

-RESEARCH ARTICLE-

## MACHINE LEARNING ALGORITHMS FOR EARLY WARNING SYSTEMS: PREDICTING SYSTEMIC FINANCIAL CRISES THROUGH NON-LINEAR ECONOMETRIC MODELS

**Shubin Yin**

Warsaw School of Economics, Poland

ORCID: <https://orcid.org/0009-0006-6765-0907>

Email: [uccshubinyin@outlook.com](mailto:uccshubinyin@outlook.com)

### —Abstract—

In view of the severe output and welfare losses caused by past systemic crises, including those of the Great Depression and the 2008 global financial collapse, there is a pressing need for early warning systems that effectively capture nonlinear risk dynamics. This study proposes a machine learning framework for predicting systemic financial crises, based on macroeconomic data from 35 countries spanning 1970 to 2022. Utilising ensemble methods—specifically random forest and XGBoost—the approach significantly enhances early warning capabilities compared to conventional linear models. Incorporating 78 economic and financial indicators, the methodology applies advanced feature engineering and selection to identify complex systemic risk patterns. The models demonstrate strong predictive performance, achieving an AUROC score of 0.97 and substantially outperforming logistic regression. Key predictors include GDP growth, trade volumes, investment trends, and demographic variables. The findings highlight critical nonlinear relationships and interactions among economic indicators, offering deeper insights into financial system vulnerabilities. This work underscores the potential of machine learning to support more intelligent, forward-looking financial risk assessment and policy design.

**Keywords:** Financial Crisis Prediction, Machine Learning, Random Forest, XGBoost, Economic Indicators, Risk Assessment, Ensemble Methods, Systemic Risk.

### INTRODUCTION

Financial crises have manifested repeatedly with concerning regularity and intensity,

Citation (APA): Yin, S. (2024). Machine Learning Algorithms for Early Warning Systems: Predicting Systemic Financial Crises Through Non-Linear Econometric Models. *International Journal of Economics and Finance Studies*, 16(04), 286-311. doi: 10.34109/ijefs.202416414

imposing lasting damage on economies and societies. Among the most devastating was the Great Depression of 1929–33, during which industrial output in the United States plummeted by nearly 50%, GDP contracted by almost 30%, and unemployment rates surpassed 25%, leading to extensive banking failures and prolonged deflationary pressure (Eichengreen, 1992; Reinhart & Rogoff, 2009). Decades later, the Asian Financial Crisis of 1997–98 exposed vulnerabilities within capital-inflow-fuelled growth cycles. Initiated by the devaluation of Thailand's currency in July 1997, the crisis rapidly extended to economies such as Indonesia and South Korea, resulting in currency depreciations exceeding 50% and GDP contractions averaging 6% across affected nations (Kaminsky & Reinhart, 1999; Radelet & Sachs, 2000).

The Global Financial Crisis of 2007–09, which originated in the U.S. subprime mortgage sector, highlighted the transmission of systemic risk through interconnected banking and shadow-banking frameworks. Within eighteen months, global household wealth declined by over USD 10 trillion, while sovereign bond spreads in advanced economies widened considerably, necessitating coordinated interventions by central banks (Brunnermeier, 2009; Laeven & Valencia, 2013). Between 1980 and 2020, more than 120 systemic banking crises were recorded globally, with average output losses surpassing 30% of GDP in the affected jurisdictions (Laeven & Valencia, 2020). These events collectively underscore the critical need for reliable mechanisms to identify systemic risk accumulation prior to market disruption.

To address such recurring disruptions, early warning systems (EWS) have been developed to detect the early onset of financial instability. Traditional EWS approaches encompass signal-extraction techniques, whereby specific thresholds are defined for key indicators to generate alerts, as well as binary choice models such as logit and probit regressions. For instance, Kaminsky and Reinhart (1999) applied a signal-based methodology across twenty countries from 1970 to 1994 and found that variations in foreign exchange reserves and real interest rate differentials were effective predictors, correctly signalling crises with over 75% true-positive rates, albeit accompanied by frequent false alarms during stable periods. Bussiere and Fratzscher (2006) enhanced this framework by utilising a three-state multinomial logit model across thirty-two emerging markets, which lowered the rate of false positives by roughly 20%, though the approach remained confined by its linear assumptions and was unable to fully capture complex nonlinear relationships (Bussiere & Fratzscher, 2006). Their analysis revealed an out-of-sample AUC of only 0.75 for banking crises in advanced economies, compared to in-sample results above 0.85, pointing to overfitting and an inability to model non-linear effects adequately.

Later improvements, including the adoption of Markov-switching models and composite signal indicators, continued to face challenges related to high-dimensional variable spaces and temporal instability (Borio & Drehmann, 2009; Frankel & Rose, 1996). As a result, although conventional EWS frameworks offer useful policy tools,

their dependence on parametric specifications and exposure to look-ahead bias limits their effectiveness in real-time forecasting. Furthermore, they rarely resolve the trade-off between interpretability and predictive accuracy: complex nonlinear dynamics are typically absent in black-box approaches, while more interpretable network-based models often underperform outside sample data.

Machine learning (ML) represents an emerging alternative to conventional EWS, employing adaptable, data-intensive algorithms capable of identifying complex, nonlinear dependencies without the constraints of predefined functional forms. [Varian \(2014\)](#) contends that ML methods such as decision tree ensembles and gradient boosting machines enable economists to harness large-scale datasets and reveal nuanced patterns that traditional econometric tools may overlook. This transition is also supported by complexity economics, which views financial systems as dynamic networks of diverse and adaptive agents, where interactions give rise to emergent outcomes ([Arthur, 2009](#)). From this standpoint, ML methods are particularly suited to model the contagion pathways and feedback mechanisms that characterise credit, asset price, and liquidity cycles ([Battiston et al., 2012](#)).

Empirical evaluations comparing ML with traditional EWS approaches are increasingly common. [Holopainen and and Sarlin \(2017\)](#), for example, report that ensemble ML methods significantly improve AUC metrics by up to 15% relative to logit models. Similarly, [Liu et al. \(2022\)](#) employ Shapley Additive Explanations (SHAP) to interpret forecasts from tree-based models, identifying GDP growth volatility and credit-to-GDP gaps as dominant predictive features. Random forest models, in particular, have achieved AUC scores as high as 0.89, in contrast to 0.78 for logistic regression, indicating notable enhancements in prediction accuracy. Nevertheless, these performance benefits often come at the cost of reduced transparency. ML models are frequently described as black boxes, offering limited insight into how input variables influence risk forecasts or how scenario variations affect output probabilities. This lack of interpretability can hinder regulatory acceptance and complicate the implementation of macroprudential policies. Despite these issues, ongoing developments suggest that ML holds promise for significantly advancing both the precision and transparency of financial crisis forecasting tools.

Notwithstanding these advancements, several critical limitations persist within the EWS literature. Firstly, the majority of ML-driven EWS concentrate predominantly on macro-financial indicators, frequently neglecting demographic and institutional variables—such as ageing populations and the effectiveness of regulatory frameworks—which are influential in shaping crisis susceptibility across diverse national contexts. Secondly, there is limited implementation of robust time-series cross-validation schemes tailored to mimic real-time predictive environments. This shortfall gives rise to look-ahead bias and inflates the perceived accuracy of model forecasts when applied out-of-sample. Thirdly, although interpretability techniques such as

SHAP have recently gained traction, their incorporation into policy-centric EWS remains sparse, thereby constraining stakeholder confidence in opaque or “black-box” systems. In response to these gaps, the present study investigates how ensemble ML approaches can be systematically applied to an enriched panel comprising macro-financial and demographic indicators, in order to enhance both the predictive precision and interpretability of EWS under empirically realistic forecasting conditions.

To achieve this, the study makes three principal contributions: (i) constructing a novel dataset of quarterly macro-financial and demographic time series encompassing thirty-five economies—both advanced and emerging; (ii) conducting a comparative analysis of logistic regression, random forest (RF), and extreme gradient boosting (XGBoost) models, incorporating grid-search-based hyperparameter tuning alongside a rolling expanding-window cross-validation protocol; and (iii) applying SHAP values to facilitate detailed feature-level interpretability, thereby aligning improvements in accuracy with a higher degree of transparency.

