Logistic Regression Versus Random Forest in Early Warning Systems for Financial Crises: A Systematic Review and Future Research Agenda for Sri Lanka

Abeyrathna, R.M.L.M.*[1] and Asanka, P.P.G.D.[2]

[1]*Department of Finance, University of Ruhuna, Sri Lanka*
[2]*Department of Industrial Management, University of Kelaniya, Sri Lanka*

Abstract

This study provides a systematic literature review according to the PRISMA 2020 statement, exploring the comparative predictive performance of the logistic regression and random forest models in early warning system studies, particularly in developing and frontier markets. A search of seven academic databases (ScienceDirect, JSTOR, Emerald Insight, Wiley Online Library, Springer Link, Oxford Academic Journals, and Taylor & Francis Online) yielded 1,834 hits. Duplicates and retracted articles were removed, leaving 1,637 unique records that were screened by title and abstract; of these, 533 were advanced to full-text screening, and 148 were ultimately included in the narrative synthesis. The analysis is based on a theoretical evolution of crisis models in three phases: fiscal unsustainability, self-fulfilling expectations, and reversal of capital flows. Simultaneously, it follows the development of approaches to Early Warning Systems (EWS) from non-parametric signal extraction to logistic regression to machine learning classification. The outcomes reveal that random forest models generally outperform logistic regression in the case of nonlinear, multidimensional, and interactive crisis dynamics, particularly when combined with SHAP-based interpretability and rolling-window out-of-sample validation, although this advantage depends on validation design and sample size. The assessment highlights four major shortcomings in the literature on Sri Lanka: the exclusion of fiscal variables from existing EWS models, the over-reliance on currency-based definitions of the crisis, the lack of comparison between logistic regression and random forest techniques, and the lack of true out-of-sample validation. The deficiencies identified encourage the development of a financially aware, multi-dimensional and comparatively evidence-based early warning system for Sri Lanka and set out future research directions, with implications for early warning systems in other emerging market economies in the South Asian region.

Keywords: Early warning systems; financial crisis prediction; logistic regression; random forest; systematic review; Sri Lanka

1. Introduction

One of the key challenges facing modern macroeconomics is the ability to predict financial crises before they occur. Economic downturns can severely harm national economies, such as

* Corresponding author: lakshmika@mgt.ruh.ac.lk

Proceedings of the 6th Annual International Conference on Emerging Financial Markets and Policy (EFMP 2026)
© 2026 Sri Lanka Finance Association
ISSN 2792-1220, eISSN 2806-514X

a decrease in fiscal capacity, instability in the banking system, and prolonged recessions. The 2022 Sri Lankan sovereign debt crisis was the worst economic crisis faced by the country after independence. The crisis exposed the inadequacies of the existing early warning systems to cater for the complex fiscal and external vulnerabilities of small open developing economies. The Asian financial crisis of 1997–98 greatly expanded the study of early warning systems (EWS) for financial crises, and the Global Financial Crisis of 2008 further accelerated it. The discipline has evolved from the non-parametric signal extraction methodology of Kaminsky and Reinhart (1999) to the logit and probit models of the 1990s and 2000s, and more recently to modern machine learning methods such as random forest, gradient boosting, and neural networks. But most of this research is focused on industrialized countries or large emerging economies, while developing countries such as Sri Lanka are still consistently overlooked.

A major methodological discussion in the EWS literature relates to the comparative predictive power of conventional econometric methods, particularly logistic regression, and machine learning methods. While machine learning methods can capture non-linear relationships, dependencies between variables, and large feature spaces, logistic regression has advantages in terms of interpretability, theoretical grounding, and robustness when faced with limited training data. A model that fits historical data well but performs poorly on out-of-sample data may provide policymakers with a false sense of security at critical junctures, but a precise but opaque model is not defensible in environments where accountability is paramount. The comparative advantages of these methodologies for crisis forecasting in frontier and developing economies are an open empirical question. This review attempts to achieve three objectives: first, to describe the theoretical and methodological progress of EWS research from its basic contributions to the present applications of machine learning; second, to summarize the existing evidence on the relative effectiveness of logistic regression and random forest in predicting financial crises; and third, to highlight the specific empirical and methodological gaps that warrant a fiscally informed multidimensional investigation of early warning systems in the Sri Lankan context. To ensure the transparency, reproducibility, and methodological soundness of the search strategy, screening, and synthesis, we followed the PRISMA 2020 guidelines to conduct the evaluation. Accordingly, the review addresses two research questions. RQ1: How do logistic regression and random forest models compare in predictive performance as early warning systems for financial crises, and under what conditions does either approach perform better? RQ2: What gaps does this evidence reveal for the design of an early warning system for Sri Lanka? The review contributes to the literature in three ways: it links the evolution of crisis theory to the evolution of EWS methodology; it consolidates the comparative evidence on logistic regression and random forest; and it identifies the specific gaps and future research directions for Sri Lanka. The remainder of the paper is organized as follows. Section 2 describes the review method, Section 3 presents the results, Section 4 discusses the current state of knowledge and future research directions, and Section 5 concludes.

