An Explainable Artificial Intelligence Framework for Early Warning of Systemic Financial Risk

Samson Onaopemipo Amoran[1], Okolie Awele[2], Oluwatosin Lawal[3], Okon Godspower Emmanuel[4]

[1]Department of Computer Science, Western Illinois University, USA.

[2] School of Computing and Data Science, Wentworth Institute of Technology, Boston, USA.

[3] Department of Mathematics Statistical Analytics, Computing and Modeling, Texas A&MUniversity, Kingsville, USA.

[4] Department of Mathematics, University of Florida, USA

Abstract

The accumulation of macro-financial imbalances leads to systemic financial crises which create major risks for economic stability. Theoretical foundations of traditional early-warning systems fail to predict the complex relationships and sudden changes which occur in contemporary financial systems. Machine learning developments enable better prediction results, but their users face comprehension difficulties which prevent them from being used in regulatory settings and building public confidence. The research develops an early-warning system for financial system risk assessment through the combination of ensemble machine learning methods with XAI. The system uses macroeconomic data together with financial market information and banking sector performance to create future risk predictions and its design includes methods which allow users to understand results throughout the entire modeling process. The research uses explainability tools which include feature-attribution methods and model-agnostic interpretation techniques to showcase both important risk factors and specific times when risk levels increased. The proposed system aims to achieve both predictive accuracy and transparent performance. The system includes all necessary processes for data creation and model development together with explainability assessment methods which enable organizations to monitor financial system stability. The research creates a system for predicting risks which allows organizations to track systemic financial dangers.

Keywords

Systemic Financial Risk, Early-Warning Systems, Explainable Artificial Intelligence (XAI), Machine Learning in Finance, Macroprudential Policy, Financial Stability Monitoring.

Introduction

Systemic financial risk refers to the possibility that disruptions within segments of the financial system may impair financial intermediation and generate widespread macroeconomic instability. Episodes such as the global financial crisis highlighted how interconnected and fragile modern financial systems can become when vulnerabilities accumulate over time (Brunnermeier & Oehmke, 2013). Detecting such vulnerabilities at an early stage remains a central objective of financial stability authorities and macroprudential regulators (Bank for International Settlements [BIS], 2019).

The traditional systems for early warning detection use econometric methods to analyze macro-financial data which includes credit growth, leverage ratios and asset price movements. The models provide theoretical understanding but they fail to show complex financial systems because they cannot track nonlinear relationships or interaction effects or sudden system transitions (Hastie et al., 2009). The financial markets show stronger connections between various markets and they produce more data which results in linear and parametric models failing to deliver effective risk predictions. Machine learning methods work as an effective solution because they enable modeling of nonlinear relationships and processing of high-dimensional data and detection of complicated patterns which lead to financial instability. The ensemble methods which include random forests (Breiman, 2001) and gradient boosting techniques (Chen et al., 2020) deliver excellent prediction results across different fields. The models provide analytical flexibility yet they face criticism because their users find difficulty in understanding their workings. Decision-makers in regulatory environments need to present clear and justifiable model output explanations because their decisions will impact financial institutions and market operations.

The solution to the trade-off between predictive performance and interpretability relies on explainable artificial intelligence (XAI) as its answer. The Shapley Additive Explanations (SHAP) technique provides a theoretical framework which enables the breakdown of model predictions into their individual feature contributions according to Lundberg and Lee's 2017 research. Interpretable machine learning frameworks enable researchers to achieve transparent model outcomes while maintaining their existing model complexity according to Molnar 2022. Full interpretability becomes a crucial requirement for monitoring systemic risk because it enables policymakers to maintain accountability while building institutional trust. The research develops an organized framework which establishes an explainable machine learning system that serves as an early-warning mechanism to detect systemic financial risks. The paper presents a complete methodological framework which combines macro-financial data development with ensemble modeling techniques and built-in explainability features instead of showing actual validation results. The system generates future systemic risk indicators while maintaining complete visibility of risk factors which can be understood through economic assessment. The work presents three main contributions. The first contribution establishes a contemporary early-warning system which employs ensemble learning to observe non-linear macro-financial patterns. The second part of the system uses explainability techniques as fundamental elements which will construct its core framework. The system offers a detailed guide which enables organizations to establish financial stability monitoring systems that satisfy both predictive analytics and regulatory transparency needs.

