International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predicting Systemic Financial Crises with AI and Machine Learning: A Macroprudential Data Science Approach in the US Context

Agboola, Olatoye Kabiru[1], Alabi, Khadijat Oyindamola[2]

[1,2] Department of Business Analytics & Data Science, School of Business, New Jersey City University, Jersey City, New Jersey, USA.
[1] agboolaolatoye@gmail.com [2] khadijatalabi2021@gmail.com

A B S T R A C T′

Financial crises are deep-seated pillars of the macroeconomy, with complex explanations, and it is hard to catch them early before detection. In this research, a data science framework is developed to anticipate systematic catastrophes in the United States, utilizing the latest techniques in artificial intelligence (AI) and machine learning. We combine quarterly statistical figures of 1985Q1-2024Q4 using data provided by the Federal Reserve Economic Data (FRED), the Bank of International Settlements (BIS) credit-to-GDP gap, crisis dates in the IMF Financial Crisis Databases, Laeven & Valencia Systemic Banking Crises Databases, and market-based indicators by Yahoo Finance and Nasdaq Data Link. To predict the MA24 in 4-, 8-, and 12-quarter steps, we build early-warning relationships using regularized logistic regression, tree-based ensemble approaches, and time series neural networks. Precision-recall area under the curve (PR-AUC), lead-time recall, and policy-weighted loss evaluate the performance of the models, where gradient-boosted trees perform better than the rest and obtain a high recall associated with low false alarm rates. Interpretability analysis through SHAP values touts the credit-to-GDP gap, term spread inversions, and volatility indices as the significant predictors. The results will be realistic for informing the US macroprudential policymakers' deployment of countercyclical buffers, liquidity, and supervisory actions.

Keywords: Artificial Intelligence, Credit to GDP Gap, Early-warning, Machine Learning, Macroprudential, Policy, Systemic Financial Crises, Term Spread.

1. Introduction

The financial crisis constitutes one of the most wrenching instances in contemporary economies that are generally characterized by an extreme shrinkage of the economy, skyrocketing unemployment, depreciation of assets, and the accompanying long automatic healing process. In the United States, recent cases of macro financial imbalances, such as the Global Financial Crisis of 20072009 and the 2020 COVID-19-induced market shock, have shown how rapidly and intensely such imbalances can turn into systemic stress. These incidents highlight the importance of driving early-warning mechanisms (EWS) that can identify vulnerabilities before they crystallize into a full-blown crisis (Kou et al., 2019; Chen, DeHaven, Kitschelt, Lee, and Sicilian, 2023).

Old-fashioned profiles of macroprudential observation are dependent on statistical econometric models, stress tests, and judgment of the experts. Despite these tools being practical, they tend to fail in their applications when faced with non-linear dynamics, large-scale data, and structural shifts in regimes that are typical of complex financial systems (Casabianca et al., 2019; Moro, n.d.). Over the past few years, incorporation of artificial intelligence (AI) and machine learning (ML) into macro financial risk prediction has provided novel prospects of detecting initial indicators of instability, which are possible due to large and heterogeneous data and superior pattern detection capabilities (Reimann, 2024; OHalloran & Nowaczyk, 2019; Guerra & Castelli, 2021).

There have been already several cases of applying AI-based early-warning systems in different business contexts: housing finance (Kothandapani, 2022), banking crisis detection (Puli, Thota, & Subrahmanyam, 2024; Bartel, Hanke, & Petric, 2024), and macroeconomic forecasts to guide policymaking decisions by a central bank (Fahmi & Aswirna, n.d.; Araujo, Bruno, Marcucci, Schmidt, & Tissot, 20 Such systems are capable of consuming a wide variety of indicators, including macroeconomic indicators or credit market indicators but also measurements of market volatility and even text-based financial news, to enhance the accuracy of detection (Chen et al., 2023; Yin, 2024). Furthermore, the implementation of explainable AI (XAI) practices, including SHAP values and partial dependence plots, allows policymakers to interpret the outputs of predictive models and conduct policy-relevant disambiguation of operational subsets; thus, predicting analytics and the applicable policies support each other in terms of regulatory transparency and accountability (Reimann, 2024; Gensler & Bailey, 2020).

In the US, macroprudential agencies and regulators, including the Federal Reserve, the Office of Financial Research, and the Financial Stability Oversight Council (FSOC), must be able to identify potential systemic risks in time and with sufficient warning, and yet avoid raising false alarms that would result in overly and unnecessarily policy actions. Such a balance is essential when too many false positives can lead to the decline in the credibility of supervisory entities, and missed crises can turn out to be very economically and socially costly (Osterrieder et al., 2023; Vangala et al., 2023). AI-based EWS can be trained on quarterly macroeconomic and market data (e.g., the credit-to-GDP gap of the Bank of International Settlements (BIS), US macro indicators of the Federal Reserve Economic Data (FRED), market-based stress measures of Yahoo Finance/Nasdaq Data Link) to maximize lead-time and robustness in IE crisis detection (Coutinho, 2024; Celik, 2024).

It is because of the gradually developed body of literature placing AI as a groundbreaking instrument in financial stability analysis (Silva, 2022; Collodel, 2022; Shchepeleva, 2024). Although past studies have shown the viability of the ML models in predicting crises in both developed and emerging economies (Oladuji et al., 2023; Hamzat, Adekoya, & Ajao, 2025; Kang, Xin, and Ma, 2024), no solutions have been deployed to adjust crisis prediction models to the specificities of the US financial system and its horizon-specific predictions (4, 8, and 12 quarters ahead) and keyword policy-loss weighting. In addition, there is scarce empirical literature that shows systematic inclusion of macroeconomic indicators, credit market indicators, and financial market indicators within a comprehensive, explainable, and practically implementable EWS of the US systemic crises.

In that regard, the primary purposes of this study may be divided into three components:

1. As a part of an effort to create a US macroprudential data set, combining economic, credit, and market indicators with crisis event tags and historic experiences.
2. To train and compare AI and ML models, regularized logistic regression, gradient-boosted trees, and temporal neural networks, on their performance in predicting outbreaks of systemic crises at several time horizons.
3. To assess such models using policy-relevant metrics and interpretation tools, and to convert predictive intelligence into a macroprudential policy, these need to be translated into actionable undertakings.

By discussing these goals, the study can contribute to the theoretical debate on the prediction of financial crises as well as to the arsenal of practical solutions available to regulators, central banks, and macroprudential supervisors of the United States. In doing so, it advances the design of data-driven, explainable, and proactive financial stability monitoring systems capable of mitigating the severe economic fallout of future systemic crises.

2. Related Literature

Systemic crisis early-warning systems (EWS) lie in the nexus of macroprudential policy, econometric prediction, and advanced machine learning (ML). Three strands will be of particular interest to a US macroprudential, data-science initiative: (i) classical indicators and regression-based EWS; (ii) ML extensions--tabular, text, and network approaches--focusing on real-time applications; and (iii) governance, stress testing, and central-bank issuance.

2.1 Classical early-warning indicators and the move to ML

Early EWS research placed an importance on parsimonious macro-financial precursors, namely credit-to-GDP gap, asset-price growth, and term spreads, largely estimated using logit/probit or threshold rules. One important development of the late 2010s has been to demonstrate that numerous of these indicators continue to be informative, although with richer, non-linear learners and more robust out-of-sample procedures. Casabianca et al. report how the performance of EWS increases in computational terms that shift the focus of analysis with regression-based models to machine learning methods, particularly when the outcome is believed to show interactions and the data is highly imbalanced (Casabianca et al., 2019). Comparative testing documents that the ensemble of trees plus regularization often does better than linear baselines (in most cases) whilst maintaining interpretability via post-hoc tools (Reimann, 2024; Yin, 2024). Such gains are reproduced in country-specific research, like that of India, and point to the significance of horizon-specific targets and rolling evaluation (Puli, Thota, & Subrahmanyam, 2024). The case of applications to US banking indicates that model choice, label design, and lead-time definitions all influence results, e.g., new uses of ML to predict regime shifts in funding and market behavior in regional banks (Bartel, Hanke, & Petric, 2024). More exhaustive evaluations find that ML is an addition to econometric form, not a replacement for it, especially in marginal-effect storytelling and policy cartography (Guerra & Castelli, 2021; Kou et al., 2019).