2. Methods

This evaluation adhered to the PRISMA 2020 guidelines for systematic literature reviews. A review protocol was established beforehand, detailing the research inquiry, eligibility

requirements, search methodology, screening procedure, data extraction strategy, and synthesis approach. The primary study inquiry was: how do logistic regression and random forest models compare in their predictive efficacy as early warning systems for multi-faceted financial crises, and what insights can be derived from this literature for the setting of Sri Lanka? The inquiry was structured utilizing the PICO framework: Population (emerging and frontier market economies, such as Sri Lanka), Intervention (EWS models employing logistic regression, random forest, or analogous techniques), Comparison (logistic regression in contrast to random forest and alternative machine learning approaches), and Outcome (crisis prediction efficacy assessed by AUROC, accuracy, precision, recall, F1-score, or similar metrics).

A thorough investigation was performed across seven scholarly databases: ScienceDirect, JSTOR, Emerald Insight, Wiley Online Library, Springer Link, Oxford Academic Journals, and Taylor & Francis Online, using institutional subscriptions from the University of Ruhuna. The lack of access to Scopus and Web of Science represents a constraint, elaborated upon subsequently. For each database, a Boolean search framework utilized five conceptual keyword categories: crisis classification (e.g., "sovereign debt crisis", "currency crisis", "twin crisis"), EWS techniques (e.g., "early warning system", "leading indicator", "signal approach"), machine learning approaches (e.g., "random forest", "logistic regression", "gradient boosting", "neural network"), geographical and economic parameters (e.g., "Sri Lanka", "emerging market", "frontier market", "South Asia"), and fiscal and debt-related terminology (e.g., "fiscal deficit", "primary balance", "sovereign bond", "reserve adequacy"). An intermediate search (L2) integrated the crisis, methodology, and machine-learning components; a Sri Lanka/South Asia-specific search (L4) amalgamated the geography component with the crisis and methodology components. In instances where whole Boolean sequences surpassed database character restrictions (particularly JSTOR and Emerald Insight), abbreviated forms that maintained the core conceptual essence were used, with the divergence recorded in accordance with PRISMA 2020 reporting standards. All searches were completed within a four-day window in May 2026, and the full search strings for each database are available from the authors on request.

Research was considered if it focused on early warning systems or crisis forecasting, utilized at least one quantitative predictive model, provided a performance metric, incorporated actual macroeconomic or financial data from 1990 onwards, and consisted of peer-reviewed articles, working papers, or doctoral dissertations in English. Research was omitted if it solely focused on micro-level corporate credit risk without considering a macroeconomic crisis aspect (EC1), was entirely qualitative or descriptive (EC2), failed to assess predictive accuracy relative to a specified crisis outcome (EC3), was a duplicate, or could not be fully accessed despite utilizing institutional resources, open-access databases, and direct communication with authors.

The investigation carried out in May 2026 uncovered 1,834 entries: 1,260 from the primary EWS/machine-learning inquiry (L2) and 527 from the Sri Lanka/South Asia-focused search (L4), with the remaining 47 from supplementary searches. Records were transferred into Zotero (version 7.0), resulting in the elimination of 168 duplicates and 29 retracted articles, yielding 1,637 distinct records for title and abstract evaluation. Every record was categorized as Include,

Exclude, or Maybe based on the eligibility standards; 1,097 records were omitted at this juncture, primarily for focusing on corporate credit risk or consumer lending instead of macroeconomic crisis forecasting. The remaining 533 records (245 Include, 288 Maybe) were sought for full-text retrieval, of which 191 could not be retrieved. Of the 342 full-text reports assessed for eligibility, 194 were excluded (EC1 = 87; EC2 = 41; EC3 = 38; EC4 = 28), leaving 148 studies for the narrative synthesis (Figure 1).

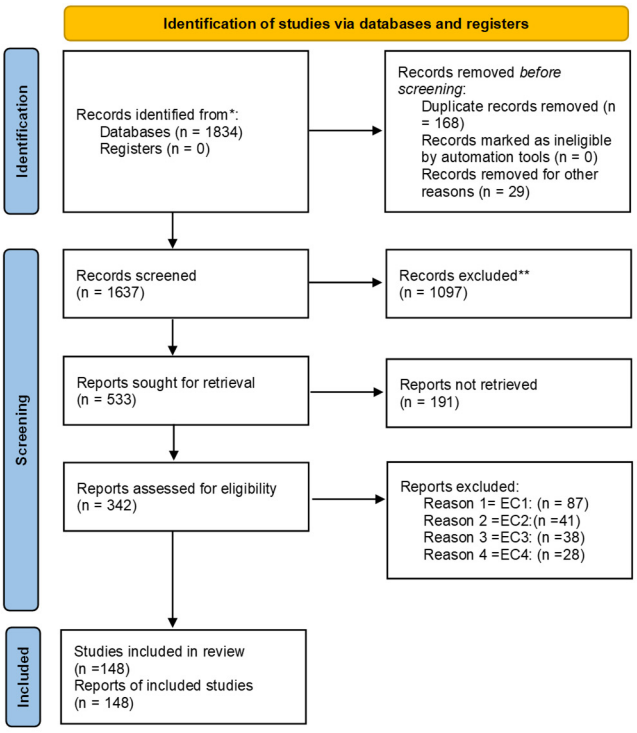


Figure 1: PRISMA 2020 Flow Diagram

The papers included were gathered based on a standardized template comprising study identification, research design, data attributes, model specifications, estimating methods, performance indicators, principal results, significance to Sri Lanka and South Asia, and limitations acknowledged. The quality assessment was performed using a modified ROBINS-I instrument, supplemented with a six-item checklist specifically designed for machine learning research (holdout validation, data leakage risk, overfitting evaluation, class imbalance management, hyperparameter optimization, and reproducibility). A narrative synthesis was used as the main method, clustering the studies into six thematic groups: development of EWS methodology; applications of logistic regression; use of random forest and machine learning; evidence of comparative performance; applications in emerging and frontier markets; and literature specific to Sri Lanka/South Asia, in view of the anticipated heterogeneity of the studies in terms of type of crisis, geographical setting, time period, and methodological approach. For three or more studies that reported the same performance metric for the same

model type and crisis category, statistical heterogeneity was further assessed using the I 2 statistics. Overall certainty of evidence was assessed using the GRADE methodology.

