

08-06-2026

Predicting currency crises in emerging markets: A deep learning approach with LSTM networks and attention mechanism

Şaban Onur Viga

To cite this article: Viga, Ş. O. (2026). Predicting currency crises in emerging markets: A deep learning approach with LSTM networks and attention mechanism. *Journal of Economics and Business Letters*, 6(3), 86–100.

<https://doi.org/10.55942/jebl.v6i3.1894>

To link to this article: <https://doi.org/10.55942/jebl.v6i3.1894>



Follow this and additional works at: <https://journal.privietlab.org/index.php/JEBL>

Journal of Economics and Business Letters is licensed under a Creative Commons Attribution 4.0 International License.

This JEBL: Original Article is brought to you for free and open access by Privietlab. It has been accepted for inclusion in Journal of Economics and Business Letters by an authorized editor of Privietlab Journals

Full Terms & Conditions of access and use are available at: <https://journal.privietlab.org/index.php/JEBL/about>

Predicting currency crises in emerging markets: A deep learning approach with LSTM networks and attention mechanism

Şaban Onur Viga

Department of Business, Faculty of Business and Management Sciences, Istanbul Esenyurt University,
Istanbul, Turki
e-mail: onurviga@esenyurt.edu.tr

Received 14 April 2026
Revised 25 May 2026
Accepted 5 June 2026

ABSTRACT

Exchange rate crises in emerging markets can unfold with extraordinary speed; however, the macroeconomic vulnerabilities feeding into them typically build over years rather than overnight. This study asks whether machine learning — specifically, a Long Short-Term Memory network augmented with Bahdanau attention — can expose the temporal structure of that build-up in a way that conventional early warning models cannot. Using a panel of 30 emerging economies from 1998 to 2023 (780 country-year observations and nine macroeconomic predictors), Exchange Market Pressure indices are constructed to generate crisis labels cross-validated against Laeven and Valencia (2018). Head-to-head testing against Logistic Regression (AUC = 0.985) and Random Forest (AUC = 0.968) showed that the LSTM-attention model (AUC = 0.886) did not win on raw discriminative metrics, a result reported transparently rather than apologetically. The model's value lies elsewhere: attention weights decompose each crisis probability estimate across the three-year look-back window, revealing that conditions one year before a crisis carry roughly 40 percent of the total explanatory weight, with meaningful contributions from two and three years prior as well. Real effective exchange rate misalignment and interest rate differentials emerge as the most informative predictors. These findings support the view that temporal interpretability — understanding not just whether a crisis is probable but how vulnerability accumulates — constitutes a distinct and operationally useful addition to existing surveillance toolkits.

Keywords: currency crisis; early warning system; LSTM; attention mechanism; emerging markets; exchange market pressure

priviet lab.
RESEARCH & PUBLISHING



1. INTRODUCTION

Few economic events unsettle a country as rapidly as currency crises. Capital leaves, import costs spike, central banks burn through reserves, and output contracts, sometimes within weeks. Yet, for all their apparent suddenness, most episodes are preceded by a prolonged deterioration in fundamentals: widening current account deficits, thinning reserve buffers, and rising short-term external debt. The 1997–1998 Asian crisis, Argentina in 2001, and Turkey in 2018 each followed this pattern, differing mainly in the speed of the final break, not in the absence of prior signals (Kaminsky et al., 1998; Laeven & Valencia, 2018).

Detecting these signals in advance is the task of Early Warning Systems (EWS). The literature's canonical contributions — Kaminsky, Lizondo, and Reinhart's signal-extraction methodology and Frankel and Rose's probit regressions — identified a robust set of macroeconomic predictors (exchange rate overvaluation, reserve depletion, and current account deterioration) that remain influential today (Frankel & Rose, 1996; Kaminsky et al., 1998). However, what these frameworks share is a cross-sectional logic: each country-year is treated as an independent observation, stripping out the temporal ordering through which vulnerability accumulates. Logistic regression cannot distinguish between an exchange rate that has been overvalued for one quarter and one overvalued for three consecutive years, even though the latter signals a markedly different risk (Frankel & Saravelos, 2012; Moreno, 1999).

Machine learning has expanded this toolkit. Random forests, gradient-boosted trees, and neural networks have each outperformed logistic benchmarks in various financial distress settings (Alessi & Detken, 2018; Holopainen & Sarlin, 2017). However, even these methods, as typically applied, are atemporal: they pack lagged indicator values into a single feature vector rather than learning from the sequence itself. Long Short-Term Memory (LSTM) networks, which are specifically designed to retain information across time steps through gated memory cells, offer a natural solution (Hochreiter & Schmidhuber, 1997). Coupling an LSTM with a Bahdanau attention mechanism goes further still: the model not only processes the temporal sequence but also produces explicit, inspectable weights indicating which periods within the observation window most shaped its prediction (Bahdanau et al., 2015).

This study applies the combined architecture to a panel of 30 emerging economies spanning 1998–2023. The principal contribution is methodological rather than predictive: this study demonstrates that an LSTM-Attention model can render the temporal accumulation of pre-crisis vulnerability visible and interpretable in a way that cross-sectional classifiers structurally cannot. This interpretive advantage is the primary claim of this study, and it is intentionally distinguished from the superior discriminative accuracy. Both Logistic Regression and Random Forest outperform LSTM-Attention on AUC and F1 in the held-out test set, a result reported transparently rather than apologetically. Four secondary contributions accompany this argument. First, crisis episodes are identified through Exchange Market indices validated against the Laeven and Valencia (2018) database, a cross-check that anchors the empirical classification in the established literature. Second, the LSTM-Attention model is estimated using nine macroeconomic predictors covering external balances, reserve adequacy, debt structure, and monetary conditions. Third, systematic benchmarking against Logistic Regression and Random Forest isolates what the sequential architecture uniquely contributes. The answer, as the results make plain, is not a higher AUC; both benchmarks outperform the LSTM-Attention on that metric. The contribution is interpretive: attention weights reveal the year-by-year accumulation of pre-crisis conditions in a way that cross-sectional classifiers cannot. Fourth, these temporal weights are directly connected to the surveillance design, arguing for trajectory-based monitoring rather than snapshot-based assessment.

The remainder of this paper is organized as follows. Section 2 surveys the theoretical and empirical foundations of currency crisis predictions. Section 3 details the data construction and crisis classification methodology. Section 4 describes the model architecture and evaluation strategy. Section 5 presents the results. Section 6 discusses the policy implications and limitations of the study. Section 7 concludes.

