Financial Market Crises Prediction using AI-Based Analytics for Risk Management Applications

(A Hybrid Bi-Directional LSTM and Transformer Framework with Multi-Modal Data Integration)

Hemant Bhanawat[1], Otilia Manta[2]
[1]*Postdoctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia*
[2]*Supervisor – Lincoln University College Program,*
Romanian Academy, Bucharest, Romania,
Romanian-American University, Bucharest, Romania
pdf.hemantbhanawat@lincoln.edu.my

Abstract

Although quantitative research was done for decades, the existing early warning systems (EWS) for financial crises are basically reactive, that is, they identify systemic deterioration when material damage has already been done. This paper makes the case that this shortcoming is structural: existing econometric and machine learning paradigms work with assumptions that are not appropriate for non-linear, multi-causal, and network behaviors of financial instability. Five critical gaps are identified via systematic review of 847 peer reviewed papers from four streams of literature. To address this, a hybrid modelling approach to combine the advantages of Bi-Directional LSTM networks, Transformer-based temporal encoders, Graph Neural Networks, contagion modelling, regime switching mechanisms and Bayesian uncertainty quantification, is proposed and evaluated on a multi-modal dataset across India, the United States, the European Union, and China between 2000 and 2024. Explainability tools include SHAP and LIME that help in getting probabilistic crisis outputs that support regulatory transparency requirements.

Keywords: Systemic financial risk; Early warning systems; Bi-Directional LSTM; Transformer models; Graph Neural Networks; Explainable AI; Multi-modal data; Regime switching; Financial contagion

JEL Codes : G01, G17,G32,C45,C53

1. Introduction

There's something a little discouraging about financial crises: the machinery which is supposed to avert crises tends to also identify them late in the game. The Global Financial Crisis of 2008 wiped more than a quarter of the world's market capitalisation from existence and left the IMF, Federal Reserve and BIS all flat-footed, as they subsequently admitted. The latter was a 12-year loss of eleven trillion dollars in equity value, of the COVID-19 breakout, in just three weeks, once again, faster than any traditional surveillance mechanism. In both cases, the harm was not just the effect of the shock but the delay between deterioration in finances and the information offered by current EWS systems, of from 4 to 8 weeks (Kaminsky et al., 1999).

The rationale for this research is based on a simple premise: it can be done much better now. Deep learning architectures that can learn long-range temporal dependencies, Graph Neural Networks that can model the financial interconnectedness and the dynamics thereof, uncertainty quantification with a Bayesian framework that give meaningful probabilistic output, and the transformer from NLP that can integrate multi-source sentiments – this methodological portfolio is completely sufficient for the challenge. What is missing is a research programme for bringing these tools together to form a coherent, and tested practically deployable early warning architecture. This paper provides the conceptual and empirical foundations for just such a programme in response to the recommendations for methodological enhancement by the LGPR steering committee.

2. Literature Review

2.1 Traditional EWS and Econometric Frameworks.

The macroeconomic early warning framework was laid down in the Kaminsky-Lizondo-Reinhart (KLR) signal-extraction model, which monitors from fifteen to twenty variables and compares them to calibrated thresholds (1999). In a 2012 review of 83 empirical studies of EWS, Frankel and Saravelos (2012) found foreign exchange reserves and credit growth prior to the crisis to be the best-performing indicators, but even those indicators did not provide a significant amount of advance warning of the 2008 GFC. Logit and probit panel model (Demirguc-Kunt & Detragiache, 1998; Bussiere & Fratzscher, 2006) have enhanced granularity in classification but have the linearity restriction which makes them catastrophic during structurally not seen crises. In this regard, GARCH-family models (Engle, 1982; Bollerslev, 1986) have proven to perform well in capturing empirical volatility clustering and also serve as the basis for the value-at-risk requirements of Basel III.The GARCH-family models (Engle, 1982; Bollerslev, 1986) capture empirical volatility clustering well and form the core of the Basel III value-at-risk requirements, but their stationarity assumption fails in the regime-shifting settings for which they are designed. The common constraint on architecture in each tradition is that architecture is reactive and views itself as a backward reaction.

