), 477–505. https://doi.org/10.1214/07-STS242

Burden, F., & Winkler, D. (2008). Bayesian regularization of neural networks. In *Artificial neural networks* (Vol. 458). Methods in Molecular Biology. Humana Press. https://doi.org/10.1007/978-1-60327-101-1_3

Cepni, O., Demirer, R., Gupta, R., & Sensoy, A. (2022). Interest rate uncertainty and the predictability of bank revenues. *Journal of Forecasting*, *41*(8), 1559–1569. https://doi.org/10.1002/for.2884

Cepni, O., Güney, I. E., & Swanson, N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, *35*(2), 555–572. https://doi.org/10.1016/j.ijforecast.2018.10.008

Cepni, O., Gupta, R., & Onay, Y. (2022). The role of investor sentiment in forecasting housing returns in China: A machine learning approach. *Journal of Forecasting*, *41*(8), 1725–1740. https://doi.org/10.1002/for.2893

Cepni, O., Gupta, R., Pienaar, D., & Pierdzioch, C. (2022). Forecasting the realized variance of oil-price returns using machine learning: Is there a role for U.S. state-level uncertainty? *Energy Economics*, *114*, 106229. https://doi.org/10.1016/j.eneco.2022.106229

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Dimitrios, A., Helen, L., & Mike, T. (2016). Determinants of non-performing loans: Evidence from Euro-area countries. *Finance Research Letters*, *18*, 116–119. https://doi.org/10.1016/j.frl.2016.04.008

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9). MIT Press. https://proceedings.neurips.cc/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, *4*(3), 197–217. https://doi.org/10.1177/096228029500400303

Gallagher, C. (2007). Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models. *Journal of the American Statistical Association*, *102*(480), 1477. https://doi.org/10.1198/jasa.2007.s238

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*(C), 1–17. https://doi.org/10.1016/0003-2670(86)80028-9

Ghosh, A. (2015). Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. *Journal of Financial Stability*, *20*, 93–104. https://doi.org/10.1016/j.jfs.2015.08.004

Giannopoulos, V., & Aggelopoulos, E. (2019). Predicting SME loan delinquencies during recession using accounting data and SME characteristics: The case of Greece. *Intelligent Systems in Accounting, Finance and Management*, *26*(2), 71–82. https://doi.org/10.1002/isaf.1456

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1, Part B), 429–450. https://doi.org/10.1016/j.jeconom.2020.07.009

Hajja, Y. (2020). Impact of bank capital on non-performing loans: New evidence of concave capital from dynamic panel-data and time series analysis in Malaysia. *International Journal of Finance and Economics*, *27*, 2921–2948. https://doi.org/10.1002/ijfe.2305

Ho, T. K. (1995). Random decision forests Tin Kam Ho perceptron training. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (pp. 278–282). IEEE.

Karadima, M., & Louri, H. (2020). Non-performing loans in the euro area: Does bank market power matter? *International Review of Financial Analysis*, *72*(October 2019), 101593. https://doi.org/10.1016/j.irfa.2020.101593

Kılıç Depren, S., & Kartal, M. T. (2020). Prediction on the volume of non-performing loans in Turkey using multivariate adaptive regression splines approach. *International Journal of Finance and Economics*, *26*(4), 6395–6405. https://doi.org/10.1002/ijfe.2126

Kohavi, R., & Wolpert, D. H. (1996). Bias plus variance decomposition for zero-one loss functions. In *Proceedings of the 13th International Conference on Machine Learning (ICML96)*. Morgan Kaufmann Publishers Inc.