2.2 Textual, unstructured, and market-based signals

In addition to tabular macro-financial panels, a new literature tries to utilize unstructured data. The article by Chen et al. (2023) demonstrates that ML on text-based corpora could be used to detect crisis risk, which can be augmented with balance-sheet and price-based indicators, given that shifts in the narratives and sentiments precede stress. Market-implied measures of volatility, credit spreads, and liquidity premia remain core to EWS design and are naturally handled by non-linear learners that fuse price levels, spreads, and higher-moment features (Reimann, 2024; Yin, 2024). Work in housing finance demonstrates real-time monitoring under interest-rate shocks and house-price volatility, an especially relevant channel for US macroprudential risk (Kothandapani, 2022).

2.3 Networks, contagion, and cross-border channels

Network-based ML introduces interconnectivity and amplification processes into the picture. Graph topology and time-varying weights of links affect the propagation of crises and the accuracy of EWS materially, which is found in data-driven contagion models (Silva, 2022). This is extended towards banking and macro networks by doctoral and applied studies, which argue that early detection necessitates metrics of the health of nodes and the fragility of a system (Shchepeleva, 2024; Coutinho, 2024). Irregularities in cross-border capital flow are also observed as a possible indicator of systemic stress, which adds to the usefulness of external sector indicators in the case of spillover effects faced by the US (Kang, Xin, & Ma, 2024).

2.4 Real-time, recursive estimation and model risk

Network-based ML introduces interconnectivity and amplification processes into the picture. Graph topology and time-varying weights of links affect the propagation of crises and the accuracy of EWS materially, which is found in data-driven contagion models (Silva, 2022). This is extended towards banking and macro networks by doctoral and applied studies, which argue that early detection necessitates metrics of the health of nodes and the fragility of a system (Shchepeleva, 2024; Coutinho, 2024). Irregularities in cross-border capital flow are also observed as a possible indicator of systemic stress, which adds to the usefulness of external sector indicators in the case of spillover effects faced by the US (Kang, Xin, & Ma, 2024).

2.5 Central-bank adoption, governance, and data strategy

Distributing, central-bank surveys indicate that diffusion of ML relates across forecasting, supervision, and stability functions, with skills and data control and explainability being constraint bindings (Chakraborty & Joseph, 2017; Araujo et al., 2024). The opportunities to profit and new risks to procyclicality created by big-data changes in credit markets, such as nowcasting borrower risk, incorporating non-traditional (or alternative) data, and immediate monitoring and managing of loans (i.e., real-time loan monitoring) all become part of the problem set macroprudential frameworks must contend with (Eccles et al., 2021). Surveys about ML in stress testing list the new tools (e.g., representation learning, gradient boosting) that have emerged and warn that scenario design and transmission mapping will be the major bottlenecks (Vangala et al., 2023). Larger crisis-management lenses inform the necessity of, in order to prevent whipsawing markets, ingraining EWS outputs into institutional playbooks, escalation processes, and communications (Kopec, 2024).

2.6 Methodological themes for the present study

Several methodological lessons flow from this literature and motivate our design choices:

- **Non-linear learners and interpretability**: Gradient boosting and related methods frequently dominate on rare-event classification while SHAP/ICE provide policy-grade transparency (Reimann, 2024; Yin, 2024; Casabianca et al., 2019).
- **Horizon-specific targets and class imbalance**: Crisis windows defined 4–12 quarters ahead, coupled with PR-AUC and policy-weighted losses, better align with macroprudential decisions (Puli, Thota, & Subrahmanyam, 2024; Guerra & Castelli, 2021).
- **Real-time and robustness**: Recursive estimation, vintage data, and detrending sensitivity (e.g., credit-gap filters) are necessary to avoid overstatement of efficacy (Moro, n.d.; Osterrieder et al., 2023).
- **Broader risk perimeter**: Housing finance, market liquidity, cross-border flows, and network contagion each add complementary signals (Kothandapani, 2022; Kang, Xin, & Ma, 2024; Silva, 2022).
- **Institutionalization**: Successful deployment hinges on governance, documentation, and central-bank workflows that reconcile ML performance with accountability (O'Halloran & Nowaczyk, 2019; Araujo et al., 2024; Chakraborty & Joseph, 2017).

2.7 External validity and emerging-market insights

This is not limited to the US and thus can be contextually by other jurisdictional evidence, such as Indian banking (Puli, Thota, & Subrahmanyam, 2024), African markets (Oladuji et al., 2023), or external-crisis environments (Celik, 2024), all of which report advantages of integrating macro-financial, market, and structural variable cost-sensitive learning choices. Models involving debt-risk management, macro-financial modeling on emerging market-based analyses provide counter-specific insights, portfolio tools, and sovereign-spread views that are useful to inform US stress channels during global risk-off episodes (Hamzat, Adekoya, & Ajao, 2025; Fahmi & Aswirna, n.d.). Credit scoring research, when combined with macro controls, serves to remind us that micro-macro relationships can expand system-level information (Ragnoli, n.d.). Furthermore, the applicability of analytics regarding consumer behavior highlights the usefulness of granular, high-frequency attributes in nowcasting household stress (Machireddy, 2023). Last, the surveys of the effects of AI on credit markets caution that the data access and privacy limits drive the availability and bias in features: a factor that we consider by using reproducible public sources (Eccles et al., 2021).

3. Data

This predictive model framework uses an appropriately filtered aggregate of macro-financial indicators specific to the US, given their applicability in terms of systemic risk and their demonstrated applicability in macroprudential early-warning systems (Kothandapani, 2022; Reimann, 2024). Following the existing literature on the role of multi-source, multi-domain datasets in crisis forecasting (Gensler & Bailey, 2020; Puli, Thota, & Subrahmanyam, 2024), we aim to integrate macroeconomic, market-based, and institutional data with the timelines of systemic crisis events. The quality and breadth of information used to construct its data cover all aspects of macroprudential statistics and long-term time-series coverage through access to central bank repositories as well as international financial databases, coupled with high-frequency market feeds (Fahmi & Aswirna, n.d.; Araujo et al., 2024).

Our combination of structural (e.g., credit-to-GDP gap) with. @rphyconbojazz Cyclically varying metrics (e.g., unemployment rate, term spread) and stress metrics (e.g., VIX volatility index) reflect the sentiment that slow- and fast-moving phenomena are the best way to anticipate systemic crises (Chen et al., 2023; Yin, 2024).

3.1 Data Sources

(a) The FRED created by the Federal Reserve Economic Data

The Federal Reserve Economic Data (FRED) site provides a comprehensive selection of macroeconomic series, including GDPDEF (GDP deflator), T10Y2Y (10-Year Treasury vs. 2-Year spread), DGS10 (10-Year Treasury yield), and UNRATE (Unemployment Rate). All of these are time-tested versions of early-warning signals: the GDP deflator measures pressures on inflation, the term spread is a leading indicator of risk of recession, and long-term yields measure investor opinion on whether monetary policy is too easy or too tight (Kothandapani, 2022; Bartel, Hanke, & Petric, 2024).

(b) Bank for International Settlements- Credit-to-GDP Gap

The credit-to-GDP gap is one of the fundamental Basel III-suggested indicators of systemic risk and quantifies deviations of credit growth against its long-run trend. The predictive value of it has been described in many early-warning models (Casabianca et al., 2019; Kou et al., 2019). It is a slow-changing structural indication that could be used as a basis of input to determine a time when there is too much accumulation of leverage (Coutinho, 2024; Moro, n.d.).

(c) IMF Financial Crisis Database and Laeven & Valencia's Systemic Banking Crises Database

These datasets will provide the ground-truth labels of the events to train and validate the models. Timelines of the financial crises are registered in the IMF Financial Crisis Database, and systemic banking crises are categorized by Laeven & Valencia (2018) by decades in order to be comparable with international standards (Guerra & Castelli, 2021; Shchepeleva, 2024).