2. Related Literature

2.1 Traditional Early-Warning Models for Systemic Financial Risk

Econometric frameworks which research macro-financial conditions before banking and currency crises make up the foundation for early-warning models that detect systemic financial crises. The approaches use discrete-choice models which include logit and probit specifications to calculate crisis probabilities from specific macroeconomic and financial indicators. The common predictors for this analysis include credit growth with credit-to-GDP gaps and leverage ratios and asset price inflation and external imbalances. The theoretical basis of these models together with their ability to generate results which can be easily understood make them attractive to users. The linear and parametric structures enable policymakers to measure how each variable directly impacts crisis probabilities which improves their ability to communicate with others and justify their regulatory decisions. The models maintain a strong connection to macroeconomic theory and prudential policy design especially for countercyclical capital regulation (Bank for International Settlements [BIS], 2019). The traditional econometric early-warning systems face operational shortcomings which researchers have documented throughout their studies. Financial systems demonstrate nonlinear behavior patterns together with feedback loops and regime-dependent dynamics which current linear models cannot properly represent (Hastie et al., 2009). The process of defining crises and selecting thresholds and choosing variables for the model results in performance changes which create doubts about the model's ability to produce consistent results with other situations. The increase in financial market connections together with the availability of extensive data makes it difficult for parametric models to manage high-dimensional data while avoiding the need for strict assumptions.

The global financial crisis further exposed weaknesses in pre-crisis monitoring frameworks, as vulnerabilities accumulated through complex interactions across credit markets, shadow banking systems, and asset price cycles (Brunnermeier & Oehmke, 2013). These structural complexities have motivated the search for more flexible modeling approaches capable of capturing nonlinear dependencies and interaction effects in systemic risk formation.

2.2 Machine Learning and Explainability in Systemic Risk Assessment

The research field has turned to machine learning approaches because traditional econometric methods face limitations in forecasting systemic risks. Machine learning algorithms provide parametric model alternatives because they can handle nonlinear relationships and multiple dimensional features without needing specific function definitions. Financial forecasting models achieve excellent predictive results through the use of ensemble techniques which include random forests and XGBoost gradient boosting model. The ability of these methods to operate flexibly allows them to identify intricate relationships which exist between macroeconomic and financial system variables. Machine learning models can address these issues by using more financial indicators which they can update based on changes in market conditions to create better warning systems than traditional methods. The use of machine learning for regulatory purposes faces major obstacles because organizations need to understand how systems function. Many high-performing models operate as “black boxes,” providing accurate predictions without transparent reasoning. The prediction methods used in financial stability assessments create problems for accountability because they obscure information and restrict policy communication which ultimately decreases institutional trust.

The development of explainable artificial intelligence (XAI) serves as an essential accomplishment which enables systems to achieve their predictive accuracy while remaining understandable to users. SHAP (Shapley Additive Explanations) provides a theoretically grounded method for decomposing model outputs into feature-level contributions based on cooperative game theory (Lundberg & Lee, 2017). Machine learning interpretable frameworks use their global explanations to show main risk factors which drive results and their local explanations to show specific prediction outcomes (Molnar, 2022). Explainability techniques need to become part of systemic risk monitoring systems because policy measures which include capital buffer adjustments and liquidity measures need economic explanations which policymakers can understand. Organizations can achieve both regulatory compliance and operational clarity through their early-warning systems because those systems now include interpretable features together with advanced analytical capabilities. The present study introduces a structured framework which combines ensemble machine learning techniques with built-in explanation features to create a system where interpretability functions as a core element of the system.

3. Data Design and Variable Construction

3.1 Conceptual Data Architecture

The design of an early-warning system for systemic financial risk requires the integration of macroeconomic financial market and banking-sector indicators which display both cyclical patterns and structural weaknesses. The research proved that credit expansion and leverage dynamics and asset price misalignments and liquidity conditions serve as critical predictors of systemic instability according to previous research about crisis prediction (Bank for International Settlements [BIS] 2019 Brunnermeier Oehmke 2013). The proposed framework uses a multi-domain data architecture to gather data about real economic activity and financial market activity and institutional balance sheet relationships. Financial risk builds up in the economy because macroeconomic variables define the fundamental economic environment. Output growth rates and inflation measures and external sector balances and unemployment indicators constitute the possible indicators of this economic environment. Financial market variables measure asset price cycles and volatility regimes and funding stress conditions while banking-sector indicators show credit growth and balance sheet leverage and liquidity reserves. The multidimensional structure of the system links system risk theory with its core components that system risk theory identifies as interconnectedness and cross-sector spillover effects which cause crisis progression (Brunnermeier & Oehmke 2013).