Konstantakis, K. N., Michaelides, P. G., & Vouldis, A. T. (2016). Non performing loans (NPLs) in a crisis economy: Long-run equilibrium analysis with a real time VEC model for Greece (2001-2015). *Physica A: Statistical Mechanics and its Applications*, *451*, 149–161. https://doi.org/10.1016/j.physa.2015.12.163

Kuzucu, N., & Kuzucu, S. (2019). What drives non-performing loans? Evidence from emerging and advanced economies during pre- and post-global financial crisis. *Emerging Markets Finance and Trade*, *55*(8), 1694–1708. https://doi.org/10.1080/1540496X.2018.1547877

Lee, J.-M., Chen, K.-H., Chang, I.-C., & Chen, C.-C. (2022). Determinants of non-performing loans, firm's corporate governance and macroeconomic factors. *International Journal of Finance and Economics*, *27*(1), 88–98. https://doi.org/10.1002/ijfe.2139

Lepage, Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika*, *58*(1), 213–217. https://doi.org/10.1093/biomet/58.1.213

Liu, L., Chen, C., & Wang, B. (2022). Predicting financial crises with machine learning methods. *Journal of Forecasting*, *41*(5), 871–910. https://doi.org/10.1002/for.2840

Louzis, D. P., Vouldis, A. T., & Metaxas, V. L. (2012). Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking and Finance*, *36*(4), 1012–1027. https://doi.org/10.1016/j.jbankfin.2011.10.012

Manz, F. (2019). Determinants of non-performing loans: What do we know? A systematic review and avenues for future research. *Management Review Quarterly*, *69*(4), 351–389. https://doi.org/10.1007/s11301-019-00156-7

Marozzi, M. (2009). Some notes on the location-scale Cucconi test. *Journal of Nonparametric Statistics*, *21*(5), 629–647. https://doi.org/10.1080/10485250902952435

Marsaglia, G., Tsang, W. W., & Wang, J. (2003). Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, *8*(18), 1–4. https://doi.org/10.18637/jss.v008.i18

Ozgur, O., Karagol, E. T., & Ozbugday, F. C. (2021). Machine learning approach to drivers of bank lending: Evidence from an emerging economy. *Financial Innovation*, *7*(1), 20. https://doi.org/10.1186/s40854-021-00237-1

Pan, Y., Xiao, Z., Wang, X., & Yang, D. (2017). A multiple support vector machine approach to stock index forecasting with mixed frequency sampling. *Knowledge-Based Systems*, *122*(C), 90–102. https://doi.org/10.1016/j.knosys.2017.01.033

Partovi, E., & Matousek, R. (2019). Bank efficiency and non-performing loans: Evidence from Turkey. *Research in International Business and Finance*, *48*(December 2018), 287–309. https://doi.org/10.1016/j.ribaf.2018.12.011

Plakandaras, V., Papadimitriou, T., & Gogas, P. (2015). Forecasting daily and monthly exchange rates with machine learning techniques. *Journal of Forecasting*, *34*(7), 560–573. https://doi.org/10.1002/for.2354

Podpiera, J., & Weill, L. (2008). Bad luck or bad management? Emerging banking market experience. *Journal of Financial Stability*, *4*(2), 135–148. https://doi.org/10.1016/j.jfs.2008.01.005

Radivojević, N., Cvijanović, D., Sekulic, D., Pavlovic, D., Jovic, S., & Maksimović, G. (2019). Econometric model of non-performing loans determinants. *Physica A: Statistical Mechanics and its Applications*, *520*(81), 481–488. https://doi.org/10.1016/j.physa.2019.01.015

Riahi, Y. M. (2019). How to explain the liquidity risk by the dynamics of discretionary loan loss provisions and non-performing loans? The impact of the global crisis. *Managerial Finance*, *45*(2), 244–262. https://doi.org/10.1108/MF-12-2017-0520

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, *7*(4), 1307–1330. https://doi.org/10.1137/0907087

Song, X.-P., Hu, Z.-H., Du, J.-G., & Sheng, Z.-H. (2014). Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. *Journal of Forecasting*, *33*(8), 611–626. https://doi.org/10.1002/for.2294

Sulong, Z., Abdullah, M., & Chowdhury, M. A. F. (2022). Halal tourism demand and firm performance forecasting: New evidence from machine learning. *Current Issues in Tourism*, 1–17. https://doi.org/10.1080/13683500.2022.2145458

Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting*, *39*(5), 769–787. https://doi.org/10.1002/for.2661

Tavana, M., Abtahi, A.-R., di Caprio, D., & Poortarigh, M. (2018). An artificial neural network and Bayesian network model for liquidity risk assessment in banking. *Neurocomputing*, *275*, 2525–2554. https://doi.org/10.1016/j.neucom.2017.11.034

Uddin, A., Tao, X., Chou, C.-C., & Yu, D. (2022). Are missing values important for earnings forecasts? A machine learning perspective. *Quantitative Finance*, *22*, 1113–1132. https://doi.org/10.1080/14697688.2021.1963825

Uddin, A., Tao, X., & Yu, D. (2021). Attention based dynamic graph learning framework for asset pricing. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 1844–1853). Association for Computing Machinery.

van der Aalst, W. M. P., Rubin, V., Verbeek, H. M. W., van Dongen, B. F., Kindler, E., & Günther, C. W. (2010). Process mining: A two-step approach to balance between underfitting and overfitting. *Software and Systems Modeling*, *9*(1), 87–111. https://doi.org/10.1007/s10270-008-0106-z

Vithessonthi, C. (2016). Deflation, bank credit growth, and non-performing loans: Evidence from Japan. *International Review of Financial Analysis*, *45*, 295–305. https://doi.org/10.1016/j.irfa.2016.04.003

Vouldis, A. T., & Louzis, D. P. (2018). Leading indicators of non-performing loans in Greece: The information content of macro-, micro- and bank-specific variables. *Empirical Economics*, *54*(3), 1187–1214. https://doi.org/10.1007/s00181-017-1247-0

Welch, I., & Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A stepwise AIC method for variable selection in linear regression. *Communications in Statistics - Theory and Methods*, *36*(13), 2395–2403. https://doi.org/10.1080/03610920701215639

Zhang, R., Tian, Z., McCarthy, K. J., Wang, X., & Zhang, K. (2022). Application of machine learning techniques to predict entrepreneurial firm valuation. *Journal of Forecasting*, *42*(2), 402–417. https://doi.org/10.1002/for.2912

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

How to cite this article: Abdullah, M., Chowdhury, M. A. F., Uddin, A., & Moudud-Ul-Huq, S. (2023). Forecasting nonperforming loans using machine learning. *Journal of Forecasting*, *42*(7), 1664–1689. https://doi.org/10.1002/for.2977

APPENDIX A

TABLE A.1 Sample summary.

Country name	Number of banks in sample
Bangladesh	32
Brazil	20
China	46
Egypt	12
India	37
Indonesia	47
Mexico	6
Nigeria	17
Pakistan	21
Philippines	16
Russia	14
South Africa	6
South Korea	11
Turkey	11
Vietnam	26
Total	322

TABLE A.2 Descriptive statistics (training and testing dataset).

Variables	*N*	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Panel A: Training dataset (75%)							
NPL	10,795	0.053	0.053	0.006	0.203	1.602	1.685
LDR	10,795	0.747	0.231	0.372	1.249	0.406	−0.430
ROA	10,795	0.004	0.003	−0.002	0.012	0.344	−0.248
LNTA	10,795	22.699	1.959	19.422	26.770	0.429	−0.453
SR	10,795	0.105	0.046	0.050	0.230	1.187	0.902
INEFF	10,795	0.460	0.363	−0.360	1.207	−0.174	0.218
DIVER	10,795	1.365	1.642	−0.626	6.526	1.853	3.122
CQ	10,795	−0.044	0.035	−0.127	−0.005	−1.065	0.155
Age	10,795	3.342	0.647	2.110	4.425	−0.111	−0.857
INF	10,795	1.590	0.896	0.337	3.700	0.772	−0.167
GDP	10,795	1.254	0.601	−0.473	2.272	−0.798	0.531
VIX	10,795	0.031	0.333	−0.331	0.833	1.132	0.371
DXY	10,795	0.003	0.036	−0.052	0.071	0.219	−0.920
USTY	10,795	0.045	0.205	−0.381	1.292	0.115	−0.021
SP500	10,795	0.022	0.055	−0.111	0.112	−0.477	−0.536
EPU	10,795	0.032	0.168	−0.266	0.376	0.300	−0.430

