Research in International Business and Finance 61 (2022) 101644

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks

Tamás Kristóf *, Miklós Virág

Institute for the Development of Enterprises Corvinus University of Budapest, Fővám tér 8, H-1093 Budapest, Hungary

ARTICLE INFO

JEL classification:
C38
C45
C53
G17
G21

Keywords:
Bank failure
Classification
Credit risk modeling
Machine learning

ABSTRACT

This article provides evidence that machine learning methods are suitable for reliably predicting the failure risk of European Union-27 banks from the experiences of the past decade. It demonstrates that earnings, capital adequacy, and management capability are the strongest predictors of bank failure. Critical and relevant field research is presented in the context of economic uncertainties arising from the COVID-19 pandemic. The results suggest that the developed models possess high predictive power, with the C5.0 decision tree model providing the best performance. The findings have policy implications for bank supervisory authorities, bank executives, risk management professionals, and policymakers working in finance. The models can be used to recognize bank weaknesses in time to take appropriate mitigating actions.

1. Introduction

The uncertainty generated by the COVID-19 pandemic has led to a general questioning of the stability of banks and banking systems. Thus, the ongoing interest in the timely recognition of potential problems faced by banks has grown significantly. Since banking systems play an important role in economic development, a banking crisis may lead to serious economic disruption within a given country. Moreover, bank failure is regarded as more severe than the failure of other business entities because of the interconnectedness and fragility of banking institutions. Accordingly, it can be argued that reliable bank failure prediction can reduce the overall potential for substantial economic problems.

From a methodological perspective, bank failure prediction does not fundamentally differ from corporate failure prediction. However, far fewer banks than registered corporations have worldwide operations. Thus, there are fewer available data and failure observations for banks, presenting considerable modeling challenges. Under such conditions, the authors have considered the importance of conducting contemporary research in this field.

This article uses a complex definition of bank failure. Beyond traditional legal failures such as bankruptcy or liquidation, the applied failure definition considers payment defaults, dissolutions with negative equity, state aid measures such as bail-ins and bailouts, and distressed mergers as bank failure events. Thus, a collective failure definition is used and includes default and state aid events. The literature has already demonstrated that predicting banking problems from the viewpoint of bail-ins and bailouts is possible using the same methods as those used to predict bank failure from traditional default events (Gerhardt, Vennet, 2017). Applying multivariate binary classification techniques is challenging since bank failure in Europe is somewhat rare.

* Corresponding author.
E-mail addresses: tamas.kristof@uni-corvinus.hu (T. Kristóf), miklos.virag@uni-corvinus.hu (M. Virág).

https://doi.org/10.1016/j.ribaf.2022.101644
Received 26 March 2021; Received in revised form 19 February 2022; Accepted 23 February 2022
Available online 25 March 2022

This article is thus positioned as a comparative empirical paper applying a selected range of descriptive variables and multivariate classification methods to provide as many reliable models as possible to predict the failure of European Union (EU)-27 banks. Similar empirical research has been performed to this point within the frameworks of several publications (see, inter alia, Poghosy and Čihák, 2009; Betz et al., 2013; Gerhardt and Vennet, 2017; Bräuning et al., 2019), although not necessarily using the same variables and methods as those in this article.

This paper contributes to the literature by applying contemporary methods to differentiate the good and bad banks in the EU-27 area by comparing logistic regression (logit) and modern machine learning (ML) approaches to better capture nonlinearities and interactions. A further contribution is made by examining drivers of bank failure with state-of-the-art ML techniques and demonstrating the advantages of ensemble classifiers to gain additional predictive performance benefits. Using novel EU-27-wide data and ML techniques, this article develops an early warning system for predicting bank failure by accounting for the experiences of the past decade.

It is further acknowledged that empirical results justifying the superiority of ensemble classifiers and neural networks (NNs) over conventional techniques in bank failure prediction are already available in the literature (see, inter alia, Tanaka et al., 2016; Le, Viviani, 2018; Bräuning et al., 2019; Appiahene et al., 2020; Shrivastava et al., 2020). It can be simultaneously observed that a relatively large share of empirical results in this field originates from the United States (US) (see, inter alia, Gogas et al., 2018; Jing and Fang, 2018; Manthoulis et al., 2020; Petropoulos et al., 2020). For that reason, it seemed appropriate to perform new empirical research in the EU to examine whether findings in the EU-27 banking area are relevant or similar.

Multivariate bank failure classification models are developed in this article by applying logit, C5.0 decision trees (DTs), and deep learning NN (DL-NN) methods. The conclusions from the results are that all three models possess high classification accuracy. However, ranked by their predictive performance, the C5.0 DT model has the best performance, the DL-NN model is second, and the logit model ranks third. Hence, this article justifies the view that ensemble DT methodology outperforms other techniques for predictive power and simultaneously reveals that the conventional logit method might also perform with a fair degree of efficiency, as similarly found by Beutel et al. (2019). Results obtained from the development of the three models demonstrate that earnings, capital adequacy, and management capability form the strongest predictors of bank failure in the EU, which also corroborates the findings of, inter alia, Le and Viviani (2018), Kolari et al. (2019), and Carmona et al. (2019).

This article is structured as follows. Section 1 provides a historical literature review of the development of bank failure prediction to identify the range of previously applied variables and methods and gain insights from previous empirical studies. Section 2 presents data collection results and the compilation of the modeling database. Financial data compiled between 2011 and 2019 were collected for 5614 European banks by formulating 32,287 bank-year observations and locating the identified failure events between 2012 and 2020. Section 3 analyzes features of input and output variable creation in line with industrial standards and presents a methodology for using ML to generate categorical variables from continuous financial ratios. Section 4 presents the model development results and provides a comparative evaluation of model performance. Multivariate classification models are subsequently developed using logit, C5.0 DT, and DL-NN methods. The concluding section summarizes the results of empirical research conducted in this article to formulate policy recommendations and indicate possible future research directions.

2. Literature review—the history of the development of bank failure prediction

Bank failure prediction represents a multivariate classification problem whereby the target variable expresses the occurrence of a failure event, and input variables are considered descriptive features that attempt to explain the failure. To underpin empirical research in this article, the prerequisite examination of previous studies is necessary when approaching the target variable and examining explanatory variables that have been deployed to predict bank failure and what multivariate classification methods have in turn been applied to develop efficient bank failure models.

The group of indicators applied in bank failure prediction is widely known as "CAMELS", incorporating *C*apital adequacy, *A*sset quality, *M*anagement capability, *E*arnings, *L*iquidity, and *S*ensitivity measures (Nguyen et al., 2020). Several financial ratios drawn from the CAMELS family have been applied in several studies to predict bank failure and subsequently provide instructive results. Throughout its application, CAMELS has also become an efficient supervisory tool for evaluating the soundness of financial institutions on a uniform basis and for identifying institutions requiring special supervisory attention (Betz et al., 2013).

As far as methodology and results of empirical studies are concerned, multivariate bank failure prediction has proceeded through a sequence of development very similar to that of corporate failure prediction (Altman et al., 2017). Initial bank failure models were based on static, single-period discriminant analysis (DA) and logit-based methods. The first notable DA-based bank failure model was introduced by Sinkey (1975) and the first logit model by Martin (1977). Multiperiod models increasingly replaced static models in the 1980s (Shumway, 2001).

Beyond multivariate classification models, the structural models of default risk in which equity and debt are viewed as contingent claims on assets have also made a significant contribution to bank failure research (Gabbi and Levich, 2019). Li (2013) argued that the defaultable yield attributable to maturity, credit spread, and duration is related to the Markov chain's finite state driven by macroeconomic conditions. Nagel and Purnanandam (2020) adapted the structural models to banks. Since typical bank assets are risky debt claims with nonlinear payoffs, the usual assumption of lognormally distributed asset values is not appropriate for banks. It was found that standard structural models with constant asset volatility can severely understate bank default risk in favorable conditions when asset values are high. Furthermore, bank equity return volatility is generally considered as much more sensitive to negative shocks to asset values than identified in standard structural models.

Default intensity models also represent a milestone in the literature in the form of new time-series methodological tools for

predicting the probability of default (PD) for banks. Siakoulis (2015) employed a duration-based approach to model the interarrival times of bank failures in the US banking system between 1934 and 2014. They found evidence of strong failure persistence along with nonmonotonic hazard rates to imply a financial contagion pattern generating a high frequency of bank failures.

Gropp et al. (2006) examined indicator properties of the distance to default (DD) model as a market-based indicator to provide early indications of financial distress and fragility in a sample of EU banks. Results suggested that market indicators reduced the extent of type II error compared to predictions based only on accounting information. Harada et al. (2013) examined movements of DD measures of failed Japanese banks to evaluate their collective predictive power of bank failure. Both DD models and DD spreads were better indicators for the deterioration of health of a failed bank than other traditional indicators.

Since the 1990s, several publications have strongly suggested that ML techniques outperform traditional statistical techniques. This has motivated researchers to compete in producing newer methods by working with the same bank failure database to achieve superior classification power. The empirical research approach used in this article has also been strongly influenced in the same regard. Among ML techniques, NN, support vector machine (SVM), and random forest (RF) methods appear to be the preferred tools for bank failure prediction. Tam and Kiang (1992) were the first authors to apply NN to bank failure prediction and found that NN outperformed other previously applied methods. Kolari et al. (2002) developed an early warning system based on logit and nonparametric trait recognition models for large US banks. Boyacioglu et al. (2009) examined NN, SVM, and multivariate statistical methods to predict the failure of Turkish banks. Results of the latter study indicate that SVM achieved the best level of accuracy.

Poghosyan and Čihák (2009) conducted research in the EU with financial data gathered between 1997 and 2007. Beyond the relevance of customary predictors based on several logit models, it was concluded that contagion effects were important when predicting EU bank failures. In particular, the PD of a bank was found to be higher if there had recently been a failure of a similarly sized bank within the same country.

Several publications applied previous techniques to more recent data on bank failures during the 2008–10 global financial crisis. Interesting conclusions were drawn regarding reasons for bank failure that had not markedly changed in comparison with earlier findings (Cole and White, 2012; Fahlenbrach et al., 2012; Wang and Cox, 2013).

The aftermath of the global financial crisis amplified the need to research additional factors to better explain and predict the financial health of banks. Credit default swaps were proven to be an effective indicator of bank financial distress during a period of systemic bank crisis across European and US banks (Avino et al., 2019). Beltratti and Stulz (2012) examined the behavior of stock price performance of banks during the global financial crisis. Results suggested that large banks with greater Tier 1 capital levels, more deposits, less exposure to US real estate markets, and less funding fragility had better stock price performance. Betz et al. (2013) developed an early warning model for predicting vulnerabilities leading to distress in European banks using bank and country-level data. Key findings of the article were that complementing bank-specific vulnerabilities with indicators for macrofinancial imbalances and banking sector vulnerabilities improved model performance and yielded useful out-of-sample predictions of bank distress. Korean experiences demonstrate that larger bank solvency risk can also be associated with increased funding costs (Aldasoro et al., 2022).