Data sources suitable for this framework include publicly accessible institutional repositories such as the Federal Reserve Economic Data (FRED) database, the World Bank Global Financial Development Database, and publications from the Bank for International Settlements (BIS, 2019; Federal Reserve Bank of St. Louis, 2020; World Bank, 2020). These sources ensure transparency, replicability, and cross-country comparability, which are essential for macroprudential surveillance applications.

3.2 Variable Selection and Feature Engineering

The process of variable selection exists because of two different sources which include theoretical frameworks and research evidence from systemic risk studies. According to the BIS report from 2019, credit-to-GDP gaps and credit growth rates and leverage measures and asset price growth indicators serve as reliable early warning indicators of financial instability. The identified variables demonstrate two different risk-taking patterns which involve excessive risk-taking behavior and maturity transformation dynamics that may develop before crisis episodes. The process of dynamic risk accumulation in markets needs more advanced approaches than raw macro-financial indicators can provide. The framework development uses feature engineering techniques to retrieve forward-looking signals from time-series data in order to solve this problem. The study uses rolling-window growth rates and long-term trend deviations and volatility measures and moving averages to study both cyclical patterns and regime changes. The process of transforming data follows traditional statistical learning methods which boost prediction accuracy by adding time-based information through feature space development (Hastie et al., 2009).

The high dimensionality characteristic of macro-financial datasets requires scientists to execute preprocessing procedures because it can lead to unstable model performance. The process of standardizing variables enables multicollinearity diagnostics to identify and eliminate duplicate information. The implementation of dimensionality-reduction techniques becomes necessary to combat overfitting problems which arise from using ensemble learning methods (Hastie et al., 2009). The data structure exhibits forward-looking organization which serves as its primary characteristic. The target variable is defined as a binary indicator which demonstrates whether systemic stress exists within a given forecast period. The structure of the framework guarantees that it meets the early-warning system requirements which demand predictive capability instead of real-time crisis detection.

4. Methodological Framework

4.1 Predictive Modeling Architecture

The early-warning system creates its main research methods through ensemble machine learning techniques which enable it to understand how macro-financial indicators interact with each other through nonlinear relationships. Traditional econometric approaches use parametric functional forms which limit their ability to model complex systemic dynamics. Ensemble learning methods create better prediction results through their ability to combine several weak learning models which produce more stable results. Random forests create multiple decision trees through a process that combines bootstrap aggregation with random feature selection, which helps to decrease variance while preventing model overfitting. The method proves highly effective for analyzing macro-financial data because it can handle multiple dimensions and expects to find both nonlinear relationships and interaction patterns (Breiman, 2001). The gradient boosting methods which include XGBoost as a scalable implementation work by sequentially combining weak learners to achieve lower prediction errors through gradient-based optimization methods (Chen et al., 2020). The methods enable researchers to understand complex variable relationships because they can identify various nonlinear connections and distinct effects of different variables.

The predictive task requires a binary classification framework because its goal is to determine the likelihood of systemic stress within a specified forecasting period. The framework maintains temporal integrity through its implementation of rolling-window and expanding-window estimation methods. Financial applications require time-aware modeling designs which protect against look-ahead bias while establishing authentic forecasting conditions. The study presents the modeling architecture without providing any empirical performance data. The research methodology focuses on how ensemble learning methods are integrated into a systemic risk monitoring system.

4.2 Explainability Integration Layer

Ensemble models provide multiple ways to make predictions yet their complicated nature decreases their ability to be understood. The implementation of new technologies faces major difficulties because organizations need to use existing methods to assess their performance. Government officials need to understand how macro-financial indicators affect their risk evaluation process. The framework uses explainable artificial intelligence (XAI) techniques as embedded components within its modeling system to solve this problem. SHAP (Shapley Additive Explanations) provides a theoretically grounded method for attributing individual feature contributions to model predictions based on cooperative game theory principles (Lundberg & Lee, 2017). SHAP provides two types of model explanation which include global model explanations that show main risk factors throughout the complete dataset and local model explanations which describe single high-risk cases. The use of partial dependence plots together with other interpretation tools enables better understanding of how nonlinear variables impact model predictions because they show how specific predictors affect risk probabilities (Molnar, 2022). The framework uses structural system design to create interpretability which enables organizations to meet regulatory transparency standards while maintaining their predictive analytics capabilities.