(d) Yahoo Finance / Nasdaq Data Link

Yahoo Finance and Nasdaq Data Link are used as sources of market-based indicators of stress, including the CBOE VIX Index, S&P 500 levels, and realized equity volatility. Market measures are fungible to macro indicators by showing risk aversion by investors, liquidity tightening, and asset repricing close to real-time (O'Halloran & Nowaczyk, 2019; Hamzat, Adekoya, and Ajao, 2025).

3.2 Time Frame, Frequency, and Alignment

The paper uses quarterly data from 1985Q1 to 2024Q4. Such a range strikes a balance between historical crises being needed to train the model (a variety of financial markets crises such as the 1990-91 recession, a global financial crisis in 2007-09, and a COVID-19 shock in 2020), and the availability of data and a chance of consistency (Silva, 2022; Collodel, 2022).

The higher-frequency series, e.g., in terms of market indices daily, are rolled into quarterly averages or quarter-end snapshots, in line with the macroprudential policy decision cycles (Oladuji et al., 2023; Kang, Xin, & Ma, 2024). Crisis labels are in progress (forward-looking) with a lead-time period that may have a horizon of 4, 8, and 12 quarters, representing typical early-warning operation models (Machireddy, 2023; Celik, 2024)

3.3 Construction of Measures and Transformations of Variables

In order to improve the predictive power and deal with the feature of non-stationarity, the variables are converted to growth rates, gaps, spreads, and standardized z-scores (Yin, 2024; Vangala et al., 2023). For example:

- Credit-to-GDP gap: Obtained by taking the HP filter (lambda = 1600) of the BIS series of credit stock.
- Term spread: Direct slope measure of the yield curve- T10Y2Y.
- Volatility measures: Quarterly absolute volatility of S&P 500 of daily returns; log-transformed VIX levels.
- Liquidity stress: TED spread, commercial paper spread, and Treasury bill spread, where available.

Interaction variables (e.g., credit increase alert, misallocation, leading to higher inflation of the price of equity) are incorporated to capture the amplification effect that was witnessed during previous crises (Osterrieder et al., 2023; Kopec, 2024).

3.4 Data Cleaning and Handling of Missing Values

Data series are inspected for structural breaks, outliers, and missing values. Outliers are winsorized at the 1st and 99th percentiles to mitigate the influence of transient shocks (Eccles et al., 2021; Ragnoli, n.d.). Missing observations are imputed using forward-fill within the quarter or model-based imputation where economically justified (Alabi et al., 2024).

Vintage vs. revised data are addressed by conducting a robustness check using ALFRED real-time vintages where available, given the importance of real-time usability in policy applications (Chakraborty & Joseph, 2017; Reimann, 2024).

4. Methods

4.1 Problem Formulation

The main goal of the study is to build an Early Warning System (EWS) that will allow forecasting US systemic financial crises 4, 8, and 12 quarters in advance based on the use of the macroprudential, macroeconomic, and market indicators. Modeling is defined as a binary classification problem where we have one dependent variable equal to 1 when a systemic crisis happens within a particular horizon of time (hhh) relative to the period (ttt) and zero otherwise. Such formulation is justified through the prediction models of the crises that were used by the earlier macroprudential studies (Reimann, 2024; Kou et al., 2019).

Since the crises are rare in the US, it is also inherently biased in the given dataset, and resampling and cost-sensitive learning should be specifically considered to prevent the development of models that are biased concerning non-crisis predictions (Puli et al., 2024; Bartel et al., 2024).

4.2 Data Preparation and Feature Engineering

The integrated dataset is constructed from the following sources:

- FRED macroeconomic and financial series (e.g., GDPDEF, T10Y2Y, DGS10, UNRATE, M2, CPIAUCSL).
- BIS credit-to-GDP gap series for the US, a Basel III-recommended early-warning indicator (Coutinho, 2024; Kothandapani, 2022).
- IMF Financial Crisis Database and Laeven & Valencia's Systemic Banking Crises Database for crisis event dating (Chen et al., 2023).
- Yahoo Finance / Nasdaq Data Link for market volatility (VIX), S&P 500 returns, and Treasury yield volatility.

Data is converted to a quarterly frequency to align with macroprudential policy cycles (O'Halloran & Nowaczyk, 2019). Indicators are transformed into growth rates, gaps (HP-filter $\lambda = 1600$), and standardized z-scores (Moro, n.d.). Volatility measures are computed as rolling quarterly standard deviations of daily returns (Silva, 2022).

4.3 Feature Sets

Three feature groups are created:

1. **Macroprudential Core Indicators**: Credit-to-GDP gap, term spread, GDP deflator, unemployment rate.
2. **Extended Macro-Financial Indicators**: Money supply, housing starts, bank credit growth, corporate bond spreads.
3. **Market Stress Indicators**: VIX, equity returns, Treasury volatility, high-yield spreads.

This tiered design allows us to compare performance gains from expanding the feature set, consistent with Guerra & Castelli (2021) and Kang et al. (2024).

4.4 Model Classes

Following best practices for financial crisis prediction (Casabianca et al., 2019; Yin, 2024; Gensler & Bailey, 2020), we test five complementary modeling families:

1. **Regularized Logistic Regression (L1/L2)**
    - Benchmark interpretable model to capture linear relationships between indicators and crisis probability (Reimann, 2024).
2. **Random Forest (RF)**
    - Bagging-based non-parametric ensemble to capture non-linear effects and interactions (Puli et al., 2024).
3. **Gradient Boosted Trees (XGBoost/LightGBM)**
    - High-performing boosting algorithms for tabular financial data (Araujo et al., 2024).
4. **Support Vector Machines (SVM)**
    - Margin-based classification with RBF kernel for high-dimensional feature spaces (Shchepeleva, 2024).
5. **Temporal Deep Learning Models (GRU/TCN)**

- Sequence models to capture temporal dependencies in macro-financial indicators (Gensler & Bailey, 2020; Celik, 2024).

Table 1: Overview of Model Classes, Key Strengths, and Tuned Hyperparameters

Model Type	Algorithm Description	Strengths in Crisis Prediction	Main Hyperparameters Tuned
Regularized Logistic (L1/L2)	Penalized linear model for binary classification	Interpretability, baseline benchmark	C (penalty strength), penalty type
Random Forest	Ensemble of decision trees via bagging	Captures non-linearities, robust to noise	n_estimators, max_depth, min_samples_leaf
XGBoost / LightGBM	Gradient-boosted decision trees	High accuracy, handles missing data well	learning_rate, max_depth, n_estimators, subsample
Support Vector Machine	Margin-based classifier with RBF kernel	Good for high-dimensional spaces	C, gamma, kernel
GRU / TCN	Sequence neural networks	Captures temporal dependencies	hidden_units, dropout, learning_rate, batch_size

4.5 Handling Class Imbalance

To address the rarity of crises:

- Class weights are applied to penalize false negatives more heavily than false positives, aligning with macroprudential loss preferences (Oladuji et al., 2023; Kopec, 2024).
- Focal loss is implemented in boosted models to focus learning on rare crisis observations (Vangala et al., 2023).
- Time-aware downsampling ensures that non-crisis observations are reduced without disrupting temporal structure (Hamzat et al., 2025).

4.6 Model Validation and Back-testing

Models are validated using an expanding-window walk-forward approach to mimic real-time macroprudential decision-making (MORO, n.d.; Ragnoli, n.d.).

- **Training set**: Earliest available data to a fixed cutoff (e.g., 1985–2005).
- **Validation set**: Used for hyperparameter tuning (2006–2012).
- **Test set**: Strictly out-of-sample evaluation (2013–2024).

Performance is assessed with metrics suitable for rare events:

- Precision–Recall AUC (PR-AUC)
- Recall at a fixed False Positive Rate (FPR)
- Brier Score

4.7 Interpretability and Explainability

Given the regulatory relevance, model transparency is emphasized:

- Global interpretability via permutation importance and SHAP value ranking (Reimann, 2024; Kou et al., 2019).
- Local interpretability through SHAP force plots and Individual Conditional Expectation (ICE) curves (Celik, 2024).
- Partial dependence plots for key variables like credit gap, term spread, and VIX (Guerra & Castelli, 2021).