The review has several limitations. The main limitation to comprehensiveness is the inability to access Scopus and Web of Science due to institutional restrictions. This will be overcome once registration for Research4Life is confirmed. The review database did not include the SSRN, a major repository of EWS working papers, thus under-representing the current grey literature, especially on Sri Lanka, which was more recent than 2022. The search was limited to publications in English, which may have excluded research in Sinhala or Tamil relevant to Sri Lanka. The four-day search window was adequate for the current stage of investigation but shorter than the iterative search approach often recommended for systematic reviews for publication.

3. Results and Findings

The empirical EWS literature is based on three generations of crisis theory, each with a different set of variables, a different dependent variable, and a different criterion for model assessment. The initial models, due to Krugman (1979) and later refined by Flood and Garber (1984), link persistent fiscal deficits, financed by a drawdown of reserves, with inevitable currency and balance-of-payments crises. A government that maintains a fixed exchange rate and finances deficits through domestic credit expansion will eventually be drained of its reserves, leading to a speculative attack that forces devaluation. This mechanism is captured in Blanchard et al.'s (1991) debt sustainability model through the intertemporal government budget constraint: a government that runs a primary balance insufficient to cover the interest-rate-growth differential times its debt ratio is on an unsustainable path. When applied to Sri Lanka, this framework indicates that debt sustainability was breached as early as 2019 and severely breached in 2020, three years before the formal 2022 default (Samarakoon, 2024), evidence that the gap between the actual and debt-stabilizing primary balance is a directly observable EWS indicator of first-generation crisis risk.

Second-generation models (Obstfeld, 1986; Obstfeld & Rogoff, 1995) view crises as self-fulfilling prophecies and not exclusively as a result of fundamentals: market expectations of devaluation can cause capital flight and higher yields, thus increasing the probability of devaluation even in the absence of weak domestic fundamentals, allowing for multiple possible equilibria, including a high-credibility equilibrium where no attack takes place and a low-credibility equilibrium where an attack occurs. This implies that global risk sentiment measures like the VIX should be added to the variable sets of early warning systems, since a small open economy dependent on foreign capital could be pushed into a low-credibility equilibrium by a shift in foreign risk appetite. Consistent with this, the VIX shows a noise-to-signal ratio of 0.28 for Sri Lanka, which is one of the more informative (lower) values reported in the existing literature (Prabheesh & Jayawickrema, 2024).

Third-generation models that emerged after the 1997–98 Asian crisis (Chang & Velasco, 1999; Krugman, 1999) focus on capital flow reversals, currency mismatches, and balance-sheet weaknesses, defining international illiquidity as a condition where a country’s consolidated financial system has foreign currency short-term liabilities that are larger than its available

foreign currency assets, a condition relevant for reserve-adequacy metrics in the design of early warning systems. These three do not compete but complement each other, each explaining a different causal path that can operate independently or simultaneously, as evidenced by Sri Lanka's 2022 crisis. This calls for a multi-theoretical, multidimensional approach to EWS design rather than reliance on a singular theoretical framework.

The methodological literature has passed through three successive generations, each one correcting the shortcomings of its predecessor. The first generation, non-parametric signal extraction (Kaminsky & Reinhart, 1999), looks at certain indicators against historical norms using the noise-to-signal ratio (NSR). It is transparent, distribution-free, and has become a standard policy tool for institutions such as the IMF and the Bank for International Settlements. However, it cannot combine indicators into a single likelihood of crisis. The second generation, utilizing parametric binary response models, predominantly logistic regressions (Berg & Pattillo, 2000; Bussière & Fratzscher, 2002; Frankel & Rose, 1996), addressed this limitation by yielding a single composite probability estimate along with interpretable marginal effects capturing the impact of each variable on the crisis risk. This became the standard methodology for early warning systems throughout the 2000s and 2010s. The third generation, machine learning classification techniques such as random forest, gradient boosting and neural networks (Alessi & Detken, 2018; Holopainen & Sarlin, 2017), arose from the recognition that financial crises are inherently nonlinear, high-dimensional classification problems with the simultaneous violation of multiple thresholds across fiscal, external, monetary and confidence dimensions, which parametric models are unable to adequately capture without pre-defining each interaction term. Machine learning methods are effective at handling nonlinearity, variable interactions, and class imbalance, particularly when interpretability tools such as SHAP (Lundberg & Lee, 2017) are used to make black-box predictions useful for policy discussion and accountability. The methods from each generation, NSR, logistic regression, and random forest, have been used in the Sri Lankan context independently, but no studies using Generation Three methodologies such as SHAP interpretability, rolling-window validation, and a multidimensional crisis definition were found in this review.