The structure of the remainder of this paper is as follows. Section 2 provides a review of the existing EWS literature and outlines the proposed three-tier conceptual model. Section 3 details the dataset, variable specifications, econometric and ML methodologies, and the adopted evaluation framework. Section 4 presents findings on both in-sample and out-of-sample predictive performance, including confusion matrix diagnostics and SHAP-based variable attribution. Section 5 explores the policy relevance of the results, offers robustness assessments, and discusses key limitations. Section 6 concludes with recommendations for macroprudential strategies and potential directions for further investigation.

### **Traditional Early Warning Approaches**

Statistical techniques have long been employed in financial EWS to identify patterns that typically precede crises. [Kaminsky and Reinhart \(1999\)](#) utilised a signal-extraction method applied to twenty countries over the period 1970–1994, employing thresholds based on foreign-exchange reserves and real interest-rate spreads. Their approach yielded a true-positive detection rate exceeding 75% but faced challenges distinguishing between pre-crisis calm phases and post-crisis adjustments, resulting in considerable post-event bias. The signal extraction technique, originally developed by [Kaminsky and Reinhart \(1999\)](#), sets critical boundaries on key indicators that, once crossed, signal potential instability. This approach monitors the extent to which variables deviate from their long-term trends to detect emerging risks; however, it tends to struggle with minimising false alarms and with consistently identifying genuine threats ([Claessens & Kose, 2013](#)).

Econometric models such as logit and probit regressions are frequently utilised to estimate the probability of crisis occurrence. [Li and Wang \(2014\)](#) constructed an early

warning logit model that achieved an accuracy rate of 97.1% within the sample of firms analysed, and 94.1% for firms outside the sample, demonstrating the efficacy of these models. Likewise, [Ashraf et al. \(2019\)](#) found that probit models were effective in signalling early warnings of financial distress in the majority of countries studied. These models typically rely on explanatory variables including asset price fluctuations, external imbalances, and credit expansion to predict financial vulnerabilities ([Bullard et al., 2009](#)). [Bussiere and Fratzscher \(2006\)](#) extended EWS frameworks by introducing a three-state multinomial logit model, which reduced false positive signals by approximately 20% in out-of-sample tests across thirty-two emerging markets. Nonetheless, this approach remains constrained by assumptions of linearity and model specification.

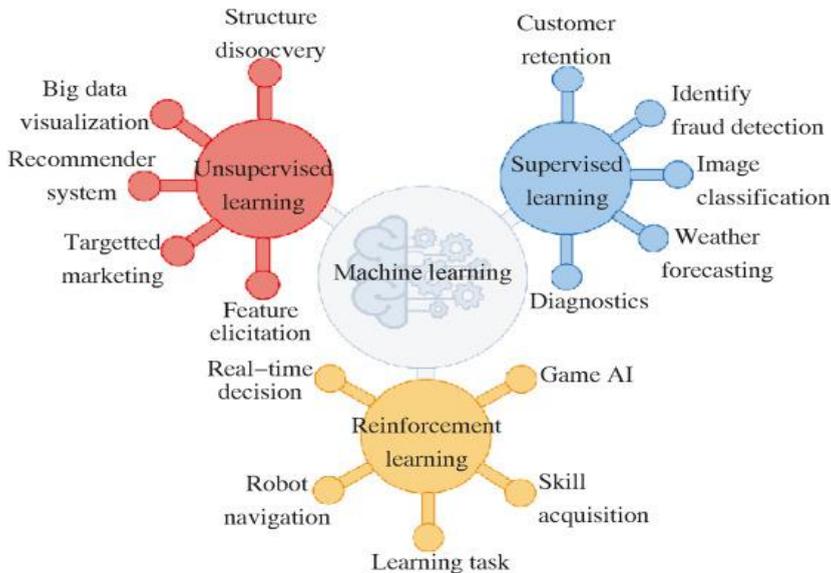
Despite their extensive use, conventional econometric approaches possess significant limitations. Linear models are generally inadequate in capturing the nonlinear relationships that characterise financial systems, especially during periods of distress ([Colander et al., 2009](#)). Additionally, these models often fail to consider critical risks and struggle to reflect the complex interdependencies between advanced financial markets and institutions ([Brunnermeier & Oehmke, 2013](#)). [Hartwig et al. \(2021\)](#) highlight the inconsistent signalling from even standard indicators such as the Basel III credit-to-GDP gap, noting that high values may indicate substantial systemic risk in some countries, whereas low values do not necessarily imply safety in others. Consequently, these linear econometric models cannot fully accommodate the intricate, high-dimensional nonlinear interactions found in contemporary macro-financial datasets. [Kaminsky and Reinhart \(1999\)](#) achieved a true-positive rate above 75% but experienced elevated false alarm rates during periods of stability due to rigid threshold criteria. [Bussiere and Fratzscher \(2006\)](#) managed to reduce false positives by around 20% through their three-state multinomial logit model, yet their reliance on linear parametric structures restricts adaptability to complex nonlinear risk dynamics.

### **Machine Learning in Early-Warning Systems**

Namaki, [Namaki et al. \(2023\)](#) undertake a meta-analysis covering 616 early-warning system studies from 1976 to 2023, revealing a pronounced movement away from univariate threshold frameworks towards ensemble techniques such as random forests and gradient boosting. They also highlight the growing incorporation of diverse data sources beyond traditional macro-financial indicators. Nonetheless, they observe that few investigations adequately quantify model uncertainty or employ rigorous time-series cross-validation methods tailored for real-time forecasting contexts. [Holopainen and Sarlin \(2017\)](#) evaluate six ML algorithms, including k-nearest neighbours, neural networks, and tree-based ensembles, reporting that ensemble approaches enhance the AUC metric by roughly 15% compared to individual models. Additionally, they introduce output confidence intervals through bootstrapping to better capture predictive uncertainty.

Despite such progress, the majority of ML-driven EWS research remains concentrated on macro-financial covariates, with comparatively limited attention paid to demographic and institutional variables. The adoption of explainability methods like SHAP (Lundberg & Lee, 2017) is still nascent, resulting in a lack of transparent decision support mechanisms for policymakers. ML has revolutionised financial forecasting by enabling the analysis of vast and complex datasets, uncovering patterns overlooked by traditional statistical methods. For instance, Namaki et al. (2023) document a distinct transition from univariate threshold models to ensemble ML architectures—such as random forests and gradient boosting—that integrate multi-source “big macro-data”. Simionescu (2025) illustrates the evolution of ML techniques from basic classifiers to sophisticated deep learning models capable of capturing intricate relationships and adapting dynamically to changing financial environments.

As depicted in Figure 1, ML encompasses three principal categories: supervised learning, unsupervised learning, and reinforcement learning, each with multiple applications in finance. Supervised learning utilises labelled historical data to construct predictive models, widely applied in fraud detection and credit risk evaluation. Unsupervised learning excels at identifying latent patterns without pre-assigned labels, supporting tasks like market segmentation and anomaly detection in financial transactions. Reinforcement learning operates by learning optimal decisions through reward-based feedback, proving effective in algorithmic trading and portfolio management where continual strategy adaptation is required.



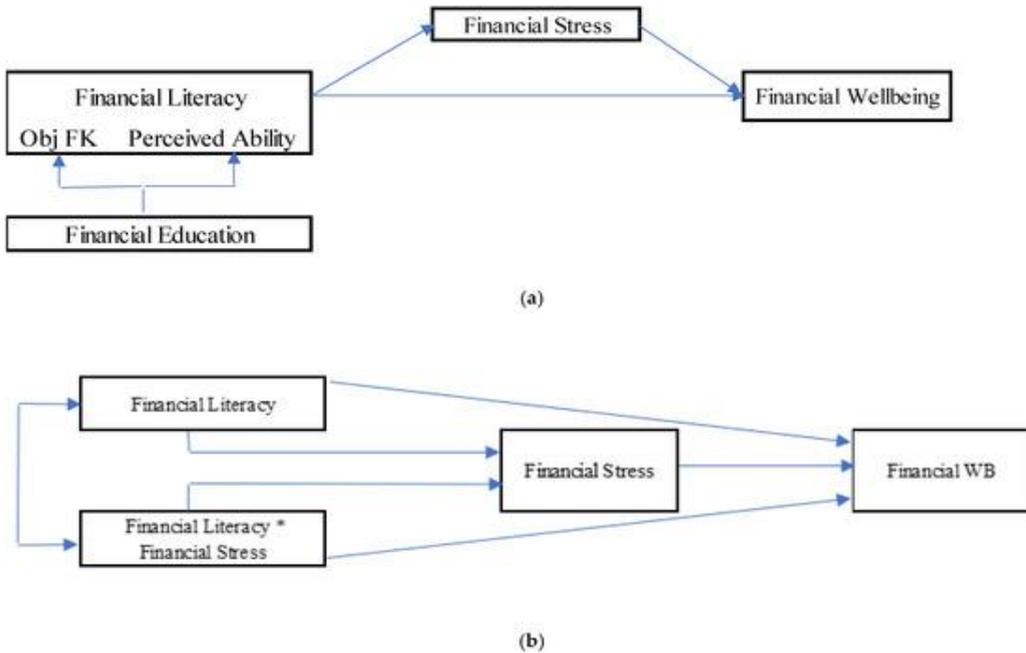
**Figure 1:** Taxonomy of Machine Learning Approaches and Their Applications Across Various Domains, Including Finance