Maghyereh and Awartani (2014) developed a hazard model for banks in the Gulf Co-operation Council countries and found that good management practices reduced the likelihood of distress. In contrast, competition and diversification were found to be bad for the financial health of banks. ElBannan (2021) examined emerging markets with multiple regressions and found that well-developed financial markets in countries with a low level of corruption reduced failure risk. Tan and Anchor (2017) concluded that the efficiency of Chinese banks was significantly affected by ownership, diversification, banking sector development, stock market development trends, and inflation and gross domestic product (GDP) growth rates. Financial distress is also related to the degree of social responsibility (Boubaker et al., 2020).

Wang et al. (2016) developed a self-organizing "neural fuzzy" inference system to predict bank failure based on the experience of 3635 US banks over a 21-year period. Experimental results of the model were encouraging in terms of both accuracy and interpretability when benchmarked against other prediction models. Tanaka et al. (2016) constructed an RF-based early warning system for predicting bank failure, with results indicating that RF outperformed conventional methods. Cox et al. (2017) employed the Cox proportional hazards model to forecast US bank failures during the 2008–2010 global financial crisis. This study provided a major contribution in outlining enduring attributes that might reduce the likelihood of bank failure.

Using logit techniques, Gerhardt and Vennet (2017) investigated the financial conditions of European banks before and after receiving state support. Data were collected from 114 banks that had benefited from government support in Europe during the crisis. Results indicated that the equity ratios, loan quality, and bank size were the main determinants of decisions to enact bailout measures. This study also revealed that bailed out banks barely improved their performance indicators in the years following government intervention, thereby demonstrating that bailouts were not singularly sufficient as a means of restoring overall bank health.

Elekdag et al. (2020) argued that the most reliable determinants regarding the performance of large eurozone banks were real GDP growth and the nonperforming loan (NPL) ratios. In most recent publications, Berger et al. (2021) found national cultural variables, Eberhardt and Presbitero (2021) commodity prices, and Chen et al. (2021) liquidity risk to respectively be important predictors of bank failure. To a certain extent managerial ability might moderate the impact of credit risk and liquidity risk on the likelihood of bank default (Abdesslem et al., 2022).

Le and Viviani (2018) compared the accuracy of traditional statistical and ML techniques to predict bank failure using a sample of 3000 US banks. Results indicated that NNs and k-nearest neighbor techniques were the most reliable methods for predicting bank failure. Audrino et al. (2019) combined a generalized logit model with mixed-data sampling to improve the accuracy of US bank failure prediction. Application of the model to data collected between 2004 and 2016 yielded substantially better results than the accuracy of classic logit models, especially for those applied to longer-term forecasting horizons. Climent et al. (2019) offered an extreme gradient

boosting (XGB) model to predict and prevent bank failure for eurozone-listed commercial banks between 2006 and 2016. Essential technical aspects can further improve the predictive accuracy of XGB that can outperform gradient boosting and adaboost techniques (Pham and Ho, 2021).

Bräuning et al. (2019) developed an early warning system for 3000 small European banks between 2014 and 2016. The C5.0 DT model achieved a 92% AUROC score from the testing sample data, thereby exceeding the 90% performance level of the benchmark logit model. Appiahene et al. (2020) combined a data envelopment analysis method with three ML approaches to evaluate bank efficiency and performance for 444 Ghanaian banks. The C5.0 DT model achieved the best performance level, followed by the RF and NN models.

Shrivastava et al. (2020) created ML-based bank failure models for Indian banks using data collected between 2000 and 2017. The synthetic minority oversampling technique was used to handle the relatively low number of failed banks. Redundant features were reduced by using the least absolute shrinkage and selection operator technique. RF and AdaBoost techniques were applied to avoid bias and overfitting and compared with the logit method to produce the best predictive model.

Manthoulis et al. (2020) explored the predictive power of attributes of US banks by describing diversification of banking operations and considered the prediction of failure in a multiperiod context. This study also introduced an enhanced ordinal classification framework consisting of multiple criteria decision analysis, statistics, ML, and ensemble methods. Results strongly indicated that both diversification attributes and ordinal classification provided superior means of prediction. Petropoulos et al. (2020) employed a series of modeling techniques to predict bank insolvency in a sample of US-based financial institutions. Results indicated that RF had superior out-of-sample and out-of-time predictive performance, while NN performed as well or nearly as well for out-of-time samples. Variables related to earnings and capital provided the largest contributions to bank failure predictions.

In contrast to the results of these relatively recent publications, Beutel et al. (2019) arrived at a provocative conclusion—while ML methods often attained a very high in-sample fit, they were outperformed by logit in terms of recursive out-of-sample evaluations. The out-of-sample predictive performance of different models for systemic banking crises was compared with a benchmark logit model using a sample of advanced economies covering the preceding 45 years. Results suggested that further enhancements to ML models would be necessary to offer a substantial value-added method of predicting bank failure, given that conventional logit models appeared to use available information fairly efficiently.

Based on an examination of the literature, it is concluded that ML methods currently dominate bank failure prediction. A wide range of input features have noticeably been applied, and CAMELS variables have generally been present in model development. In the initial era of development history, relatively simpler models were produced, but more complex models have emerged in recent years. The most recent studies primarily conducted in the US have revealed outstanding results achieved by NN and ensemble DT classifiers, especially those using XGB, RF, and C5.0 methods. However, logit has remained popular both as a benchmark and a standalone method and is relatively efficient. In risk management, expert rating systems are usually applied rather than statistical models to evaluate the riskiness of banks. This is primarily due to the difficulties of data collection and the rarity of bank failure events.

The problems identified in the preceding narrative collectively constitute a research gap and act as a basis for conducting original empirical research applied to banks based in the EU-27 area. This is necessary to examine whether the findings of previous studies are relevant to European banks and provide new empirical models to efficiently predict bank failure. By considering favorable experience with NN and ensemble DT methods, an advanced DL-NN method and the widely used C5.0 DT method have been selected for this purpose. A logit model has also been developed as a benchmark method. A selected range of CAMELS variables have been applied in model development as they were demonstrated to be effective model variables in the literature. Therefore, the most important empirical research objective of this study is to locate the most relevant variables and the best classification method to predict the failure of EU-27 banks.

3. Data collection and construction of the database

To conduct empirical research, the financial data of several EU-27 banks were collected from Moody's Analytics BankFocus database, a comprehensive global source of historical banking data. When commencing the data collection process, data were available for 11,628 EU-27 banks, including duplicated records due to the parallel availability of consolidated/unconsolidated financial data and financial reports compiled by using different accounting standards such as international financial reporting standards (IFRS) or local generally accepted accounting principles (GAAP). The extracted data were screened in a manner consistent with research goals. Bank records were discarded in the following cases:

- where 2011–2019 financial data were not available;
- where bank classification or status was unknown;
- where a central bank appeared in the database;
- where unconsolidated financial data were used but consolidated financial data were available (this was also applied if data were used to handle the duplicated records of the same entities);
- local GAAP financial data were used when IFRS financial data were available (this was also applied if data were used to handle the duplicated records of the same entities).

During the screening process, 5614 unduplicated entities were extracted from the initial sample of 11,628 banks. A total of 32,287 bank-year observations between 2011 and 2019 were thus incorporated into the modeling database. The national and financial year breakdown of the database by bank-year observations is summarized in Table 1.

The relative rarity of bank failure events in Europe is strongly acknowledged. Hence, the application of multivariate classification techniques is somewhat challenging. To collect a comprehensive range of failure events beyond traditional legal failures such as bankruptcy or liquidation, the failure definition used in this study also considered payment defaults and dissolution with negative equity as bank failure events as well as state aid measures and distressed mergers such as bail-ins and bailouts as bank failure events. A collective failure definition is accordingly used to include default and state aid events.

Based on the experience of the authors, any bank failure in Europe is a notable media event even if it occurs with a small bank, thus underlining its relatively rare occurrence. Therefore, it is somewhat unlikely that a single bank failure could be discretely hidden in the remit of the EU-27 area and not be detected by referral to professional databases, credit risk agencies, and specialized business media.

A failure event as a modeling target variable was defined temporally from when the failure was announced and ended with the officially recognized date of the failure. Bank-year observations were denoted as one (1) as a binary target variable when the failure event occurred and zero (0) when no failure occurred, both in the subsequent year. Data for failure events were collected from the following sources:

- records of bankruptcy, reorganization, liquidation, payment default, and dissolution due to negative equity status obtained from BankFocus;
- records of state aid measures, distressed mergers, bail-ins and bailouts from the European Commission, and publicly available market data sources (i.e., Reuters, Forbes, BBC).

A total of 303 failed bank-year observations were identified, thereby constituting a somewhat imbalanced database with a 0.938% failure rate. Table 2 summarizes the distribution of failure events in the database.

22 EU-27 bank failure events were identified by examining the compendium reports of Moody's, Standard & Poors, and Fitch credit rating agencies between 2012 and 2020. However, all failure events were verified as having already been identified from previously compiled sources.

For further calculations, the reason for failure was not differentiated, and hence, all 303 instances represented the entire failure data class. Table 3 breaks down all 303 failed observations by failure event, with nearly half the cases (144) classified as bail-ins and bailouts.

Table 4 summarizes the distribution of the 303 failed observations by EU-27 country by year. Based on the cross-tabulation analysis, it can be argued that both the country and the failure year possess a statistically significant relationship with the occurrence of the failure event in the modeling database. The Pearson chi-squared value between the country and the binary target variable is 606.274 ($p = 0.000$) and between the failure year and the binary target variable is 67.039 ($p = 0.000$).

The time effect is present primarily for two reasons. First, a notable number of bailouts occurred in 2014, especially in Italy, France,

Table 1
Breakdown of the modeling database by EU-27 country and financial year.

	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
AT	557	544	657	547	615	564	524	488	463	4959
BE	23	28	38	35	36	36	36	29	28	289
BG	12	12	23	23	21	20	20	20	18	169
CY	6	7	15	28	32	33	32	31	32	216
CZ	13	14	30	33	33	31	32	30	28	244
DE	299	579	1744	1717	1691	1627	1559	1495	1294	12,005
DK	33	33	59	58	76	76	78	77	70	560
EE	2	2	9	9	10	10	11	10	10	73
ES	31	26	124	121	124	121	121	120	114	902
FI	6	6	24	28	54	176	176	184	176	830
FR	218	248	318	328	322	313	301	288	268	2604
GR	9	9	10	9	10	10	10	10	10	87
HR	25	25	37	37	35	32	31	26	26	274
HU	13	13	25	25	22	24	23	24	21	190
IE	16	16	19	20	19	35	35	35	32	227
IT	538	540	585	591	571	537	462	444	420	4688
LT	6	6	8	8	6	6	5	5	4	54
LU	67	67	86	81	78	77	71	70	52	649
LV	6	6	14	14	15	16	14	16	15	116
MT	4	4	10	13	16	13	14	13	13	100
NL	28	33	41	41	46	40	35	36	26	326
PL	31	33	69	83	104	140	137	133	118	848
PT	12	11	31	115	126	123	123	118	117	776
RO	9	9	22	25	24	25	23	23	23	183
SE	22	27	89	82	93	98	100	78	67	656
SI	8	8	17	20	16	14	14	13	12	122
SK	11	12	19	19	18	16	16	15	14	140
Total	2005	2318	4123	4110	4213	4213	4003	3831	3471	32,287

This table presents the number of bank-year observations per country and financial year retrieved from BankFocus records to construct the modeling database. In line with the predefined filtering criteria, 32,287 records met the requirements and were applied for further modeling steps.