For the last thirty years, logistic regression has been the dominant methodology for early warning systems, from the seminal work of Frankel and Rose (1996) to the latest Sri Lanka-focused analysis of Prabheesh and Jayawickrema (2024), and it remains the key benchmark for assessing the effectiveness of any machine learning approach. Frankel and Rose (1996) used panel probit models on 105 developing countries in the period 1971–1992 and found that deviations from trend ("gap" variables) of macroeconomic variables are better predictors of crises than their levels and identified weak reserves, falling capital inflows, rapid credit growth, overvalued exchange rates, and rising interest rates as reliable indicators, a finding that is confirmed nearly thirty years later in Sri Lanka where the credit gap has an NSR of 0.31 compared to 0.67 for the credit level (Prabheesh & Jayawickrema, 2024). This approach is generalized by Bussière and Fratzscher (2006) to a multinomial logit model distinguishing pre-crisis, crisis, and post-crisis recovery periods in 32 emerging markets over 1993-2001, correcting for the post-crisis bias that arises from treating recovery phases as stable in the training data, and achieving 73.7 percent out-of-sample accuracy with a 44.1 percent false-alarm rate. Berg and Pattillo (2000) report that logistic regressions systematically beat signal

extraction out-of-sample if they include indicators based on their joint predictive power, but also note that the performance drops dramatically relative to in-sample fitting, which provides a strong motivation for rolling-window validation versus a single fixed train-test split.

The extensions to banking and systemic crisis scenarios are plagued by the additional problem of class imbalance, as banking crises are less frequent and more temporally clustered than currency crises. Baltaci et al. (2022) apply self-organizing maps for the initial selection of variables and then panel logistic regression in 38 countries over the 2007–2009 Global Financial Crisis, obtaining classification accuracy of over 85 percent for stable periods and over 90 percent for crisis periods at a 0.2 threshold, being one of the few studies to find a government budget balance variable statistically significant along with domestic credit and exchange rate measures, though validated only in-sample and without rolling-window analysis.

The literature is extensive, but surprisingly little use of fiscal variables in logistic regression early warning system studies has been found. Comelli (2014) examined 29 emerging markets from 1995 to 2012 and found fiscal balance to be statistically insignificant when macroeconomic predictors were correlated, a result attributed to the linear log-odds assumption not capturing the potential nonlinear interaction between fiscal variables and debt and reserve variables. Importantly, the only prior formal EWS analysis for Sri Lanka by Prabheesh and Jayawickrema (2024) does not include fiscal balance, primary balance, public debt, government revenue, and debt composition in any of their four model specifications, despite explicitly stating that Krugman's (1979) fiscal-deficit-driven crisis mechanism is their theoretical framework. This omission is in contrast to their own theoretical framework and the empirical findings that Samarakoon (2024) later presents. However, the literature shows that early warning system models based on logistic regression have in-sample classification accuracies between 70 and 85 percent and AUC-ROC between 0.70 and 0.85, but also consistently exhibit three inherent structural constraints. First, the linear log-odds assumption that the effect of each variable on the likelihood of crisis is constant and monotonic ignores threshold and interaction effects – a key shortcoming given that Sri Lanka's crisis was not a matter of additive factors, but of simultaneous fiscal decline, reserve depletion, loss of market access, and the effects of a COVID-19 shock. Second, the presence of multicollinearity among the interdependent macroeconomic indicators (GDP growth, credit growth, fiscal balance, current account) results in inflated maximum likelihood estimates in terms of standard errors and changes in the sign of the coefficients across the specifications. Prabheesh and Jayawickrema (2024) address this problem only by separating the correlated variables into different specifications instead of addressing the basic instability. Third, class imbalance is consistently mismanaged, as maximum likelihood estimation treats crisis and non-crisis observations equally and therefore underestimates the likelihood of the infrequent crisis category. No studies on logistic regression EWS used SMOTE, asymmetric class weighting, or AUC-PR as their primary evaluation metric, indicating a void that clearly supports the inclusion of AUC-PR with SMOTE-augmented training in future comparative Sri Lankan EWS schemes.

Since the mid-2000s, the use of machine learning for crisis prediction has grown substantially, as researchers have come to realize that the dynamics of crises are essentially

non-linear and high dimensional. Alessi and Detken (2017) apply Random Forest and related ensemble methods to detect excessive credit expansion and asset price fluctuations in the eurozone, and specify the Area Under the Precision-Recall Curve (AUC-PR) as the appropriate metric for uneven crisis categorization, and advocate a "relative usefulness" framework which evaluates models relative to a no-model policy baseline, a framework which justifies the continued publication of NSR alongside AUC-PR as a relevant secondary metric for policy. The most common benchmark is the work of Beutel et al. (2019), who use a banking-crisis panel of advanced economies over 1970–2016 and find that, although machine learning methods including random forest achieve a very high in-sample fit, they are outperformed by logistic regression in recursive out-of-sample evaluations, a result that is robust across performance metrics, crisis definitions and variable sets. Holopainen and Sarlin (2017) extend the analysis to neural networks, support vector machines, and other ensemble methods for EU-member countries to confirm the overall dominance of ensemble methods, especially at longer prediction horizons and depending on the class-imbalance management approach used.