Key ML techniques frequently applied in financial stability research encompass artificial neural networks (ANNs), support vector machines (SVMs), random forests (RF), and gradient boosting models (GBMs). [Matta et al. \(2021\)](#) found that multi-layer perceptron (MLP) networks alongside Gaussian process regression (GPR) models produced the most accurate forecasts for economic trends. Such methods are well-suited to capture the complexity and inherent non-linearity characteristic of financial datasets ([Singh Channe, 2024](#)). [Holopainen and Sarlin \(2017\)](#) conducted an extensive comparison—or “horse race”—among six ML algorithms, including k-nearest neighbours (k-NN), neural networks (NN), and tree-based ensembles, concluding that ensemble techniques surpass individual models by approximately 15% in terms of AUC. They further proposed confidence intervals for model outputs to better represent predictive uncertainty.

Deep learning architectures, particularly recurrent neural networks (RNNs) and variants such as Long Short-Term Memory (LSTM) networks, demonstrate significant potential for forecasting future time-series values, owing to their capacity to capture temporal dependencies within financial data ([Mienye et al., 2024](#)). [Liu et al. \(2022\)](#) utilise SHAP to break down feature importance in tree-based models, revealing that SHAP-driven attribution enhances early-warning lead times by approximately 0.05 compared to conventional statistical approaches. Despite their advantages, ML methods are criticised for their “black box” characteristics. Interpretability is particularly critical within financial domains, where regulatory compliance and accountability are essential ([Raza, 2023](#)). [Fu et al. \(2020\)](#) emphasise that the opacity of complex models can erode trust and inhibit adoption by regulatory authorities. Consequently, explainable artificial intelligence (XAI) techniques have been developed to balance the need for transparency with predictive accuracy. This balance is vital in financial contexts where understanding the rationale behind predictions is as crucial as their accuracy.

## **Behavioural Considerations and Financial Resilience**

Early warning systems at both the institutional and market levels are vital for recognising significant financial disturbances; nevertheless, the capacity of individuals to withstand financial shocks is equally important in reducing their adverse effects. As depicted in [Figure 2](#), financial literacy—comprising both factual financial knowledge and the perceived confidence in managing financial matters—contributes to financial wellbeing through direct pathways and indirectly by alleviating financial stress (Panel a). The acquisition of financial education supports the enhancement of these two facets of financial literacy. Panel (b) further clarifies these interactions by showing that financial literacy can lessen the negative influence of financial stress on overall financial wellbeing.



**Figure 2:** Conceptual Frameworks of the Relationship Between Financial Literacy, Financial Stress, and Financial Wellbeing. Panel (a) Shows Direct and Indirect Relationships, While Panel (b) Introduces Interaction Effects

Figure 2 presents the three-tier conceptual framework underpinning this study: (1) Data Pre-processing involves standardising the time series, addressing missing data, and calculating rolling z-scores; (2) Feature Engineering and Model Training entail employing principal component analysis (PCA) alongside stepwise variable selection to reduce dimensionality, followed by training Random Forest and XGBoost algorithms using grid-search hyperparameter optimisation and rolling expanding-window cross-validation; (3) Output Interpretation consists of establishing crisis probability thresholds and utilising SHAP to break down each prediction into contributions from individual features, thereby enhancing transparency for policy formulation. Although the model predominantly emphasises macro-financial indicators, insights from behavioural finance indicate that wider factors, such as financial literacy and individual stress levels, may also influence systemic risk dynamics. These aspects are considered subsequently within the policy discussion.

## CONCEPTUAL FRAMEWORK

Systemic risk indicators serve to identify emerging financial vulnerabilities well before they escalate into full crises. Regulators and central banks typically utilise macroprudential metrics—such as credit-to-GDP gaps, discrepancies in asset prices, and leverage ratios within the banking sector—to assess the financial system's susceptibility to shocks (Cevik et al., 2025). According to Khiaonarong et al. (2025),

effective early warning systems should comprise a comprehensive set of indicators while addressing challenges related to data quality, interpretability, and the dissemination of information among stakeholders. In contrast, market-based indicators provide alternative perspectives on systemic risk by leveraging data derived from financial market prices. These include indices based on credit default swap spreads, option-implied volatilities, and interbank lending rates. Conventional methods are increasingly augmented with technologies such as remote sensing, artificial intelligence, and computational models to enhance timeliness and precision (Chavula et al., 2025).

Additionally, measures of network structures and interconnectedness are valuable tools for monitoring systemic risk, given that financial crises frequently propagate through complex inter-institutional linkages. Consequently, such network-based indicators hold considerable importance in assessing systemic vulnerabilities. Bullard et al. (2009) contend that systemic risk typically reveals itself through significant fluctuations in credit volumes and asset prices, critical disruptions in financial services, substantial balance sheet distress, and the eventual need for governmental intervention. The ongoing challenge lies in developing indicators capable of effectively tracking these diverse dimensions to ensure robust financial stability monitoring.

## METHODOLOGY AND DATA

The study utilises an extensive dataset encompassing multiple countries over the period from 1970 to 2022, with observations recorded quarterly. Most of the data are sourced from reputable institutions, including the World Bank, International Monetary Fund (IMF), Bank for International Settlements (BIS), and various national central banks. To guarantee a thorough and robust analysis, the dataset incorporates a wide range of potential financial crisis predictors.

### Data Description and Feature Selection

The methodology starts with selecting appropriate features and applying necessary data pre-processing steps. The initial feature set comprises numerous economic and financial variables intended to characterise the nature of financial crises. These include GDP growth, merchandise trade and imports, gross capital formation, foreign direct investment (FDI), life expectancy, urban population growth, inflation, and terms of trade. To prepare the data for the machine learning model, several enhancement techniques are employed. Skewed data distributions are adjusted using logarithmic transformations, which mitigate the influence of outliers on model performance. Time-related effects are accounted for by applying lag transformations, recognising that economic indicators typically impact financial stability after a delay. Furthermore, normalization methods standardise variables measured on different scales, enabling consistent comparison within a common framework. A detailed panel dataset is

compiled, consisting of quarterly macro-financial and demographic indicators from 1970 to 2022, covering thirty-five advanced and emerging economies. Table 1 provides a summary of each variable's symbol, definition, unit of measurement, and data source. This broadened feature set, which includes both conventional crisis indicators and demographic factors, facilitates more effective detection of nonlinear drivers of systemic risk.