Table 2
Composition of the modeling database by failure status.

	Nonfailed	Failed	Total	Failure rate
Bank-year observations	31,984	303	32,287	0.938%

This table analyzes the ratio of failed bank-year observations to the total records in the modeling database. The 0.938% failure ratio demonstrates that the long-run failure ratio of EU-27 banks can be regarded as limited.

Table 3
Breakdown of the failed bank-year observations by failure event.

Failure event	Bank-year observations
Bail-in / Bailout	144
Bankruptcy / Reorganization	44
Default of payment	7
Dissolution with negative equity	19
Liquidation	89
Total	**303**

This table analyzes the distribution of the 303 failed bank-year observations across identified failure types. It can be seen that 47.5% of the cases (144 failures) can be linked to state aid measures (bail-ins or bailouts) and 29.4% (89 failures) to liquidation. Payment default as a reason for bank failure is negligible in the case of EU-27 banks.

Table 4
Breakdown of the failed observations by EU-27 country and failure year.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
AT	4	3	4	3	6	5	4	4	1	34
BE	1		1		1	1	2	1		7
BG	1	2	3	3	1					10
CY	2	1	1		1	1				6
CZ			1	1	1					3
DE	3	2	4	4	4	4	2	7	2	32
DK	1			1	2	2				6
EE						1	1			2
ES	1	5	6	4	1	2				19
FR	2	1	12	3	2	1	1	1	1	24
GR	3	3	2	2	1	1				12
HR				2	2	3	1			8
HU	1		1	1			1			4
IE	1	2	3	1	2	1				10
IT	10	9	11	9	10	11	4	2	1	67
LT			1							1
LU			2	1	2	1	1	1		8
LV			1	1	2	1	2	2	3	12
NL	1		1	1	2				2	7
PL	1						1	1		3
PT	2	1	1	3	2	1	1			11
SE		1	1			1			1	4
SI	3	4	4	2						13
Total	37	34	60	42	42	37	21	19	11	303

This table analyzes the distribution of the 303 failed bank-year observations across countries and failure years. Figures indicate that most bank failure cases (67) took place in Italy over the analyzed period. An outstanding number of bank failures (60) occurred in 2014 because of bailouts, especially in Italy, France, and Spain. Surprisingly, the smallest number of failures occurred in 2020, indicating that to that point, the EU-27 banks had evaded the negative impacts of the COVID-19 crisis. By relating the failed cases to the total observations presented in Table 1, it can be concluded that in total, Greek banks possess the highest failure ratio within the EU-27 area.

and Spain. Second, an extremely low number of bank failures occurred in Europe during 2019 and 2020 despite the COVID-19 pandemic. It can thus be concluded that as far as failure events are concerned, the negative effects of COVID-19 have largely not affected EU-27 banks to this point.

The distribution of failure events among EU-27 countries shows interesting results. The database retrieved from BankFocus indicates that no bank failures have occurred in the past decade in Romania, Slovakia, and Malta. By contrast, Germany, Finland, and Poland possess significantly lower-than-average bank failure statistics. The bank-year observations indicate that Greece has the highest bank failure rate within the EU-27 area. However, it is notable that BankFocus data strongly rely on data provided by banks resident in

the EU-27 area.

4. Variable creation and data transformation

The variables selected for this study are extensively used in bank risk management practices and incorporate features of each component of the CAMELS group of indicators. Financial ratios for the bank-year observations were extracted from the BankFocus database. In cases where BankFocus data contained missing values for one or more variables, these were replaced by median imputation. Consideration of modeling variables for which too many values were missing (e.g., dividend payout ratio) in the BankFocus database was thus avoided. Table 5 summarizes the range of applied input variables in accord with the rationale of the CAMELS classification system.

Table 6 presents basic statistics for applied input variables as continuous ratios before categorization. As observed from certain minimum and maximum values, the database contains outliers that were handled by categorization, as subsequently described.

Various previous empirical studies have conclusively demonstrated that logit and NN methods could provide more reliable failure prediction accuracy if the range of input variables were categorized according to significantly different empirical failure rates in the created categories. Traditional methods might have difficulties in capturing effective information (Duan et al., 2021). In this study, categorization was conducted using univariate chi-squared automatic interaction detection (CHAID) DT analysis. In addition to its favorable added value of enhanced classification power, CHAID operates well in handling outliers and nonlinear and nonmonotonous relationships, which are also present in the modeling database. Thus, CHAID was used as an ML method to create new variables in this study. Critical intervals were estimated ratio by ratio by CHAID, whereby optimal binning points were determined using the chi-squared homogeneity test. Significantly different categories were therefore generated relative to failure behavior.

Variables were also discarded from further modeling steps as a feature selection method in which CHAID could not create reasonable categories. This aspect affected net charge-off / average gross loans, net charge-off / net income before loan loss provision, and nonoperating items & taxes / average assets ratios.

In the subsequent step, membership of the created categories was denoted by flag variables set at one (1) or zero (0). Table 7 presents an analysis of features and decision rules relating to the created flag variables.

5. Results and discussion—the developed models

To enable model validation and evaluate data not used in model development, the database was divided into training and testing subsets by applying random sampling techniques resulting in a 70/30% distribution. Table 8 summarizes the characteristics of the two subsets.

To ensure that random sampling partitioning did not result in inconsistency, failure statistics of the generated categorical variables (decision rules) in the training and testing subsets were analyzed and are summarized in Table 9.

Several empirical studies in recent years have examined the problem of imbalanced classification that could influence ML algorithm performance (i.e., Hasanin et al., 2019). This is primarily due to the possibility that ML algorithms may completely ignore the minority class, which is the category that researchers tend to be most concerned with in terms of failure prediction. In most cases, the proposed solution is oversampling (Shrivastava et al., 2020), which usually has a positive impact on ML model performance but might increase the likelihood of overfitting since it tends to produce exact copies of cases in the minority class.

This study was not intended to apply oversampling or undersampling techniques to handle problems of low failure rates in the

Table 5
Input variables applied according to the CAMELS classification.

Capital adequacy	Asset quality
Tier 1 ratio	Loan loss reserves / Gross loans
Total capital ratio	Loan loss provisions / Net interest revenue
Equity / Total assets	Loan loss reserves / Nonperforming loans
Equity / Liabilities	Nonperforming loans / Gross loans
Equity / Net loans	Impaired loans / Equity
Management capability	**Earnings**
Cost to income	Return on Average Assets (ROAA)
Noninterest expenses / Average assets	Return on Average Equity (ROAE)
Other operating income / Average assets	Recurring earning power
Nonoperating items & taxes / Average assets	Net interest margin
Nonoperating items / Net income	Net interest revenue / Average assets
Net loans / Total assets	Pre-tax operating income / Average assets
Liquidity	**Sensitivity**
Liquid assets / Deposit & short-term funding	Interbank Ratio
Equity / Deposit & short-term funding	Net charge-off / Average gross loans
Net loans / Deposit & short-term funding	Net charge-off / Net income before loan loss provision
Liquid assets / Total deposit & borrowings	Unreserved impaired loans / Equity
Net loans / Total deposit & borrowings	

This table presents the grouping of the applied input variables using the CAMELS classification system, which incorporates Capital adequacy, Asset quality, Management capability, Earnings, Liquidity, and Sensitivity indicators.

Table 6
Basic statistics for applied input variables.

Input variable	Minimum	Maximum	Mean	Median	Std. Dev.	Skewness	Kurtosis
Loan loss reserves / Gross loans	−124.7	182.9	4.1	3.2	7.1	7.1	92.8
Loan loss provisions / Net interest revenue	− 86,587.9	131,300.0	28.1	6.5	1223.8	47.4	6125.3
Loan loss reserves / Nonperforming loans	−99.6	344,884.6	106.7	51.0	2474.5	104.0	12,926.5
Nonperforming loans / Gross loans	0.0	3389.1	7.7	7.5	21.6	120.8	18,576.4
Net charge-off / Average gross loans	− 113,000.0	544.7	−6.1	0.0	744.6	−135.7	19,060.4
Net charge-off / Net income before loan loss provision	− 195,715.0	147,373.6	−16.9	−16.9	1924.1	−47.6	6826.0
Impaired loans / Equity	− 21,688.5	59,560.0	49.3	22.2	601.9	60.6	5342.9
Unreserved impaired loans / Equity	− 16,443.6	45,789.3	27.1	16.6	381.3	68.2	7417.2
Tier 1 ratio	−21.0	3155.0	18.0	15.5	22.8	87.0	11,368.7
Total capital ratio	−20.3	1086.9	20.2	17.5	16.6	19.3	752.8
Equity / Total assets	− 14,554.5	99.9	9.7	9.4	116.3	−108.5	12,768.7
Equity / Net loans	− 22,196,817.6	14,072,250.0	1449.7	16.0	166,784.6	−46.4	11,597.0
Equity / Deposit & short-term funding	−2364.7	13,990,200.0	1789.6	11.3	96,991.4	107.7	14,114.1
Equity / Liabilities	−159.2	278,658.4	34.9	10.4	1648.1	154.7	25,559.7
Net interest margin	− 1,661,541.7	2261.0	−48.5	1.9	9247.0	−179.7	32,286.2
Net interest revenue / Average assets	−51.8	142.0	2.0	1.8	2.5	12.8	460.7
Other operating income / Average assets	−54.6	427.0	2.2	1.0	9.0	16.5	434.2
Noninterest expenses / Average assets	−1399.2	428.0	3.5	2.3	11.4	−52.3	7346.0
Pre-tax operating income / Average assets	−1344.8	291.5	0.7	0.6	8.3	−128.1	21,077.0
Nonoperating items & taxes / Average assets	−1417.5	78.7	−0.4	−0.2	10.6	−125.1	15,920.4
Return on Average Assets (ROAA)	−4677.9	291.5	0.2	0.3	27.5	−156.9	26,141.7
Return on Average Equity (ROAE)	−4064.6	1765.6	2.6	3.4	37.3	−53.8	5498.7
Nonoperating items / Net income	− 8,134,000.0	56,799.0	−367.9	−4.1	45,440.9	−177.7	31,796.4
Cost to income	−9835.5	44,953.3	74.6	70.6	319.6	97.3	12,795.1
Recurring earning power	−199.7	291.9	1.1	0.8	4.2	17.2	1427.2
Interbank Ratio	0.0	861,644,680.3	76,922.0	80.1	5,325,068.9	141.1	21,852.4
Net loans / Total assets	0.0	99.9	57.9	61.0	20.4	−0.8	0.6
Net loans / Total deposit & borrowings	0.0	113,250.0	77.1	71.4	671.9	153.6	25,230.0
Net loans / Deposit & short-term funding	0.0	33,937,000.0	1960.6	73.4	204,731.9	148.6	23,786.1
Liquid assets / Deposit & short-term funding	0.0	15,262,531.6	2036.5	17.6	141,533.6	98.2	9967.7
Liquid assets / Total deposit & borrowings	0.0	595,200.0	111.5	14.9	5359.5	84.7	7931.7

This table analyzes descriptive statistics of applied financial ratios as continuous variables before categorization. It can be seen from various excessive minimum and maximum values that outliers are present in the database, which was deemed necessary for handling in subsequent modeling steps.

database or enhance the classification power of ML methods. This approach was primarily due to sampling bias concerns and efforts to realistically estimate the probability of failure (PF) values obtained from the models. Hence, the results of this study indicate that traditionally applied methods arrive at reliable models with high classification accuracy, even if employed in a low failure rate database.

