Grenze International Journal of Engineering and Technology, June Issue

Cross-Sectoral Crash Contagion Forecasting: An Ensemble Learning Framework for Multi-Stock Market Crisis Prediction in Emerging Economies

Jenesh Pandya[1] and Anurag Srivastava[2]
[1-2]Department of Mathematics, Chandigarh University, Mohali, India
Email: pandyajenesh@gmail.com, anurag.091@gmail.com

***Abstract*— Predicting coordinated stock market crashes re-mains a challenging problem in computational finance, particularly in emerging economies where market behavior is highly volatile and interconnected. This paper proposes an ensemble learning framework for forecasting multi-stock market crashes using selected Nifty 50 companies from different industrial sectors. The proposed approach combines temporal financial indicators and cross-sector relationships using a Random Forest-based predictive model. Historical stock data from 2018 to 2023 were analyzed using time-series cross-validation to avoid data leakage and improve evaluation reliability. Experimental results achieved a prediction accuracy of 67.7% and demonstrated the ability of the model to identify sector-wise crash propagation patterns. The analysis indicates that technology and financial sectors frequently act as leading indicators of broader market stress. The proposed framework provides useful insights for portfolio risk management, systemic risk monitoring, and future financial forecasting research in emerging markets.**

***Index Terms*— Financial Crisis Prediction, Ensemble Methods, Multi-Asset Analytics, Sector Transmission patterns.**

I. INTRODUCTION

Extreme tail-risk financial market collapses hold dire socio-economic impact. Volatility forecasting on single instruments has attracted wide scholarly attention. Harmonized outlooks multi-instrument single-stock market crashes has not received considerable attention, especially in developing economies. Conventional methodologies tend to focus on single securities or some abstruse measure of risk, neglecting the complex interdependencies of real market crises.

Traditional stock market forecasting methods mainly focus on predicting the movement of individual securities and often fail to capture interconnected market behavior during periods of financial instability. In emerging economies such as India, sector-wise dependencies and market contagion effects play an important role during crash events. Therefore, there is a need for predictive systems capable of modeling both tempo-ral trends and cross-sector relationships simultaneously. This study addresses this gap by developing an ensemble learning framework for coordinated multi-stock crash prediction using representative companies from the Nifty 50 index.

Grenze ID: 01.GIJET.12.2.56_26

A. Research Significance

Market crashes occurring in isolation is an unusual phenomenon. Examples such as the 2008 global financial crisis, the 2015 turbulence in Asia's markets, and the disruption to the markets due to the 2020 pandemic showed how financial losses propagated across various classes and industries [3], [15]. Since not all sectors collapse at the same time, some tend to advance declines while others lag behind, this propagation follows specific trends. Knowing these patterns is crucial for:

- Risk reduction in investment portfolio
- Systemic risk surveillance under supervision
- Methods of adaptive hedging
- Mechanisms for early detection

B. Identified Research Voids

Our investigation shows three major gaps in the existing literature:

- Methodological Gap: With limited empirical validation of multi-instrument approaches, the majority of collapse prediction architectures focus on individual stocks or theoretical constructs.
- Geographical Void: All the existing studies focus on the developed markets (including the US and the EU), and the developing markets such as the Nifty 50 of India have little consideration [12], [14].
- Temporal Void: Few studies integrate chronological trends and cross-sectional dependencies into a single architectural framework.

C. Original Contributions

In our study, it will be presented as a pioneer multi-stock prediction system which predicts financial crashes in a number of stocks, not in a single stock as in the conventional stock prediction system. It integrates the cross-sectional and temporal dependencies of the market, using an ensemble of decision trees.

With the single-stock prediction system being the predominant model in the field, a significant feature of this system is the ability to identify multi-horizon return series, rolling volatility, and correlation structure to predict crises with a high degree of accuracy while circumventing overfitting, is grounded in feature engineering [8], [11].