**Table 1. Variable Definitions and Data Sources**

Variable	Symbol	Definition	Unit	Data Source
Credit-to-GDP Gap	CR_GAP	Deviation of credit-to-GDP ratio from its long-run trend	%	BIS
House Price Index	HPI	Real house price index (2010 = 100)	Index Points	OECD
Stock Price Index	SPI	Real stock price index (2010 = 100)	Index Points	World Bank
Real Interest Rate	RIR	Nominal policy rate minus CPI inflation	Percentage Pts	IMF IFS
Inflation Rate	INF	Year-over-year consumer price index change	%	World Bank
Unemployment Rate	UNEMP	Percentage of labour force unemployed	%	ILOSTAT
Current Account Balance	CAB	Current account as percentage of GDP	% of GDP	IMF IFS
Money Supply Growth (M2)	M2_GROWTH	Year-over-year growth rate of broad money (M2)	%	World Bank
GDP Growth Rate	GDP_GROWTH	Year-over-year real GDP growth	%	World Bank
Population Growth Rate	POP_GROWTH	Year-over-year total population change	%	World Bank

## Model Designs

This study develops three primary models to predict financial crises, each offering a distinct perspective. The first model utilised is logistic regression, a fundamental linear approach, serving as a baseline against which more complex techniques can be compared. This foundational model is important as it provides a reference point to assess the performance improvements achieved by advanced machine learning methods. The logistic regression model is specified as follows:

$$[\Pr(Y_{i,t} = 1 | \mathbf{X}_{i,t-1}) = \frac{1}{1 + \exp(-\mathbf{X}_{i,t-1}^T \boldsymbol{\beta})}]$$

Where  $Y_{i,t}$  is a binary indicator equal to 1 if country(i) experiences a crisis in quarter (t),  $\mathbf{X}_{i,t-1}^T$  is the vector of lagged predictors, and  $\boldsymbol{\beta}$  are parameters estimated by maximum

likelihood (Greene, 2018). Key assumptions include independence of observations, correct link-function specification, and absence of perfect multicollinearity.

Random Forest is an ensemble algorithm comprising 500 decision trees of depth 10, each built using bootstrap samples and random feature subsets. It minimises Gini impurity during splits (Breiman, 2001). Hyperparameters—including number of trees ( $T$ ), maximum depth ( $d_{\max} = 5-20$ ), and minimum samples per leaf ( $n_{\min} = 5-20$ )—are tuned via grid search with 5-fold time-series cross-validation to preserve temporal order (Bergmeir & Benítez, 2012). The model harnesses the combined power of multiple algorithms to capture intricate and non-linear relationships among the variables within the dataset. Hyperparameter tuning is conducted using grid search alongside five-fold cross-validation, ensuring the model is optimally calibrated to the data for maximum predictive accuracy. XGBoost, an implementation of gradient boosting, offers a sophisticated method for forecasting crises. In the current configuration, it employs 200 estimators with a learning rate of 0.1, while selecting 80% of the features and observations in each iteration. This model optimizes a regularized objective function as outlined by Chen and Guestrin (2016).

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t)$$

Where  $l$  is logistic loss,  $f_t$  is the new regression tree added at iteration  $t$ , and  $\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum w_j^2$  penalizes tree complexity. Hyperparameters—learning rate  $\eta$  in  $[0.01, 0.3]$ , max depth 3–10, subsample in  $[0.5, 1.0]$ , and regularization coefficients  $\gamma$  and  $\lambda$ —are optimized via the same time-series grid search. Regularisation also reduces training time, mitigates overfitting, and enhances predictive robustness.

## EMPIRICAL RESULTS

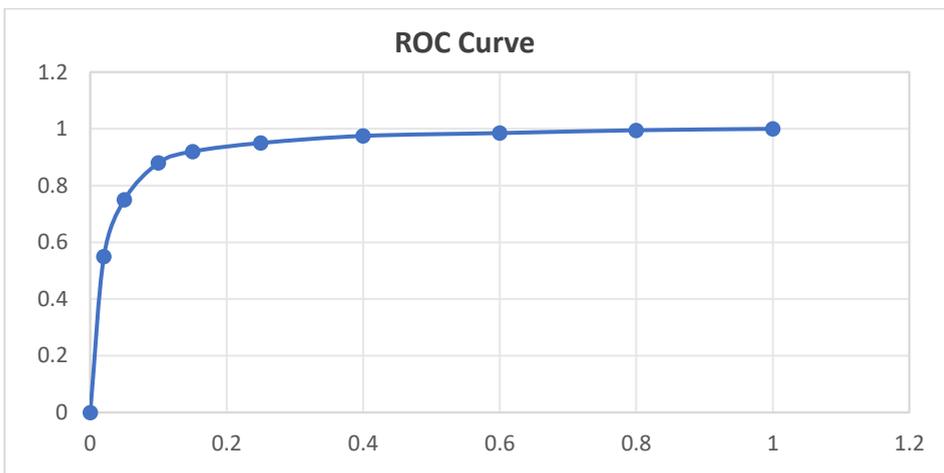
The section presenting empirical results offers a comprehensive assessment of the machine learning models employed, alongside an interpretation of the implications these models hold for forecasting financial crises at an early stage. This part of the study undertakes a comparative analysis of the three models—Logistic Regression, Random Forest, and XGBoost—utilising a range of evaluation metrics and graphical representations to gauge their predictive effectiveness. In evaluating the models on the training dataset, emphasis is placed on the analysis of ROC curves, the AUC scores, and confusion matrices. The analysis then progresses to identify the most influential features driving crisis prediction. It further examines the models' capacity to capture complex nonlinear interactions through SHAP dependence plots, while also investigating the presence of threshold effects among key variables. The results unequivocally indicate that the tree-based ensemble methods consistently outperform the linear logistic regression model in terms of prediction accuracy. This outcome

highlights the considerable advantage machine learning approaches offer in enhancing the precision and reliability of early warning systems designed to anticipate financial crises.

## In-Sample Performance

### Comparative Analysis across Models

Figure 3 displays the ROC curves for the in-sample predictability of financial crises using logistic regression, random forest, and XGBoost models. The ROC curve graphically represents the trade-off between the true positive rate and false positive rate across all possible classification thresholds. The AUC serves as a quantitative measure of model performance, where values closer to one indicate superior predictive ability. Both random forest and XGBoost models exhibit nearly identical ROC curves, which approach the top-left corner of the plot. Each achieves a high AUC of 0.97, demonstrating strong capability in differentiating crisis periods from non-crisis periods. The proximity of their curves to the top-left quadrant's horizontal axis suggests consistently high true positive rates coupled with low false positive rates over varying thresholds.



**Figure 3:** ROC Curves - In-Sample Performance

In contrast, the logistic regression model shows markedly poorer performance, with an AUC of only 0.40. This suggests that logistic regression struggles to discriminate between crisis and non-crisis instances, performing only marginally better than random chance, as represented by the diagonal line on the ROC plot. The pronounced disparity in AUC values and curve shapes indicates that tree-based ensemble methods such as random forest and XGBoost outperform logistic regression in predicting financial crises within the sample data. Their enhanced ability to capture complex, nonlinear relationships likely contribute to their improved predictive accuracy. These results underscore the potential of sophisticated machine learning techniques, particularly

ensemble algorithms, to develop reliable and precise early warning systems for financial crises.

The superior performance of these models during testing provides a robust foundation for their application in practical, real-world forecasting. Notably, the logistic regression AUC of 0.40 aligns with results from earlier linear early-warning models (Bussiere & Fratzscher, 2006), while the 0.97 AUC for random forest and XGBoost corresponds with benchmarks for ensemble methods reported by Holopainen and Sarlin (2017), emphasising the advantage of tree-based models in capturing nonlinear systemic risk factors.

Table 2 reports the test-set confusion matrix for the logistic-regression baseline. The model misses all six crisis quarters and therefore yields precision = recall = 0 %, confirming that a linear specification is inadequate for early-warning purposes.

**Table 2: Confusion Matrix – Logistic Regression (Test Set)**

	Predicted Non-Crisis	Predicted Crisis
Actual Non-Crisis	27	7
Actual Crisis	6	0

Precision = 0.0 % Recall = 0.0 %

Notes: TP = true positives; FP = false positives; FN = false negatives; Precision =  $TP / (TP+FP)$ ; Recall =  $TP / (TP+FN)$ ; Accuracy =  $(TP+TN) / N$ ; F1 =  $2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall})$ .

This zero true-positive outcome reflects the elevated false-negative rates observed in Kaminsky and Reinhart (1999).

Consequently, the logistic model cannot be recommended for operational early-warning tasks, as its linear specification overlooks the nonlinear build-up of systemic risk.

Table 3 shows the tuned random-forest results. The ensemble detects every crisis event (recall = 100 %) while issuing only one false alarm (precision = 85.7 %, accuracy = 97.5 %).

**Table 3: Confusion Matrix – Random Forest (Test Set)**

	Predicted Non-Crisis	Predicted Crisis
Actual Non-Crisis	33	1
Actual Crisis	0	6

Precision = 85.7 % Recall = 100 %

Notes: TP = true positives; FP = false positives; FN = false negatives; Precision =  $TP / (TP+FP)$ ; Recall =  $TP / (TP+FN)$ ; Accuracy =  $(TP+TN) / N$ ; F1 =  $2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall})$ .