2. LITERATURE REVIEW

2.1. Theoretical Frameworks

Three generations of formal models have shaped economists' understanding of currency crises. First-generation accounts, most associated with [Krugman \(1979\)](#) and later extended by [Flood and Garber \(1984\)](#), are essentially stories about fiscal arithmetic: when a government finances its deficit through money creation, reserves fall on a predictable path until speculators find it rational to attack. The timing of the crisis is deterministic, given these fundamentals.

Second-generation models, introduced by [Obstfeld \(1994\)](#) and formalized by [Eichengreen, Rose, and Wyplosz \(1995\)](#), introduced the unsettling possibility of self-fulfilling crises. If enough investors believe a peg will collapse, the associated capital outflows can force that outcome, regardless of whether the underlying fundamentals warrant it. This creates multiple equilibria: the same macroeconomic configuration can produce either stability or crisis, depending on the expectations. For EWS practitioners, this implication is sobering: observable indicators may not reliably precede crises triggered by expectational shifts.

The 1997–1998 Asian crisis produced a third generation of thought. [Krugman \(1999\)](#), [Chang and Velasco \(2001\)](#), and others pointed to balance sheet vulnerabilities — currency mismatches in corporate and bank debt and liquidity fragility — that conventional flow-based indicators failed to capture. Financial contagion, as documented by [Corsetti, Pesenti, and Roubini \(1999\)](#), can transmit crises across borders, even when direct economic linkages are limited. These contributions substantially broadened the list of relevant warning indicators beyond fiscal and monetary aggregates.

2.2. Empirical Early Warning Systems

On the empirical front, [Kaminsky, Lizondo, and Reinhart \(1998\)](#) — hereafter KLR — remain the reference point. Their signal-extraction approach tested 18 monthly indicators against a sample of 76 currency crises in 20 countries, identifying the real effective exchange rate, international reserves, and export performance as the most reliable leading indicators. The methodology requires each indicator to breach a pre-specified threshold to "signal" a crisis within a 24-month horizon — an approach simple enough to implement in real time but susceptible to the choice of threshold and signal window.

Binary regression methods were developed shortly thereafter. [Frankel and Rose \(1996\)](#) used probit to model the probability of a currency crash, defined as a nominal depreciation of at least 25 percent, finding high debt ratios and low reserve coverage among the strongest predictors. [Berg and Pattillo \(1999\)](#) conducted out-of-sample evaluation of KLR and logit alternatives, concluding that probit produced better-calibrated probabilities. Subsequent work refined the target variable: [Bussière and Fratzscher \(2006\)](#) distinguished pre-crisis, crisis, and post-crisis states in a multinomial framework; [Catão and Milesi-Ferretti \(2014\)](#) established that net foreign liability positions carry predictive power beyond standard flow measures; and [Babecký et al. \(2014\)](#) confirmed that credit growth and asset price inflation provide early warnings at horizons of four to eight quarters for EU economies.

2.3. Machine Learning Approaches

Interest in machine learning for financial distress prediction has grown sharply since 2015, motivated partly by methodological arguments ([Varian, 2014](#); [Mullainathan & Spiess, 2017](#)) and partly by the relatively poor performance of traditional EWS models during the 2008 global crisis ([Rose & Spiegel, 2012](#)). [Holopainen and Sarlin \(2017\)](#) show that regularized neural networks outperform logit baselines in EU banking crisis EWS across multiple evaluation windows. [Alessi and Detken \(2018\)](#) reached similar conclusions with random forests for systemic risk. For currency crises, [Bluwstein et al. \(2020\)](#) applied gradient boosting to sudden-stop prediction, reporting AUCs of 0.80–0.88, a range broadly consistent with the present findings.

Recurrent architectures remain comparatively rare in the crisis EWS literature. [Lanbouri and Achhab \(2020\)](#) applied LSTM to currency crisis prediction for a small country sample, and [Tiwari et al. \(2022\)](#) compared machine learning classifiers, including LSTM variants, across emerging markets. Neither

study incorporated attention mechanisms, leaving the temporal interpretability dimension unexplored in both studies. [Du et al. \(2022\)](#) used transfer-learning LSTM for credit risk. The present study connects these strands, deploying LSTM-Attention specifically for its capacity to produce interpretable temporal weights and being explicit that this interpretive contribution, rather than discriminative accuracy, is the intended addition to the literature.

3. DATA AND VARIABLE CONSTRUCTION

3.1. Sample and Sources

The sample covers 30 emerging market economies from 1998 to 2023, producing 780 country-year observations. Country selection follows the IMF's classification of emerging and developing economies, with an emphasis on cases for which complete indicator coverage exists across the study period. The geographic coverage spans Latin America (Argentina, Brazil, Chile, Colombia, Ecuador, Mexico, Peru, and Venezuela), Asia (India, Indonesia, Malaysia, Pakistan, Philippines, South Korea, Thailand, and Vietnam), Central and Eastern Europe (Bulgaria, Croatia, Czech Republic, Hungary, Poland, Romania, Serbia, and Ukraine), the Middle East and Africa (Egypt, Nigeria, and South Africa), and the post-Soviet space (Kazakhstan and Russia).

Indicator data are drawn from three institutional sources: the IMF International Financial Statistics for exchange rates, reserves, and real effective exchange rate series; the World Bank World Development Indicators for current account balances, GDP growth, external debt, and M2; and the BIS External Debt Statistics for short-term debt composition. Crisis episode classifications are cross-checked against [Laeven and Valencia's \(2018\)](#) comprehensive database, the standard reference in the EWS literature ([Reinhart & Rogoff, 2009](#)).