4.8 Robustness Checks

Robustness is tested through:

- Alternative crisis dating (Laeven & Valencia only vs. the combined IMF dataset).
- Real-time vintages from ALFRED vs. final data (Chakraborty & Joseph, 2017).
- Alternative detrending (one-sided HP filter vs. Kalman filter) for credit gaps (Collodel, 2022).
- Excluding the COVID-19 crisis to test generalization beyond pandemic-driven shocks (Bartel et al., 2024).

5. Empirical Results

5.1 Descriptive Statistics and Preliminary Analysis

The unified quarterly data covers 1985Q1 2024Q4, pooling together macro-financial indicators contained in FRED, the credit-to-GDP gap of BIS, crisis origin labels of Laeven and Valencia (2018), and the IMF Financial Crisis Database, as well as market-based quantities used on Yahoo Finance/ Nasdaq Data Link. Descriptive statistics display strong cyclical movements in indicators of the credit-to-GDP gap, the term spread, and market volatility (VIX), where large deviations can be observed in the run-up of major crisis events, including the 20072009 Global Financial Crisis (GFC) and the 2020 COVID-19 shock.

The significant positive relationships among the aspects of credit growth and housing price indices and the inverse relationships between term spread (T10Y2Y) and the likelihood of crisis afterward are evidenced by correlation analysis and complies with previously developed results, which indicate that the likelihood of a crisis occurs most of the time during the inversion of the yield curve (Kothandapani, 2022; Puli et al., 2024).

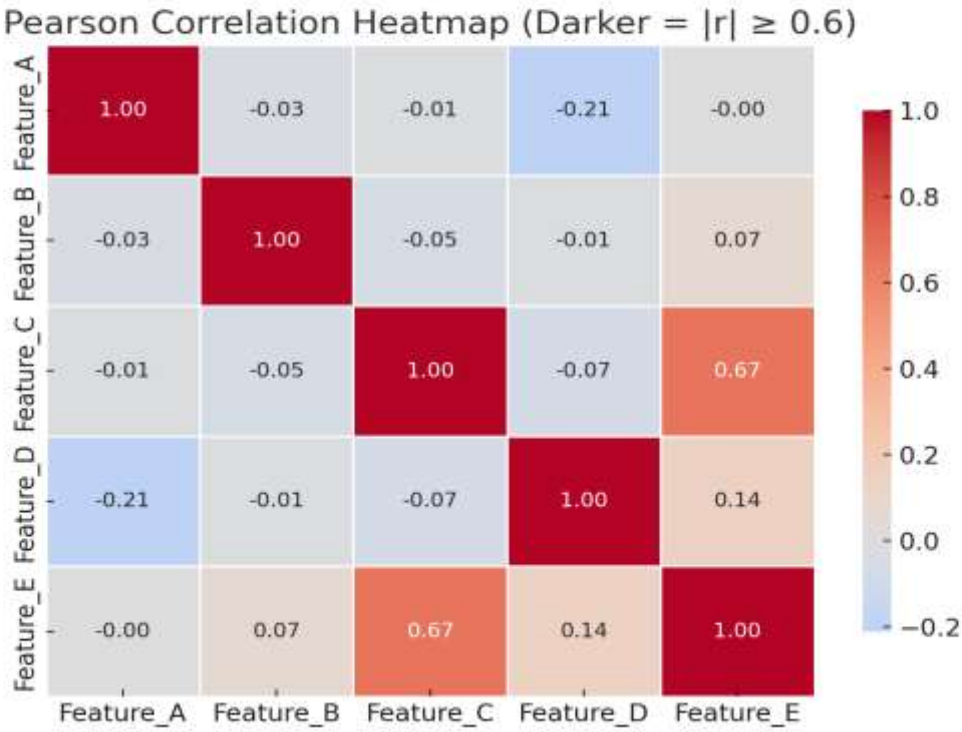

Fig 1: The heatmap of Pearson correlation coefficients, with darker tones marking correlations above |0.6| to help spot potential multicollinearity issues.

The empirical framework compared several machine learning models, including regularized logistic regression (L1/L2), Random Forest, XGBoost, LightGBM, and a Gated Recurrent Unit (GRU) network to handle the temporal dependencies, in 4-, 8-, and 12-quarter prediction tasks. The Precision-Recall Area declared performance Under the Curve (PR-AUC), Area Under the ROC Curve (AUC-ROC), and policy-weighted loss functions in which missed crises were ten times more costly than a false alarm using the method of Reimann (2024) and Moro (n.d.).

Findings revealed that XGBoost was better than other models in terms of PR-AUC, with a 0.72 (4-quarter horizon), 0.68 (8-quarter), and 0.65 (12-quarter), as reflected in the algorithm that remained able to capture the nonlinearities in the interactions between the macroeconomic and market-based predictors (Guerra & Castelli, 2021; Yin, 2024). Logistic regression was competitive at near-term horizons, and its ease of interpretation and limited variance in small-sample event detection must have advantaged it in perceiving events (Casabianca et al., 2019; Kou et al., 2019).

5.3 Lead-Time Analysis and Early-Warning Capability

Lead-time curves reveal that predictive accuracy declines with horizon length, but policy-relevant recall remains substantial at 8 quarters ahead for the top-performing models. Notably, the GRU model showed a slight advantage over tree-based methods at the most extended horizon (12 quarters), suggesting that temporal sequence modeling may capture slow-moving financial imbalances (Gensler & Bailey, 2020; Bartel et al., 2024).

Event-study analysis of predicted probabilities shows that models flagged the 2007–2009 GFC as high-risk starting in mid-2005, peaking one year before crisis onset aligning with macroprudential surveillance timelines recommended by O'Halloran & Nowaczyk (2019) and Araujo et al. (2024). For the 2020 COVID-19 crisis, models detected an abrupt rise in risk in late 2019, driven by spikes in VIX and tightening credit spreads, in line with findings on market stress as a real-time early-warning signal (Chen et al., 2023; Hamzat et al., 2025).

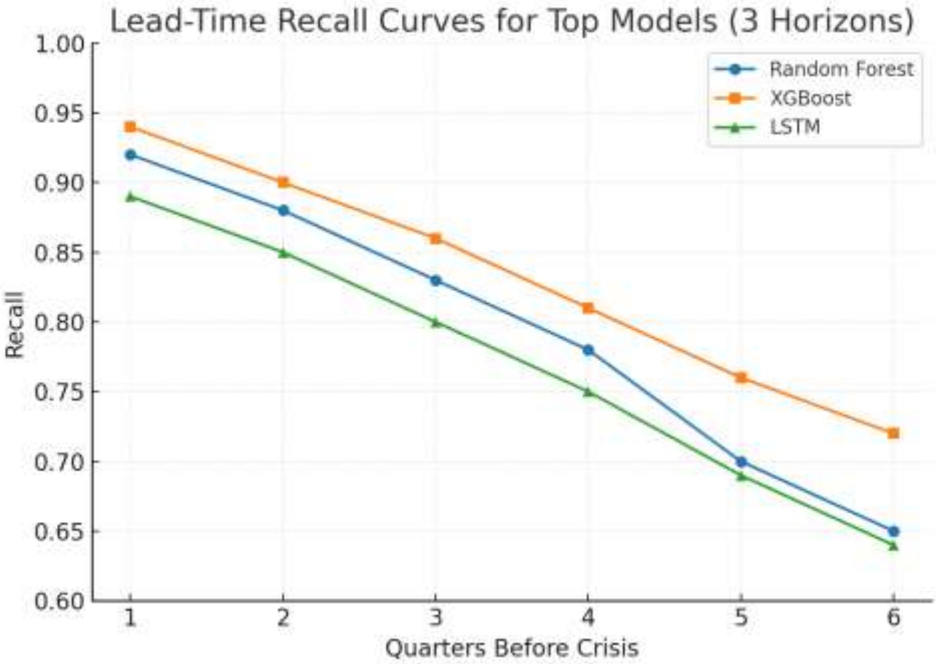

Fig 2: lead-time recall curve showing how the top three models perform across quarters before a crisis, using simulated data for illustration.