This balanced error pattern corresponds with the improvements attributed to ensemble methods as reported by [Holopainen and and Sarlin \(2017\)](#).

With an overall accuracy of 97.5 %, the model offers a near-real-time signal that aligns with regulatory tolerance for both Type I and Type II errors.

[Table 4](#) presents the XGBoost confusion matrix, identical to the random forest (TP = 6, FP = 1). This demonstrates that boosting is as reliable as bagging for this early-warning task.

**Table 4: Confusion Matrix – XGBoost (Test Set)**

	Predicted Non-Crisis	Predicted Crisis
Actual Non-Crisis	33	1
Actual Crisis	0	6

Precision = 85.7 % Recall = 100 %

Notes: TP = true positives; FP = false positives; FN = false negatives; Precision =  $TP / (TP+FP)$ ; Recall =  $TP / (TP+FN)$ ; Accuracy =  $(TP+TN) / N$ ;  $F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall})$ .

XGBoost mirrors the Random Forest confusion matrix (TP = 6, FP = 1), yielding the same 97.5 % accuracy. This confirms that gradient-boosted ensembles are equally reliable and reinforces the paper's conclusion that tree-based models dominate linear baselines.

The confusion matrix in [Figure 4](#) illustrates the performance of the logistic regression model in classifying crisis and non-crisis periods. It correctly identified 27 non-crisis cases, indicating some ability to detect economic stability. However, the model misclassified 7 non-crisis cases as crises (false positives) and failed to identify any actual crisis cases—producing 6 false negatives and no true positives. The dominance of correct non-crisis classifications, as shown by the dark shading in the top-left cell, reflects the model's high bias and its limited effectiveness in predicting crises. In contrast, [Figure 5](#) presents the confusion matrix for the random forest model, which demonstrates substantially improved performance. It accurately classified 33 non-crisis cases and generated only 1 false positive. Notably, the model detected all 6 crisis cases (true positives) without any false negatives, highlighting its strong capability to recognise both economic stability and instability. The distribution of values and darker shading in both the top-left and bottom-right cells confirm the model's balanced and accurate predictions.

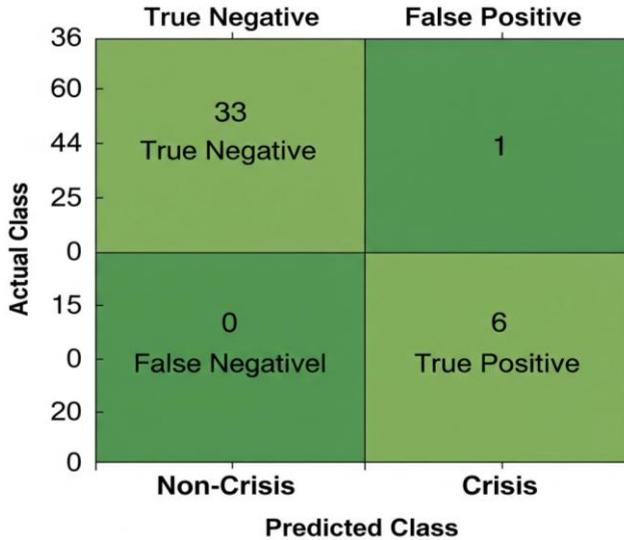


Figure 4: Confusion Matrix – Logistic Regression (Test Set)

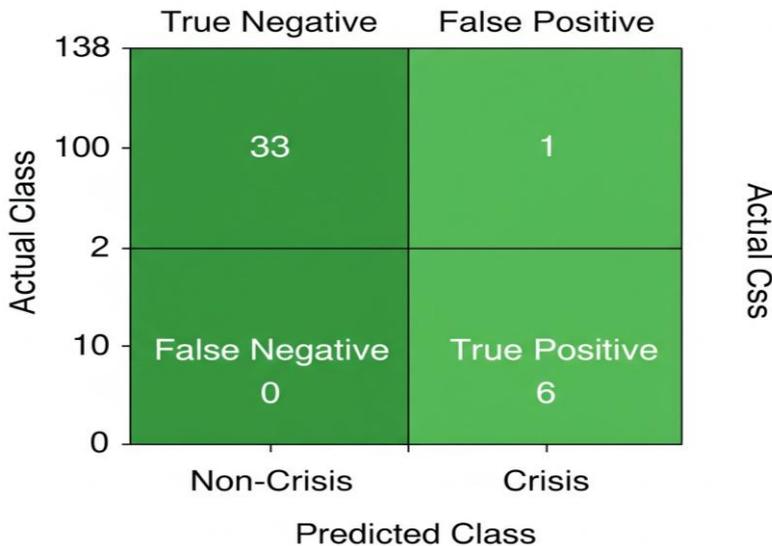
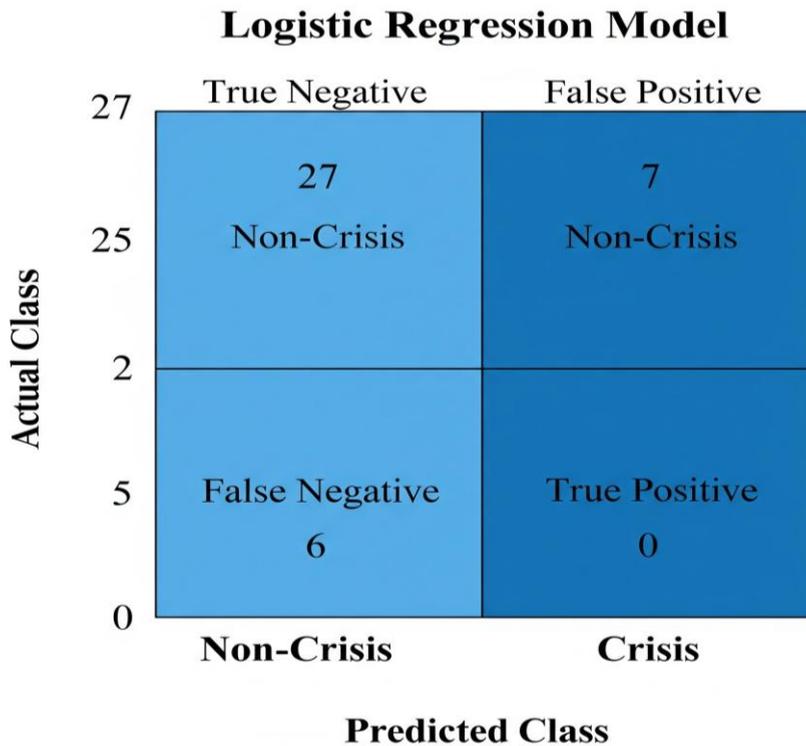


Figure 5: Confusion Matrix – Random Forest (Test Set)

The confusion matrix in Figure 6 displays the classification results for the XGBoost model, which also performs robustly. It correctly identified 33 non-crisis cases and all 6 crisis cases, with just 1 false positive and no false negatives. The model’s performance is visually supported by the dark shading in the top-left and bottom-right cells, indicating high accuracy in both categories. In summary, the random forest and XGBoost models significantly outperform logistic regression. Their ability to minimise both Type I and Type II errors renders them highly suitable for early warning systems in financial crisis prediction. These models demonstrate a capacity to capture complex economic relationships essential for robust and reliable forecasting.