The models were developed using logit, C5.0 DT, and DL-NN techniques, and their classification power was evaluated by the widely applied AUROC performance indicator.

5.1. The logit model

Since logit methodology is well recognized in failure prediction studies (see, inter alia, Lizares and Bautista, 2021), no detailed description is deemed necessary within the framework of this study. The logit model was developed using the forward stepwise method relative to flag input variables. The entry criterion was specified by the 0.05 Wald test, and the removal criterion was set at 0.1. Table 10 summarizes the design of the 11-variate logit model whereby it can be concluded that the model variables and definitions of their parameters meet all statistical and professional requirements. Based on the results of the Wald test and β coefficients, the strongest variable for explaining bank failure is the favorable category of the return on average equity (ROAE) ratio, followed by the unfavorable category of the equity / total assets ratio, and then the most favorable category of the loan loss provision / net interest revenue ratio.

The result of the Omnibus chi-squared test is 872.735 ($p = 0.000$), thus confirming the statistical significance of the model. The Hosmer–Lemeshow test resulted in a chi-squared value of 9.904 and a p-value of 0.272. As typically found in most goodness of fit tests, small p-values (usually below 5%) would infer that the model does not have a good fit, which is not the case in this study. The −2 log-likelihood test resulted in a value of 1555.840, again indicating a relatively good model fit. The AUROC results of 93.3% for the training set and 92.7% for the testing set indicate remarkable predictive power. Fig. 1 illustrates the ROC curves for the logit model. Notably, the development of a logit model using the original continuous variables was also attempted, but results were unfavorable by comparison.

5.2. The C5.0 DT model

The C5.0 DT method is an advanced version of the commonly applied C4.5 DT version with additional boosting capability used to

Table 7
Critical intervals and failure statistics of categorized variables.

Categorized variable (decision rule)	Result of decision rule (1: if true, 0: if not true)					
	0			1		
	Nonfailed	Failed	Failure rate	Nonfailed	Failed	Failure rate
Loan_loss_reserves_Gross_loans ≤ 4.095	14,489	263	1.783%	17,495	40	0.228%
4.095 < Loan_loss_reserves_Gross_loans ≤ 7.854	20,603	161	0.775%	11,381	142	1.232%
7.854 < Loan_loss_reserves_Gross_loans	28,876	182	0.626%	3108	121	3.747%
Loan_loss_provision_Net_interest_revenue ≤ 25.698	6240	218	3.376%	25,744	85	0.329%
25.698 < Loan_loss_provision_Net_interest_revenue ≤ 48.929	28,814	243	0.836%	3170	60	1.858%
48.929 < Loan_loss_provision_Net_interest_revenue	28,914	145	0.499%	3070	158	4.895%
Loan_loss_reserves_NPL ≤ 39.283	25,568	262	1.014%	6416	41	0.635%
39.283 < Loan_loss_reserves_NPL ≤ 50.995	17,576	118	0.667%	14,408	185	1.268%
50.995 < Loan_loss_reserves_NPL	20,824	226	1.074%	11,160	77	0.685%
NPL_Gross_loans ≤ 7.664	15,736	277	1.730%	16,248	26	0.160%
7.664 < NPL_Gross_loans ≤ 15.008	19,353	149	0.764%	12,631	154	1.205%
15.008 < NPL_Gross_loans	28,879	180	0.619%	3105	123	3.810%
Impaired_loans_Equity ≤ 101.584	3113	116	3.592%	28,871	187	0.644%
101.584 < Impaired_loans_Equity	28,871	187	0.644%	3113	116	3.592%
Unreserved_impaired_loans_Equity ≤ 51.631	3112	117	3.623%	28,872	186	0.640%
51.631 < Unreserved_impaired_loans_Equity	28,872	186	0.640%	3112	117	3.623%
Tier1_ratio ≤ 11.596	28,868	206	0.709%	3116	97	3.019%
11.596 < Tier1_ratio	3116	97	3.019%	28,868	206	0.709%
Total_capital_ratio ≤ 13.400	28,860	197	0.678%	3124	106	3.282%
13.400 < Total_capital_ratio	3124	106	3.282%	28,860	197	0.678%
Equity_Total_assets ≤ 5.309	28,887	173	0.595%	3097	130	4.029%
5.309 < Equity_Total_assets	3097	130	4.029%	28,887	173	0.595%
Equity_Net_loans ≤ 8.894	28,858	203	0.699%	3126	100	3.100%
8.894 < Equity_Net_loans	3126	100	3.100%	28,858	203	0.699%
Equity_Dep&ST_funding ≤ 6.582	28,855	203	0.699%	3129	100	3.097%
6.582 < Equity_Dep&ST_funding ≤ 23.612	6318	139	2.153%	25,666	164	0.635%
23.612 < Equity_Dep&ST_funding	28,795	264	0.908%	3189	39	1.208%
Equity_Liabilities ≤ 5.632	28,884	175	0.602%	3100	128	3.965%
5.632 < Equity_Liabilities ≤ 19.653	6285	171	2.649%	25,699	132	0.511%
19.653 < Equity_Liabilities	28,799	260	0.895%	3185	43	1.332%
Net_interest_margin ≤ 0.911	28,863	197	0.678%	3121	106	3.285%
0.911 < Net_interest_margin ≤ 1.316	28,794	259	0.891%	3190	44	1.361%
1.316 < Net_interest_margin	6311	150	2.322%	25,673	153	0.592%
Net_interest_revenue_AvgAssets ≤ 0.819	28,870	187	0.644%	3114	116	3.591%
0.819 < Net_interest_revenue_AvgAssets ≤ 1.241	28,800	252	0.867%	3184	51	1.577%
1.241 < Net_interest_revenue_AvgAssets	6298	167	2.583%	25,686	136	0.527%
Non_interest_expenses_AvgAssets ≤ 3.231	6278	180	2.787%	25,706	123	0.476%
3.231 < Non_interest_expenses_AvgAssets ≤ 4.728	28,802	256	0.881%	3182	47	1.456%
4.728 < Non_interest_expenses_AvgAssets	28,888	170	0.585%	3096	133	4.119%
Pretax_op_income_AvgAssets ≤ 0.102	28,925	136	0.468%	3059	167	5.177%
0.102 < Pretax_op_income_AvgAssets	3059	167	5.177%	28,925	136	0.468%
ROAA ≤ 0.013	28,975	76	0.262%	3009	227	7.015%
0.013 < ROAA	3009	227	7.015%	28,975	76	0.262%
ROAE ≤ 0.149	28,994	65	0.224%	2990	238	7.373%
0.149 < ROAE	2990	238	7.373%	28,994	65	0.224%
NonOp_Items_Net_income ≤ 0.000	8797	169	1.885%	23,187	134	0.575%
0.000 < NonOp_Items_Net_income ≤ 6.128	26,319	231	0.870%	5665	72	1.255%
16.128 < NonOp_Items_Net_income	28,852	206	0.709%	3132	97	3.004%
Cost_to_Income ≤ 91.832	3126	102	3.160%	28,858	201	0.692%
91.832 < Cost_to_Income	28,858	201	0.692%	3126	102	3.160%
Recurring_earning_power ≤ 0.198	28,879	178	0.613%	3105	125	3.870%
0.198 < Recurring_earning_power	3105	125	3.870%	28,879	178	0.613%
Interbank_ratio ≤ 106.762	12,826	90	0.697%	19,158	213	1.100%
106.762 < Interbank_ratio	19,158	213	1.100%	12,826	90	0.697%
Net_loans_Total_Assets ≤ 29.649	28,823	235	0.809%	3161	68	2.106%
29.649 < Net_loans_Total_Assets ≤ 65.030	16,000	144	0.892%	15,984	159	0.985%
65.030 < Net_loans_Total_Assets ≤ 74.347	25,575	255	0.987%	6409	48	0.743%
74.347 < Net_loans_Total_Assets	25,554	275	1.065%	6430	28	0.434%
Net_loans_Total_Dep&Borr ≤ 46.207	28,821	237	0.816%	3163	66	2.044%
46.207 < Net_loans_Total_Dep&Borr ≤ 81.736	9582	104	1.074%	22,402	199	0.880%
81.736 < Net_loans_Total_Dep&Borr	25,565	265	1.026%	6419	38	0.589%
Net_loans_Dep&ST_funding ≤ 39.255	28,812	247	0.850%	3172	56	1.735%
39.255 < Net_loans_Dep&ST_funding ≤ 92.878	9560	125	1.291%	22,424	178	0.788%
92.878 < Net_loans_Dep&ST_funding	25,596	234	0.906%	6388	69	1.069%
Liquid_assets_Dep&ST_funding ≤ 60.365	3162	66	2.045%	28,822	237	0.816%

(*continued on next page*)

Table 7 (*continued*)

Categorized variable (decision rule)	Result of decision rule (1: if true, 0: if not true)					
	0			1		
	Nonfailed	Failed	Failure rate	Nonfailed	Failed	Failure rate
60.365 < Liquid_assets_Dep&ST_funding	28,822	237	0.816%	3162	66	2.045%
Liquid_assets_Total_Dep&Borr ≤ 12.434	22,417	185	0.819%	9567	118	1.218%
12.434 < Liquid_assets_Total_Dep&Borr ≤ 39.644	12,734	179	1.386%	19,250	124	0.640%
39.644 < Liquid_assets_Total_Dep&Borr	28,817	242	0.833%	3167	61	1.890%

This table presents the results of variable categorization in the form of decision rules generated by CHAID decision trees. Critical intervals of the financial ratios were estimated by univariate CHAID decision trees, ratio by ratio, with optimal binning points determined with the chi-squared homogeneity test. A true or false answer applies to each decision rule, from which flag (1/0) variables were created item by item. The differences between failure ratios per decision rule are worth analyzing, especially in cases where the failure ratio is substantially higher or lower than the 0.938% failure rate characterizing the entire database. The application of categorized variables generally contributes to better model performance and stability.

Table 8
Breakdown of the database after partitioning.