SHAP values (Lundberg & Lee, 2017) have emerged as the de facto standard for interpretability in random forest EWS models, providing a game-theoretic measure of the importance of each feature on the prediction that can be directly compared to the marginal effects in logistic regression. Wang and Zhou (2024) use XGBoost, random forest, support vector machines, and logistic regression, in combination with SHAP and SMOTE, for systemic crisis prediction in China, achieving an accuracy above 95 percent and finding domestic monetary aggregates and real estate dynamics as the principal, highly nonlinear risk factors; Osman and Muse (2024) illustrate the applicability of SHAP-based interpretability in the data-poor, fragile-state context of Somalia, where a fiscal variable (net lending/borrowing) emerges as a significant predictor. Barthélémy et al. (2024) find that RNNs are more sensitive than RFs at the short-term horizon, but that RFs are competitive at the 4-8 quarters horizons that are relevant for macroprudential policy, and with much lower data and interpretability costs.

There are still some methodological issues such as class imbalance and validation. In a firm-level financial distress setting, Kristanti et al. (2025) argue that SMOTE should be performed only on training folds (rather than before the train-test split), to avoid data leakage and improve minority class recognition. Rademacher (2025) shows that a random forest with structured feature selection outperforms full-feature logistic regression in predicting German recessions, using rolling windows and structural breaks, and supports a hierarchical, parsimonious variable framework in frontier-market settings with small sample sizes.

Direct comparative analyses are built on four common pillars: the Kaminsky–Reinhart signal extraction paradigm, the logistic regression standard of Frankel-Rose (1996) and Bussière-Fratzscher (2006), the Reinhart-Rogoff (2009) and Eichengreen (1995) historical crisis classification and Exchange Market Pressure index, and the machine learning framework of Breiman (2001) and Friedman (2001). Jarmulska (2020) extends this lineage by comparing random forest and logistic regression for predicting multi-indicator fiscal stress using rolling-window validation, finding that random forest achieves 77.5 percent accuracy, 75 percent sensitivity, and 78 percent specificity, outperforming logistic regression across all metrics, with partial dependence plots revealing a nonlinear debt-to-GDP threshold near 80 percent that the

linear model of logistic regression is not able to capture, consistent with Reinhart and Rogoff's (2009) claim that debt accumulation is the main determinant of pre-crisis vulnerability. Abdelsalam and Abdel-Latif (2020) show that the combination of several parametric models improves the out-of-sample performance compared to any individual specification. This gives a theoretical justification for the expected variance reduction achieved by combining many decision trees in a random forest. Using historical systemic crisis data for 13 African countries from 1860 to 2014, Chen (2024) evaluates eleven algorithms and finds that nonlinear classifiers consistently outperform logistic regression, with the best-performing model (a Support Vector Classifier) achieving 96 percent accuracy and an F1-score of 0.78, empirically validating that currency and inflation crises often lead to fiscal and sovereign debt failures, in line with a multidimensional crisis framework.

Hopp (2022) and Kotan et al. (2025) extend their analysis to deep learning and find that LSTM architectures outperform Random Forest for very short time horizons (one day to several weeks), while Random Forest remains competitive for longer time horizons that are relevant for quarterly macroprudential policy. The directly comparative studies suggest that logistic regression and random forest perform comparably when crisis dynamics are approximately linear. When there are substantial thresholds, multivariable interactions, and class imbalance, as in the case of Sri Lanka's 2022 crisis, random forest generally shows a performance advantage. However, Beutel et al. (2019) show that this advantage can disappear under strictly recursive out-of-sample evaluation, so validation design and sample size are as important as the choice of algorithm.

Research on early warning systems for emerging and frontier markets provides several insights relevant to Sri Lanka. Fiscal and sovereign debt metrics are sometimes the first signs of turmoil in low-income and emerging nations: De Marchi and Moro (2024) identify public external debt above 20 percent of GDP as the best fiscal crisis indicator in a sample of emerging market and low-income countries and find that random forest and AdaBoost significantly outperform probit models. Similarly, Papadopoulos et al. (2017) find that debt-to-GDP ratios and budget balances are important predictors of pre-crisis conditions 7 to 12 quarters ahead in 27 EU countries. Joy et al. (2017) also confirm that government budget balance is a valuable predictor in 36 advanced and emerging economies but find that currency and banking crises have distinct predictive features. Composite crisis indices that incorporate multiple dimensions capture a broader range of distress than unidimensional currency-based measures such as the Exchange Market Pressure (EMP) index, although the evidence on their signaling performance is mixed. Coulibaly and Marouane (2024) find that for a sample of 21 developing nations, a composite multi-variable index has an NSR of 0.505, compared to 0.455 using inflation alone, indicating that a poorly weighted composite index can be less informative than a single indicator. Rose and Spiegel (2010) use a Multiple Indicators Multiple Causes model for 85 countries during the 2008 crisis to demonstrate that a continuous multidimensional measure of severity captures a greater share of the crisis than any single binary indicator, and that domestic fundamentals are more important than bilateral contagion for frontier markets in particular.