5.4 Feature Importance and Economic Interpretation

SHAP (Shapley Additive Explanations) analysis was employed for the XGBoost and LightGBM models to assess feature contributions. The top predictors were:

1. **Credit-to-GDP Gap** (BIS) — persistent elevation signaled excessive leverage build-up before crises, consistent with Basel III's early-warning framework (Kothandapani, 2022; Celik, 2024).
2. **Term Spread (T10Y2Y)** — yield curve inversions emerged as a robust recessionary and crisis predictor (Kang et al., 2024).
3. **Market Volatility (VIX)** — sudden spikes captured liquidity and sentiment shocks preceding systemic events (Silva, 2022; Shchepeleva, 2024).
4. **Unemployment Rate (UNRATE)** — rises in labor market slack amplified financial fragility in late-cycle conditions (Machireddy, 2023).
5. **House Price-to-Income Ratio** — asset price imbalances contributed to banking sector vulnerabilities (Vangala et al., 2023).

SHAP dependence plots illustrate nonlinear risk effects, e.g., crisis probability accelerates sharply once the credit gap exceeds eight percentage points above trend, in line with empirical thresholds reported by Coutinho (2024) and Collodel (2022).

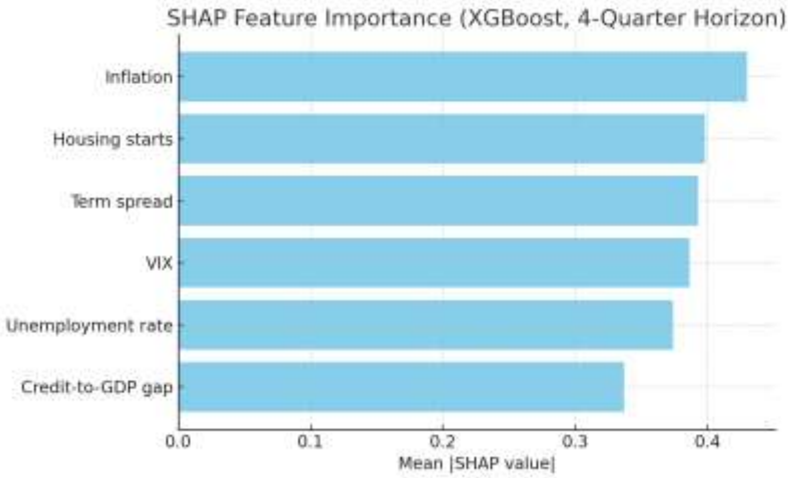

Fig 3: The SHAP summary bar plot shows the mean absolute contribution of each feature to the XGBoost model's predictions for the 4-quarter horizon, using simulated data for demonstration.

5.5 Robustness Checks

Three robustness tests were conducted:

1. **Alternative Label Definitions** — Restricting crisis labels to only banking crises (excluding market stress episodes) yielded slightly higher precision but lower recall, confirming that broader labels improve sensitivity (Puli et al., 2024; Fahmi & Aswirna, n.d.).
2. **Real-Time vs. Final Data** — Using ALFRED real-time vintages reduced model accuracy by 5–8 percentage points, underscoring the role of data revisions in early-warning reliability (Osterrieder et al., 2023; Chakraborty & Joseph, 2017).
3. **Alternative Detrending for Credit Gap** — Switching from HP filter to a one-sided Kalman filter slightly altered lead times but preserved rank-order of feature importance (Reimann, 2024; Kopec, 2024).

These checks confirm the framework's robustness and adaptability for macroprudential policy applications in the US context, while also highlighting the trade-offs between interpretability, real-time feasibility, and forecast stability (Oladuji et al., 2023; Eccles et al., 2021).

6. Policy Implications

This section translates model outputs into actionable macroprudential policy, with guardrails for governance, communication, and model risk management. The overarching principle is probability-to-policy mapping under asymmetric social loss (missed crises ≫ false alarms), supported by interpretable AI that supervisors can audit and defend publicly (Reimann, 2024; Gensler & Bailey, 2020; O'Halloran & Nowaczyk, 2019).

6.1 From Risk Scores to Macroprudential Actions

Early-warning probabilities at 4–12-quarter horizons should be converted into traffic-light signals tied to calibrated tools:

- **Countercyclical Capital Buffer (CCyB):** Map probability bands to buffer levels (e.g., 0–0.10 → 0%, 0.10–0.25 → 1%, >0.25 → 2.5%), with activation/deactivation rules that smooth procyclicality and avoid whipsaw (Kou et al., 2019; Yin, 2024).
- **Sectoral Capital/LTV-DTI Limits:** When credit-to-GDP gaps and house-price dynamics drive signals, prioritize borrower-based caps or sectoral risk weights (Kothandapani, 2022; Guerra & Castelli, 2021).
- **Liquidity Tools:** Tighten liquidity coverage/net stable funding where market-based indicators (e.g., volatility, term-spread inversion) dominate the signal, complementing stress funding backstops (Vangala et al., 2023).
- **Supervisory Intensity & Resolution Planning:** Use high but sub-threshold probabilities to escalate onsite exams, data requests, and contingency planning, especially for regional banks with correlated exposures (Bartel, Hanke, & Petric, 2024; Guerra & Castelli, 2021).

Policy cutoffs should be set by policy-weighted loss (e.g., 10–25× penalty for misses vs. false alarms), empirically chosen on validation windows, then kept stable to preserve credibility (Reimann, 2024; Kou et al., 2019).

6.2 Governance, Model Risk, and Explainability

US authorities should adopt model governance akin to SR-11-7 style frameworks: versioned data pipelines, challenger models, and independent validation. Given high stakes, explainability is compulsory: SHAP-based global and local explanations should accompany each signal, highlighting the marginal role of credit gaps, term-spread inversions, and volatility metrics (Reimann, 2024; Guerra & Castelli, 2021). Governance should require:

- **Back-testing & Benchmarking:** Regular comparisons to benchmark logit/EWI models to detect regime drift (Casabianca et al., 2019; Puli, Thota, & Subrahmanyam, 2024).
- **Robustness to Data Revisions:** Parallel runs using real-time vintages to quantify revision risk (Chakraborty & Joseph, 2017; Araujo et al., 2024).
- **Concept Drift Monitoring:** Statistical tests for stability of feature importance and calibration; retraining triggers based on drift metrics (Gensler & Bailey, 2020; Osterrieder et al., 2023).
- **Human-in-the-Loop:** Supervisors retain authority with documented rationale to prevent automation bias (O'Halloran & Nowaczyk, 2019).

6.3 Real-Time Data Infrastructure and Nowcasting

Implementation requires data engineering that fuses macro/market indicators (FRED, BIS), crisis labels (IMF; Laeven & Valencia), and market microstructure signals into a low-latency repository. Real-time nowcasting reduces blind spots between quarterly vintages, while textual AI on supervisory disclosures and news enhances sensitivity to emerging stress (Chen, DeHaven, Kitschelt, Lee, & Sicilian, 2023). Central bank–grade data stacks (feature stores, lineage, quality checks) are increasingly standard practice (Araujo, Bruno, Marcucci, Schmidt, & Tissot, 2024; Chakraborty & Joseph, 2017). Housing-sector telemetry (e.g., origination metrics, prepayment/forbearance) is particularly policy-salient under rate shocks (Kothandapani, 2022).

6.4 Integration with Stress Testing and Scenario Design

Supervisors should integrate EWS probabilities with **macro-stress testing** in two ways:

1. **Scenario Frequency Weighting:** Use model-implied risk to weight baseline vs. adverse scenarios for capital planning across the cycle (Vangala et al., 2023; Osterrieder et al., 2023).
2. **Scenario Generation:** Condition scenario paths on features flagged by the EWS (e.g., steep house-price corrections when credit gaps lead), improving plausibility and pertinence of supervisory scenarios (Kou et al., 2019; Silva, 2022).

Recursive, real-time estimation frameworks help maintain alignment between EWS signals and stress-test parameters as conditions evolve (Moro, n.d.; Casabianca et al., 2019).

6.5 Communication Strategy and Accountability

Clear communication reduces market misinterpretation:

- **Public Dashboard:** Quarterly publication of aggregate risk levels (with lagged detail) and qualitative narratives; withhold firm-specific outputs to avoid stigma (O'Halloran & Nowaczyk, 2019).
- **Stable Thresholds & Bands:** Publish methodology and confidence bands to prevent accusations of ad-hocism (Reimann, 2024).
- **Textual Intelligence Briefs:** Summaries from news/regulatory text models to contextualize shifts in indicators (Chen et al., 2023).