Partition	Nonfailed	Failed	Total	Failure rate
Training subset	22,330	215	22,545	0.953%
Testing subset	9654	88	9742	0.903%
Total	**31,984**	**303**	**32,287**	**0.938%**

This table presents the distribution of the database after partitioning the bank-year observations into training and testing subsets completed by random sampling. To validate the performance of the bank failure prediction models, it was essential to separate observations into a testing subset not used in the phase of model development. As a result of partitioning, 70% of the bank-year observations (22,545) made up the training subset, and the remaining 30% (9742) made up the testing subset.

enhance its performance and thus render it to be a suitable choice for failure prediction (Pang and Gong, 2009). It uses information gain as a splitting criterion to grow trees in the model. The lowest-level categories are effectively pruned according to the specified criteria to prevent overtraining and excessive tree depth. A winnowing process is conducted as an initial modeling step. After producing the first tree, the C5.0 method iterates weights using the AdaBoost technique, and weighted decision rules are generated from subsequent iteration results. The boosting technique builds multiple DTs during the training process. The final probability estimate for each instance is derived from the mean of probabilities provided by the multiple trees.

No conclusive result was obtained using CHAID split categorical variables after several experiments had been conducted. Accordingly, the original continuous ratios represented the range of input variables, and 100 boosted DTs were constructed as ensemble classifiers in the training subset. Pruning severity and a limitation of minimum records per subsequent tree branch were applied to prevent overtraining. The process determined the extent to which the generated DTs were pruned, with the value set to 50. The minimum records per subsequent tree branch value was set to 30.

A set of related models are created when a boosted C5.0 DT model is developed. For this study, 100 different rulesets were generated from 100 boosted trees. To use the model in practice, all 100 rulesets must be run on the data of a bank to result in 100 different PF values, and the mean of the 100 PF values provides the final estimation. The model rule browser for a boosted C5.0 model shows the list of models in hierarchical order to interpret results, the estimated accuracy of each model, and the overall accuracy of the ensemble of all boosted models.

It would be impossible to present all 100 DTs individually within the framework of a single study. In summary, the tree depth value typically varied between 4 and 12, and the range of variables and splitting values displayed strong diversity. In general, the resulting financial ratio structures provided multisided relationship systems with a means to explain and estimate bank failure in different forms, which presents a considerable advantage over conventional approaches.

The resulting AUROC score was 99.7% in the training subset and 93.7% in the testing subset, indicating performance superior to that of the logit model. The difference in classification accuracy between the two subsets might indicate overtraining, but the strong accuracy attributed to the testing subset indicates outstanding model performance and is thus strongly convincing. Fig. 2 illustrates the ROC curves of both the training and the testing subsets.

It is somewhat uncontestable that programming 100 rulesets is a more complicated task than calculating PF values from just one logit function. However, since the C5.0 model provides classification accuracy superior to that of logit, the decision of which model to select based on their respective advantages and disadvantages is up to the end user.

5.3. The DL-NN model

DL is a branch of ML, which operates through a series of nonlinear processing units comprising multiple layers for feature transformation and extraction. It possesses several layers of NN to perform the ML process (Schmidhuber, 2015). The first layer of the NN model processes input data and passes information to the second layer. The second layer then processes the information further by

Table 9
Failure statistics of the categorized variables (decision rules) in the training and the testing subsets.

Categorized variable (decision rule)	Training subset Number of observations (if decision rule is true)			Testing subset Number of observations (if decision rule is true)		
	Total	Failed	Failure rate	Total	Failed	Failure rate
Loan_loss_reserves_Gross_loans ≤ 4.095	12253	29	0.237%	5282	11	0.208%
4.095 < Loan_loss_reserves_Gross_loans ≤ 7.854	8089	104	1.286%	3434	38	1.107%
7.854 < Loan_loss_reserves_Gross_loans	2203	82	3.722%	1026	39	3.801%
Loan_loss_provision_Net_interest_revenue ≤ 25.698	18048	57	0.316%	7781	28	0.360%
25.698 < Loan_loss_provision_Net_interest_revenue ≤ 48.929	2272	40	1.761%	958	20	2.088%
48.929 < Loan_loss_provision_Net_interest_revenue	2225	118	5.303%	1003	40	3.988%
Loan_loss_reserves_NPL ≤ 39.283	4486	31	0.691%	1971	10	0.507%
39.283 < Loan_loss_reserves_NPL ≤ 50.995	10191	129	1.266%	4402	56	1.272%
50.995 < Loan_loss_reserves_NPL	7868	55	0.699%	3369	22	0.653%
NPL_Gross_loans ≤ 7.664	11365	18	0.158%	4909	8	0.163%
7.664 < NPL_Gross_loans ≤ 15.008	8976	107	1.192%	3809	47	1.234%
15.008 < NPL_Gross_loans	2204	90	4.083%	1024	33	3.223%
Impaired_loans_Equity ≤ 101.584	20289	131	0.646%	8769	56	0.639%
101.584 < Impaired_loans_Equity	2256	84	3.723%	973	32	3.289%
Unreserved_impaired_loans_Equity ≤ 51.631	20323	131	0.645%	8735	55	0.630%
51.631 < Unreserved_impaired_loans_Equity	2222	84	3.780%	1007	33	3.277%
Tier1_ratio ≤ 11.596	2217	69	3.112%	996	28	2.811%
11.596 < Tier1_ratio	20328	146	0.718%	8746	60	0.686%
Total_capital_ratio ≤ 13.400	2269	70	3.085%	961	36	3.746%
13.400 < Total_capital_ratio	20276	145	0.715%	8781	52	0.592%
Equity_Total_assets ≤ 5.309	2238	94	4.200%	989	36	3.640%
5.309 < Equity_Total_assets	20307	121	0.596%	8753	52	0.594%
Equity_Net_loans ≤ 8.894	2233	70	3.135%	993	30	3.021%
8.894 < Equity_Net_loans	20312	145	0.714%	8749	58	0.663%
Equity_Dep&ST_funding ≤ 6.582	2258	73	3.233%	971	27	2.781%
6.582 < Equity_Dep&ST_funding ≤ 23.612	18088	117	0.647%	7742	47	0.607%
23.612 < Equity_Dep&ST_funding	2199	25	1.137%	1029	14	1.361%
Equity_Liabilities ≤ 5.632	2242	94	4.193%	986	34	3.448%
5.632 < Equity_Liabilities ≤ 19.653	18120	91	0.502%	7711	41	0.532%
19.653 < Equity_Liabilities	2183	30	1.374%	1045	13	1.244%
Net_interest_margin ≤ 0.911	2248	74	3.292%	979	32	3.269%
0.911 < Net_interest_margin ≤ 1.316	2209	31	1.403%	1025	13	1.268%
1.316 < Net_interest_margin	18088	110	0.608%	7738	43	0.556%
Net_interest_revenue_AvgAssets ≤ 0.819	2234	81	3.626%	996	35	3.514%
0.819 < Net_interest_revenue_AvgAssets ≤ 1.241	2229	34	1.525%	1006	17	1.690%
1.241 < Net_interest_revenue_AvgAssets	18082	100	0.553%	7740	36	0.465%
Non_interest_expenses_AvgAssets ≤ 3.231	18067	92	0.509%	7762	31	0.399%
3.231 < Non_interest_expenses_AvgAssets ≤ 4.728	2234	26	1.164%	995	21	2.111%
4.728 < Non_interest_expenses_AvgAssets	2244	97	4.323%	985	36	3.655%
Pretax_op_income_AvgAssets ≤ 0.102	2229	114	5.114%	997	53	5.316%
0.102 < Pretax_op_income_AvgAssets	20316	101	0.497%	8745	35	0.400%
ROAA ≤ 0.013	2224	164	7.374%	1012	63	6.225%
0.013 < ROAA	20321	51	0.251%	8730	25	0.286%
ROAE ≤ 0.149	2225	172	7.730%	1003	66	6.580%
0.149 < ROAE	20320	43	0.212%	8739	22	0.252%
NonOp_Items_Net_income ≤ 0.000	16300	89	0.546%	7021	45	0.641%
0.000 < NonOp_Items_Net_income ≤ 6.128	4033	52	1.289%	1704	20	1.174%
16.128 < NonOp_Items_Net_income	2212	74	3.345%	1017	23	2.262%
Cost_to_Income ≤ 91.832	20286	144	0.710%	8773	57	0.650%
91.832 < Cost_to_Income	2259	71	3.143%	969	31	3.199%
Recurring_earning_power ≤ 0.198	2261	82	3.627%	969	43	4.438%
0.198 < Recurring_earning_power	20284	133	0.656%	8773	45	0.513%
Interbank_ratio ≤ 106.762	13515	153	1.132%	5856	60	1.025%
106.762 < Interbank_ratio	9030	62	0.687%	3886	28	0.721%
Net_loans_Total_Assets ≤ 29.649	2240	49	2.188%	989	19	1.921%
29.649 < Net_loans_Total_Assets ≤ 65.030	11352	112	0.987%	4791	47	0.981%
65.030 < Net_loans_Total_Assets ≤ 74.347	4467	32	0.716%	1990	16	0.804%
74.347 < Net_loans_Total_Assets	4486	22	0.490%	1972	6	0.304%
Net_loans_Total_Dep&Borr ≤ 46.207	2260	48	2.124%	969	18	1.858%
46.207 < Net_loans_Total_Dep&Borr ≤ 81.736	15806	141	0.892%	6795	58	0.854%
81.736 < Net_loans_Total_Dep&Borr	4479	26	0.580%	1978	12	0.607%
Net_loans_Dep&ST_funding ≤ 39.255	2249	41	1.823%	979	15	1.532%
39.255 < Net_loans_Dep&ST_funding ≤ 92.878	15850	126	0.795%	6752	52	0.770%
92.878 < Net_loans_Dep&ST_funding	4446	48	1.080%	2011	21	1.044%
Liquid_assets_Dep&ST_funding ≤ 60.365	20306	172	0.847%	8753	65	0.743%

(*continued on next page*)

Table 9 (*continued*)

Categorized variable (decision rule)	Training subset Number of observations (if decision rule is true)			Testing subset Number of observations (if decision rule is true)		
	Total	Failed	Failure rate	Total	Failed	Failure rate
60.365 < Liquid_assets_Dep&ST_funding	2239	43	1.921%	989	23	2.326%
Liquid_assets_Total_Dep&Borr ≤ 12.434	6770	84	1.241%	2915	34	1.166%
12.434 < Liquid_assets_Total_Dep&Borr ≤ 39.644	13527	88	0.651%	5847	36	0.616%
39.644 < Liquid_assets_Total_Dep&Borr	2248	43	1.913%	980	18	1.837%

This table demonstrates that random sampling-based partitioning does not result in sampling bias, as failure statistics of the categorical variables (decision rules) are not significantly different in the training and testing subsets. It is in line with expectations that the 88 failed observations in the testing subset cannot distribute naturally in exactly the same form between the ratio intervals as the 215 failed observations in the training subset; however, it does not lead to a consistency problem.

Table 10
Main features of the logit model.