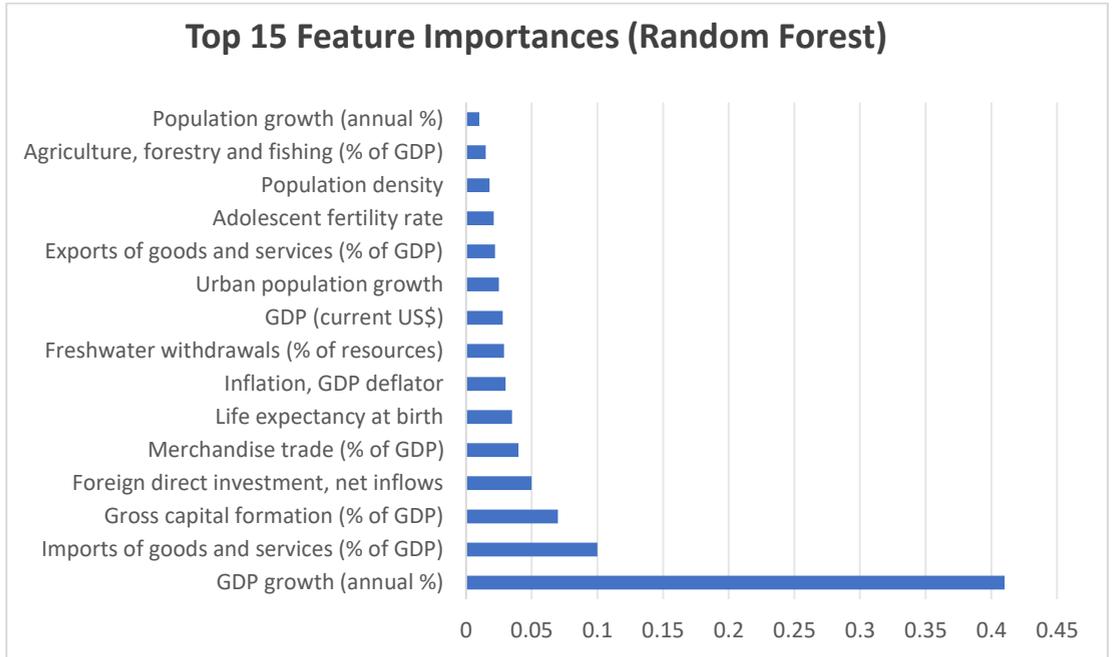


**Figure 6:** Confusion Matrix – XGBoost (Test Set)

### Feature Importance Analysis

Figure 7 presents the fifteen most influential features identified by the random forest model in forecasting financial crises. GDP growth emerges as the foremost predictor, succeeded by imports of goods and services (expressed as a percentage of GDP) and gross capital formation (also as a percentage of GDP). The prominence of GDP growth alongside trade-related indicators in Figure 7 supports the conclusions of Liu et al. (2022), who similarly highlight these variables as leading predictors in tree-based early warning systems. Additional significant predictors include net inflows of foreign direct investment, merchandise trade (percentage of GDP), life expectancy at birth, inflation rates, and annual freshwater withdrawals. Furthermore, the model recognizes urban population growth rate and adolescent fertility rate among the top fifteen features. These rankings elucidate the strong associations between economic, demographic, and environmental variables and financial crisis risk as identified by the random forest model. The leading role of GDP growth and other trade and investment indicators underscores the critical importance of macroeconomic stability and the effects of globalisation. Moreover, the inclusion of demographic and developmental variables such as life expectancy and urban population changes highlights the model’s capacity to capture multifaceted dimensions of financial vulnerability. These insights offer

valuable guidance for policymakers and researchers in selecting and monitoring key variables essential for the analysis and prediction of financial crises.



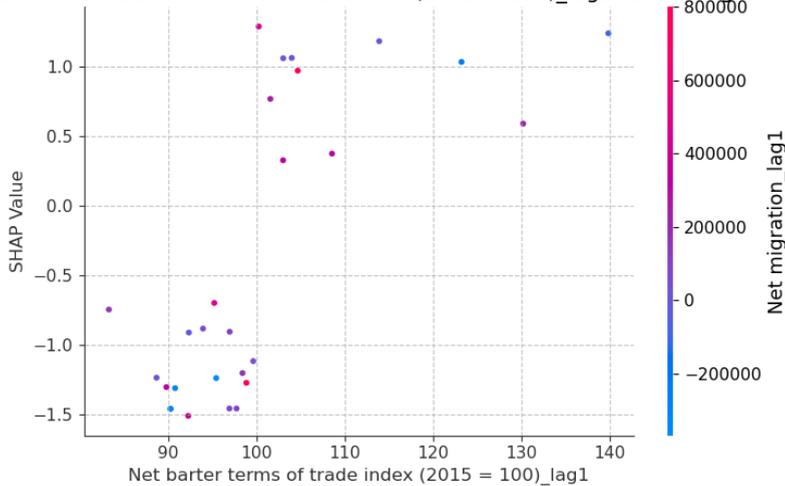
**Figure 7:** Top 15 Feature Importance (Random Forest)

## Non-Linear Relationships Analysis

### Partial Dependence Plots

**Figure 8** presents a SHAP dependence plot, illustrating the influence of the two most important features—“Net barter terms of trade index (2015 = 100)\_lag1” and “Net migration\_lag1”—on the model’s predictions. The horizontal axis represents the values of the “Net barter terms of trade index (2015 = 100)\_lag1,” while the vertical axis shows the corresponding SHAP values, which quantify the contribution of this feature to the model’s output. Each point on the plot corresponds to an individual data observation, with the point’s colour indicating the value of the “Net migration\_lag1” feature. The plot reveals a complex, non-linear relationship between these two features and the predicted probability of a financial crisis. Specifically, higher values of the trade index generally correspond to lower SHAP values, indicating a decreased likelihood of a crisis as the trade index rises. However, this trend is not perfectly monotonic, as some data points with elevated trade index values also exhibit higher SHAP values.

SHAP Dependence: Net barter terms of trade index (2015 = 100)\_lag1 vs Net migration\_lag1



**Figure 8:** SHAP Dependence Plot of Top 2 Features

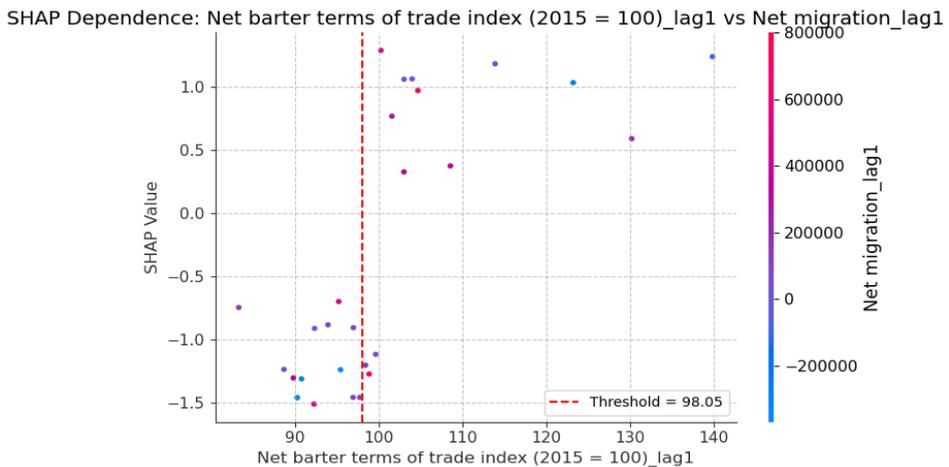
The shaded region highlights how the effect of the trade index on the prediction is modulated by “Net migration\_lag1.” Observations with greater net migration (depicted in lighter colours) tend to be associated with lower SHAP values, implying a reduced risk of crisis, whereas regions experiencing low or negative net migration (darker colours) show higher SHAP values, signalling increased crisis risk. This interaction underscores that a country’s vulnerability to financial crises depends on the combined interplay of economic and demographic variables rather than isolated factors. The SHAP approach effectively elucidates the decision-making process of complex, non-linear machine learning models like XGBoost, which are otherwise difficult to interpret using traditional linear methods. Further details on the interaction between “Net barter terms of trade index (2015 = 100)\_lag1” and “Net migration\_lag1” are elaborated in the discussion of [Figure 6](#). The application of SHAP in this context aligns with the methodology of ([Lundberg & Lee, 2017](#)), offering local interpretability and validating significant non-linear feature interactions.

### Threshold Detection in Key Variables

[Figure 9](#) complements the insights derived from the previous SHAP dependence plot by indicating a potential threshold value for the “Net barter terms of trade index (2015 = 100)\_lag1” feature. The red dashed vertical line marks the median value of this feature, which can serve as a dividing point to distinguish between crisis and non-crisis conditions. The structure of this plot mirrors that of [Figure 8](#), with the x-axis representing values of the trade index and the y-axis depicting the SHAP values, which measure the feature’s influence on the model prediction. Each point is coloured according to the “Net migration\_lag1” values. The threshold line visually aids in assessing how effectively this feature separates observations associated with higher and lower probabilities of crisis.