(Continues)

TABLE A.2 (Continued)

Variables	*N*	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Panel B: Testing dataset (25%)							
NPL	3599	0.050	0.049	0.006	0.203	1.710	2.352
LDR	3599	0.767	0.222	0.372	1.249	0.217	−0.435
ROA	3599	0.004	0.003	−0.002	0.012	0.718	0.340
LNTA	3599	23.158	1.921	19.422	26.770	0.194	−0.578
SR	3599	0.105	0.049	0.050	0.230	1.240	0.719
INEFF	3599	0.439	0.346	−0.360	1.207	−0.003	0.323
DIVER	3599	1.494	1.688	−0.626	6.526	1.728	2.572
CQ	3599	−0.043	0.032	−0.127	−0.005	−1.039	0.388
Age	3599	3.462	0.577	2.110	4.425	−0.091	−0.409
INF	3599	1.264	0.948	0.337	3.700	1.164	0.493
GDP	3599	0.939	0.868	−0.473	2.272	−0.169	−0.948
VIX	3599	0.042	0.296	−0.331	0.833	1.202	1.349
DXY	3599	0.008	0.030	−0.037	0.071	0.367	−0.366
USTY	3599	0.226	0.642	−0.381	1.292	0.822	−1.007
SP500	3599	0.029	0.078	−0.111	0.112	−0.746	−0.935
EPU	3599	0.018	0.165	−0.266	0.376	0.131	−0.157

Note: SD represents standard deviation; *N* represents number of observations.