Model variable (flag value = 1)	β	Standard error	Wald test	p-value	Exp (β)
Loan_loss_provision_Net_interest_revenue ≤ 25.698	−0.998	0.184	29.411	0.000	0.368
50.995 < Loan_loss_reserves_NPL	−0.367	0.180	4.150	0.042	0.693
NPL_Gross_loans ≤ 7.664	−1.285	0.270	22.587	0.000	0.277
Tier1_ratio ≤ 11.596	0.625	0.177	12.466	0.000	1.869
5.309 < Equity_Total_assets	−0.962	0.160	36.114	0.000	0.382
Net_interest_revenue_AvgAssets ≤ 0.819	0.595	0.236	6.378	0.012	1.813
1.241 < Net_interest_revenue_AvgAssets	−0.754	0.221	11.669	0.001	0.470
4.728 < Non_int_expenses_AvgAssets	0.845	0.160	27.957	0.000	2.328
0.149 < ROAE	−2.204	0.192	132.370	0.000	0.110
16.128 < NonOp_Items_Net_income	0.798	0.162	24.279	0.000	2.221
Interbank_ratio ≤ 106.762	0.423	0.165	6.525	0.011	1.526
Constant	−2.110	0.290	52.958	0.000	0.121

This table summarizes the contents and the statistical testing of the logit model developed in the training subset (N = 22,545). Since each input variable is quantified at a 1/0 flag level, the signs and values of the regression parameters can be directly associated with bank failure. The strongest predictor of bank failure is the least favorable category of ROAE, followed by the worst category of equity to total assets.

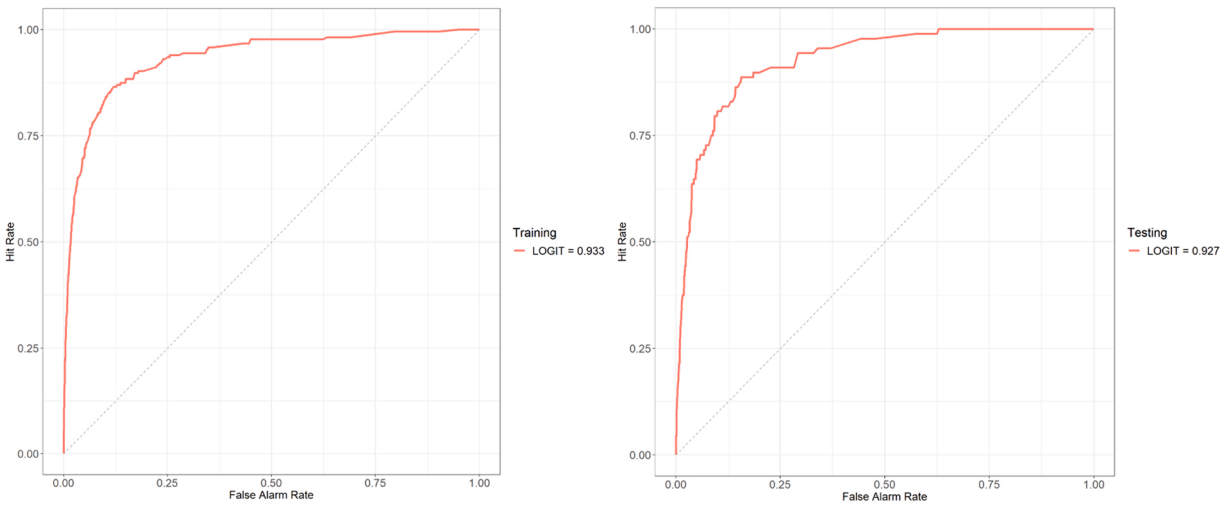

Fig. 1. ROC curves of the logit model in the training and the testing subsets.

adding additional information and passes it on to the next layer. This process continues through all layers of the DL network until the desired result is attained.

A recursive architecture form was applied in this study, with network training parameters chosen to ensure an entirely thorough search of the space to locate the best possible model. The training was accomplished by applying gradient enhancement and automatic learning rate adjustment during stochastic gradient descent.

Several parameters controlled the training process. Principally, the alpha value expressing the momentum terms used in updating weights used during training was initially set at 0.95 to ensure that weight changes occurred in a consistent direction. The eta learning

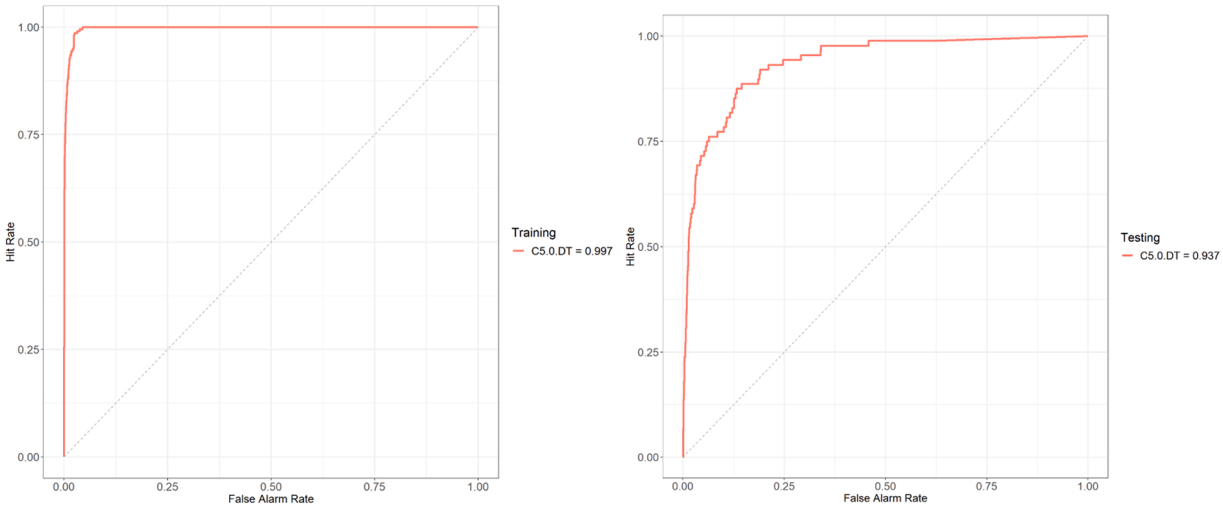

Fig. 2. ROC curves of the C5.0 DT model in the training and the testing subsets.

rate controls the extent to which weights were adjusted at each updating point and changed as training proceeded. During training, eta began at an initial eta point, decreased to low eta values, then was reset to high eta values, and subsequently reverted to low eta values. The two final stages were repeated until training was completed with the initial eta set to 0.4, the high eta to 0.1, the low eta to 0.01, and the eta decay rate to 30.

Three hidden layers were enabled to be used for the DL simulations. In line with the flag variables derived from the CHAID categorization presented in Table 7, the theoretical maximum number of input neurons was set at 69. The theoretical number of neurons in each hidden layer was also maximized at a value of 69 to maintain the structure within a manageable size. Several networks with multiple structures were trained and pruned in a parallel sequence to arrive at the final structure, indicating the best classification power. The process began with a relatively large network with the weakest units represented as neurons with nonrelevant impact in the hidden and input layers pruned as training proceeded.

In the final model, the input layer contained 34 neurons, the first hidden layer had 30 neurons, and the second hidden layer had 17 neurons, with the output layer acting as the failure event containing one single neuron. Despite the deeper architecture formed throughout the training process, no third hidden layer was required to locate the best model. The time used for model building was 7 h and 52 min, demonstrating that training a DL-NN algorithm is somewhat inefficient in terms of time and resource use.

The "black-box" problem of the NN was resolved by iterating the relative importance of input neurons in relation to the model variables. Sensitivity analysis of input fields was performed after the network had been trained to provide information about which input fields were most important for predicting the occurrence of the failure event. Table 11 summarizes the relevant input variables in decreasing order of their importance in the DL-NN model.

AUROC scores of 93.2% in the training set and 93.0% in the testing set were recorded, indicating a strong level of classification power. Fig. 3 presents ROC curves for the DL-NN model.

The radial basis function (RBF) network model was attempted as an alternative method to test whether traditional NNs could perform as well as the DL-NN technique. The number of RBF clusters in the model was set to 20, corresponding to the size of the hidden layer, with the number of cycles for which the network continued to train if no improvement occurred to indicate persistence was set to 50. The alpha value was set to 0.95, and the eta value remained constant and was computed based on the first two cycles. The RBF overlapping parameter was set to 5, and the input layer contained 34 flag variables. After multiple trials, the AUROC score recorded for the best RBF network was 91.1% in both the training and the testing subsets, thus indicating generally strong predictive power. However, performance was not as strong as that produced by the logit, C5.0, and DL-NN models.

Construction of DL-NN and RBF network models was also attempted with the original continuous variables; however, each trial resulted in AUROC scores below 90%.

5.4. Model evaluation and results

From the results, it can reasonably be argued that all models developed in this study demonstrate favorable predictive power and that each is suitable for bank failure prediction. Based on ROC curve illustrations (Fig. 4), it would be difficult to decide on the best model performance, since the curves follow similar paths and intersect at multiple points. Each curve commences with a high gradient and effectively screens out more than 50% of the failed observations at the point where the PF percentiles originate. The C5.0 model seems to outperform the other models somewhat in terms of locating 60–70% of bank failure cases; however, competition between models subsequently assumes a relatively equal basis. Table 12 analyzes the AUROC indicators.

Furthermore, it can be concluded that all three models possess a high level of classification accuracy. In terms of AUROC scores, the C5.0 model exhibited the best performance, with the DL-NN and logit models coming in second and third, respectively.

Table 11
Model variables (input neurons) and their relative importance in the DL-NN model.

Input variable (flag value = 1)	Relative importance
0.149 < ROAE	0.0078966
ROAE ≤ 0.149	0.0064387
16.128 < NonOp_Items_Net_income	0.0048464
Net_interest_revenue_AvgAssets ≤ 0.819	0.0043283
1.241 < Net_interest_revenue_AvgAssets	0.0042112
4.728 < Non_interest_expenses_AvgAssets	0.0041322
5.309 < Equity_Total_assets	0.0040199
15.008 < NPL_Gross_loans	0.0039941
Loan_loss_provision_Net_interest_revenue ≤ 25.698	0.0032047
Cost_to_Income ≤ 91.832	0.0030136
ROAA ≤ 0.013	0.0029452
3.231 < Non_interest_expenses_AvgAssets ≤ 4.728	0.0027769
91.832 < Cost_to_Income	0.0026964
Unreserved_impaired_loans_Equity ≤ 51.631	0.0025733
11.596 < Tier1_ratio	0.0025219
Equity_Total_assets ≤ 5.309	0.0024739
Tier1_ratio ≤ 11.596	0.0023727
106.762 < Interbank_ratio	0.0022776
NPL_Gross_loans ≤ 7.664	0.0022770
51.631 < Unreserved_impaired_loans_Equity	0.0022721
25.698 < Loan_loss_provision_Net_interest_revenue ≤ 48.929	0.0020865
Net_interest_margin ≤ 0.911	0.0020842
50.995 < Loan_loss_reserves_NPL	0.0020575
0.911 < Net_interest_margin ≤ 1.316	0.0018776
39.283 < Loan_loss_reserves_NPL ≤ 50.995	0.0018015
NonOp_Items_Net_income ≤ 0.000	0.0017956
0.013 < ROAA	0.0017432
48.929 < Loan_loss_provision_Net_interest_revenue	0.0017322
Interbank_ratio ≤ 106.762	0.0016935
Liquid_assets_Dep&ST_funding ≤ 60.365	0.0016857
Non_interest_expenses_AvgAssets ≤ 3.231	0.0016840
1.316 < Net_interest_margin	0.0014721
60.365 < Liquid_assets_Dep&ST_funding	0.0013197
Loan_loss_reserves_NPL ≤ 39.283	0.0007726

When interpreting an NN model, resolution of the black-box problem is generally encountered, particularly in iterating relative importance to the model variables after the network has been trained. Relative importance values have been determined within the framework of this study using sensitivity analysis to provide information on which input fields are the most important in predicting failure events.