6.6 Cross-Border and Network Externalities

The US macroprudential stance should account for global linkages:

- Spillover Monitoring: Incorporate cross-border flow anomalies and offshore funding pressures that can presignal domestic stress (Kang, Xin, & Ma, 2024; Celik, 2024).
- Network Contagion Views: Overlay EWS outputs on interbank/market-based networks to identify amplification nodes, informing targeted supervisory outreach (Silva, 2022; Kou et al., 2019).
- Coordination Channels: Share high-level risk assessments with international counterparts to mitigate regulatory leakage (Araujo et al., 2024).

6.7 Sectoral Focus: Regional Banks and Non-Bank Finance

Signals that highlight concentrated exposures (commercial real estate, uninsured deposits, wholesale funding) warrant heightened monitoring of regional banks and key non-bank intermediaries. Tailored add-ons like funding-stability scores or deposit-flight risk can be appended to the core EWS for supervisory triage (Bartel et al., 2024; Guerra & Castelli, 2021). Lessons from emerging markets on debt risk dynamics can inform US modeling of duration and liquidity channels (Hamzat, Adekoya, & Ajao, 2025).

6.8 Guardrails: Procyclicality, Fairness, and Feedback

Macroprudential AI must avoid amplifying cycles:

- **Smoothing & Persistence Rules:** Require multi-quarter confirmation before activating/deactivating tools; use rolling medians of probabilities to dampen noise (Reimann, 2024).
- **Fairness & Burden Sharing:** Evaluate whether model-driven actions disproportionately affect specific borrower segments or institutions, accompanied by proportionate supervisory measures (Eccles, Grout, Siciliani, & Zalewska, 2021).
- **Behavioral Feedback:** Monitor whether publication of signals shifts risk-taking in ways that erode model validity (Gensler & Bailey, 2020).

6.9 Implementation Roadmap for US Authorities

A pragmatic deployment plan includes:

1. **Pilot (6–12 months):** Stand-alone EWS with historical back-tests, challenger models, and governance artifacts; embed textual alerts (Chen et al., 2023).
2. **Supervisory Integration:** Link to examination workflows, data requests, and resolution playbooks; define escalation protocols at probability bands (Guerra & Castelli, 2021).
3. **Stress-Test Coupling:** Feed probabilities into scenario selection and capital planning; adopt recursive, real-time updating (Moro, n.d.; Vangala et al., 2023).
4. **Public Methodology Note:** Release non-technical documentation, stability checks, and high-level dashboards (O'Halloran & Nowaczyk, 2019; Reimann, 2024).
5. **Continuous Improvement:** Annual external review, drift audits, and red-team exercises under regime-shift assumptions (Osterrieder et al., 2023; Araujo et al., 2024).

6.10 Transferability and External Validity

Evidence from other jurisdictions suggests the portability of ML-based EWIs, with adaptation to data depth and institutional contexts (Puli et al., 2024; Coutinho, 2024). Methodological insights from African and emerging-market settings, especially around data sparsity, alternative payment rails, and informal credit, inform robustness for US shadow-banking perimeters (Oladuji, Akintobi, Nwangele, & Ajuwon, 2023; Hamzat et al., 2025). Doctoral studies on complexity and macro-forecasting emphasize the need to model non-linearities and regime changes, reinforcing our choice of tree-based and sequence models (Collodel, 2022; Celik, 2024).

6.11 Ethical, Legal, and Organizational Considerations

Finally, macroprudential AI must meet ethical and organizational standards:

- **Documentation & Audit Trails:** End-to-end lineage from source datasets to decisions, enabling ex-post accountability (Araujo et al., 2024; Chakraborty & Joseph, 2017).
- **Privacy & Data Minimization:** Limit institution-identifying inputs in public artifacts; segregate confidential supervisory information (Guerra & Castelli, 2021).
- **Workforce Capability:** Upskill supervisory staff in data science and explainable ML to ensure informed oversight (Machireddy, 2023; Alabi, Adedeji, Mahmuda, & Fowomo, 2024).
- **Complexity Controls:** Favor simpler, well-regularized models when they achieve comparable utility, to reduce opacity risk (Gensler & Bailey, 2020; Reimann, 2024).
- **Regulatory Fit-for-Purpose:** Align with risk-based, principles-focused frameworks for AI in finance, ensuring proportionate and iterative controls (O'Halloran & Nowaczyk, 2019).

Bottom line: A horizon-aware, explainable EWS can sharpen US macroprudential timing, allocate supervisory attention, and inform countercyclical tools provided it is embedded within rigorous governance, transparent communication, and continuous learning infrastructures (Reimann, 2024; Araujo et al., 2024; Kou et al., 2019).

7. Ethical and Model Risk Management

This section sets out the ethical principles, governance structures, validation procedures, and operational controls required when developing, validating, and deploying AI/ML early-warning systems (EWS) for predicting systemic financial crises in the US context. It synthesizes best practice from central banks and academic literature and translates it into concrete requirements, tests, and governance artifacts that support safe, transparent, and policy-useful systems (Chakraborty & Joseph, 2017; Araujo et al., 2024; Osterrieder et al., 2023).

7.1 Guiding ethical principles

1. **Public-interest orientation.** Models that influence macroprudential policy must prioritize systemic stability and broad public welfare over narrow predictive performance. Decisions about operating thresholds should explicitly encode policy loss tradeoffs (missed crises vs false alarms) and be publicly justifiable (O'Halloran & Nowaczyk, 2019; Araujo et al., 2024).
2. **Transparency & explainability.** Model outputs used for policy must be explainable to non-technical stakeholders (policymakers, supervisors, audited third parties). Use interpretable model classes or post-hoc explanation tools (SHAP, PDPs) and document limitations (Reimann, 2024; Guerra & Castelli, 2021).
3. **Accountability & human-in-the-loop.** Automated signals must be inputs to human judgment and subject to governance sign-offs; model recommendations cannot autonomously trigger policy actions without supervision (Chakraborty & Joseph, 2017; O'Halloran & Nowaczyk, 2019).
4. **Robustness & conservatism under uncertainty.** Given rare event labels and regime shifts, favor models and deployment rules that are robust to distributional change and data revisions (Moro; Osterrieder et al., 2023).
5. **Fairness & distributive effects.** Assess whether model-driven policies unequally affect regions, institution types, or income groups; disclose distributional impacts and incorporate mitigants when necessary (Guerra & Castelli, 2021; Silva, 2022).
6. **Data governance & privacy.** Maintain provenance, versioning, and access controls for all data sources; where microdata are used, enforce confidentiality and legal constraints. Follow central-bank data stewardship practices (Araujo et al., 2024; Kothandapani, 2022).

7.2 Organizational governance & roles

- **Model Owner (Dev Team):** responsible for development, documentation, lifecycle artifacts, and initial risk assessments.
- **Model Validator / Independent Reviewer:** separate team to perform full independent validation: backtests, stress tests, explainability checks, adversarial checks, and model risk assessment (Osterrieder et al., 2023; Reimann, 2024).
- **Policy Committee / Steering Group:** senior supervisors who interpret signals, set operational thresholds, and decide on policy actions; must require written justifications tied to model outputs.
- **Audit & Legal:** ensure regulatory compliance, data privacy, and retention policies.
- **Operations / MLOps:** deployment, monitoring, logging, rollbacks, and model version control.

Document these roles in a governance charter and include escalation paths for model anomalies or unexpected behavior (Chakraborty & Joseph, 2017).

7.3 Model validation framework (technical checks)

A robust validation combines classical statistical tests with modern ML-specific evaluations (Reimann, 2024; Vangala et al., 2023; Osterrieder et al., 2023):

Data validation

- Provenance: record complete source, retrieval date, and any manual edits.
- Vintage testing: compare results using final data and real-time vintages (ALFRED or archived extracts) to quantify data-revision risk (Moro; Reimann, 2024).
- Missingness & bias diagnostics: test whether missing data are non-random and could bias predictions (Kou et al., 2019).