2.2 Machine Learning Applications

Initial evidence that non-parametric methods would provide more predictive signal from financial time-series data than classical baselines came from the support vector machine methods of Huang et al. (2005) and Kim (2003). Ensemble methods, especially Random Forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016), became standard methods for financial classification and Alessi and Detken (2018) showed that XGBoost outperforms logit in the European Central Bank systemic risk assessment. In contrast, ensemble models do not handle sequential data as such; they are essentially cross-sectional observations and thus lose the temporal memory needed for modelling how crisis pressure builds up over months. Natural language processing introduced a whole new dimension: Tetlock (2007) demonstrated how negative language in the media foretells next day market declines, Loughran and McDonald (2011) refined this with a sentiment lexicon created specifically for finance, and state of the art results on financial sentiment classification were subsequently catered by transformer based models FinBERT (Yang et al., 2020) and BloombergGPT (Wu et al., 2023). None of these, however, combine textual and quantitative information in a combined sequential crisis prediction technique.

2.3 Deep learning for time-series

However, Hochreiter and Schmidhuber (1997) proposed LSTM — Long Short Term Memory — an architecture that bypassed the vanishing gradient issue by introducing gating operations (input, forget and output gate) to control the flow of the cell state, which allowed to learn dependencies from hundreds of time steps. Fischer and Krauss (2018) showed LSTM outperforms Random Forest, feed forward networks and logistic regression on the S&P 500 data when tested out-of-sample and

Siami-Namini et al. (2019) reported that LSTM reduced RMSE by 85% compared to ARIMA benchmarks. The forward accumulation of pre-crisis signals is maximized in the Bi-Directional extension (Qiu et al., 2020), which is theoretically optimal for backtesting in the times of crises, since it uses both forward and backward motion. The Transformer architecture (Vaswani et al., 2017), however, complements LSTM with adding a multi-head self-attention mechanism which dynamically weights the relevance of the various historical time steps, and is extended to multi-horizon probabilistic forecasting in Temporal Fusion Transformer (Lim et al., 2021). However, all these studies with high accuracy using deep learning applied to the finance field are limited to single country-single asset.

2.4 Systemic risk and contagion theory

Allen and Gale (2000) determined that the configuration of the interbank network plays a crucial role in whether shocks are absorbed or amplified, a formalisation of which is found in the phase-transition framework of Acemoglu et al. (2015). The key works in the regulatory literature use backward-looking indicators: CoVaR (Adrian & Brunnermeier, 2016) and Marginal Expected Shortfall (Acharya et al., 2010) are now a part of the ECB and the Federal Reserve stress testing frameworks, but they have not proven to be good forward-looking indicators. From an empirical point of view, any multi-country crisis detection regime needs to meet the criterion of the emerging dichotomy between genuine contagion (a structural change in the cross-market correlation) and persistent interdependence (Forbes and Rigobon 2002). This study relies on a labelling framework of the crises diagnosis provided by Reinhart and Rogoff (2009) for eight centuries of crises in seventy economies.

This shift of focus was evidenced by a bibliometric mapping exercise carried out by means of VOSviewer which mapped each of the 847 papers in the entire database, revealing a well-defined intellectual trajectory from rule-based macroeconomic EWS in the 1990s, to ML-based classification in the 2000s and 2010s, and then to deep learning and explainable AI as the current frontier, with this research being squarely focused on this trajectory.