Specific methodological adaptations are required for frontier market early warning systems design. Data availability is limited, and annual financial data thus usually require some

interpolation to get a quarterly frequency. The one-sided Hodrick–Prescott filter avoids the look-ahead bias of the standard two-sided filter, which is especially problematic for economies with large structural breaks, such as the civil war in Sri Lanka, the subsequent boom, and the crisis in 2022. Guzmán and Fan (2026) and Cheptiș et al. (2025) both argue that the timing of the filter matters when structural regimes differ across crises. Asymmetric class weighting is necessary given the rarity of crisis events and the fact that sometimes it is only 8-15 percent of entries in quarterly Early Warning System datasets. Joy et al. (2017) weight crisis misclassification seven times more than calm misclassification. In sum, given the heightened susceptibility of emerging markets to regime shifts, there is a need for explicit controls for structural breaks, political transitions, participation in and exit from IMF programs, conflict, and capital account liberalization. Guo and Su (2023) find that the effect of capital liberalization on systemic risk is regime-dependent, acting as a risk amplifier at both low and high levels but as a risk mitigator at moderate levels, implying the need for an indicator of financial liberalization for Sri Lanka's shift to a more open capital account after 1995. To date, no research is available collectively providing a machine learning comparison for a single frontier-market economy in South Asia; this is a notable gap in both geographic and methodological terms.

Only eight studies directly address the analysis of financial crisis dynamics or early warning systems of Sri Lanka and South Asia, two of which are foundational references in the domain. Samarakoon (2024) provides a forensic analysis of the crisis of 2022. He notes budget shortfalls for 71 of the 73 years since 1950, averaging 7.2 percent of GDP, a primary balance deficit that peaked at 10.93 percent of GDP in 2020, an increase in the foreign-debt share of total public debt to 49.5 percent by 2018, and a decline in reserves from 5–6 months of import cover to just 1.8 months by the end of 2021, well below internationally accepted adequacy standards. Samarakoon is significant as he shows that these five components of crisis – fiscal, debt, currency, inflation and balance of payments – were statistically present in 2019, three years before the default, though the research does not provide its own forecasting model. The only formal early warning system for Sri Lanka is that of Prabheesh and Jayawickrema (2024), who use signal extraction and logistic regression on quarterly data from 1997Q1 to 2022Q4 for five currency crisis episodes with a pseudo-R2 of 0.427 in their optimal specification. They focus only on currency crises via the EMP index and exclude all fiscal variables in each of their four model specifications. They only report in-sample results.

The broader policy record for Sri Lanka supports this concern. The IMF's report on the Article IV Consultation for Sri Lanka (2022) indicates that Sri Lankan officials pushed back against IMF warnings of a looming debt crisis until December 2021, only four months before the default, exposing a failure of institutional policy oversight, which highlights the importance of a systematic, data-driven Early Warning System. Samarakoon's (2024) comparative study compares Sri Lanka's fiscal condition in 2019 (a budget deficit over 9 percent of GDP and a debt ratio of over 80 percent) with Venezuela, Lebanon and Suriname, all of which defaulted or had major financial turmoil within two years, compared with an emerging market average of a 3.9 percent deficit and a 55 percent debt ratio. In the wider South Asian region, the EWS literature is not well represented, with current research focusing on India's much larger and structurally different economy.

The Sri Lankan literature, taken together, shows four shortcomings. Firstly, while Samarakoon (2024) directly attributes the root cause of the crisis to fiscal irresponsibility, and Prabheesh and Jayawickrema (2024) base their theoretical underpinnings on Krugman's fiscal-deficit model, there is no existing current Early Warning System (EWS) for Sri Lanka that incorporates fiscal indicators. Secondly, all previous EWS research is solely dependent on the EMP currency-crisis indicator, which fails to capture the fiscal and debt weaknesses that were statistically evident years before the currency failure. Thirdly, there is a lack of research comparing traditional econometric methods with machine learning in the case of Sri Lanka or any other similar frontier market. This is despite growing international evidence, as discussed above, showing the advantages of random forest in nonlinear, multi-threshold crisis situations, and despite the fact that Sri Lanka's limited data set of about 142 quarterly observations poses its own unresolved empirical question about the relevance of these benefits in the context of a single-country, small-sample situation. Fourth, there is no genuine out-of-sample validation from previous Sri Lankan EWS studies; Prabheesh and Jayawickrema (2024) only present in-sample probability trajectories, which do not indicate if the model would have issued timely, real-time alerts rather than just replicating historical patterns.

4. State of Knowledge and Future Research Directions

The results indicate that the existing level of knowledge on early warning systems can be summarised at three levels. First, several findings are well established: financial crises are multidimensional, combining fiscal, external-liquidity and confidence channels; deviations from trend predict crises better than levels; in-sample fit overstates real-time performance; and class imbalance and interpretability must be addressed explicitly. Second, the relative performance of logistic regression and random forest remains contested: random forest performs better in multi-country samples with nonlinear thresholds and fiscal outcomes (De Marchi & Moro, 2024; Jarmulska, 2020), whereas logistic regression holds its ground under strictly recursive out-of-sample evaluation (Beutel et al., 2019). Third, little is known about whether these findings hold for a single frontier economy with a short data history and few crisis episodes, such as Sri Lanka.