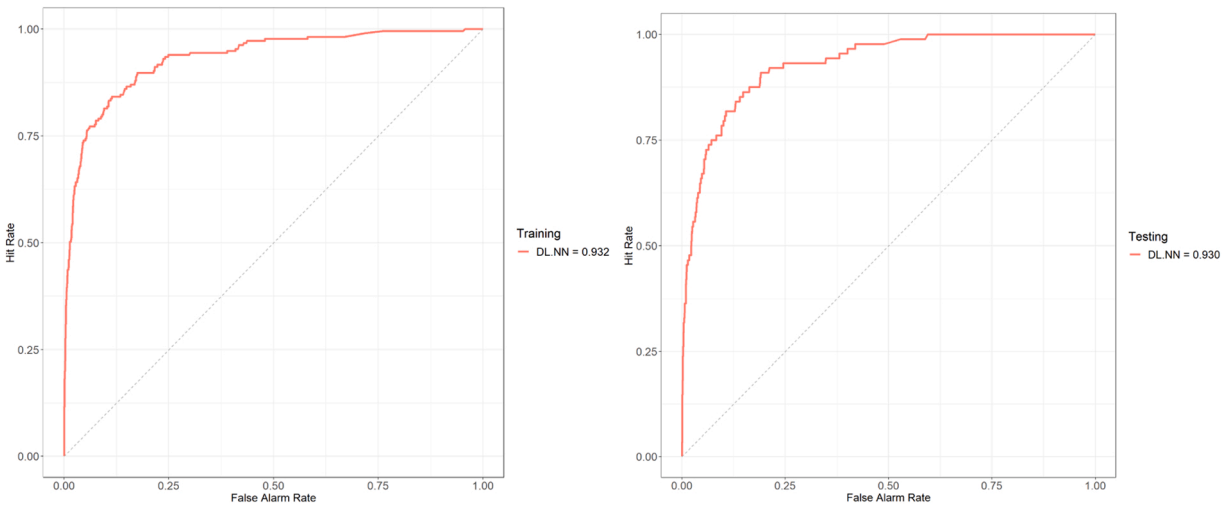

Fig. 3. ROC curves of the DL-NN model in the training and the testing subsets.

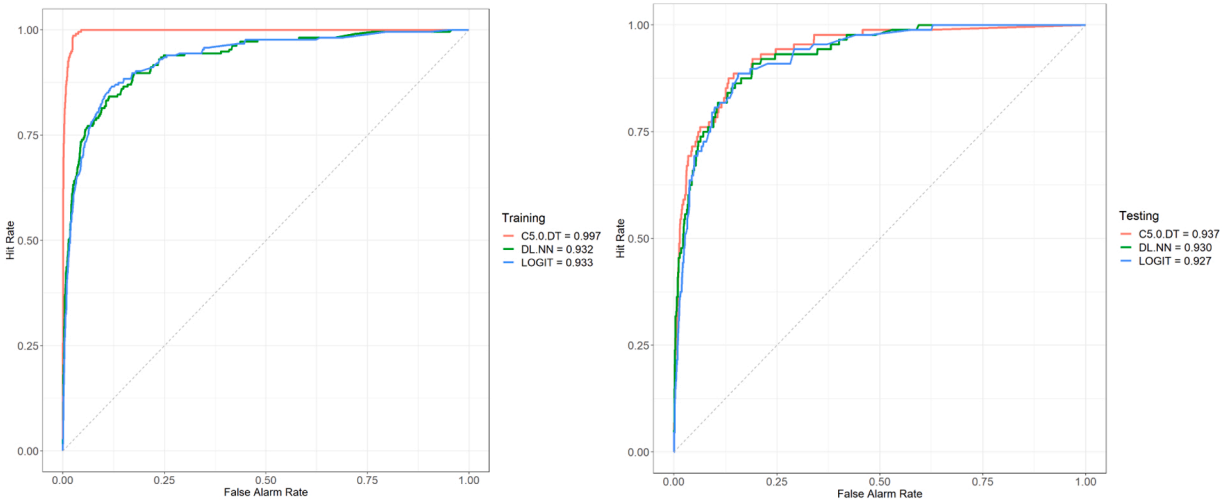

Fig. 4. ROC curves of the three models in the training and the testing subsets.

6. Conclusion

Analysis conducted in this study is highly applicable to current economic uncertainty associated with the COVID-19 pandemic in terms of increased focus on bank failure prediction and its attendant consequences. Given this contextual background, the study also presents highly important and relevant contemporary research in this field.

The comprehensive literature review in this study primarily concludes that historically bank failure prediction has in methodological terms followed a similar development path to that of corporate failure prediction. Moreover, bank failure prediction essentially originated with smaller sample-based linear models, more recently arriving at the use of contemporary ML methods and their creative combinations. It can thus be concluded that ML methods currently dominate bank failure prediction. The most recent studies, primarily conducted in the US, have revealed outstanding results using the NN and ensemble DT classifier methods. However, logit methods have remained popular and have generally been found to be reliably efficient. A research gap was identified in the literature as motivation for this study to perform original empirical research applied to banks based in the EU-27 area. This aimed to examine whether findings are relevant for EU-27 banks and provide new empirical models to accomplish efficient bank failure prediction.

Relevant data of EU-27 banks were extracted from BankFocus records to conduct empirical research. In this process, the financial data of 5614 banking entities were extracted for the 2011–2019 period, resulting in 32,287 bank-year observations. Moreover, 303 bank failure events were identified between 2012 and 2020, resulting in a 0.938% failure rate in the database. Input and output variables were created in line with industry best practices. The CHAID technique was used to generate categorical variables from continuous financial ratios. This was necessary to improve model performance and stability, handle outliers, and manage nonlinear and nonmonotonous relationships.

Subsequently, multivariate classification models were developed using logit, C5.0 DT, and DL-NN methods. It can ultimately be concluded that all three models possess high classification accuracy based on the AUROC scores in the testing subset. Moreover, by ranking the AUROC scores, the C5.0 DT model demonstrates the best performance level (0.937), the DL-NN model the second-best (0.930), and the logit model the third-best (0.927). Based on the results, this study justifies using the ensemble DT method because it outperformed the other techniques for predictive power. However, this study simultaneously reveals that the conventional logit method might also perform fairly efficiently, given that its relatively high accuracy of 92.7% provides little methodological cause for concern. Hence, the EU-based empirical results of this study corroborate earlier findings from the US.

Regarding the relevance of input variables used to predict bank failure, noticeable experiences in the EU-27 area are similar to those of previous empirical studies conducted in the US. The logit model's earnings and capital adequacy ratios were the strongest

Table 12
AUROC based model performance indicators of the three models.

Model	AUROC	Standard error	Asymptotic significance	Asymptotic 95% confidence interval	
				Lower bound	Upper bound
Logit	0.927	0.013	0.000	0.902	0.951
C5.0	0.937	0.013	0.000	0.912	0.962
DL-NN	0.930	0.013	0.000	0.906	0.955

This table summarizes the AUROC indicators achieved by the three models in the testing subset, demonstrating the differences in predictive power. It can be concluded that all models display favorable predictive power, and each is suitable for bank failure prediction. However, when ranking the AUROC scores, it can be determined that C5.0 indicates the best performance, DL-NN the second, and logit the third, respectively.

predictors, followed by asset quality. Similarly, when the DL-NN model was applied, earnings provided the highest contribution, followed by management capability and capital adequacy indicators. Interpretation of the C5.0 DT model was rendered somewhat more complicated by the different rulesets generated by the 100 boosted trees. By evaluating the uppermost variable per tree, capital adequacy ratios dominated 36, earnings ratios 24, management capability indicators 19, and asset quality indicators 9 rulesets, respectively. It can thus be concluded that liquidity and sensitivity ratios were present in the operational background in all deployed models.

The empirical results of this study have policy implications for bank supervisory authorities, bank executives, risk management professionals, and policymakers working in finance. The developed models can be used to recognize bank weaknesses in time to take appropriate mitigating actions. The article also contributes to the literature by examining the drivers of bank failure through state-of-the art ML techniques and demonstrating the benefits of ensemble DTs for developing additional predictive power.

Research conducted in this study provides pointers to possible future research directions which may be formulated in the accomplishment of panel data econometric modeling. Research results may also be extended to global banking data and other ML-based techniques to enrich the overall methodological experience. They may also be used to conduct research applicable to the economic impact of the COVID-19 pandemic when measurable bank failure events are deemed to have been derived from it.

CRediT authorship contribution statement

Tamás Kristóf: Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Miklós Virág:** Conceptualization, Investigation, Validation, Writing – original draft.

Declarations of interest

None.

Funding

This research project did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abdesslem, R.B., Chkir, I., Dabbou, H., 2022. Is managerial ability a moderator? The effect of credit risk and liquidity risk on the likelihood of bank default. Int. Rev. Financ. Anal. 80, 102044 https://doi.org/10.1016/j.irfa.2022.102044.

Aldasoro, I., Cho, C.H., Park, K., 2022. Bank solvency risk and funding cost interactions: Evidence from Korea. J. Bank. Financ. 134 (1), 106348 https://doi.org/10.1016/j.jbankfin.2021.106348.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E., Suvas, A., 2017. Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-Score model. J. Int. Financ. Manag. Account. 28 (2), 131–171. https://doi.org/10.1111/jifm.12053.

Appiahene, P., Missah, Y.M., Najim, U., 2020. Predicting bank operational efficiency using machine learning algorithm: Comparative study of decision tree, random forest, and neural networks. Adv. Fuzzy Syst. 2020 (5), 8581202 https://doi.org/10.1155/2020/8581202.

Audrino, F., Kostrov, A., Ortega, J.-P., 2019. Predicting U.S. bank failures with MIDAS logit models. J. Financ. Quant. Anal. 54 (6), 2575–2603. https://doi.org/10.1017/S0022109018001308.

Avino, D.E., Conlon, T., Cotter, J., 2019. Credit default swaps as indicators of bank financial distress. J. Int. Money Financ. 94 (C), 132–139. https://doi.org/10.1016/j.jimonfin.2019.03.001.