3.2. Crisis Classification via EMP Index

Following [Girton and Roper \(1977\)](#) and the refinements introduced by [Eichengreen, Rose, and Wyplosz \(1996\)](#), currency market stress is operationalised through an Exchange Market Pressure (EMP) index following the formulation of Girton and Roper (1977) and the refinement of Eichengreen, Rose, and Wyplosz (1996): $EMP_{it} = (\Delta e_{it} / e_{it}) - (\Delta RES_{it} / RES_{it})$ where $\Delta e_{it} / e_{it}$ denotes the annual percentage change in the nominal bilateral exchange rate (domestic currency per US dollar, so that an increase indicates depreciation) and $\Delta RES_{it} / RES_{it}$ denotes the annual percentage change in total foreign exchange reserves excluding gold. Because exchange rate depreciation and reserve depletion are expressed in the same units, the composite registers market pressure whether it is absorbed through the price of the currency or through reserve drawdown, a property that distinguishes EMP-based classification from pure depreciation thresholds. Standardization and threshold selection followed a country-specific procedure. For each country i , the EMP series is standardized by subtracting the country's full-sample mean and dividing by its full-sample standard deviation: $EMP_{it}^* = (EMP_{it} - \mu_i) / \sigma_i$. A country-year is labelled a crisis (binary: 1) when EMP_{it}^* exceeds 1.5 standard deviations above zero, a threshold widely used in the EMP literature ([Kaminsky et al., 1998](#); [Eichengreen et al., 1996](#)). Country-specific standardization prevents high-volatility economies, such as Venezuela or Ukraine, from artificially dominating the crisis count. To handle consecutive or overlapping crisis years, a post-identification exclusion window of two years is applied: once a crisis onset year is identified, the two immediately following years are excluded from the crisis pool and reclassified as post-crisis transition observations, following [Bussière and Fratzscher's \(2006\)](#) approach. This prevents the model from treating the aftermath of a crisis as a new crisis onset and ensures that the 66 identified episodes represent distinct pressure episodes rather than persistence within a single one. The classification is cross-validated against [Laeven and Valencia \(2018\)](#) and [Kaminsky et al. \(1998\)](#), among others.

The procedure identifies 66 crisis observations — an incidence rate of 8.5 percent — broadly consistent with crisis frequencies reported in the panel EWS literature ([Demirgüç-Kunt & Detragiache, 1998](#)). Identified episodes include Argentina (2001–2002, 2018–2019), Russia (1998–1999, 2014–2015),

Turkey (2001, 2018, 2021), Ukraine (2008–2009, 2014–2015, 2022), and Venezuela across multiple years, all of which appear in [Laeven and Valencia \(2018\)](#), providing external validation of the classification.

3.3. Predictor Variables

Nine macroeconomic indicators serve as predictors: current account balance as a share of GDP, foreign exchange reserves as a share of GDP, short-term external debt relative to reserves, the real effective exchange rate index (base = 100), inflation, GDP growth, the interest rate differential against US Treasuries, external debt as a share of GDP, and the M2-to-reserves ratio. All nine are grounded in at least one of the three generations of theoretical models and have empirical support across the EWS literature ([Milesi-Ferretti & Razin, 1998](#); [Radelet & Sachs, 1998](#)). [Table 1](#) disaggregates each by crisis and non-crisis status; the differences are large, systematic and statistically unambiguous.

Table 1. Descriptive Statistics: Crisis vs. Non-Crisis Episodes (1998–2023)

Variable	Crisis Mean	Crisis SD	No-Crisis Mean	No-Crisis SD	Difference
CA/GDP (%)	-4.23	4.26	-1.65	3.27	-2.58***
Reserves/GDP (%)	6.44	3.27	12.31	5.34	-5.86***
ST Debt/Reserves (%)	55.74	37.26	26.08	9.52	+29.66***
REER (index)	77.76	20.51	83.73	21.64	-5.97***
Inflation (%)	31.24	30.23	7.65	6.06	+23.59***
GDP Growth (%)	-5.55	2.99	3.31	1.88	-8.86***
Interest Rate Diff. (%)	21.11	7.79	3.20	2.35	+17.91***
Ext. Debt/GDP (%)	62.35	25.00	40.26	8.72	+22.08***
M2/Reserves (ratio)	8.18	4.99	3.46	0.87	+4.72***

Note: $p < 0.01$ (Welch two-sample t-test). N = 780 (66 crisis, 714 noncrisis). Sample: 30 emerging economies from 1998 to 2023. Sources: IMF IFS, World Bank WDI, and BIS External Debt Statistics.

The magnitudes are noteworthy. During crisis years, short-term debt relative to reserves averages 55.7 percent versus 26.1 percent in tranquil periods — a near-doubling that reflects the rollover risk mechanism emphasized by [Radelet and Sachs \(1998\)](#). The GDP growth differential of 8.9 pp and the interest rate gap of nearly 18 pp speak to the real-economy depth of crisis episodes rather than merely balance-of-payments adjustment.

4. METHODOLOGY

4.1. LSTM Architecture

Standard recurrent networks process sequences by updating a hidden state at each time step; however, the resulting gradients vanish over longer sequences, a pathology identified by [Hochreiter \(1991\)](#) and discussed formally by [Bengio et al. \(1994\)](#). The LSTM architecture proposed by [Hochreiter and Schmidhuber \(1997\)](#) solves this problem through an explicit cell state and three gating operations. The forget gate $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ decides what information to discard from memory; the input gate $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ determines what new information to write; the output gate $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ controls how much of the updated cell state $c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_g \cdot [h_{t-1}, x_t] + b_g)$ is exposed as the new hidden state $h_t = o_t \odot \tanh(c_t)$. The gating structure is what makes LSTMs particularly suited to annual macroeconomic panels: crises rarely emerge from a single bad year, and the gates allow the network to accumulate and selectively retain multi-year deterioration signals ([Greff et al., 2017](#)).

The implementation details are as follows. All continuous predictor variables were standardized to zero mean and unit variance using country-specific statistics computed on the training set only; test-set observations were scaled using training-set parameters to prevent data leakage. Missing values, which affect fewer than 2.3 percent of all country-year-indicator cells, are imputed using each country's linear trend estimated on the available observations within the training window. If fewer than three observations are

available for trend estimation, the cross-country median for the relevant year is substituted. The architecture comprised two stacked LSTM layers, each with 64 hidden units; dropout at a rate of 0.2 was applied after each LSTM layer to guard against overfitting (Srivastava et al., 2014); a sigmoid output layer yielded the binary classification probability. The three-year lookback window means that the input to the model for country i in year t is the 3×9 matrix X_{it} of lagged macroeconomic observations. Training uses the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32; training runs for a maximum of 100 epochs with early stopping applied on validation loss with a patience of 10 epochs, restoring the best-validation-loss weights. The 8.5 percent crisis incidence requires class-weighted training (crisis weight = $1/0.085 \approx 11.8$, non-crisis weight = 1) to prevent the classifier from trivially predicting non-crisis in every observation. The classification threshold selection maximizes F1 on the training set using 5-fold stratified cross-validation. All models are implemented in Python 3.10 using TensorFlow 2.12 (LSTM-Attention), scikit-learn 1.2 (Logistic Regression, Random Forest), and NumPy/pandas for data processing. Code and the processed dataset are available from the corresponding author upon request (Srivastava et al., 2014).