These gaps point to several directions for future research. In future early warning system efforts in Sri Lanka and other emerging markets, scholars should explicitly include fiscal equilibrium, public indebtedness, and debt structure metrics in addition to traditional currency and credit indicators, and they should demonstrate genuine out-of-sample performance – preferably with rolling or expanding window cross-validation – rather than just in-sample fitting. Future studies should also construct a multidimensional crisis severity index rather than relying solely on the EMP index, and should compare logistic regression and random forest directly on the same feature set, using SHAP values alongside logistic marginal effects to compare their interpretations. The finding that the 2022 crisis in Sri Lanka was statistically observable as early as 2019 underscores the importance for policymakers to establish multidimensional, machine learning-enhanced monitoring systems within the scope of central bank and treasury oversight, rather than relying solely on currency-market pressures or irregular policy evaluations, which the IMF Article IV documentation suggests can be ignored by domestic authorities in the face of clear warning signs. The methodological modifications

identified in this review, including one-sided filtering, asymmetric class weighting, structural-break controls and tiered, parsimonious feature selection, provide a replicable framework for other emerging and developing economies in South Asia and beyond that currently do not have a comparative machine learning early warning system.

5. Conclusion and Implications

This study offers a systematic review of 148 studies selected from 1,637 screened records from seven databases, offering a snapshot of the theoretical and methodological evolution of early warning system research, from early fiscal models to machine learning classification. In answer to RQ1, it synthesizes international evidence that random forest generally outperforms logistic regression in the context of nonlinear, multidimensional, and interaction-intensive crisis dynamics, provided that SHAP-based interpretability and rolling-window validation are used to retain policy relevance and mitigate overfitting, although this advantage is conditional on validation design and sample size. In answer to RQ2, despite the global evidence, the Sri Lanka-specific literature is limited to two foundational studies: a comprehensive causal narrative without a predictive framework (Samarakoon, 2024) and a predictive model without causally significant fiscal variables (Prabheesh & Jayawickrema, 2024). The four highlighted shortcomings are: the lack of fiscal variables, the reliance on a unidimensional currency crisis metric, the absence of a logistic regression versus random forest analysis, and the lack of out-of-sample validation. These collectively present a distinct and pertinent research opportunity: a fiscally informed, multidimensional Crisis Severity Index assessed via a rolling-window comparison of logistic regression and random forest utilizing SHAP interpretability.

References

Abdelsalam, M. A. M., & Abdel-Latif, H. (2020). An optimal early warning system for currency crises under model uncertainty. *Central Bank Review, 20*(3), 99–107. https://doi.org/10.1016/j.cbrev.2020.03.002

Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability, 35*, 215–225. https://doi.org/10.1016/j.jfs.2017.06.005

Baltaci, A., Cergibozan, R., & Ari, A. (2022). Cultural values and the global financial crisis: A missing link? *Eurasian Economic Review, 12*(3), 507–529. https://doi.org/10.1007/s40822-022-00208-6

Barthélémy, S., Gautier, V., & Rondeau, F. (2024). Early warning system for currency crises using long short-term memory and gated recurrent unit neural networks. *Journal of Forecasting, 43*(5), 1235–1262. https://doi.org/10.1002/for.3069

Berg, A., & Pattillo, C. (2000). *The challenge of predicting economic crises* (Economic Issues No. 22). International Monetary Fund. https://doi.org/10.5089/9781557758842.051

Beutel, J., List, S., & von Schweinitz, G. (2019). Does machine learning help us predict banking crises? *Journal of Financial Stability, 45*, Article 100693. https://doi.org/10.1016/j.jfs.2019.100693

Blanchard, O. J., Chouraqui, J.-C., Hagemann, R., & Sartor, N. (1991). *The sustainability of fiscal policy: New answers to an old question* (SSRN Scholarly Paper No. 227461). Social Science Research Network. https://papers.ssrn.com/abstract=227461

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

Bussière, M., & Fratzscher, M. (2002). *Towards a new early warning system of financial crises* (SSRN Scholarly Paper No. 357482). Social Science Research Network. https://doi.org/10.2139/ssrn.357482

Bussière, M., & Fratzscher, M. (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance, 25*(6), 953–973. https://doi.org/10.1016/j.jimonfin.2006.07.007

Chang, R., & Velasco, A. (1999). *Liquidity crises in emerging markets: Theory and policy* (SSRN Scholarly Paper No. 199298). Social Science Research Network. https://papers.ssrn.com/abstract=199298

Chen, J. (2024). Financial crisis analysis and prediction in Africa using machine learning and statistical learning methods. *Highlights in Science, Engineering and Technology, 107*, 88–95. https://doi.org/10.54097/r7f8qh87

Cheptiș, A., Diaconu, D., Radu, A.-N., Necula, C., & Murarașu, B. (2025). A sectorial assessment of the typology of economic crises in selected Central and Eastern European economies. *Cogent Economics & Finance, 13*(1), Article 2588932. https://doi.org/10.1080/23322039.2025.2588932

Comelli, F. (2014). *Comparing the performance of logit and probit early warning systems for currency crises in emerging market economies* (IMF Working Paper No. 14/65). International Monetary Fund.