Beltratti, A., Stulz, R., 2012. The credit crisis around the globe: why did some banks perform better? J. Financ. Econ. 105 (1), 1–17. https://doi.org/10.1016/j.jfineco.2011.12.005.

Berger, A.N., et al., 2021. The effects of cultural values on bank failures around the world. J. Financ. Quant. Anal. 56 (3), 945–993. https://doi.org/10.1017/S0022109020000150.

Betz, F., Oprica, S., Peltonen, T., Sarlin, P., 2013. Predicting distress in European banks. Working Paper Series No. 1597 / October 2013, European Central Bank, Frankfurt. ⟨https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2338998⟩.

Beutel, J., List, S., von Schweinitz, G., 2019. Does machine learning help us predict banking crises? J. Financ. Stab. 45 (C), 100693 https://doi.org/10.1016/j.jfs.2019.100693.

Boubaker, S., Cellier, A., Manita, R., Saeed, A., 2020. Does corporate social responsibility reduce financial distress risk? Econ. Model. 91 (8), 835–851. https://doi.org/10.1016/j.econmod.2020.05.012.

Boyacioglu, M.A., Kara, Y., Baykan, Ö.K., 2009. Predicting bank financial failures using neural network, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. Expert Syst. Appl. 36 (2), 3355–3366. https://doi.org/10.1016/j.eswa.2008.01.003.

Bräuning, M., Malikkidou, D., Scalone, S., Scricco, G. , 2019. A new approach to early warning systems for small European banks. Working Paper Series No 2348 / December 2019. European Central Bank, Frankfurt. ⟨https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2348~351ba1be4c.en.pdf⟩.

Carmona, P., Climent, F., Momparler, A., 2019. Predicting failure in the US banking sector: an extreme gradient boosting approach. Int. Rev. Econ. Financ. 61 (C), 304–323. https://doi.org/10.1016/j.iref.2018.03.008.

Chen, W.-D., Chen, Y., Huang, S.-Y., 2021. Liquidity risk and bank performance during financial crises. J. Financ. Stab. 56 (C), 100906 https://doi.org/10.1016/j.jfs.2021.100906.

Climent, F., Momparler, A., Carmona, P., 2019. Anticipating bank distress in the Eurozone: an extreme gradient boosting approach. J. Bus. Res. 101 (C), 885–896. https://doi.org/10.1016/j.jbusres.2018.11.015.

Cole, R.A., White, L.J., 2012. Déjà vu all over again: the causes of U.S. commercial bank failures this time around. J. Financ. Serv. Res. 42 (1), 5–29. https://doi.org/10.1007/s10693-011-0116-9.

Cox, R.A.K., Kimmel, R.K., Wang, G.W.Y., 2017. Proportional hazards model of bank failure: evidence from USA. J. Econ. Financ. Stud. 5 (3), 35–45. https://doi.org/10.18533/jefs.v5i3.290.

Duan, Y., Goodell, J.W., Li, H., Li, X., 2021. Assessing machine learning for forecasting economic risk: evidence from an expanded Chinese financial information set. Financ. Res. Lett., 102273 https://doi.org/10.1016/j.frl.2021.102273.

Eberhardt, M., Presbitero, A., 2021. Commodity prices and banking crises. J. Int. Econ. 131, 103474 https://doi.org/10.1016/j.jinteco.2021.103474.

ElBannan, M.A., 2021. On the prediction of financial distress in emerging markets: what matters more? Empirical evidence from Arab spring countries. Emerg. Mark. Rev. 47 (June), 100806 https://doi.org/10.1016/j.ememar.2021.100806.
Elekdag, S., Malik, S., Mitra, S., 2020. Breaking the bank? A probabilistic assessment of euro area bank profitability. J. Bank. Financ. 120 (C), 105949 https://doi.org/10.1016/j.jbankfin.2020.105949.
Fahlenbrach, R., Prilmeier, R., Stulz, R.M., 2012. This time is the same: using bank performance in 1998 to explain bank performance during the recent financial crisis. J. Financ. 67 (6), 2139–2185. https://doi.org/10.1111/j.1540-6261.2012.01783.x.
Gabbi, G., Levich, R., 2019. Controlling risks to ensure financial stability and reducing volatility. J. Int. Financ. Manag. Account. 30 (3), 183–187. https://doi.org/10.1111/jifm.12107.
Gerhardt, M., Vennet, R.V., 2017. Bank bailouts in Europe and bank performance. Financ. Res. Lett. 22, 74–80. https://doi.org/10.1016/j.frl.2016.12.028.
Gogas, P., Papadimitriou, T., Agrapetidou, A., 2018. Forecasting bank failures and stress testing: a machine learning approach. Int. J. Forecast. 34 (3), 440–455. https://doi.org/10.1016/j.ijforecast.2018.01.009.
Gropp, R., Vesala, J., Vulpes, G., 2006. Equity and bond market signals as leading indicators of bank fragility. J. Money Credit Bank. 38 (2), 399–428. https://doi.org/10.1353/mcb.2006.0032.
Harada, K., Ito, T., Takahashi, S., 2013. Is the distance to default a good measure in predicting bank failures? A case study of Japanese major banks. Jpn. World Econ. 27, 70–82. https://doi.org/10.1016/j.japwor.2013.03.007.
Hasanin, T., Khoshgoftaar, T.M., Leevy, J.L., Seliya, N., et al., 2019. Examining characteristics of predictive models with imbalanced big data. J. Big Data 6 (69), 1–21. https://doi.org/10.1186/s40537-019-0231-2.
Jing, Z., Fang, Y., 2018. Predicting US bank failures: a comparison of logit and data mining models. J. Forecast. 37 (2), 235–256. https://doi.org/10.1002/for.2487.
Kolari, J., Glennon, D., Shin, H., Caputo, M., 2002. Predicting large US commercial bank failures. J. Econ. Bus. 54 (4), 361–387. https://doi.org/10.1016/S0148-6195(02)00089-9.
Kolari, J., López-Iturriaga, F.J., Sanz, I.P., 2019. Predicting European banks stress tests: survival of the fittest. Glob. Financ. J. 39 (C), 44–57. https://doi.org/10.1016/j.gfj.2018.01.015.
Le, H.H., Viviani, J.-L., 2018. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Res. Int. Bus. Financ. 44 (C), 16–25. https://doi.org/10.1016/j.ribaf.2017.07.104.
Li, W., 2013. A default model under macroeconomic conditions. J. Fixed Income 23 (2), 98–113. https://doi.org/10.3905/jfi.2013.23.2.098.
Lizares, R.M., Bautista, C.C., 2021. Corporate financial distress: the case of publicly listed firms in an emerging market economy. J. Int. Financ. Manag. Account. 32 (1), 5–20. https://doi.org/10.1111/jifm.12122.
Maghyereh, A.I., Awartani, B., 2014. Bank distress prediction: empirical evidence from the Gulf Cooperation Council countries. Res. Int. Bus. Financ. 30 (C), 126–147. https://doi.org/10.1016/j.ribaf.2013.07.001.
Manthoulis, G., Michalis, S., Constantin, Z., Emilios, G., 2020. An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks. Eur. J. Oper. Res. 282 (2), 786–801. https://doi.org/10.1016/j.ejor.2019.09.040.
Martin, D., 1977. Early warning of bank failure: a logit regression approach. In: J. Bank. Financ., 1, pp. 249–276. https://doi.org/10.1016/0378-4266(77)90022-X.
Nagel, S., Purnanandam, A., 2020. Banks' risk dynamics and distance to default. Rev. Financ. Stud. 33 (6), 2421–2467. https://doi.org/10.1093/rfs/hhz125.
Nguyen, A.H., Nguyen, H.T., Pham, H.T., 2020. Applying the CAMEL model to assess performance of commercial banks: empirical evidence from Vietnam. Banks Bank. Syst. 15 (2), 177–186. https://doi.org/10.21511/bbs.15(2).2020.16.
Pang, S.-L., Gong, J.-Z., 2009. C5.0 classification classification algorithm and application on individual credit evaluation of banks. Syst. Eng. Theory Pract. 29 (12), 94–104. https://doi.org/10.1016/S1874-8651(10)60092-0.
Petropoulos, A., Siakoulis, V., Stavroulakis, V., Vlachogiannakis, N.E., 2020. Predicting bank insolvencies using machine learning techniques. Int. J. Forecast. 36 (3), 1092–1113. https://doi.org/10.1016/j.ijforecast.2019.11.005.
Pham, X.T.T., Ho, T.H., 2021. Using boosting algorithms to predict bank failure: an untold story. Int. Rev. Econ. Financ. 76 (November), 40–54. https://doi.org/10.1016/j.iref.2021.05.005.
Poghosyan, T., Čihák, M., 2009. Distress in European banks: an analysis based on a new data set. IMF Working Paper WP/09/9. International Monetary Fund, Washington. ⟨https://www.imf.org/external/pubs/ft/wp/2009/wp0909.pdf⟩.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61 (1), 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
Shrivastava, S., Jevanthi, P.M., Singh, S., 2020. Failure prediction of Indian banks using SMOTE, Lasso regression, bagging and boosting. Cogent Econ. Financ. 8 (1), 1729569 https://doi.org/10.1080/23322039.2020.1729569.
Shumway, T., 2001. Forecasting bankruptcy more accurately: a simple hazard model. J. Bus. 74 (1), 101–124. https://doi.org/10.1086/209665.
Siakoulis, V., 2015. Modeling bank default intensity in the USA using autoregressive duration models. MPRA Paper. University Library of Munich, Munich, p. *64526*. ⟨https://mpra.ub.uni-muenchen.de/64526/1/MPRA_paper_64526.pdf⟩.
Sinkey, J.F., 1975. A multivariate statistical analysis of the characteristics of problem banks. J. Financ. 30 (1), 21–36. https://doi.org/10.1111/j.1540-6261.1975.tb03158.x.
Tam, K.Y., Kiang, M.Y., 1992. Managerial applications of neural networks: the case of bank failure predictions. Manag. Sci. 38 (7), 926–947. https://doi.org/10.1287/mnsc.38.7.926.
Tan, Y., Anchor, J., 2017. The impacts of risk-taking behaviour and competition on technical efficiency: evidence from the Chinese banking industry. Res. Int. Bus. Financ. 41 (C), 90–104. https://doi.org/10.1016/j.ribaf.2017.04.026.
Tanaka, K., Kinkyo, T., Hamori, S., 2016. Random forests-based early warning system for bank failures. Econ. Lett. 148 (C), 118–121. https://doi.org/10.1016/j.econlet.2016.09.024.
Wang, D., Quek, C., See, N.G., 2016. Bank failure prediction using an accurate and interpretable neural fuzzy inference system. AI Commun. 29 (4), 477–495. https://doi.org/10.3233/AIC-160702.
Wang, G.W.Y., Cox, R.A.K., 2013. Risk taking by US banks led to their failures. Int. J. Financ. Serv. Manag. 6 (1), 39–59. https://doi.org/10.1504/IJFSM.2013.052877.