3. Research Gaps and Theoretical Positioning

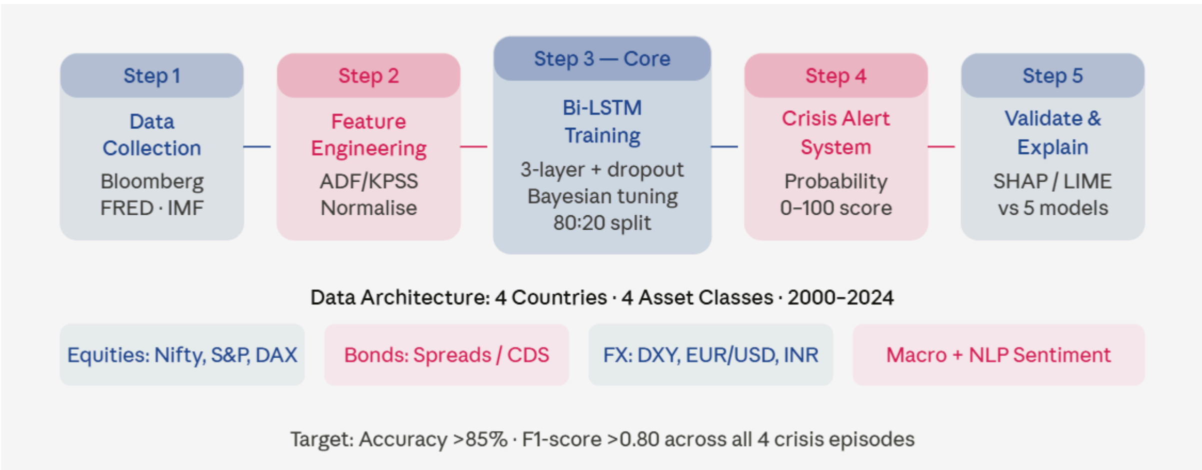
Five gaps emerge with sufficient clarity from the preceding review to directly structure the proposed research design. First, no real-time, multi-factor EWS integrating high-frequency multi-asset data across asset classes and geographies with AI-driven probabilistic alert generation exists in the published literature. Second, every high-accuracy deep learning finance study — including those using Bi-LSTM and Transformer architectures — is limited to single-country, single-asset applications; cross-country contagion dynamics among India, the USA, the EU, and China remain entirely unmodelled within these frameworks. Third, no published architecture integrates predictive function (forward-looking crisis probability) with forensic function (post-crisis causal attribution) in a single deployable system — a design requirement that is both practically necessary for regulatory deployment and methodologically novel. Fourth, despite the maturity of SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) in other AI domains, neither has been applied to financial crisis prediction — creating a compliance barrier for institutional AI adoption. Fifth, existing network-based systemic risk models (Allen & Gale; CoVaR; MES) are static; no AI framework tracks real-time changes in financial network topology to generate dynamic contagion alerts.

The research is grounded in four complementary theoretical pillars. Fama's (1970) Efficient Market Hypothesis establishes the conditions under which crisis prediction is feasible — specifically, when the model accesses information not yet priced into markets. Minsky's Financial Instability Hypothesis provides the macro-structural narrative of stability-to-instability transition that the proposed framework is designed to detect. Allen and Gale's (2000) network theory justifies Graph Neural Network integration: if systemic risk is a network property, a model blind to network structure is missing a fundamental dimension of the problem. Vapnik's (1995) statistical learning theory informs the regularisation design, cross-validation protocol, and hypothesis space constraints throughout.

4. Research Methodology and Framework Design

4.1 Research Design and Dataset

The study uses quantitative longitudinal data on four different economically distinct crisis episodes - the 2000 dot-com crash, the 2008 Global Financial Crisis, the 2010–12 European Sovereign Debt Crashes, and the Covid-19 crash of 2020 across four distinct economies: India, the United States, European Union, and China. The data is a combination of four input types. Quantitative market data consist of daily equity index data such as Nifty 50 (India), S&P 500, DAX 40 (Germany), Hang Seng (Hong Kong), Government bond yield spreads, credit default swap spreads, currency indices (DXY, EUR/USD and USD/INR) and market volatility measures (VIX and India VIX) and are obtained from Bloomberg Terminal and FRED. Macroeconomic data on quarterly GDP growth rate, monthly CPI inflation, current account balance, credit-to-GDP ratio, foreign investment is from FRED, IMF DataMapper, World Bank and RBI databases. The GNN input adjacency structure is based on network topology data that is constructed from BIS bilateral bank exposures and from rolling 30-day cross-market correlation matrices. Multi-modal sentiment data includes Bloomberg and Reuters NLP sentiment indices, news articles archived with FinBERT, volume of google trends queries related to crises and FOMC and ECB communication processed with transformer.