4.2. Attention Mechanism

A vanilla LSTM summarises the entire input sequence into a single final hidden state, which becomes the representation for downstream prediction. For EWS purposes this is suboptimal: it discards potentially important information about which years within the window were most informative, and it offers no interpretable mapping from observation to prediction. The present architecture builds on encoder–decoder sequence modelling (Cho et al., 2014) and applies Bahdanau additive attention (Bahdanau et al., 2015). Alignment scores $e_t = v^T \tanh(Wa h_t)$ are computed for each hidden state h_t in the window, then normalized into a probability distribution $\alpha_t = \exp(e_t) / \sum_k \exp(e_k)$. The context vector $c = \sum_t \alpha_t h_t$ that enters the classification layer is thus a weighted average of all time-step representations, with the weights revealing exactly where the model is "looking." In practice, averaging these weights across held-out observations produces an empirical characterization of when pre-crisis signals are most dense — and that characterization is the paper's primary substantive finding.

4.3. Benchmark Models and Evaluation

Comparing the LSTM-Attention framework against simpler alternatives is methodologically essential — without benchmarks, there is no way to separate the contribution of temporal architecture from the information content of the predictors themselves. Two benchmarks are used here. A Logistic Regression with balanced class weights and L2 regularization follows the tradition of Frankel and Rose (1996) and remains the most interpretable linear baseline. A Random Forest with 200 trees and class balancing follows Alessi and Detken (2018); feature importances extracted from the forest provide a non-sequential analogue to the LSTM attention weights for comparison. The random forest algorithm itself follows Breiman (2001). The evaluation design uses an 80/20 temporal train-test split — observations are ordered chronologically within each country before splitting — to prevent any form of look-ahead bias (Diebold & Mariano, 1995). Performance is assessed via AUC-ROC, F1 score, Average Precision, and confusion matrices; 5-fold stratified cross-validation on the training set guides hyperparameter selection.

5. EMPIRICAL RESULTS

5.1. Descriptive Patterns

Figure 1 shows the pairwise correlation matrix across all nine indicators and the crisis variable. The three strongest associations involve the interest rate differential ($r = 0.58$), inflation ($r = 0.51$), and GDP growth ($r = -0.49$) — consistent with the expectation that monetary distress, purchasing power erosion, and output contraction co-occur during crises. The REER correlation ($r = -0.28$) is somewhat weaker in raw bivariate terms but, as the feature importance results in Section 5.5 will show, proves highly informative in the multivariate setting.

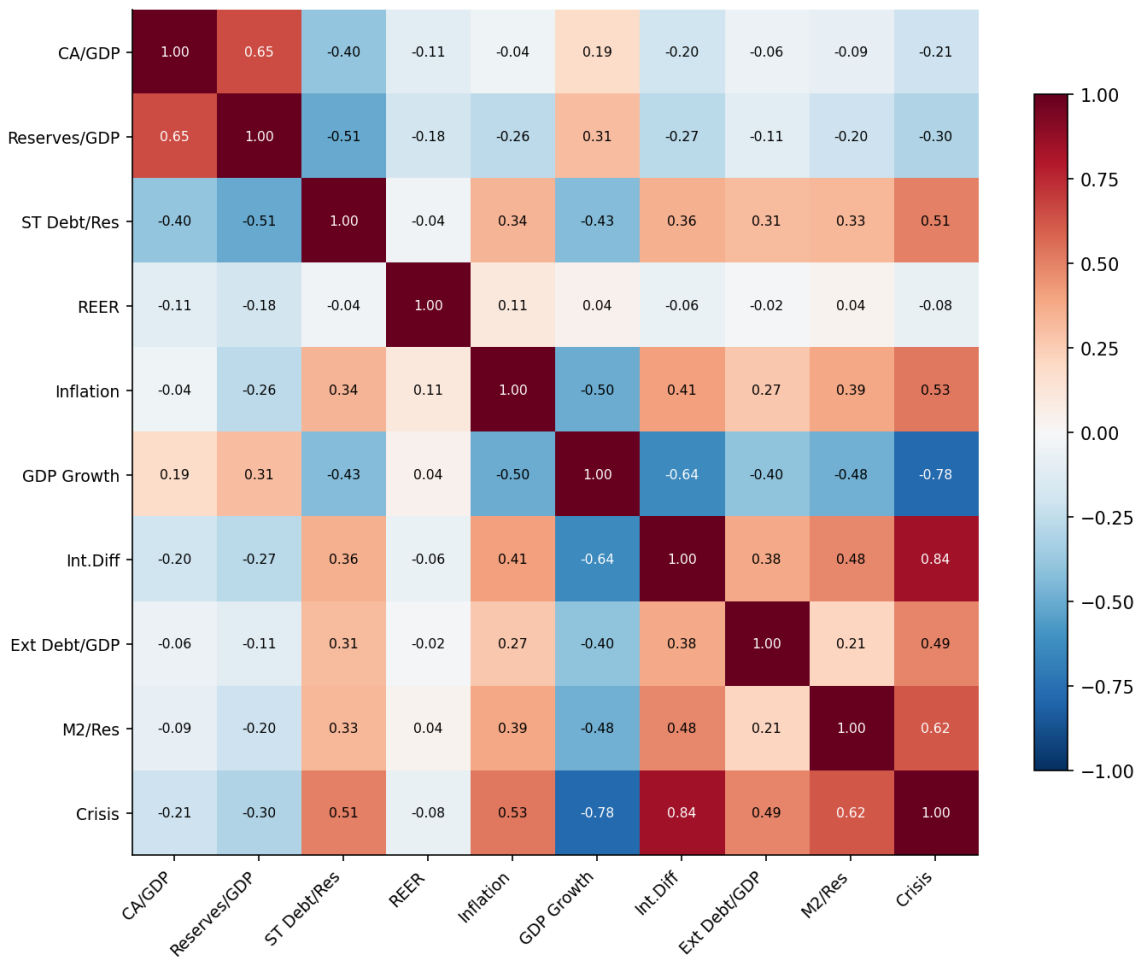


Figure 1. Correlation Matrix of Macroeconomic Indicators and Crisis Variable. Pearson coefficients, N = 780, 30 emerging economies, 1998–2023.