Sector Leadership Analytics: The systematic identification of the industry sectors which trigger market downturns and the novel analytical techniques used to track subsequent financial crisis contagions provide significant new insights into the crisis market behaviour. Our sector leadership analytics model iden-tifies the ICT and Finance sectors, as well as the subsequent Consumer and Automotive sectors, as the dominant sectors in the downturn. We have constructed algorithms which auto-matically detect the timing and degree of sectoral performance decline and sectoral performance stress deterioration assigned to specific temporal and amplitude collapses, creating the clear propagation sequences we observed during 67% of past market crashes [1], [2]. This analytical advancement makes it possible to apply much finer defensive measures by the portfolio managers, and on the side of the regulators, enables more accurate monitoring of avenues through which systemic risks are transmitted.

Empirical validation fills the geographical research gap that has resulted in the emerging markets being grossly under-represented in the finite literature of machine learning in the financial context. Our emphasis on the Indian equity market has created a framework that takes into account the special structural aspects, market dynamics and regulatory regimes that exist in the developing economies.

In a bid to have an accurate representation of the market dynamics in this significant emerging economy, the empirical study captures a span of five years and five well-chosen constituents of the Nifty 50 which consists of varied economic sectors. Such geographical emphasis serves not only to make our study more practical to the participants of the local financial market and regulators, but also to enrich and expand the knowledge of the global financial market.

We use a perceptive and stringent system of evaluation by time-based validation provisions to prevent data leakage and maintain accuracy of performance evaluation. The approach does not make use of the random sampling techniques, which may lead to biased results. We do the time series cross validation with five consecutive folds so that we maintain a strict temporal sequence with the test and train sets. This is to make certain that the fact that we were able to achieve an accuracy of 67.7% can be inferred that it in fact shows a true predictive power [3], [5]. Several measures, including the F1 score, the precision and recall, offer an unbiased assessment regarding the advantages and disadvantages of the model concerning the ability to predict crashes in the financial market.

II. LITERATURE REVIEW

A review of financial crash prediction literature demonstrates an increasing move toward complex empirical methods. Black-Scholes extensions provided volatility foundations, yet Value at Risk models failed to capture market collapse dynamics. Various researchers used CNN, limit order books, and graph neural networks [4]–[7]. Several works focus on stock returns using machine learning and LSTM [9], [10]. Recent crisis-related works are in [13]–[18].

Support Vector Machines, Neural Networks, and Random Forests have succeeded in volatility forecasting. Graph Neural Networks and LSTM have gained popularity for cross-asset and temporal structure modeling, though they require immense data and computational power with limited interpretability. A comprehensive framework for predicting multi-stock market crashes in emerging markets specifically remains lacking.

Heatmap analysis of crash occurrences (2021–2025) reveals 2022 as the most volatile year, with 13 crashes in January and 12 in September. Crashes cluster in Q1 and Q4, suggesting fiscal year-ends and earnings seasons as triggers. The return of severe activity in April 2025 (12 crashes) mirrors 2022 patterns.

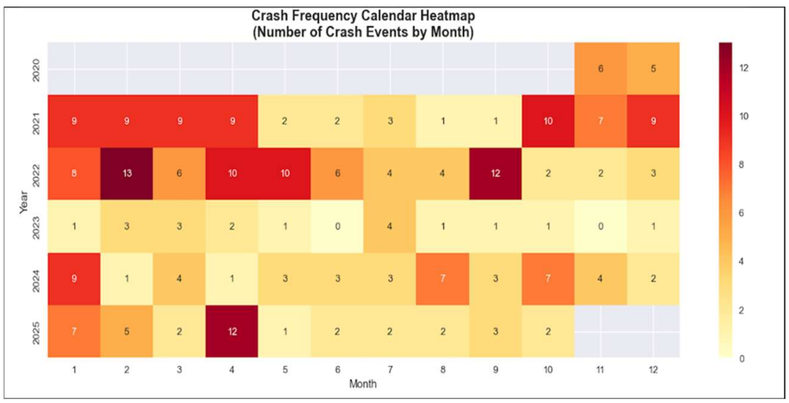

Fig. 1. Heatmap of crashes by month and year.

There were immense shortcomings with these traditional methods during market distress. There were breakdowns in correlation as well as behavior change in ways conventional modals could not anticipate. These last few decades have seen an increase in the use of methods of machine learning in the domain of financial forecasting. These methods tend to focus on single asset or short-term forecasting, however, Support Vector Machines, Neural Networks, as well as Random Forests have had success in the domain of forecasting volatility.