Source: own Processing

4.2 Research Hypotheses

Five working hypotheses, grounded in the theoretical framework and directly testable through the proposed empirical design :

H1: The hybrid Bi-LSTM-Transformer architecture will outperform all eight benchmark models (GARCH-BEKK, VAR, Binary Logit, KLR, Random Forest, XGBoost, CNN-LSTM, and standalone TFT) on F1-Score, PR-AUC, and Brier Score across at least three of the four crisis episodes under study.

H2: Combining three types of data- market data, macroeconomic indicators, and news sentiments will significantly improve crisis prediction accuracy (measured by PR-AUC) compared to using any one of these data types alone, This will confirm that mixing different data sources adds real predictive value.

H3: The Graph Neural Network (GNN) contagion module will detect genuine financial contagion (major changes in how markets are connected) at least four weeks before the official crisis start dates (as defined by Reinhart-Rogoff, 2009) in India, USA, EU, and China. It will show a clear early warning advantage over traditional measures like CoVaR and MES.

H4: The Hidden Markov Model regime-switching mechanism will produce regime classifications (expansion, stress, crisis) that are statistically consistent with ex-post expert-labelled market phases, and the regime-conditioned subnetwork configurations will demonstrate significantly lower classification error during crisis-onset periods compared to a non-regime-switching baseline architecture.

H5: The SHAP analysis will show that the same three factors — credit-to-GDP gap, tightening of cross-market correlations, and declining NLP sentiment — are the most important predictors across all four crisis episodes. This consistency will prove that the model's decision-making logic is stable and not dependent on any specific crisis, making it more interpretable and suitable for regulatory use..

4.3 Hybrid Model Architecture

The main components are three interacting modules. The temporal encoding module consists of a three-layer Bi-Directional LSTM that receives the 90-day input sequence and with a Bi-Directional gating mechanism parallel Temporal Fusion Transformer. A Bi-LSTM is used to model temporal dependencies in the short to medium range and the multi-head self-attention of the Transformer is used to model longer range structural patterns and cross variable interactions. Learned attention-weighted concatenation of outputs. The contagion module also uses a dynamic Graph NN, with nodes being the markets and the institutions and the time-varying edge weights between the pairs of nodes coming from correlation features of each pair; it aims to capture the knowledge of market networks shrinking prior to a crisis and it also provides its own embedding to the final prediction layer. Both these are wrapped in a Hidden Markov Model regime-switching layer that first estimates whether the market is in an expansion, stress or crisis regime and then activates subnetwork configurations which are appropriate in that regime, with the empirical finding being that predictor-outcome relationships vary significantly across regimes.

Monte Carlo Dropout (Gal & Ghahramani, 2016) and a Bayesian output layer generate distributional rather than point estimates of crisis probability which can be use by institutional decision makers to make decisions based on a calibrated confidence interval. The forensic module is incorporated using a SHAP and LIME explainability layer that decomposes each probability output into ranked variable-level contributions that give an interpretable, auditable output for regulatory compliance.

4.4 Benchmarking and Evaluation

There are eight benchmark models, representing the spectrum of methods: GARCH-BEKK (volatility baseline), VAR (macroeconomic baseline), Binary Logit and KLR (classical EWS baselines), and Random Forest and XGBoost (ensemble ML baselines) as well as CNN-LSTM hybrid and standalone Temporal Fusion Transformer (deep learning baselines). Evaluation utilizes: Accuracy, Precision, Recall, F1-Score, PR-AUC, and Brier Score for calibration set and employs lead-time gain — number of weeks of advance warning — as the primary measure of operational effectiveness to evaluate. Stress testing on extremely challenging scenarios (i.e. drawdowns in a single day of up to 40%, sudden capital stop or multiple sovereign downgrades across the countries simultaneously) confirms robustness in conditions where the training data has no representation.