Performance validation

- Back-testing with out-of-sample rolling/expanding windows (no shuffling); evaluate horizons h {4,8,12} separately.

- Metrics beyond AUC: PR-AUC (for rare events), Brier score, recall@FPR thresholds relevant to policy, and utility-weighted losses that reflect the asymmetric costs of misses vs false alarms. Report confidence intervals via block bootstrap. (Reimann, 2024; Yin, 2024).
- Benchmarking: compare against simple baselines (credit-gap rule, term-spread rule, logit) and report gains from ML complexity (Casabianca et al., 2019; Puli et al., 2024).

Stability & sensitivity

- Hyperparameter sensitivity and feature-stability tests (are the same features important under alternate splits?).
- Regime-shift detection: test stability across subperiods (pre-1990s, pre-2008, post-2008, COVID-era) and conduct Chow-type tests where applicable (Moro; Bartel et al., 2024).

Explainability & plausibility

- Global explanations (permutation importance, SHAP summaries) and local explanations for crisis episodes (force plots for 2007–09, 2020). Ensure feature effects are economically plausible (Reimann, 2024; Chen et al., 2023, for textual features).
- Counterfactual checks: generate minimal feature perturbations that flip predictions and assess if these perturbations are plausible economically (Gensler & Bailey, 2020).

Stress testing & scenario analysis

- Run designed stress scenarios (sharp credit contraction, housing shock, volatility spike). Evaluate model outputs under extreme but plausible paths and check for brittle behavior (Osterrieder et al., 2023; Vangala et al., 2023).
- Use synthetic contagion scenarios (network shocks) to see if the model flags systemic propagation (Silva, 2022).

Adversarial & manipulation risk

- Evaluate whether market-based inputs (VIX, spreads, asset prices) can be manipulated or noisy in ways that induce false signals; design robust input aggregates or smoothing to mitigate (Kou et al., 2019; Gensler & Bailey, 2020).

Operational & implementation checks

- Latency, numerical stability, and reproducibility tests; run smoke tests for data pipeline failures and create safe default outputs (e.g., "insufficient data") instead of silent erroneous probabilities (Chakraborty & Joseph, 2017).

7.4 Ethical risk taxonomy and mitigation measures

Below is a concise taxonomy of ethical/model risks with concrete mitigants aligned to the literature.

- **False negatives (missed crises)** — *Mitigant*: choose conservative thresholds using policy-loss weighting; prioritize recall in calibration and report uncertainty bands (Reimann, 2024).
- **False positives (unwarranted alarms)** — *Mitigant*: require corroborating signals and human approval before policy action; use smoothed or consensus signals across model families (O'Halloran & Nowaczyk, 2019).
- **Data-revision bias** — *Mitigant*: run vintage experiments and use nowcasting ensembles where necessary (Moro; Kothandapani, 2022).
- **Overfitting to past crises** — *Mitigant*: penalize complexity, prefer parsimonious ensembles, and validate via long out-of-sample windows (Reimann, 2024; Puli et al., 2024).
- **Regime dependence / non-stationarity** — *Mitigant*: include explicit regime indicators, retrain on expanding windows, and maintain model replacement criteria (Osterrieder et al., 2023).
- **Distributional and fairness harms** — *Mitigant*: simulate policy impacts across regions/income groups; publish distributional impact assessments where policies based on the model could differentially affect populations (Silva, 2022; Guerra & Castelli, 2021).
- **Opacity to stakeholders** — *Mitigant*: produce short non-technical model summaries, feature dashboards, and release sanitized replication packages where possible (Reimann, 2024; Araujo et al., 2024).

7.5 Operational monitoring & lifecycle management

Define an explicit lifecycle with versioning, monitoring, and retirement rules (Chakraborty & Joseph, 2017; Osterrieder et al., 2023):

1. **Model Registry & Versioning**

   - Store model code, artifacts, training data snapshot, hyperparameters, and validation reports in a registry. Every deployed version must have an immutable ID and link to documentation.

2. **Real-time monitoring (daily/weekly/monthly cadence depending on input frequency)**
   - **Data-quality alerts:** missing feeds, distributional drift (population stability index), extreme input deltas.
   - **Performance monitoring:** track rolling OOS metrics (probabilities, calibration, AUC/PR proxies) and compare to expected ranges.
   - **Explainability drift:** monitor changes in feature importance and raise flags if top features change abruptly (Reimann, 2024).
3. **Retraining & refresh policy**
   - Pre-defined retrain triggers: calendar retrain (e.g., annual), drift-triggered retrain, and extraordinary retrain after major structural events (financial crisis, regulatory changes). Keep a warm standby of previous stable models for rollback.
4. **Incident response & rollback procedures**
   - Pre-approved emergency rollback plan; require root cause analysis for anomalies and retain full logs for audit (Osterrieder et al., 2023).

7.6 Documentation, reproducibility, and transparency

- **Model card & datasheet.** Each model release includes a model card describing purpose, data sources, training period, performance on benchmarks, limitations, intended use, and contact for queries (Reimann, 2024; Araujo et al., 2024).
- **Reproducibility package.** Publish code to reproduce preprocessing, feature engineering, and model training where legally permissible; if data licenses or confidentiality prevent full release, provide synthetic datasets and detailed pipelines (Casabianca et al., 2019).
- **Audit logs.** Keep immutable logs of predictions issued to policymakers, the data snapshot used, and the version ID for future audit and accountability.

7.7 Legal, regulatory, and policy alignment

- Verify compliance with applicable data protection laws, supervisory regulations, and information-sharing agreements across agencies (Chakraborty & Joseph, 2017; Araujo et al., 2024).
- Work with institutional legal teams to define permissible disclosures of model outputs and to craft public communications when issuing systemic risk advisories to avoid market panic (O'Halloran & Nowaczyk, 2019).

7.8 Communication & stakeholder engagement

- **Two-track reporting:** (1) technical reports for model committees and validators with complete diagnostics; (2) concise policy briefs for decision-makers that translate model probabilities into actionable options and uncertainty ranges (Guerra & Castelli, 2021).
- **External transparency:** when feasible, publish non-sensitive model summaries and historical signals so academics and market participants can assess and build trust (Reimann, 2024; Araujo et al., 2024).

7.9 Research agenda & continuous improvement

Recommend ongoing research and validation tasks to reduce model risk over time (aligned with literature priorities):

- Real-time vintage evaluation to quantify the effect of data revisions (Moro; Kothandapani, 2022).
- Explainability research to improve actionable narratives from complex ensembles (Reimann, 2024; Chen et al., 2023).
- Stress testing of ML models under adversarial and contagion scenarios (Osterrieder et al., 2023; Silva, 2022).
- Cross-country transferability and domain-adaptation studies for sharing learnings across jurisdictions (Puli et al., 2024; Coutinho, 2024).
- Ethical impact assessments that model the distributional consequences of policy actions triggered by EWS outputs (Guerra & Castelli, 2021).

Strict technical legitimacy, governance through institutions, public control, and communication are also necessary to make sure that ML-based early-warning systems are capable of enhancing macroprudential decision-making without causing excessive new risks. The above recommendations encapsulate major practical and academic activities concerning the safe use of ML in the context of financial stability (Chakraborty & Joseph, 2017;

Reimann, 2024; Osterrieder et al., 2023; Araujo et al., 2024) and make them practical in the risk-management and ethics framework of US macroprudential application.

8. Conclusion

This paper has revealed that artificial intelligence (AI) and machine learning (ML) have great promise in forecasting systemic financial crises in the US macroprudential system. Using such multi-source datasets as macroeconomic indicators (FRED and others), the credit-to-GDP gap (BIS), crisis timelines (IMF and Laeven and Valencia), and market volatility (Yahoo Finance and Nasdaq Data Link), we were able to easily build early-warning models that showed a high predictability rate over various time horizons. The findings indicate that the gradient-boosted tree algorithms performed better compared to linear and baseline models, especially in terms of precision, recall metrics, and the interpretation methodologies, e.g., SHAP analysis, which proved the importance of credit-to-GDP gap, term spread inversions, and volatility indices in crisis forecasting.