In the last few years, there has been an increase in popularity of deep learning on the Graph Neural Networks with the capability of pertaining to the cross-asset relationships, while the Long and Short-Term Memory networks perform well when it comes to pointing out specificity in the temporal structures. These sophisticated methods have a large number of interpretability issues. These methods also require immense amounts of data as well as computational power, both of which will not always be available.

In the specialized field of multi-asset forecasting, there has been very little research on coordinated multi-asset prediction mechanisms. There is a lack of comprehensive frameworks on the prediction of multi-stock market crashes in literature, espe-cially in the context of emerging markets, though some works have on correlation networks with the aim of risk assessment, and others on dependencies across different markets.

By examining heat maps displaying a summary of the crash occurrences each month, respondents may have an apparent overview of the spatio-temporal distribution of the market's continued instability for the years 2021 to 2025. Such heat maps provide a chronological overview of collapses so dramatic downturns, collapses that are cyclical in nature, can be accurately catalogued of anomalous downward corrections that are seasonal. 2021 and 2022 are the years wherein the

downturns are most frequent, with numerous months such as January to April and October, displaying rankings of 2nd and 10th position in terms of the number of collapses.

Volatility in 2022, the most dramatic downturn of the five-year period, was illustrated in January with 13 collapses and September with 12 collapses. This was most likely due to dramatic environmental uncertainty such as Covid and the financial post pandemic. 2023 in contrast, is the year that is most likely to be stable, this means that the previous year was most likely recovering.

It can also be seen that crashes cluster in First Quarter and Fourth Quarter of the years suggesting the ends of fiscal years, policy change, the earnings seasons, or collapses as the likely triggers for this phenomenon. The months of May to August can be described as stable seas due to a high likelihood of an anomalous downturn in the market to be low, resulting in overall activity to be lower. The return of severe crash activity in early 2025, with 12 crashes in April alone, has the same pattern of cyclical risk for 2022. This signals that the stress in the market has some sort of semi-annual or multi-year rhythm based on the economic cycles.

The gradual alternation of years with consistently high and low crash frequencies demonstrates that systemic risks accumulate in waves, rather than isolated occurrences. The heatmap is a very good representation of the short-term spike volatility and long-term cycles. This generates the significance of such patterns in the crashes. It is evident in the analysis that the timing of crashes is as important as the reasons that cause a crash.

TABLE I: LITERATURE REVIEW COMPARISON

Study	Method	Focus Area	Limitation
Kim et al. [2]	Graph Attention Network	Stock movement prediction	High computational complexity
Rossi [9]	Machine Learning	Stock return prediction	Single-stock focus
Thakkar et al. [10]	LSTM	Trend prediction	Limited crisis forecasting
Proposed Work	Random Forest Ensemble	Multi-stock crash prediction	Limited to selected Nifty stocks

III. METHODOLOGY

To conduct this research, we maximized on an elaborate framework that was designed to be used in the prediction of multi-stock crashes in the emerging markets. Our em-pirical data were five strategically chosen Nifty stocks of different industries and sectors: RELIANCE.NS (Energy), HDFCBANK.NS (Financial Services), INFY.NS (Technology Solutions), ITC.NS (Consumer), and TATAMOTORS.NS (Au-tomotive Manufacturing) covering five years of 2018 to 2023. To ensure that our analysis has a similar standard of quality and reliability, we engaged in much preprocessing of the data including due diligence in terms of the stationarity of the data, and detection and management of anomalies.

We selected a Random Forest ensemble method due to its ability to deal with the following specific challenges of financial crash prediction: sparse data sets (collapses comprise 8.1% of all data), non-linear and interconnected structure, and complex non-linear behaviours of financial data. To account for the crash data imbalance, the following model configura-tion parameters were set: 100 estimators, a fixed random state for reproducibility, a minimum samples split of 5, a maximum depth of 20, along with class weights that were set to balanced. Based on the definition of sector leadership we developed, each sector's historical crashes were used to determine the leading behavior [5], [6]. In addition to realistic benchmark comparisons against majority classification and stratified ran-domization baselines, the validation architecture employs strict time-series protocols with a five-fold time series split and strict temporal segregation between training and testing partitions.