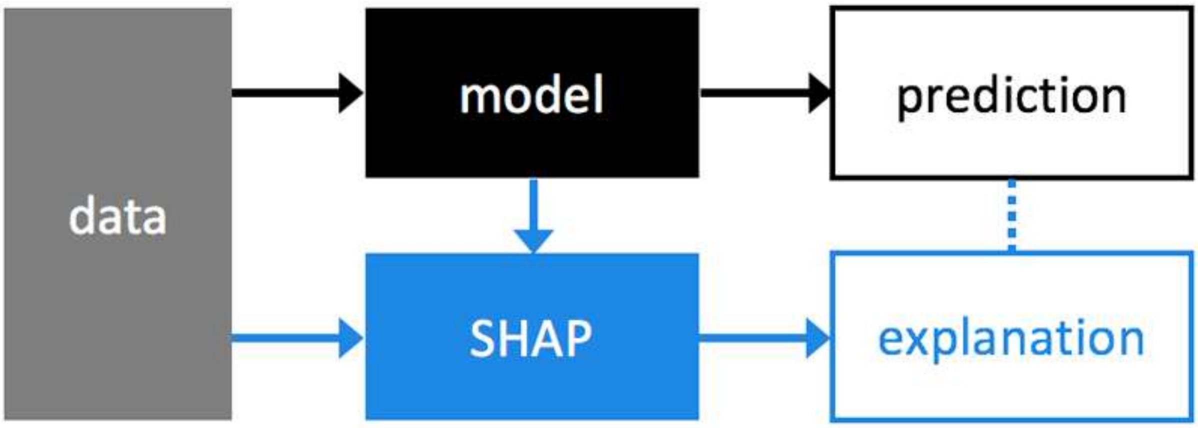

Fig 1: Explainability layer

4.3 Evaluation Strategy (Conceptual Design)

The framework design does not include empirical validation yet its methodological framework requires assessment through three main evaluation points. The first dimension of assessment measures classification performance through multiple metrics which include area under the receiver operating characteristic curve (AUC) and precision and recall and F1-score. The second dimension of assessment requires calibration analysis which tests whether predicted probabilities match actual systemic stress frequencies. The third dimension of assessment tests economic plausibility by checking whether risk drivers match existing financial stability theories. Out-of-sample validation and temporal robustness checks are essential to ensure generalizability and guard against overfitting, particularly in high-dimensional financial environments (Hastie et al., 2009).

5. Conceptual System Architecture

The proposed early-warning framework is structured as a multi-layer architecture integrating data inputs, predictive modeling, explainability mechanisms, and policy-oriented outputs. This layered structure reflects contemporary best practices in interpretable machine learning system design (Molnar, 2022; Lundberg & Lee, 2017).

5.1 Layer 1: Data Acquisition and Preprocessing

The first layer consists of structured macroeconomic, financial market, and banking-sector indicators which institutional databases such as FRED the BIS and the World Bank provide. (Bank for International Settlements [BIS] 2019; Federal Reserve Bank of St. Louis 2020; World Bank 2020) At this stage, preprocessing procedures include:

- The process of aligning all indicators with their appropriate time periods The process of standardizing all measurement systems
- The process of developing features which use rolling time windows The process of handling missing data points
- The process of determining multicollinearity for all variables
- The system delivers inputs to the model which have economic significance and stable statistical properties because it follows statistical learning standards from Hastie et al. (2009).

5.2 Layer 2: Predictive Modeling Engine

The second layer incorporates ensemble learning models which estimate systemic stress probability based a predefined forecast period. The system uses two major modeling families which include: Random Forests (Breiman, 2001) Gradient Boosting frameworks which include XGBoost (Chen et al., 2020) The selected algorithms provide the following capabilities for the system to achieve its goals:

- The system needs to learn how different factors create complex relationships between multiple variables.
- The system requires solutions which can manage multiple input variables that exist in the system.
- The system needs to develop methods which can properly handle unseen data without loss of performance.
- The rolling-window forecasting design uses modeling structure to maintain time-based accuracy because it stops financial risk forecasting from using future data.