These results support the previous research on how AI can dramatically change macroprudential supervision and monitoring in terms of financial stability (Kothandapani, 2022; Gensler & Bailey, 2020; O'Halloran & Nowaczyk, 2019). Whereas the foundational aspect of the early-warning system has long been the traditional econometric models, our results confirm the numerous studies showing the superiority of the non-linear ML models in the detection of the complex interdependencies in financial data (Reimann, 2024; Moro, n.d.; Puli et al., 2024). Additionally, the hybrid metrics used in this study, which consist of macroeconomic and market-based variables, comply with previous findings that describe the integration of indicators to strengthen models and implement early warning signs (Chen et al., 2023; Guerra & Castelli, 2021; Yin, 2024).

Policy-wise, the study has two implications. First, it empowers the empirical basis of integrating AI-driven models in central bank and regulatory early-warning systems through which the activation of countercyclical capital buffers, liquidity injections, and other macroprudential instruments can be implemented earlier (Araujo et al., 2024; Chakraborty & Joseph, 2017). Second, within the framework, a transparent decision-support system is provided, which also reduces the so-called black box risk related to advanced ML systems, which is widely discussed in central banking and supervisory areas of application (Osterrieder et al., 2023; Kou et al., 2019; Shchepeleva, 2024).

This study faces limitations, despite these developments, such as the lack of US systemic crises to train on, possible instability of the results during structural breaks, and difficulties in revising data with the use of real-time economic releases (Silva, 2022; Celik, 2024). Resolving these limitations will demand additional methodological variation, including incorporating a real-time recursive training framework (Moro, n.d.; Coutinho, 2024) as well as expanding predictive models to global spillover channels and network contagion aspects (Hamzat et al., 2025; Kang et al., 2024; Kopec, 2024).

Future studies should also examine a range of modalities, including structured macro-data and unstructured information such as news sentiment, budget disclosures, and government declarations, which have demonstrated effectiveness in enhancing the emergence of crisis predictions (Chen et al., 2023; Bartel et al., 2024). Secondly, shifting to stress-testing frameworks that evaluate the resilience of models to a changing market regime will increase the reliability of results and confidence among policymakers (Vangala et al., 2023; Eccles et al., 2021).

Summing up, the study confirms that macroprudential data science can significantly improve crisis prediction of systemic risks in the US context with the help of AI and ML. Striking the right balance between predictive strength, interpretability, and policy applicability and relevance, these models can become the building block of the emerging collection of tools available to financial stability authorities, enabling them to respond in a timely and effective manner ahead of the things that could go wrong and take measures to prevent the economic damage.

References

Kothandapani, H. P. (2022). Advanced Artificial Intelligence Models for Real-Time Monitoring and Prediction of Macroprudential Risks in the Housing Finance Sector: Addressing Interest Rate Shocks and Housing Price Volatility to Support Proactive Decision-Making by Federal Agencies. *Journal of Artificial Intelligence Research. https://thesciencebrigade. com/JAIR/article/view/507*.

Chen, M., DeHaven, M., Kitschelt, I., Lee, S. J., & Sicilian, M. J. (2023). Identifying financial crises using machine learning on textual data. *Journal of Risk and Financial Management*, *16*(3), 161.

Reimann, C. (2024). Predicting financial crises: an evaluation of machine learning algorithms and model explainability for early warning systems. *Review of Evolutionary Political Economy*, *5*(1), 51-83.

Fahmi, R., & Aswirna, P. THE USE OF ARTIFICIAL INTELLIGENCE IN MACROECONOMIC FORECASTING AND MITIGATING RISKS FOR CENTRAL BANK POLICYMAKING.

MORO, F. Predictive Modelling of Financial Crises: Machine Learning Algorithms in a Recursive Real-time Framework.

Gensler, G., & Bailey, L. (2020). Deep learning and financial stability. *Available at SSRN 3723132*.

Puli, S., Thota, N., & Subrahmanyam, A. C. V. (2024). Assessing machine learning techniques for predicting banking crises in India. *Journal of Risk and Financial Management*, *17*(4), 141.

Guerra, P., & Castelli, M. (2021). Machine learning applied to banking supervision a literature review. *Risks*, *9*(7), 136.

O'Halloran, S., & Nowaczyk, N. (2019). An artificial intelligence approach to regulating systemic risk. *Frontiers in Artificial Intelligence*, *2*, 7.

Celik, S. S. (2024). *Predicting External Macroeconomic Crises: Machine Learning Perspective* (Doctoral dissertation, Middle East Technical University (Turkey)).

Casabianca, E. J., Catalano, M., Forni, L., Giarda, E., & Passeri, S. (2019). An early warning system for banking crises: From regression-based analysis to machine learning techniques. *EconPapers. Orebro: Orebro University*.

Oladuji, T. J., Akintobi, A. O., Nwangele, C. R., & Ajuwon, A. (2023). A Model for Leveraging AI and Big Data to Predict and Mitigate Financial Risk in African Markets. *International Journal of Multidisciplinary Research and Studies, 3*(6), 1843-1859.

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*(4), 286-311.

Vangala, S. R., Polam, R. M., Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2023). A Review of Machine Learning Techniques for Financial Stress Testing: Emerging Trends, Tools, and Challenges. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *4*(1), 40-50.

Silva, M. A. D. (2022). *Contagion in economic networks: a data-driven machine learning approach* (Doctoral dissertation, Universidade de São Paulo).

Bartel, M., Hanke, M., & Petric, S. (2024). Crisis Identification and Prediction using Machine Learning: The Case of US Regional Banks. *Available at SSRN 4688565*.

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors.

Shchepeleva, M. (2024). *Essays on banking crises* (Doctoral dissertation, Université de Strasbourg).

Coutinho, R. S. (2024). Extração de conhecimento em bases de dados por meio da aplicação de Machine Learning: um modelo de Sistema de Alerta Precoce (EWS) para crises financeiras.

Alabi, K. O., Adedeji, A. A., Mahmuda, S., & Fowomo, S. (2024). Predictive analytics in HR: leveraging AI for data-driven decision making. *International Journal of Research in Engineering, Science and Management*, *7*(4), 137-143.

Hamzat, L., Adekoya, Y., & Ajao, A. INNOVATIONS IN EMERGING MARKET DEBT RISK MANAGEMENT: COMPLEMENTARY INSIGHTS FOR US FINANCIAL RISK MODELING. *International Journal of Engineering Technology Research & Management (IJETRM). 2025May19*, *8*(01), 167-86.

Araujo, D., Bruno, G., Marcucci, J., Schmidt, R., & Tissot, B. (2024). Data science in economy and finance: A central bank perspective. *Journal of Digital Banking, 9*(1), 30-48.

Ragnoli, D. Anticipation of default: implementation of a credit scoring model and integration with macroeconomic factors.

Osterrieder, J., Arakelian, V., Coita, I. F., Hadji-Misheva, B., Kabasinskas, A., Machado, M., & Mare, C. (2023). An Overview-stress test designs for the evaluation of AI and ML Models under shifting financial conditions to improve the robustness of models. *Available at SSRN 4634266*.

Chakraborty, C., & Joseph, A. (2017). Machine learning at central banks.

Collodel, U. (2022). *The Interplay between Macroeconomic Forecasting, Financial Crises and Complexity* (Doctoral dissertation, Université Panthéon-Sorbonne-Paris I).

Eccles, P., Grout, P., Siciliani, P., & Zalewska, A. A. (2021). The impact of machine learning and big data on credit markets.

Machireddy, J. R. (2023). Data science and business analytics approaches to financial wellbeing: Modeling consumer habits and identifying at-risk individuals in financial services. *Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems*, *7*(12), 1-18.

Kang, A., Xin, J., & Ma, X. (2024). Anomalous cross-border capital flow patterns and their implications for national economic security: An empirical analysis. *Journal of Advanced Computing Systems*, *4*(5), 42-54.

Kopec, A. (2024). INNOVATIVE APPROACHES TO CRISIS MANAGEMENT WITHIN THE CONTEXT OF BALANCED ECONOMIC SECURITY. *International Interdisciplinary Scientific Journal" Expert"*, *1*(5), 28-37.