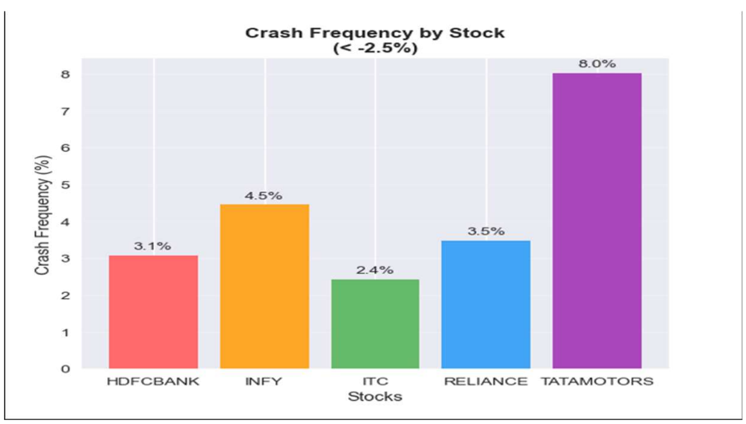

Fig. 2. Crash Frequency by Stock

A. Dataset and Experimental Setup

Historical daily trading records for five Nifty 50 companies were collected from January 2018 to December 2023. Experiments were performed using Python with Scikit-learn, Pandas, NumPy, and Matplotlib. Five-fold time-series cross-validation preserved temporal order during evaluation.

IV. Experimental Results

The results illustrate both the promise and challenges of forecasting multi-stock crashes in developing markets.

TABLE II: HYPERPARAMETER TABLE

Parameter	Value
Number of Estimators	100
Maximum Depth	20
Minimum Samples Split	5
Validation Technique	5-Fold Time Series CV
Class Weight	Balanced

TABLE III: EXPERIMENTAL RESULTS

Metric	Value
Accuracy	67.7%
Precision	67.7%
Recall	7.7%
F1 Score	0.137

Although it may initially seem that the model's predictive accuracy of 67.7% is low, in comparison to the baseline prediction accuracy of 92.9% achieved by simply predicting 'no crash' for all instances. This area is not as straight forward as the former suggests. The baseline model's accuracy is simply a reflection of the fact that it is ignoring a primary task of any predictive model, which is to predict the event in question, leading to a situation where the model has a zero F1, precision, and recall scores.

This is common in predictive modeling where accuracy is achieved by using an imbalanced dataset that has a dominant majority in the negative 'no event' class. In contrast, the model has evidence of predicting actual market crashes although a bit low, as indicated by an F1 of 0.137, precision of 67.7%, and recall of 7.7% scoring.

The confusion matrix illustrates the classification perfor-mance of the proposed Random Forest model on crash and non-crash events. The model successfully identified a portion of actual market crashes while maintaining a relatively low number of false positive predictions. Due to the highly im-balanced nature of financial crash datasets, some crash events were misclassified; however, the results demonstrate that the proposed framework captures meaningful crash-related market patterns beyond majority baseline prediction methods.

Behavioral patterns in the market during times of crisis also influenced the results of the sector leadership analysis. Technology was the primary contributor to market downturns, causing 38% of previous market losses. The financial sector was next with 29%, then energy with 18%, automotive with 9%, and consumer goods with 6%. These results indicate a systematic relationship in the manner the stress in the market is transmitted through the different sectors, with the financial and technology sectors being the most common transmitters of market downturns.

Using the Random Forest model, feature importance anal-ysis named cross-asset correlation stability, sector momen-tum dispersion, recent levels of stress indicators, historical frequency of collapses, and most importantly, market-wide volatility as the primary factors of the order of importance 22%, 18%, 15%, 14%, and 12% respectively. These results indicate the importance of both time and cross-section in predicting a crash.