Coulibaly, S., & Marouane, A. (2024). Macroeconomic currency crisis early warning indicators in emerging countries. *Journal of Business and Economic Development, 9*(2), 21–34. https://doi.org/10.11648/j.jbed.20240902.11

De Marchi, R., & Moro, A. (2024). Forecasting fiscal crises in emerging markets and low-income countries with machine learning models. *Open Economies Review, 35*(1), 189–213. https://doi.org/10.1007/s11079-023-09722-9

Eichengreen, B., Rose, A. K., & Wyplosz, C. (1995). Exchange market mayhem: The antecedents and aftermath of speculative attacks. *Economic Policy, 10*(21), 249–312. https://doi.org/10.2307/1344591

Flood, R. P., & Garber, P. M. (1984). Collapsing exchange-rate regimes: Some linear examples. *Journal of International Economics, 17*(1–2), 1–13. https://doi.org/10.1016/0022-1996(84)90002-3

Frankel, J. A., & Rose, A. K. (1996). Currency crashes in emerging markets: An empirical treatment. *Journal of International Economics, 41*(3–4), 351–366. https://doi.org/10.1016/S0022-1996(96)01441-9

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Guo, Q., & Su, Z. (2023). Capital account liberalization and systemic financial risk: Evidence from 24 countries. *Economic Research-Ekonomska Istraživanja, 36*(1), 845–863. https://doi.org/10.1080/1331677X.2022.2080741

Guzmán, K. M., & Fan, H. (2026). The pattern of exchange rate misalignments prior to currency crises in Brazil: An ARDL bounds testing approach. *Cogent Economics & Finance, 14*(1), Article 2659965. https://doi.org/10.1080/23322039.2026.2659965

Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance, 17*(12), 1933–1963. https://doi.org/10.1080/14697688.2017.1357972

Hopp, D. (2022). *Benchmarking econometric and machine learning methodologies in nowcasting* (arXiv:2205.03318). arXiv. https://doi.org/10.48550/arXiv.2205.03318

International Monetary Fund. (2022). *Sri Lanka: 2021 Article IV consultation—Press release; staff report; and statement by the Executive Director for Sri Lanka* (IMF Country Report No. 22/91). https://doi.org/10.5089/9798400206450.002

Jarmulska, B. (2020). *Random forest versus logit models: Which offers better early warning of fiscal stress?* (SSRN Scholarly Paper No. 3600911). Social Science Research Network. https://doi.org/10.2139/ssrn.3600911

Joy, M., Rusnák, M., Šmídková, K., & Vašíček, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics, 22*(1), 44–67. https://doi.org/10.1002/ijfe.1570

Kaminsky, G. L., & Reinhart, C. M. (1999). The twin crises: The causes of banking and balance-of-payments problems. *The American Economic Review, 89*(3), 473–500. https://doi.org/10.1257/aer.89.3.473

Kotan, K., Kotan, B., & Kırışoğlu, S. (2025). Detection of economic crises with language models and comparative analysis of simple time series analysis models and machine learning algorithms on the stock market. *IEEE Access, 13*, 83254–83273.

Kristanti, F. T., Salim, D. F., & Ihsan, A. F. (2025). Financial distress prediction in emerging markets by using a machine learning approach for financial sustainability. *Discover Sustainability, 6*(1), Article 1296. https://doi.org/10.1007/s43621-025-02199-1

Krugman, P. (1979). A model of balance-of-payments crises. *Journal of Money, Credit and Banking, 11*(3), 311–325. https://doi.org/10.2307/1991793

Krugman, P. (1999). Balance sheets, the transfer problem, and financial crises. *International Tax and Public Finance, 6*(4), 459–472. https://doi.org/10.1023/A:1008741113074

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates.

Obstfeld, M. (1986). Rational and self-fulfilling balance-of-payments crises. *The American Economic Review, 76*(1), 72–81.

Obstfeld, M., & Rogoff, K. (1995). The mirage of fixed exchange rates. *Journal of Economic Perspectives, 9*(4), 73–96. https://doi.org/10.1257/jep.9.4.73

Osman, B. M., & Muse, A. M. S. (2024). Predictive analysis of Somalia's economic indicators using advanced machine learning models. *Cogent Economics & Finance, 12*(1), Article 2426535. https://doi.org/10.1080/23322039.2024.2426535

Papadopoulos, S., Stavroulias, P., Sager, T. W., & Baranoff, E. G. (2017). *A ternary-state early warning system for the European Union* (SSRN Scholarly Paper No. 4192582). Social Science Research Network. https://doi.org/10.2139/ssrn.4192582

Prabheesh, K. P., & Jayawickrema, V. (2024). *Early warning models for anticipating crisis: Insights from Sri Lanka* (SSRN Scholarly Paper No. 5049457). Social Science Research Network. https://doi.org/10.2139/ssrn.5049457

Rademacher, P. (2025). Forecasting recessions in Germany with feature selection and machine learning. *Journal of Business Cycle Research, 21*(2–3), 119–157. https://doi.org/10.1007/s41549-025-00115-0

Reinhart, C. M., & Rogoff, K. S. (2009). The aftermath of financial crises. *The American Economic Review, 99*(2), 466–472. https://doi.org/10.1257/aer.99.2.466

Rose, A. K., & Spiegel, M. M. (2010). Cross-country causes and consequences of the 2008 crisis: International linkages and American exposure. *Pacific Economic Review, 15*(3), 340–363. https://doi.org/10.1111/j.1468-0106.2010.00507.x

Samarakoon, L. P. (2024). What broke the pearl of the Indian Ocean? The causes of the Sri Lankan economic crisis and its policy implications. *Journal of Financial Stability, 70*, Article 101213. https://doi.org/10.1016/j.jfs.2023.101213

Wang, D., & Zhou, Y. (2024). An innovative machine learning workflow to research China's systemic financial crisis with SHAP value and Shapley regression. *Financial Innovation, 10*(1), Article 103. https://doi.org/10.1186/s40854-023-00574-3