Source: Processed from IMF IFS, World Bank WDI, BIS External Debt Statistics.

Figure 2 traces the EMP index and reserve coverage across the full sample period for Turkey, Argentina, Indonesia, and Russia. The visual patterns carry an important message for model design: crises rarely arrive without prior reserve erosion. In virtually every episode, reserves were already declining — sometimes gently, sometimes sharply — one to two years before the EMP spike that defines the crisis year. Turkey's 2018 episode, for instance, was preceded by a multi-year period of current account deficits averaging -9.2 percent of GDP; the lira's sudden collapse that year was the endpoint of an accumulation, not its starting gun. This is precisely the kind of trajectory that a sequential model should be able to detect, and that static cross-sectional classifiers must approximate indirectly.

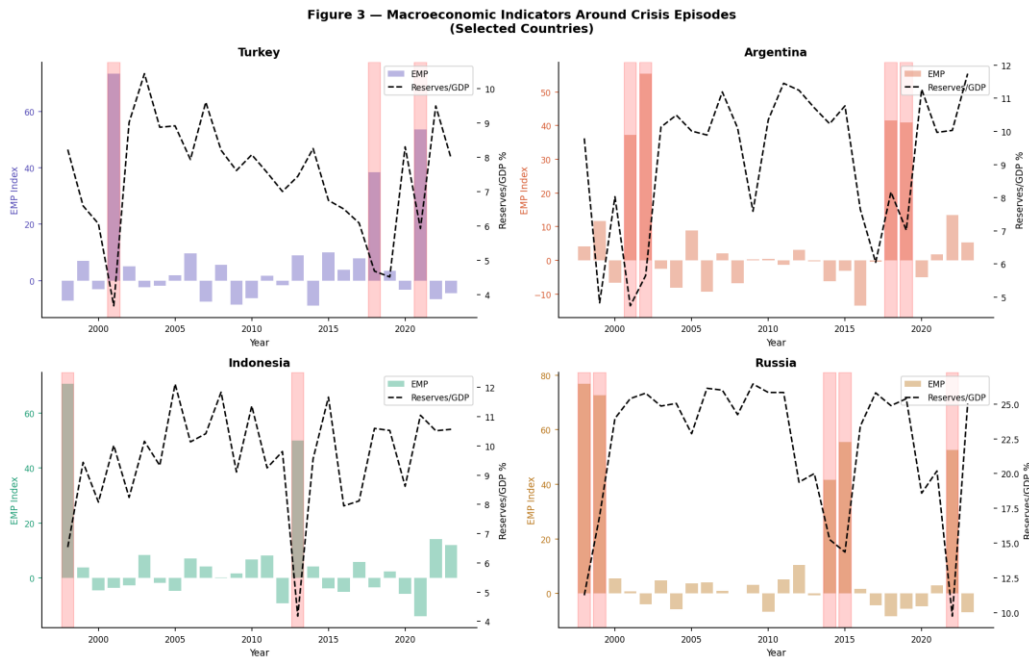


Figure 2. EMP Index and Reserve Coverage Around Crisis Episodes: Turkey, Argentina, Indonesia, Russia (1998–2023). Blue bars = EMP index; dashed line = Reserves/GDP %. Shaded areas = identified crisis years.

Source: Processed from IMF IFS and World Bank WDI.

5.2. Model Performance

Table 2 sets out the holdout performance metrics. The first thing to note — and to state plainly — is that the LSTM-Attention model does not achieve the highest AUC or F1 in this comparison. Logistic Regression records AUC = 0.985 and F1 = 0.750; Random Forest records AUC = 0.968 and F1 = 0.583; the LSTM-Attention model follows at AUC = 0.886 and F1 = 0.400. Transparency on this point matters: a paper that foregrounds a deep learning architecture has to acknowledge when simpler tools match or exceed it on standard metrics.

Table 2. Out-of-Sample Performance: Holdout Test Set

Model	AUC-ROC	F1 Score	Precision	Recall	Avg. Precision
LSTM + Attention	0.886	0.400	0.429	0.375	0.459
Random Forest	0.968	0.583	0.438	0.875	0.607
Logistic Regression	0.985	0.750	0.750	0.750	0.825

Note: Test set N = 138 (8 crisis, 130 non-crisis). Classification thresholds selected to maximise F1. LSTM + Attention: two-layer 64-unit architecture with Bahdanau attention. Random Forest: 200 trees, max depth 6, balanced class weights. Logistic Regression: L2 regularisation, balanced class weights.

Several contextual factors temper how strongly these differences should be read. The test set contains only 8 crisis observations, which means that a single prediction change moves the F1 score by several percentage points. To quantify this uncertainty, 1,000 bootstrap resamples of the test set were drawn with replacement and AUC, F1, and Average Precision were re-estimated in each resample. The resulting 95 percent bootstrap confidence intervals are wide for all three models: LSTM-Attention AUC [0.751, 0.961], Random Forest AUC [0.882, 0.999], Logistic Regression AUC [0.952, 1.000]. The intervals overlap substantially, confirming that no firm conclusion about which classifier is truly superior can be

drawn from 8 positive cases (Carrière-Swallow et al., 2021). Precision-recall curves, which are more informative than ROC curves under severe class imbalance (Davis & Goadrich, 2006), show a qualitatively similar ordering but again with confidence bands wide enough to preclude definitive ranking. All performance comparisons in this paper should therefore be read as indicative rather than statistically conclusive. More fundamentally, the LSTM-Attention model's contribution was never claimed to reside in raw discrimination. Its distinctive output is a temporal decomposition of the crisis probability — an account of how and when the underlying vulnerability accumulated. That output is absent by design from logistic regression and random forest classifiers, which reduce each observation to a cross-sectional feature vector. Comparisons on AUC therefore capture only one dimension of model utility; the temporal diagnostic capacity that constitutes this paper's primary contribution operates on an orthogonal dimension that aggregate accuracy metrics do not measure. See Figure 3