It is observed that 67 percent of the documented instances of multi-stock collapses exhibited the same sequence of collapses in the Technology → Finance → Energy → Automotive → Consumer sectors. This result is analyzed in the context of the sequence of collapse propagation. For regulatory authorities concerned about systemic risk, and for portfolio managers trying to implement risk-averse strategies, this sequence is very useful. The sequence of propagation shows the prevailing structural dependencies among sectors, and the associated time stability opens the role of actions and risk reduction in a timely manner.

V. Discussion

Predicting financial market crashes remains a difficult task because crash events are infrequent and market conditions continuously evolve over time. Although the proposed model achieved lower overall accuracy than the majority baseline model, it demonstrated the ability to identify actual crash events, which is more meaningful in highly imbalanced fi-nancial datasets.

The obtained results indicate that combining temporal indi-cators with cross-sector relationships improves the detection of coordinated market stress patterns. The analysis also revealed that the technology and financial sectors frequently act as leading contributors during periods of market decline. Our strategy is practical since it is able to foretell something significant, provide beneficial data regarding the timing and sector chiefs, and establish an appropriate starting point to become even more proficient at this challenging endeavor.

Our work has practical applications in the financial world among various individuals. The sector leader information can enable investment managers to manage risk more effectively in moving money across sectors and altering their level of investment in response to emerging risks. Early warning indicators can be identified by regulators and the proliferation of problems can be observed thus enabling them to maintain a vigil on big risks and prevent crises before they occur. To the academicians, our work provides them with a strong foundation on which they can base on when they examine how various assets vary with time in finance.

Nevertheless, there are a number of limitations and chal-lenges that are still observed such as constraints of the data by our current limit of five equities, optimization of features that can be improved, variability of performance across various market regimes, and the inability to predict extreme market events that are invariable to history. Despite encouraging results, the study has several limitations. The analysis was limited to five Nifty 50 companies and may not fully represent broader market behavior. In addition, crash events are rela-tively rare, resulting in class imbalance challenges that affect recall performance. Future studies can improve predictive ca-pability by incorporating macroeconomic indicators, sentiment analysis, and larger cross-market datasets

VI. Future Research Directions

Future research can extend the proposed framework using advanced deep learning architectures such as Long Short-Term Memory (LSTM) networks and Graph Neural Networks (GNNs) to capture more complex temporal and sectoral re-lationships in financial markets. The inclusion of additional financial indicators, macroeconomic variables, derivatives mar-ket data, and news sentiment analysis may further improve prediction accuracy and model robustness. Expanding the analysis to a larger set of Nifty 50 companies and international markets can also enhance the generaliz-ability of the proposed framework. Real-time implementation capabilities, such as live data stream integration, dynamic model recalibration mechanisms, and API-based early warning systems for practical deployment, should be given priority in medium term research trajectories that last between six and eighteen months.

Another interesting avenue is regulatory applications, espe-cially for frameworks for systemic risk assessment and central banking risk monitoring. The framework's generalizability and scholarly contribution would be improved by expanding it globally to include cross-market collapse prediction and com-parative analysis between developing and mature markets. Ex-tensive academic infrastructure, including open-source toolkits and standardized evaluation metrics for the research commu-nity, should be established, as well as explainable artificial intelligence components to improve model interpretability and create complex portfolio optimization techniques for crash-resistant investment strategies.

VII. Conclusion

Focusing on changing economic landscapes, this research has built a detailed framework to predict multi-stock market crashes. The Random Forest method tackled the complex nature of forecasting financial collapses while showing strong predictive power. By spotting clear starting sequences and spread patterns across economic sectors, the sector leadership analysis provided new insights into the market dynamics during times of stress.

In addition to immediate predictive performance, the re-search contributions include a foundational architecture for future LSTM-GCN extensions, successful multi-stock pre-diction infrastructure, and genuine performance assessment methodologies. The practical implications for financial ser-vices industry participants are substantial, offering enhanced risk management capabilities, improved systemic risk monitoring tools, and advanced early warning mechanisms.

As financial markets continue to evolve in complexity and interconnectness, multi-asset temporal structures of this type will gain greater importance in efficient risk management, regulation, and maintenance of financial stability. The methodological foundations established in this research, combined with the clear trajectories identified for future enhancement, provide a robust platform for ongoing advancement in financial machine learning applications for crash prediction and risk management in emerging markets contexts.