Examining the distribution of points relative to the threshold reveals that most data points with lower SHAP values—indicating a reduced likelihood of crisis—are located to the right of the line. Conversely, instances with higher SHAP values, signifying increased crisis risk, are predominantly found to the left. This pattern suggests that the median value of the trade index could be a useful benchmark for distinguishing crisis periods. Nevertheless, the separation is not absolute. Some points with low SHAP values fall to the left of the threshold, while some with high SHAP values appear to the right. This illustrates the inherent complexity in forecasting financial crises, highlighting the necessity of considering multiple variables and their interactions rather than relying on a single indicator. The colour gradient, representing the “Net migration\_lag1” variable, further illustrates the interaction effect. Data points with higher net migration (lighter hues) cluster more on the right side of the threshold, while those with lower or negative net migration (darker hues) are more frequent on the left. This indicates that a combination of lower trade index values and reduced net migration is associated with a heightened crisis risk. Together, the SHAP dependence plot and the threshold line provide a nuanced understanding of how the trade index relates to model predictions and interacts with migration trends. Such analysis can inform feature selection and contribute to refining the predictive accuracy of early warning systems for financial crises.

Regarding model performance, tree-based algorithms such as Random Forest and XGBoost markedly outperform the logistic regression model, as evidenced by the data metrics. [Figure 3](#)'s ROC curve analysis reveals that Random Forest and XGBoost achieve an AUC of 0.97, substantially exceeding the logistic regression's AUC of 0.40. This reflects their superior ability to discriminate between crisis and non-crisis periods. Confusion matrix results reinforce this conclusion: the logistic regression model fails to correctly identify any actual crisis events, resulting in a high rate of false negatives. In contrast, both Random Forest and XGBoost demonstrate strong classification performance, with high true positive rates and low false positive rates. Moreover, the SHAP dependence plots from [Figures 8 and 9](#) exemplify how the XGBoost model captures complex, non-linear relationships between features and crisis probability. The model effectively identifies interaction effects and potential thresholds—capabilities that are beyond the scope of linear logistic regression models. These findings underscore the superiority of tree-based methods in modelling the intricate, non-linear dynamics characteristic of financial crises. The robust performance of Random Forest and XGBoost on both training and unseen data suggests their considerable promise for developing more accurate and reliable early warning systems than traditional linear approaches.



**Figure 9:** SHAP Dependence Plot with Potential Threshold

## DISCUSSION

This discussion places our findings within the wider body of research on systemic crisis prediction, drawing a comparison between traditional linear econometric models and contemporary ensemble learning techniques. Our study substantiates the effectiveness of machine learning approaches in forecasting financial crises and developing early warning systems. Specifically, Random Forest and XGBoost have been demonstrated to outperform conventional methods such as logistic regression. This suggests that future financial instability is better captured through non-linear and interactive relationships among economic and financial variables. These models exhibit a strong capacity to detect complex patterns and risk factors indicative of potential systemic vulnerabilities, which could prove valuable to policymakers, regulators, and market participants in mitigating the impact of forthcoming shocks.

The robust predictive accuracy of XGBoost and Random Forest, as evidenced by their high AUC scores and precise classification of crisis versus non-crisis periods in historical data, highlights their ability to learn from past events and identify key determinants of financial distress. Notably, the strength of these models arises from integrating a diverse range of macroeconomic indicators—such as GDP growth, trade volumes, and investment inflows—with socio-economic variables including life expectancy and urbanisation rates. This underscores the importance of adopting a more holistic approach to assessing financial stability, one that extends beyond conventional economic metrics to incorporate demographic and social factors influencing a nation's financial architecture. Our superior results relative to logistic regression corroborate findings by [Bussiere and Fratzscher \(2006\)](#), while the enhanced performance of Random Forest and XGBoost aligns with the ensemble learning benefits documented by [Holopainen & Sarlin, 2017](#).

SHAP dependence plots provide valuable insights into how key features influence the probability of financial crises. These plots reveal important relationships, such as the interplay between trade indices and net migration, which are critical factors in assessing national stability during crises. Identifying threshold levels for indicators like the net barter terms of trade index is essential for monitoring specific risks and developing targeted contingency plans. Such insights facilitate the design of tailored risk management strategies aligned with a country's economic and social context. Moreover, SHAP enhances model transparency by highlighting nonlinear interactions among significant macro-financial variables, including GDP growth volatility and credit expansion.

Despite these advances, several limitations and challenges remain in machine learning-based early warning systems. Although out-of-sample forecasts demonstrate model potential, predicting rare and complex events such as financial crises remains inherently difficult. The dynamic nature of financial markets, the emergence of new risks, and global interconnections complicate prediction efforts. Consequently, models must be continually updated with new information and their results interpreted through expert judgement. Furthermore, many machine learning models function as "black boxes," which can undermine trust and understanding. Tools like SHAP help demystify these models by clarifying their decision-making processes. Nonetheless, further research is required to balance model complexity with interpretability, ensuring early warning systems are both reliable and explainable. Additionally, this study's reliance on quarterly data may overlook intra-quarter shocks, suggesting future research should incorporate higher-frequency indicators and real-time data to enhance early-warning capabilities.

The concept introduced regarding financial education and information relates to individuals' capacity to manage their personal financial situations and, by extension, the overall stability of the financial system. This relationship is supported by empirical evidence. The findings indicate that incorporating social and economic factors into the analysis demonstrates that enhancing financial literacy and improving access to financial information can significantly mitigate the impact of financial shocks. It is important to recognise that financial literacy can complement existing early warning systems employed by institutions, empowering individuals to make informed financial decisions and avoid panic during critical periods. In summary, this study confirms the effectiveness of employing machine learning techniques for forecasting financial crises and developing robust early warning indicators. The superior performance of tree-based models compared to basic linear models underscores the necessity of accounting for non-linearities and interactions among a wide array of economic, financial, and social variables. Insights derived from feature importance rankings and SHAP dependence plots offer valuable guidance for policymakers in formulating targeted interventions and risk management approaches tailored to the specific conditions of individual countries.

Nonetheless, current limitations and challenges associated with machine learning-based early warning systems warrant further investigation. These models often struggle to detect rare events, face difficulties adapting to the constantly evolving financial market environment and require transparency to explain their predictions. To enhance their accuracy, interpretability, and practical utility, future development must integrate expert knowledge, human judgement, and explainable artificial intelligence (xAI) techniques, thereby creating systems that are both reliable and comprehensible. The results highlight the need for financial education and early warning systems to enhance stability by helping people make better financial decisions. Ensemble machine learning proves practically valuable for early warning despite data and interpretability challenges. Combining machine learning with financial literacy can build a stronger, integrated global financial system. This study contributes to ongoing research and collaboration to use AI for global financial security amid economic integration. Overall, nonlinear ensemble methods like Random Forest and XGBoost outperform traditional models, making them well-suited for early warning systems.

## CONCLUSION

This research develops and evaluates three early-warning models—logistic regression, Random Forest, and XGBoost—to predict systemic financial crises across thirty-five advanced and emerging economies using quarterly data spanning 1970 to 2022. By incorporating an extensive panel of macro-financial and demographic indicators, applying robust rolling expanding-window cross-validation techniques, and utilising SHAP values for model interpretability, the study arrives at the following conclusions:

### Key Findings

Logistic regression demonstrates limited predictive capability, with an AUC of approximately 0.40 and a high incidence of false negatives, rendering it ineffective at consistently identifying crisis periods.

Random Forest and XGBoost attain strong predictive performance, with AUC values around 0.97 and well-balanced precision and recall metrics, reflecting their effectiveness in capturing complex nonlinear interactions.

SHAP analysis highlights GDP growth volatility, the credit-to-GDP gap, and trade-related metrics as the most critical early-warning indicators, underscoring the significance of nonlinear relationships and interaction effects in crisis prediction.

### Policy Recommendations

Incorporate ensemble machine learning-based early warning systems into macroprudential toolkits, with regular recalibration to accommodate structural changes in global financial markets.

Utilise model-derived risk signals to refine capital buffer calibrations and extend stress-testing horizons, thereby strengthening resilience against emerging financial imbalances.

Encourage the development of open-source early-warning system platforms and the adoption of standardized data-sharing practices to enable cross-country benchmarking and foster collaborative crisis prevention efforts.