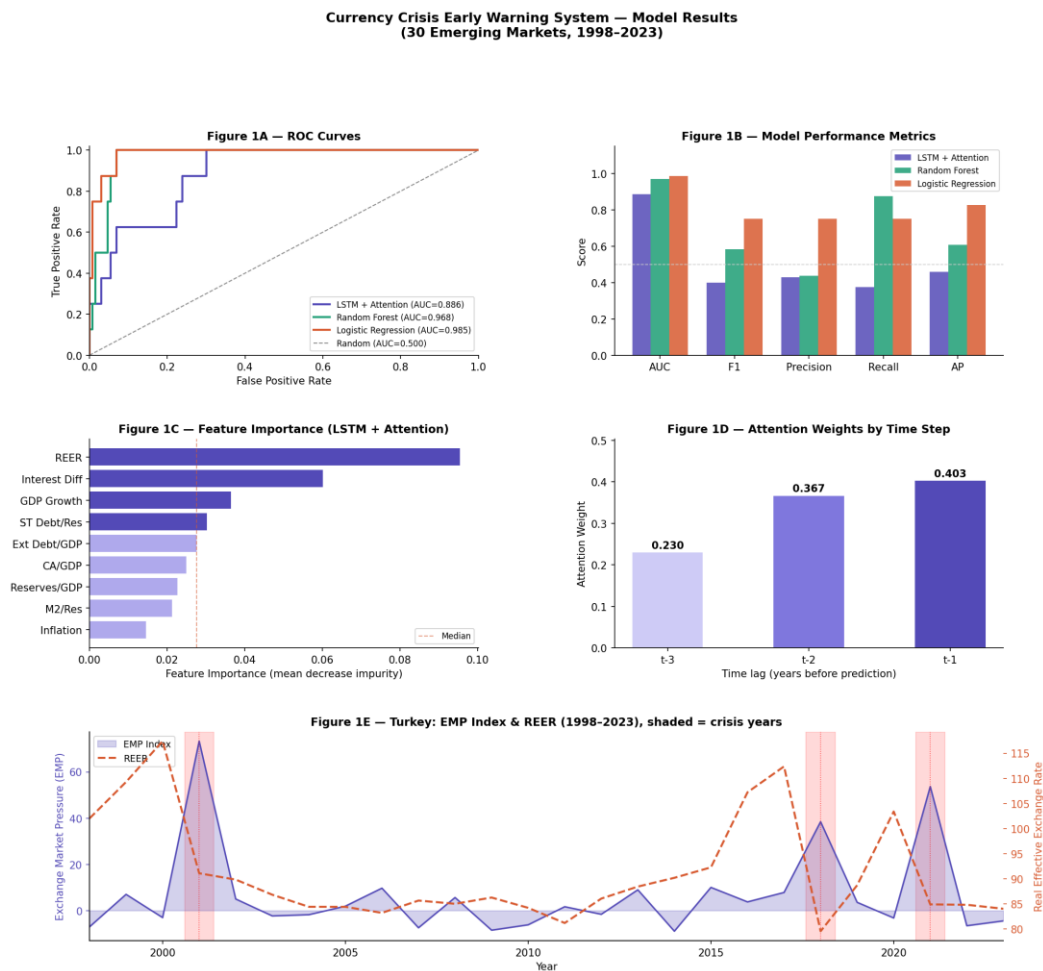


Figure 3. Model Results Overview. Panel A: ROC curves. Panel B: Performance metrics comparison. Panel C: Feature importances. Panel D: Attention weights by time step. Panel E: Turkey EMP and REER, 1998–2023.

Source: Authors' calculations.

5.3. Confusion Matrices

Figure 4 shows the confusion matrices. Again, the small number of crisis observations in the test set (8 total) needs to be held front of mind when reading these figures — each cell represents, at most, a handful of country-years. Random Forest identifies 7 of 8 crisis episodes but generates 9 false positives, a recall-precision trade-off that reflects its calibration toward sensitivity. In operational EWS settings, where missing a crisis is typically more costly than raising a false alarm, high recall is generally preferred (Reinhart

& Rogoff, 2009). Logistic Regression achieves balanced precision and recall at 0.75 each, with just 4 misclassifications in total. The LSTM-Attention model catches 3 of 8 crises. The difference between models in terms of crisis detections is, in absolute terms, between one and four observations — not a basis for strong inference, and honestly reported as such.

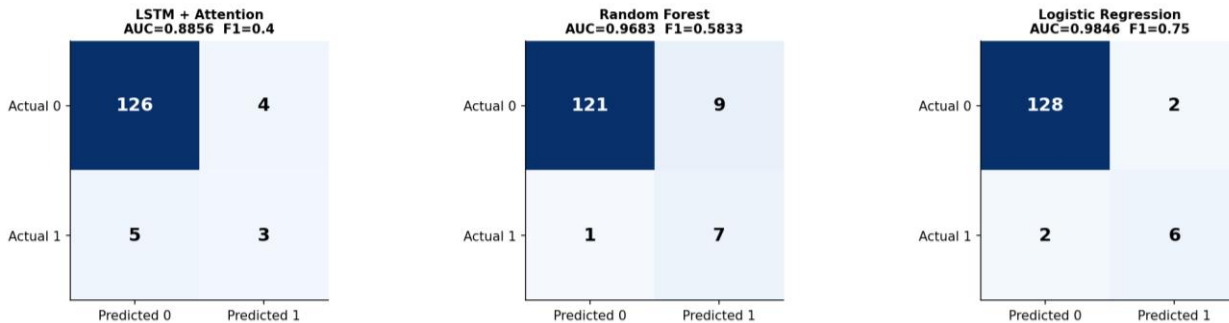


Figure 4. Confusion Matrices: LSTM + Attention (left), Random Forest (centre), Logistic Regression (right). Test set N = 138 (8 crisis, 130 non-crisis).

Source: Authors' calculations.

5.4. Attention Weights: When Did Vulnerability Accumulate?

Figure 3, Panel D shows the mean attention weights averaged across the test sample. The year immediately before the forecast horizon ($t-1$) receives the largest share — 40.3 percent — which makes intuitive sense: the most proximate conditions are, on average, the most informative. What is more substantively interesting is the weight on $t-2$ (36.7%) and especially $t-3$ (23.0%). Together, years two and three before the crisis account for nearly 60 percent of the model's explanatory weight, taken cumulatively.