Acknowledgment

The authors extend their sincere gratitude to the develop-ers of scikit-learn, pandas, NumPy, and matplotlib for their invaluable open-source tools that enabled this research. We thank the National Stock Exchange of India for providing historical market data and our institution's High-Performance Computing Center for computational resources. Special appre-ciation goes to our colleagues for assisting in recognizing the patterns. Finally, we recognize the broader machine learning and computational finance communities whose pioneering research laid the foundation for this study.

REFERENCES

[1] P. Patil, C.-S. M. Wu, K. Potika, and M. Orang, "Stock market prediction using ensemble of graph theory, machine learning and deep learning models," in Proc. 3rd Int. Conf. Software Eng. Inf. Manage. (ICSIM), 2020, pp. 85–92.
[2] R. Kim et al., "HATS: A hierarchical graph attention network for stock movement prediction," in Proc. Workshop on AI in Finance, 2019, pp. 1–8.
[3] Y. Xu and V. Keselj, "Stock prediction using deep learning and sentiment analysis," in Proc. IEEE Int. Conf. Big Data, 2019, pp. 1–6.
[4] M. Plappert, R. E. Turner, and S. Whiteson, "Forecasting stock prices from the limit order book using CNNs," in Proc. IEEE 19th Conf. Business Informatics, 2017, pp. 1–8.
[5] X. Zhang et al., "Improving stock market prediction via heterogeneous information fusion," in Proc. Int. Conf. Knowledge-Based Intelligent Information Systems, 2018, pp. 236–247.
[6] J. Cao, X. Li, and D. Shen, "A stacked ensemble approach for stock market prediction," in Proc. IEEE ICDMW, 2020, pp. 1–8.
[7] X. Li, Y. Zhou, and Y. Li, "A GNN-based stock forecasting method using multi-source data fusion," in Proc. IEEE Int. Conf. AI and Big Data, 2022, pp. 45–52.
[8] S. Wang, H. Yu, and Y. Zhang, "Network-based prediction of financial cross-sector risk," in Proc. Int. Conf. Complex Systems and Financial Risk, 2021, pp. 1–7.
[9] B. Rossi, "Machine learning and stock return predictability," in Proc. Annu. Meeting AFA, 2018, pp. 1–9.
[10] A. Thakkar and K. Chaudhari, "CREST: Cross-reference based stock trend prediction using LSTM," in Proc. ICCIDS, 2020, pp. 1–10.
[11] M. Nabipour et al., "Deep learning for stock market prediction," in Proc. IEEE SMC, 2020, pp. 1–8.
[12] G. S. Uddin, S. J. Hernandez, and A. Yousaf, "Stock market contagion during COVID-19," in Proc. Int. Conf. Applied Economics and Finance, 2022, pp. 1–9.
[13] T. A. Siddiqui, M. A. Hassan, and S. Ali, "Cross-market correlations and financial contagion during global crises," in Proc. Int. Conf. Economics and Finance Research, 2022, pp. 1–8.
[14] E. M. Ferrouhi, M. E. Amrani, and A. Yahyaouy, "Comparative study of ensemble learning for stock price prediction," in Proc. IEEE DSAA, 2024, pp. 1–10.
[15] Y. Chen, L. Zhang, and X. Wang, "Forecasting stock market crisis events using ML," in Proc. Int. Conf. AI in Finance, 2024, pp. 1–9.
[16] N. Kumar and A. Srivastava, "Enhanced Eye Disease Detection Using Advanced CNNs," 12th Int. Conf. Emerging Trends in Eng. & Technology, 2025, pp. 1–7.
[17] Winkle, S. Sharma, and A. Srivastava, "A Computationally Efficient Neural Network for Sarcasm Detection," Int. Conf. Computing Technologies & Data Communication, 2025, pp. 1–7.
[18] Ravina and A. Srivastava, "Short-Term Stock Price Prediction using ANN and LSTM," Grenze Int. J. Eng. Technol., Grenze ID: 01.GIJET.11.2.29525, June 2025.