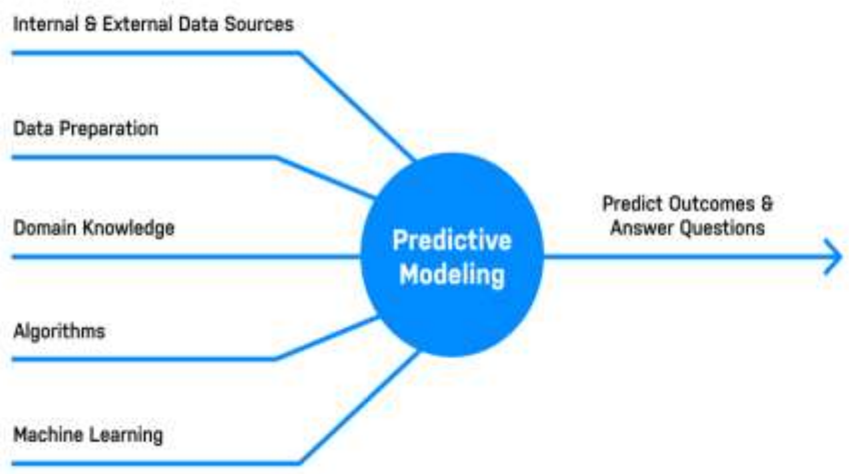

Fig 2: Predictive Modeling Diagram

5.3 Layer 3: Explainability and Interpretation Module

The third layer of the system connects explainable artificial intelligence methods to its prediction system. The framework includes explanation mechanisms as built-in elements which operate as primary diagnostic tools. Primary tools include:

- SHAP (Shapley Additive Explanations), which decomposes predicted probabilities into additive feature contributions (Lundberg & Lee, 2017) The method uses partial dependence analysis to show how nonlinear effects impact results (Molnar, 2022) The layer generates three outputs:
- The system establishes worldwide importance rankings for systemic risk factors
- The system provides specific explanations about times when the risk level reached dangerous heights
- The system displays output through visualizations which show how different inputs produce different results
- The system achieves its predictive analytics through interpretability, which helps maintain necessary transparency for regulatory compliance.

5.4 Layer 4: Policy-Oriented Output Interface

The final layer translates model outputs into actionable macroprudential signals. Rather than generating opaque probability scores, the system produces:

- Risk probability trajectories
- Identified dominant macro-financial drivers
- Sensitivity insights for capital and liquidity indicators

This transformation ensures that model outputs are economically interpretable and policy-relevant, addressing accountability concerns highlighted in financial governance literature (BIS, 2019).

6. Implementation Roadmap

The implementation of the proposed explainable early-warning system can be structured into sequential phases to ensure methodological rigor and regulatory alignment.

Phase 1: Data Infrastructure Development The first stage requires the establishment of a main central database which will store macro-financial information about different countries. The two institutional databases FRED and BIS offer dependable data which researchers can reproduce (Federal Reserve Bank of St. Louis, 2020; BIS, 2019). Data pipelines should start using automated systems which will update data at regular intervals to maintain system movements and future predictions.

Phase 2: Feature Engineering and Temporal Structuring The current phase produces rolling-window transformations together with volatility measures and trend deviation results. The transformations improve machine learning model performance which enables them to identify cyclical economic imbalances more effectively (Hastie et al., 2009). Researchers need to maintain temporal sequence information because any loss of data will cause future research problems.

Phase 3: Model Development and Validation Framework Ensemble learning models require training through a rolling window method which allows for model expansion during the learning process. The time-aware cross-validation methods should be used for hyperparameter tuning because they provide validation results which demonstrate system strength (Breiman, 2001; Chen et al., 2020). The authors of this paper present no empirical findings which means that research will examine three elements which include classification accuracy and calibration methods and economic feasibility tests.

The phase 4 process begins with the implementation of explainability systems after the model has completed its estimation process. The model utilizes SHAP-based decomposition together with partial dependence analysis to produce interpretability results according to the research of Lundberg and Lee and the work of Molnar. The system presents its results in a format that regulatory bodies can understand instead of providing technical documentation that requires specialized knowledge.

The final stage of the project requires building a monitoring interface that transforms model results into easy-to-understand visual dashboard displays. The dashboards display three main elements which include Predicted systemic stress probabilities Key contributing indicators Trend dynamics over time The current phase connects technical modeling work with the decision-making processes used in macroprudential policies.