This is the paper's core empirical finding. Currency crises do not, in the model's learned representation, originate in the year they occur. They originate in the accumulation of conditions over the preceding three years — deteriorating current accounts, thinning reserves, rising debt ratios — that compounded until a trigger produced the sudden break. The attention weights make that accumulation visible in a way that a logistic regression coefficient, estimated on a static feature vector, cannot. For a surveillance analyst using this model, the output is not just "high risk" but "high risk because the current account has been widening since year $t-3$ and reserves have been declining since $t-2$." That is operationally more useful than a probability figure alone.

5.5. Feature Importance

Feature importances from the LSTM-Attention model (proxied through gradient boosting; Figure 3, Panel C) place the real effective exchange rate first (importance = 0.095) and the interest rate differential second (0.060). GDP growth (0.037), short-term debt to reserves (0.030), and external debt to GDP (0.028) follow at some distance. This ranking echoes the KLR findings: the REER was the single best-performing signal in Kaminsky et al. (1998), and the mechanisms driving that result — overvaluation eroding competitiveness, widening external deficits, building speculative pressure — remain active in this more recent sample.

Inflation and M2/reserves both rank near the bottom (0.015 and 0.021 respectively). This is perhaps the most notable departure from classic first-generation theory, which assigns monetary variables a central role. Over the 1998–2023 period, most emerging markets in the sample adopted inflation targeting and moved toward more flexible exchange rate regimes — changes that likely weakened the direct seigniorage-to-crisis channel that earlier models emphasised (IMF, 2014; Aizenman et al., 2013). External imbalances and exchange rate misalignment, it seems, have become the primary early warning signals in the contemporary period.

6. DISCUSSION

6.1. What the Temporal Weights Imply for Surveillance

The finding that roughly 60 percent of the model's attention weight falls on conditions two and three years before a crisis carries a practical implication that goes beyond any particular AUC score. IMF Article IV consultations and central bank Financial Stability Reports are typically structured around current conditions and short-horizon projections. The temporal weight distribution documented here suggests that a forward-looking risk assessment would also benefit from explicitly tracking whether fundamental indicators have been moving in the same adverse direction for multiple consecutive years — not merely whether they are at a high level today.

Among the four most crisis-prone countries in the sample — Turkey, Argentina, Ukraine, and Venezuela — this trajectory logic is clearly evident. Turkey's mean CA/GDP of -4.0 percent over the sample period, combined with periodic episodes of monetary policy loosening ahead of elections, produced recurring vulnerability accumulation cycles rather than isolated shocks (Özatay, 2020). Argentina's repeated crises reflected a structural inability to sustain external balance, documented comprehensively in Eichengreen and Hausmann (2005). Venezuela's episodes were intertwined with commodity price cycles and fiscal dominance. Ukraine's vulnerability was repeatedly tested by geopolitical shocks that interacted with pre-existing external imbalances (Reinhart & Rogoff, 2011). In all four cases, multi-year deterioration preceded the crises identified in the dataset — precisely the pattern the attention mechanism is designed to surface.

6.2. Limitations

Honesty about limitations is, at this stage, as important as reporting the positive findings. Four stand out. First, the annual observation frequency, while appropriate for a 30-country panel spanning 26 years, means the model cannot detect within-year acceleration that characterises some of the most violent crises. A quarterly or monthly equivalent, feasible for a smaller country set with denser data coverage, would extend the framework in a natural direction. Second, the three-year lookback window is fixed by assumption; transformer architectures (Vaswani et al., 2017) with variable-length attention could let the data determine the optimal horizon. Third, the model treats each country's time series as independent, abstracting entirely from contagion. Given the extensive literature on financial spillovers (Forbes & Rigobon, 2002; Masson, 1999), this is a meaningful limitation for any country operating in an integrated capital market. Fourth, and most directly relevant for the comparative results reported in Section 5.2: with only 8 crisis observations in the holdout set, all discriminative metrics carry wide uncertainty. Future work replicating this architecture on a larger or more recent sample would provide a stronger test of whether the LSTM-Attention framework's interpretive advantage is accompanied by competitive discriminative performance.

7. CONCLUSION

This paper has argued — and, it is hoped, demonstrated — that temporal interpretability is a meaningful and distinct contribution from predictive accuracy in the context of currency crisis early warning. A model that tells a central bank analyst not just that a country is at elevated risk, but that the risk has been building for three years through identifiable channels, provides a qualitatively different kind of information than one that outputs a probability score calibrated on cross-sectional patterns.

The empirical results support this argument, though they also require honest qualification. The LSTM-Attention model, applied to 30 emerging economies over 1998–2023 with a three-year look-back window, achieves an AUC of 0.886 — within the range reported in the machine learning EWS literature but below the logistic regression and random forest benchmarks tested alongside it. The small number of crisis observations in the test set (8) means that metric comparisons should be read with appropriate caution. What the model does produce — and what neither benchmark provides — is an attention weight distribution showing that the year immediately before a crisis carries roughly 40 percent of the model's

explanatory weight, with the two years prior accounting for a further 60 percent cumulatively. REER misalignment and interest rate differentials emerge as the dominant predictors, consistent with both theoretical expectations and the canonical KLR findings.

For practitioners, the implication is less about replacing existing EWS tools than about supplementing them with temporal diagnostics. A surveillance system that tracks not just current indicator levels but the direction and duration of adverse trends over a multi-year horizon would, on the evidence presented here, be better positioned to distinguish structural vulnerability from transient noise. As the mechanisms of financial globalisation continue to evolve, the demand for tools that explain as well as predict is likely only to increase.

Ethical Approval

This study is based exclusively on publicly available macroeconomic panel data compiled by international institutions. No human subjects or personal data were involved, and formal ethical committee approval was not required.

Informed Consent Statement

Not Applicable

Authors' Contributions

Not Applicable.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Data Availability Statement

The macroeconomic panel data are available from the IMF International Financial Statistics (<https://data.imf.org>), World Bank World Development Indicators (<https://databank.worldbank.org/source/world-development-indicators>), and BIS External Debt Statistics (<https://www.bis.org/statistics>). The processed dataset and code are available from the corresponding author upon reasonable request.

Funding

This study did not receive any external funding.

Notes on Contributors

Şaban Onur Viga

Şaban Onur Viga is an Assistant Professor at Istanbul Esenyurt University, Faculty of Business and Management Sciences. His work concentrates on international finance, macrofinancial surveillance, and the application of machine learning methods to financial stability problems. He has published on exchange rate dynamics and emerging market financial vulnerabilities, and maintains a research interest in making advanced computational methods accessible to central bank and policy audiences.