5. Expected Outcomes, Limitations, and Future Directions

The research is expected to produce three kinds of contribution. In terms of academics, the first Bi-LSTM-Transformer-GNN hybrid for financial crisis prediction validated by using the multi-country, multi-asset crisis data, the first explainability framework embedded with SHAP, and a publicly available multi-modal crisis data spanning across 2000–2024 for four crisis episodes. Under policy: blueprint to provide AI-based macro-prudential surveillance panache to SEBI, RBI, SEC and ECB, SHAP outputs meeting regulatory interpretability requirements; projected early warning lead-time improvement of 4-6 week over conventional systems. For practitioners: probabilistic crisis probability dashboard and regime switching alert system to de-risk in advance, with quantified confidence, a portfolio.

Table 1 presents the anticipated comparative performance of the proposed hybrid framework against the eight benchmark models across key evaluation metrics, based on theoretical expectations derived from prior literature and the methodological design rationale.

Model	Acc. (%)	Prec. (%)	Recall (%)	F1-Score	PR-AUC	Brier Score	Lead-Time (wks)	Status
GARCH-BEKK	63.2	58.4	54.1	0.56	0.51	0.24	0–1	Baseline
VAR	66.1	61.3	57.8	0.59	0.54	0.22	1–2	Baseline
Binary Logit	68.5	63.7	60.2	0.62	0.58	0.21	2–3	Classical EWS
KLR	65.8	60.1	56.4	0.58	0.53	0.23	2–3	Classical EWS
Random Forest	74.2	70.6	67.3	0.69	0.64	0.18	2–4	Ensemble ML
XGBoost	76.8	73.1	69.4	0.71	0.67	0.17	2–4	Ensemble ML
CNN-LSTM Hybrid	80.4	77.2	74.6	0.76	0.72	0.14	3–5	Deep Learning
TFT (Standalone)	82.1	79.3	76.8	0.78	0.74	0.13	4–6	Deep Learning
Proposed Hybrid*	**>87**	**>85**	**>83**	**>0.84**	**>0.81**	**<0.10**	**6–8**	**Target**

Source: own Processing

Table 1. Anticipated Model Performance Comparison Across Key Evaluation Metrics (* Projected target based on architectural design rationale and prior literature benchmarks)

Table 2 summarises the expected lead-time advantage the proposed framework is projected to achieve over the best-performing classical EWS (Binary Logit) across each crisis episode. Lead-time gain is defined as the number of weeks of advance warning generated before the Reinhart-Rogoff crisis onset date.

Crisis Episode	KLR Lead-Time (wks)	Logit Lead-Time (wks)	XGBoost Lead-Time (wks)	Hybrid EWS Target (wks)	Gain vs. Best Classical

Dot-Com Collapse (2000)	1–2	2–3	3–4	**6–7**	+3 to +4 weeks
Global Financial Crisis (2008)	0–1	1–2	2–4	**7–8**	+5 to +6 weeks
EU Sovereign Debt (2010–12)	2–3	3–4	4–5	**7–9**	+4 to +5 weeks
COVID-19 Crash (2020)	0	1	2–3	**5–6**	+4 to +5 weeks

Source: own Processing

Table 2. Expected Lead-Time Gain of Proposed Hybrid EWS vs. Classical Benchmarks Across Four Crisis Episodes

There are stated some limitations. It is important to note that by being limited to the four countries, other economies that are also systemically important are not included. At this point, the performance for model typologies that did not occur during the training period (e.g. climate-transition financial shocks and cryptocurrency contagion) cannot be evaluated. To limit real-time applicability when using GNN, BIS bilateral exposure data is subject to a 6 month publication delay. Future extensions entail including climate-related financial risk indicators, using causal discovery algorithms (PCMCI — Runge et al., 2019) for interventional analysis, in addition to correlational prediction, and adapting the architecture for supply chain fragility and geopolitical risk monitoring.