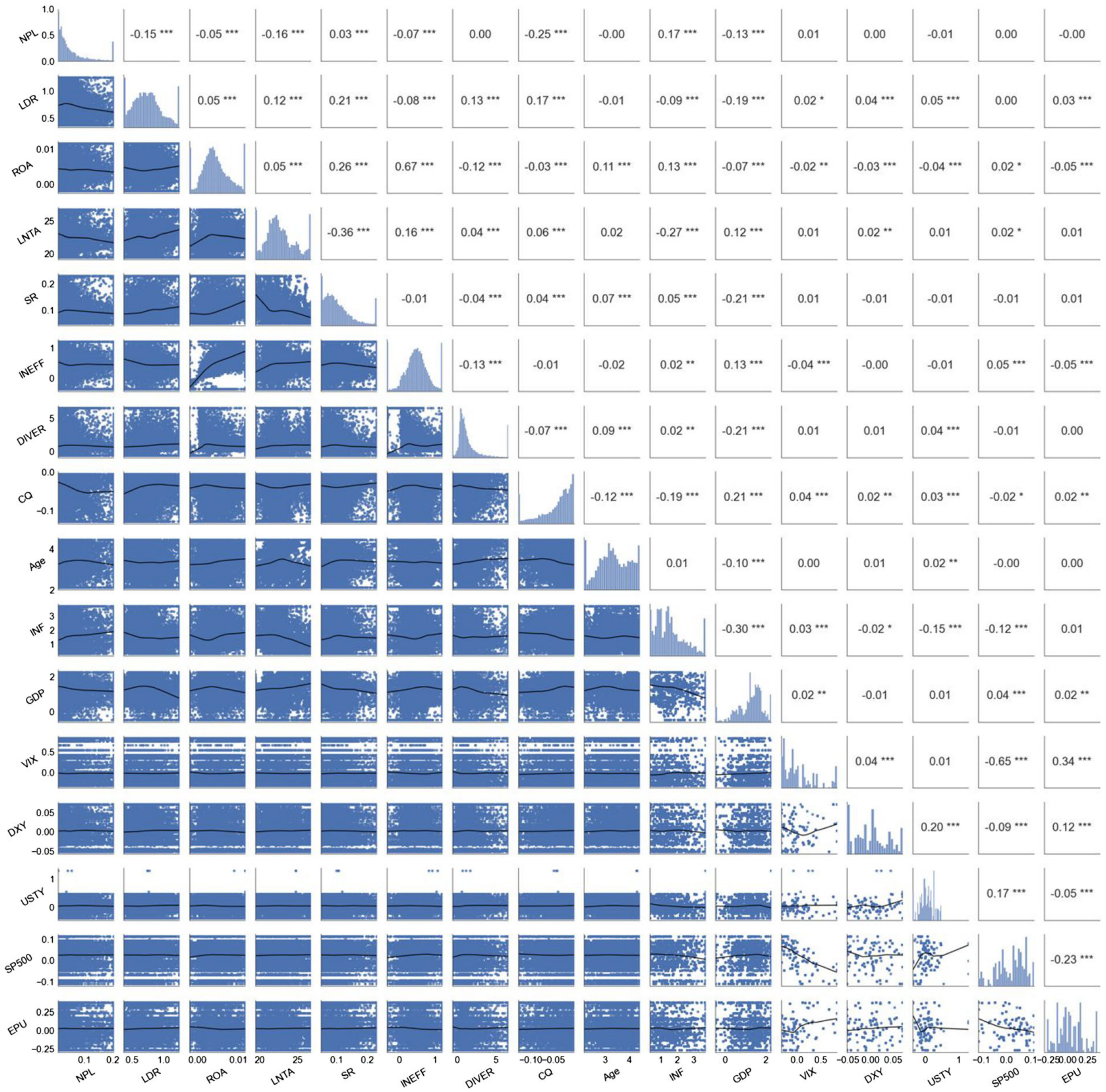

FIGURE A.1 Correlation matrix and distribution plot of training sample. ***, **, and * indicate correlation is significant at 1%, 5%, and 10% significance level.