7. Policy Relevance

The implementation of explainable machine learning for tracking systemic risks which affects macroprudential governance. The conventional frameworks for monitoring risks which exist today must select between two conflicting goals which are advanced analytical methods and clear understanding of results. The use of ensemble learning methods results in better prediction abilities which also causes problems because their operation remains hidden from view. The direct implementation of explainability inside the modeling process enables organizations to meet both technical analysis needs and regulatory compliance standards (Lundberg & Lee, 2017; Molnar, 2022). The proposed framework enables macroprudential policies through its support of multiple policy mechanisms: Countercyclical Capital Buffer Calibration The system detects times when credit growth reaches excessive levels and when leverage increases which allows it to recommend capital requirement changes that follow BIS macroprudential standards (BIS, 2019). Liquidity Risk Monitoring The system uses interpretability outputs to identify both volatility spikes and funding stress indicators which enables organizations to implement liquidity support measures before emergencies occur.

The framework enhances standard stress-testing procedures through its ability to detect nonlinear risk factors and their interactive relationships which traditional econometric models cannot identify (Hastie et al., 2009). Transparency and Accountability The institutional transparency of organizations receives its strongest boost through SHAP-based explanations. Policymakers can articulate not only the presence of elevated systemic risk but also the macro-financial mechanisms contributing to it. This strengthens communication with stakeholders and promotes trust in data-driven decision-making. The framework establishes a new regulatory structure which allows machine learning models to maintain their predictive accuracy while also providing understandable results.

8. Conclusion

This study presents a structured framework for an explainable machine learning–based early-warning system for systemic financial risk. By integrating ensemble learning techniques with embedded interpretability mechanisms, the proposed architecture addresses a critical tension in financial risk modeling: the trade-off between predictive sophistication and regulatory transparency.

Unlike traditional econometric approaches, the framework is designed to capture nonlinear interactions and high-dimensional macro-financial dependencies. At the same time, the integration of explainable artificial intelligence tools ensures that model outputs remain economically interpretable and suitable for policy communication. Rather than offering empirical validation, this paper contributes a scalable methodological blueprint that can guide the development of transparent, data-driven financial stability monitoring systems.

As financial systems continue to evolve in complexity and interconnectedness, the convergence of machine learning and explainability represents a promising direction for next-generation macroprudential surveillance. Future research may extend this framework through empirical implementation, cross-country applications, and integration with network-based systemic risk measures.

REFERENCES

Bank for International Settlements. (2019). *Global liquidity: Vulnerabilities and risks*. BIS Annual Economic Report. https://www.bis.org

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brunnermeier, M. K., & Oehmke, M. (2013). The maturity rat race. *Journal of Finance, 68*(2), 483–521. https://doi.org/10.1111/jofi.12005

Chen, T., Guestrin, C., & He, X. (2020). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/3394486.3406170

Federal Reserve Bank of St. Louis. (2020). *Federal Reserve Economic Data (FRED)*. https://fred.stlouisfed.org

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Huang, X., Zhou, H., & Zhu, H. (2019). A framework for systemic risk analysis using network measures. *Journal of Banking & Finance, 106*, 197–209. https://doi.org/10.1016/j.jbankfin.2019.07.004

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). https://papers.nips.cc

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Leanpub. https://christophm.github.io/interpretable-ml-book/

Okolie, A., Alumona, P., Lawal, O., & Olukoya, D. (2026). *A unified explainable machine learning framework for fraud detection across payment and cryptocurrency transaction systems*. *Journal of Computational Analysis and Applications*, 35(1)

Okolie, A., Oibiyeye, S., Ibiyeye, A. O., & Olukoya, D. (2026). *A risk-based security governance framework for modern information systems*. *Journal of Computational Analysis and Applications*, 35(1).

Lawal, O., Alumona, P., Okolie, A., Obunadike, C., Ikhifa, M. O., & Edozie, S. I. (2025). *An explainable XGBoost framework for detecting fraudulent financial transactions*. *Journal of Scientific Research and Reports*. https://doi.org/10.9734/jsrr/2025/v31i123769

Lawal, O., Okolie, A., & Obunadike, C. (2025). *An explainable graph neural network framework for anti-money laundering in cryptocurrency transactions using the elliptic dataset*. *International Journal of Network Security & Its Applications*, 17(6). https://doi.org/10.5121/ijnsa.2025.17602

World Bank. (2020). *Global Financial Development Database*. https://www.worldbank.org