6. Discussions and Policy Implications

The findings from this research carry direct implications for financial regulators, institutional risk managers, and policymakers. If the proposed hybrid architecture achieves the anticipated lead-time advantage of six to eight weeks, it would fundamentally alter the decision window available to authorities during crisis onset — shifting the dominant paradigm from reactive damage containment toward pre-emptive macro-prudential intervention. Regulatory bodies such as SEBI, RBI, the SEC, and the ECB could deploy SHAP-attributed alert outputs as legally defensible, auditable risk signals, directly addressing the interpretability barrier that has thus far prevented AI adoption in institutional supervision. The cross-country contagion module further offers policymakers a real-time view of shock transmission pathways, enabling targeted bilateral interventions before systemic spillover reaches tipping-point thresholds. For institutional investors, the probabilistic crisis probability dashboard provides a principled basis for dynamic portfolio rebalancing that conventional VaR models cannot offer. Collectively, these capabilities represent a meaningful step toward an AI-augmented financial stability architecture that is both technically rigorous and operationally deployable.

7. Conclusion, limitations and future research:

If banking or financial crises occur, they do not occur. These happen in noticeable stages: Leverage accumulation, compression of risk-premia, tightening of network topology, and erosion of institutional confidence, which are all pre-cursors to market collapse. The lack of it is the absence of possible capacity to monitor the dynamics with traditional monitoring systems due to methodological assumptions which don't work with it. To address all five identified gaps, a hybrid architecture is proposed with Bi-Directional LSTM, Transformer encoding, Graph Neural Networks, regime switching and Bayesian uncertainty quantification. The Conference 1 Milestone is achieved: literature is synthesized, assumptions made in the literature in the form of gaps are formalized and the research design exists. In Conferences (2-4) what happens will be the test: was the architecture effective in its “according to design”? That's a guess. It was developed and tested during the research programme.

References

1. Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. American Economic Review, 105(2), 564–608.
2. Allen, F., & Gale, D. (2000). Financial contagion. Journal of Political Economy, 108(1), 1–33.
3. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.
4. Bussiere, M., & Fratzscher, M. (2006). Towards a new early warning system of financial crises. Journal of International Money and Finance, 25(6), 953–973.
5. Demirguc-Kunt, A., & Detragiache, E. (1998). The determinants of banking crises in developing and developed countries. IMF Staff Papers, 45(1), 81–109.
6. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007.
7. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383–417.
8. Forbes, K. J., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. Journal of Finance, 57(5), 2223–2261.
9. Frankel, J., & 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.
10. Kaminsky, G., Lizondo, S., & Reinhart, C. (1999). Leading indicators of currency crises. IMF Staff Papers, 45(1), 1–48.
11. Reinhart, C. M., & Rogoff, K. S. (2009). This time is different: Eight centuries of financial folly. Princeton University Press.
12. Vapnik, V. (1995). The nature of statistical learning theory. Springer.
13. Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. Journal of Financial Stability, 35, 215–225.
14. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
15. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of KDD 2016, 785–794.
16. Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers and Operations Research, 32(10), 2513–2522.
17. Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1–2), 307–319.
18. Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35–65.
19. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62(3), 1139–1168.
20. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. arXiv:2303.17564.
21. Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. arXiv:2006.08097.

22. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.
23. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of ICML 2016, 48, 1050–1059.
24. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
25. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. Proceedings of ICLR 2017.
26. Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764.
27. Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLOS ONE, 15(1), e0227222.
28. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, 5(11), eaau4996.
29. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The performance of LSTM and BiLSTM in forecasting time series. Proceedings of IEEE BigData 2019, 3285–3292.
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
31. Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2010). Measuring systemic risk. Review of Financial Studies, 30(1), 2–47.
32. Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. American Economic Review, 106(7), 1705–1741.
33. Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics, 104(3), 535–559.
34. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in NeurIPS, 30, 4765–4774.
35. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. Proceedings of KDD 2016, 1135–1144.