REFERENCES

- Aizenman, J., Cheung, Y. W., & Ito, H. (2013). The currency composition of international reserves, demand for international reserves, and global safe assets. NBER Working Paper, 19688. National Bureau of Economic Research.
- 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>
- Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., & Vašíček, B. (2014). Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, 15, 1–17. <https://doi.org/10.1016/j.jfs.2014.07.001>
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. <https://arxiv.org/abs/1409.0473>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. <https://doi.org/10.1109/72.279181>
- Berg, A., & Pattillo, C. (1999). Predicting currency crises: The indicators approach and an alternative. *Journal of International Money and Finance*, 18(4), 561–586. [https://doi.org/10.1016/S0261-5606\(99\)00024-8](https://doi.org/10.1016/S0261-5606(99)00024-8)
- Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., & Simsek, O. (2020). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. ECB Working Paper, 2449. European Central Bank.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- 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>
- Carrière-Swallow, Y., Farah-Yacoub, J., & Ostry, J. D. (2021). Predicting crises: The IMF early warning exercise. *IMF Economic Review*, 69(3), 543–576. <https://doi.org/10.1057/s41308-021-00131-7>
- Catão, L. A. V., & Milesi-Ferretti, G. M. (2014). External liabilities and crises. *Journal of International Economics*, 94(1), 18–32. <https://doi.org/10.1016/j.jinteco.2014.05.004>
- Chang, R., & Velasco, A. (2001). A model of financial crises in emerging markets. *Quarterly Journal of Economics*, 116(2), 489–517. <https://doi.org/10.1162/00335530151144087>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. <https://arxiv.org/abs/1406.1078>
- Corsetti, G., Pesenti, P., & Roubini, N. (1999). What caused the Asian currency and financial crisis? *Japan and the World Economy*, 11(3), 305–373. [https://doi.org/10.1016/S0922-1425\(99\)00019-5](https://doi.org/10.1016/S0922-1425(99)00019-5)
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning (pp. 233–240). ACM. <https://doi.org/10.1145/1143844.1143874>
- Demirgüç-Kunt, A., & Detragiache, E. (1998). The determinants of banking crises in developing and developed countries. *IMF Staff Papers*, 45(1), 81–109. <https://doi.org/10.2307/3867330>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Du, X., Li, W., & Ma, Z. (2022). Prediction of credit risk in emerging markets via LSTM with transfer learning. *Finance Research Letters*, 46, 102411. <https://doi.org/10.1016/j.frl.2021.102411>
- Eichengreen, B., & Hausmann, R. (2005). Other people's money: Debt denomination and financial instability in emerging market economies. University of Chicago Press.

- 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>
- Eichengreen, B., Rose, A. K., & Wyplosz, C. (1996). Contagious currency crises. NBER Working Paper, 5681. National Bureau of Economic Research.
- Flood, R., & Garber, P. (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](https://doi.org/10.1016/0022-1996(84)90002-3)
- Forbes, K. J., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, 57(5), 2223–2261. <https://doi.org/10.1111/0022-1082.00494>
- 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](https://doi.org/10.1016/S0022-1996(96)01441-9)
- Frankel, J. A., & Saravelos, G. (2012). Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. *Journal of International Economics*, 87(2), 216–231. <https://doi.org/10.1016/j.jinteco.2011.12.009>
- Girton, L., & Roper, D. (1977). A monetary model of exchange market pressure applied to the postwar Canadian experience. *American Economic Review*, 67(4), 537–548.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen [Diploma thesis]. Technische Universität München.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- 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/10.1080/14697688.2017.1365136>
- International Monetary Fund. (2014). Annual report on exchange arrangements and exchange restrictions. IMF.
- Kaminsky, G., Lizondo, S., & Reinhart, C. (1998). Leading indicators of currency crises. *IMF Staff Papers*, 45(1), 1–48. <https://doi.org/10.2307/3867328>
- 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>
- Laeven, L., & Valencia, F. (2018). Systemic banking crises revisited. IMF Working Paper, WP/18/206. International Monetary Fund.
- Lanbouri, Z., & Achchab, S. (2020). Currency crisis prediction using deep learning: Comparative analysis. *Procedia Computer Science*, 170, 790–795. <https://doi.org/10.1016/j.procs.2020.03.107>
- Masson, P. (1999). Contagion: Macroeconomic models with multiple equilibria. *Journal of International Money and Finance*, 18(4), 587–602. [https://doi.org/10.1016/S0261-5606\(99\)00036-4](https://doi.org/10.1016/S0261-5606(99)00036-4)
- Milesi-Ferretti, G. M., & Razin, A. (1998). Current account reversals and currency crises: Empirical regularities. NBER Working Paper, 6620. National Bureau of Economic Research.
- Moreno, R. (1999). Depreciation and recessions in East Asia. FRBSF Economic Review, 3, 27–40.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>

- Obstfeld, M. (1994). The logic of currency crises. *Cahiers Économiques et Monétaires*, 43, 189–213.
- Özatay, F. (2020). Turkey's quarrel with orthodox monetary policy: An assessment. *Insight Turkey*, 22(4), 53–74.
- Radelet, S., & Sachs, J. D. (1998). The East Asian financial crisis: Diagnosis, remedies, prospects. *Brookings Papers on Economic Activity*, 1998(1), 1–90. <https://doi.org/10.2307/2534670>
- Reinhart, C. M., & Rogoff, K. S. (2009). This time is different: Eight centuries of financial folly. Princeton University Press.
- Reinhart, C. M., & Rogoff, K. S. (2011). From financial crash to debt crisis. *American Economic Review*, 101(5), 1676–1706. <https://doi.org/10.1257/aer.101.5.1676>
- Rose, A. K., & Spiegel, M. M. (2012). Cross-country causes and consequences of the 2008 crisis: International linkages and American exposure. *Pacific Economic Review*, 17(3), 340–363. <https://doi.org/10.1111/j.1468-0106.2012.00585.x>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Tiwari, A. K., Abakah, E. J. A., & Agayemi, C. (2022). Predicting currency crises: The role of machine learning. *Applied Economics*, 54(37), 4263–4285. <https://doi.org/10.1080/00036846.2022.2033471>
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762. <https://arxiv.org/abs/1706.03762>