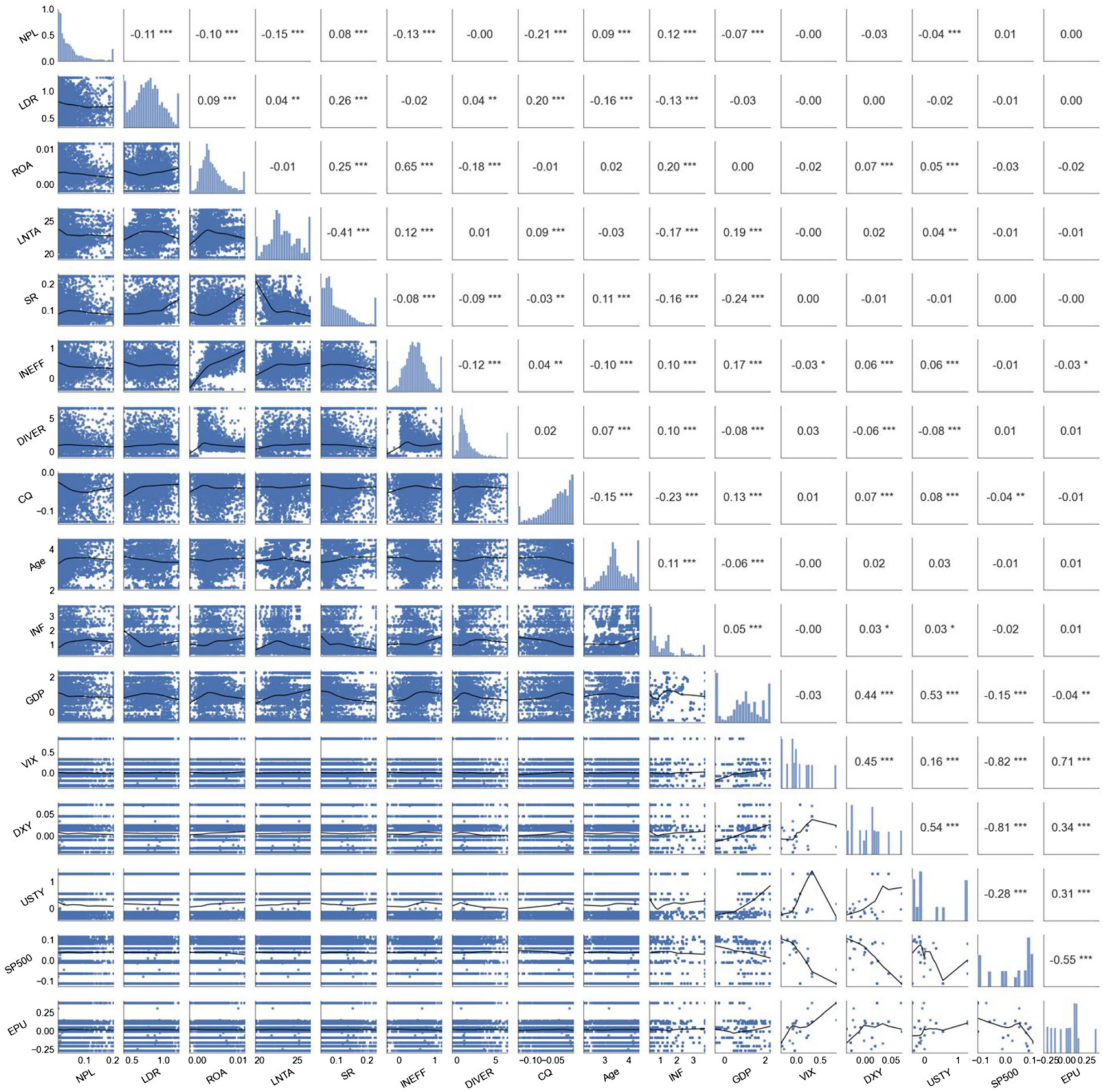

FIGURE A.2 Correlation matrix and distribution plot of testing sample. ***, **, and * indicate correlation is significant at 1%, 5%, and 10% significance level.

AUTHOR BIOGRAPHIES

Mohammad Abdullah is a PhD research fellow and GRA at the Faculty of Business and Management in Universiti Sultan Zainal Abidin, Malaysia. He has published several scholarly articles on emerging financial issues in different international journals and has expertise in econometrics modeling, machine learning, sentiment analysis, deep learning, big data analytics, and so on. His research interests include corporate finance, ESG, behavioral finance, banking, and FinTech.

Mohammad Ashraful Ferdous Chowdhury is a postdoctoral research fellow at IRC for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum and Minerals, Saudi Arabia. He is also a faculty member (on leave) of the Department of Business Administration, Shahjalal University of Science and Technology, Bangladesh. He holds PhD and MSc degree in Islamic finance from INCEIF, Malaysia. Prior to that, he obtained MBA in Finance and Banking and BBA in Finance from SUST. His research interests include financial economics, institutional economics, banking and finance, and Islamic finance. Dr. Chowdhury published more than 45 articles in ISI/SSCI/SCI/SCOPUS indexed journals.

Ajim Uddin is an assistant professor of Financial Technology at Martin Tuchman School of Management (MTSM), New Jersey Institute of Technology (NJIT). In the broadest sense, his research interests are machine learning and data mining with application to finance. He is currently working on nonlinear tensor factorization and network representation for financial markets. Primarily his focus is on modeling dynamic changes in network structures and incorporating network information into traditional asset pricing models using spectrum analysis and graph neural networks.

Syed Moudud-Ul-Huq is a professor of the Department of Accounting and Founder Chairman of Accounting & Former Chairman of the Department of Business Administration & Former Dean, Faculty of Business Studies at Mawlana Bhashani Science and Technology University. His primary research interests are financial economics, risk management, econometric modeling, corporate governance, financial institutions, and so on. He has published several scholarly articles in different international journals.