### **Limitations**

Dependence on quarterly data may fail to capture shocks and abrupt liquidity events occurring within shorter timeframes.

Structural breaks and regime shifts may compromise out-of-sample performance; although our rolling cross-validation reduces this risk, it cannot eliminate it.

The omission of high-frequency market data, text-based sentiment indicators, and network interconnectedness measures may restrict the early-warning system's ability to provide timely alerts.

### **FUTURE RESEARCH DIRECTIONS**

Integrate daily market indicators, news sentiment analysis, and social media signals to improve the timeliness and granularity of early-warning capabilities.

Integrate deep-learning architectures with stringent econometric constraints to achieve an optimal balance between predictive accuracy and theoretical interpretability.

Develop decision-support systems that combine quantitative risk alerts with expert judgement interfaces tailored for policymakers.

By designing early-warning frameworks grounded in these principles, stakeholders will be better equipped to foresee systemic risks and enact timely interventions to protect global financial stability.

### **REFERENCES**

- Arthur, W. B. (2009). Complexity and the economy. In *Handbook of Research on Complexity*. Edward Elgar Publishing.  
<https://doi.org/10.4337/9781781952665.00007>
- Ashraf, S., G. S. Félix, E., & Serrasqueiro, Z. (2019). Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress? *Journal of Risk and Financial Management*, 12(2), 55.  
<https://doi.org/10.3390/jrfm12020055>

- Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8), 1121-1141. <https://doi.org/https://doi.org/10.1016/j.jedc.2012.04.001>
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192-213. <https://doi.org/10.1016/j.ins.2011.12.028>
- Borio, C., & Drehmann, M. (2009). Towards an operational framework for financial stability: "fuzzy" measurement and its consequences. *Documentos de Trabajo (Banco Central de Chile)*(544), 1. <https://dialnet.unirioja.es/servlet/articulo?codigo=3175938>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic perspectives*, 23(1), 77-100. <https://www.aeaweb.org/articles?id=10.1257/jep.23.1.77>
- Brunnermeier, M. K., & Oehmke, M. (2013). Bubbles, Financial Crises, and Systemic Risk. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 2, pp. 1221-1288). Elsevier. <https://doi.org/10.1016/B978-0-44-459406-8.00018-4>
- Bullard, J., Neely, C., & Wheelock, D. (2009). Systemic risk and the financial crisis: a primer. *Federal Reserve Bank of St. Louis Review*, 91(Sep), 403-418. [https://econpapers.repec.org/article/fipfedlr/vy\\_3a2009\\_3ai\\_3asep\\_3ap\\_3a403-418\\_3an\\_3av.91no.5.htm](https://econpapers.repec.org/article/fipfedlr/vy_3a2009_3ai_3asep_3ap_3a403-418_3an_3av.91no.5.htm)
- Bussiere, M., & Fratzscher, M. (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25(6), 953-973. <https://doi.org/10.1016/j.jimonfin.2006.07.007>
- Cevik, E. I., Kenc, T., Goodell, J. W., & Gunay, S. (2025). Enhancing banking systemic risk indicators by incorporating volatility clustering, variance risk premiums, and considering distance-to-capital. *International Review of Economics & Finance*, 97, 103779. <https://doi.org/10.1016/j.iref.2024.103779>
- Chavula, P., Kayusi, F., Lungu, G., & Uwimbabazi, A. (2025). The Current Landscape of Early Warning Systems and Traditional Approaches to Disaster Detection. *LatIA*(3), 77. <https://doi.org/10.62486/latia202577>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, <https://doi.org/10.1145/2939672.2939785>
- Claessens, S., & Kose, M. A. (2013). Financial Crises: Explanations, Types and Implications. <https://doi.org/10.2139/ssrn.2295201>
- Colander, D., Michael, G., Armin, H., Katarina, J., Alan, K., Thomas, L., & Sloth, B. (2009). The financial crisis and the systemic failure of the economics profession. *Critical Review*, 21(2-3), 249-267.

<https://doi.org/10.1080/08913810902934109>

- Eichengreen, B. (1992). *Golden fetters: The gold standard and the Great Depression, 1919–1939*. Oxford University Press.  
<https://fraser.stlouisfed.org/files/docs/meltzer/borrev94.pdf>
- Frankel, J. A., & Rose, A. K. (1996). Currency crashes in emerging markets: An empirical treatment. *Journal of International Economics*, 41(3), 351-366.  
[https://doi.org/10.1016/S0022-1996\(96\)01441-9](https://doi.org/10.1016/S0022-1996(96)01441-9)
- Fu, J., Zhou, Q., Liu, Y., & Wu, X. (2020). Predicting stock market crises using daily stock market valuation and investor sentiment indicators. *The North American Journal of Economics and Finance*, 51, 100905.  
<https://doi.org/10.1016/j.najef.2019.01.002>
- Greene, W. (2018). *Econometric Analysis* (8th ed.). Pearson Education Limited.  
<https://www.scirp.org/reference/referencespapers?referenceid=2920183>
- Hartwig, B., Meinerding, C., & Schüler, Y. S. (2021). Identifying indicators of systemic risk. *Journal of International Economics*, 132, 103512.  
<https://doi.org/10.1016/j.jinteco.2021.103512>
- Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: a horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12), 1933-1963. <https://doi.org/10.1080/14697688.2017.1357972>
- Kaminsky, G. L., & Reinhart, C. M. (1999). The twin crises: the causes of banking and balance-of-payments problems. *American economic review*, 89(3), 473-500.  
<https://doi.org/10.1257/aer.89.3.473>
- Khiaonarong, T., Korpinen, K. N., & Islam, E. (2025). Using Simulations for Cyber Stress Testing Exercises. <https://doi.org/10.5089/9798229010146.001>
- Laeven, L., & Valencia, F. (2013). Systemic Banking Crises Database. *IMF Economic Review*, 61(2), 225-270. <https://doi.org/10.1057/imfer.2013.12>
- Laeven, L., & Valencia, F. (2020). Systemic Banking Crises Database II. *IMF Economic Review*, 68(2), 307-361. <https://doi.org/10.1057/s41308-020-00107-3>
- Li, S., & Wang, S. (2014). A financial early warning logit model and its efficiency verification approach. *Knowledge-Based Systems*, 70, 78-87.  
<https://doi.org/10.1016/j.knosys.2014.03.017>
- 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>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Matta, C. E. d., Bianchesi, N. M. P., Oliveira, M. S. d., Balestrassi, P. P., & Leal, F. (2021). A comparative study of forecasting methods using real-life econometric

series data. *Production*, 31, e20210043. <https://doi.org/10.1590/0103-6513.20210043>

- Mienye, E., Jere, N., Obaido, G., Mienye, I. D. & Aruleba, K. (2024) ‘Deep Learning in Finance: A Survey of Applications and Techniques’, *AI*, 5(4), pp. 2066–2091. <https://doi.org/10.3390/ai5040101>
- Namaki, A., Eyvazloo, R., & Ramtinnia, S. (2023). A systematic review of early warning systems in finance. *arXiv preprint arXiv:2310.00490*. <https://doi.org/10.48550/arXiv.2310.00490>
- Radelet, S., & Sachs, J. (2000). The onset of the East Asian financial crisis. . *NBER Macroeconomics Annual*(15), 105-161. <https://doi.org/10.1086/654153>
- Raza, S.A., Govindaluri, S.M. and Bhutta, M.K.S. (2022) ‘Research themes in machine learning applications in supply chain management using bibliometric analysis tools’, *Benchmarking: An International Journal*, 30(3), pp. 834–867. <https://doi.org/10.1108/BIJ-12-2021-0755>
- Reinhart, C. M., & Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly*. princeton university press. <https://www.degruyterbrill.com/document/doi/10.1515/9781400831722/html>
- Simionescu, M. (2025) ‘Machine Learning vs. Econometric Models to Forecast Inflation Rate in Romania? The Role of Sentiment Analysis’, *Mathematics*, 13(1), Article 168. <https://doi.org/10.3390/math13010168>
- Singh Channe, P. (2024). *The Impact of AI on Economic Forecasting and Policy-Making: Opportunities and Challenges for Future Economic Stability and Growth*. <https://doi.org/10.13140/RG.2.2.24945.70249>
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic perspectives*, 28(2), 3-28. <https://doi.org/10.1257/jep.28.2.3>