Journal of Computer Technology and Applied Mathematics
Journal Home: http://jctam.suaspress.org | CODEN: JCTAAD
Vol. 2, No. 6, 2025 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

Funding-Rate Driven Routing in Financial Prediction Systems

CHENG, Xueyi [1*]

[1] Duke University, USA

***** *CHENG, Xueyi is the corresponding author, E-mail: Frances.cheng17@gmail.com***

Abstract: Financial forecasting models tend to be based on past market action trends. While such methods are very good under steady-state conditions, financial markets tend to change abruptly due to a change in liquidity, macroeconomic news, or leverage imbalances. Such models trained on past price action are too sluggish in responding to regime changes and therefore make the wrong and not to be relied upon predictions at the worst possible moment when accurate advice is most required. To address this challenge, this paper proposes a funding-rate driven routing model that dynamically switches between forecasting models depending on inferred market states. Funding rates and spot–perpetual basis signs are used to detect volatility and sentiment changes that cause shock regimes. In ordinary times, a light univariate LSTM is used to ensure forecasting efficacy. As volatility rises, the system pushes forecasting to a multi-source deep learning model that incorporates price, volume, and sentiment background. The architecture uses an edge–cloud hybrid execution strategy that compromises on latency, privacy, and scalability. This work demonstrates that, through the integration of regime detection and model switching, one can obtain more accurate forecasting without sacrificing computational responsibility. The model introduces an additional more adaptive, context-aware methodology for real-world financial forecasting environments that must operate under continuously varying risk levels.

Keywords: Funding Rate, LSTM, Deep Learning, Edge–Cloud Execution.

Disciplines: Big Data Technology. **Subjects:** Data Analytics.

DOI: https://doi.org/10.70393/6a6374616d.333332 **ARK:** https://n2t.net/ark:/40704/JCTAM.v2n6a05

1 INTRODUCTION AND MOTIVATION

Financial markets follow no single pattern. Prices respond not only to their own previous trajectories but also to the behavior, expectations, and risk appetites of market participants. Such behaviors do change suddenly and in unexpected ways. Prediction systems that presuppose temporal regularity are then intrinsically susceptible. Therefore, the task of financial time-series prediction increasingly depends on ascertaining not only what patterns exist, but also when particular models are true.

Traditional forecasting methods, such as ARIMA, assume linear dynamics and stationarity that rarely hold in modern electronic trading. Machine learning methods, particularly Long Short-Term Memory networks (LSTM), demonstrate superior modeling capacity by learning long-range nonlinear price dependencies[1]. Yet, they inherit a structural weakness: the inability to recognize regime disruptions without explicit contextual triggers. If the environment that produced past data no longer applies, the LSTM continues to predict the continuation of outdated patterns. This lack of adaptability may lead to magnified error when markets are most unpredictable.

The last decade has seen an explosion of regime diversity. Automated trading machines, derivative leverage structures, and macroeconomic imbalances have all conspired to accelerate volatility transmission. The extreme one-day moves can be triggered by small catalysts that cascade through computerized liquidity. Such shocks are likely to be accompanied by leverage imbalance, which cannot be observed if models are monitoring only pricing information. Even when price direction is quiet, underlying risk conditions can destabilize insidiously.

Here lies the significance of funding-rate signals. In perpetual futures markets, funding rates measure who must pay financing cost: long traders or short traders. When funding rates spike, one side of the market is aggressively stacked. Such imbalance historically precedes violent corrections because forced liquidation cascades occur when the majority's trade fails. Funding-rate deviation is thus a leading indicator of market risk, not a lagging one.

Spot – perpetual basis presents a complementary perspective. When futures prices deviate significantly from spot asset value, speculative expectations temporarily overpower fundamental pricing. This divergence often marks

Journal of Computer Technology and Applied Mathematics
Journal Home: http://jctam.suaspress.org | CODEN: JCTAAD
Vol. 2, No. 6, 2025 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

anticipation of future volatility, even before volume increases.

Motivated by this insight, this research proposes to elevate funding-rate behavior to a central supervisory role in financial forecasting workflows. Instead of waiting for volatility to explicitly appear in price changes, the system reacts proactively to the root signals of instability.

Equally important is the engineering challenge of delivering reliable prediction services under resource and infrastructure constraints. Edge devices such as trading terminals and brokerage risk controllers require immediate responses with minimal latency[2]. Meanwhile, advanced machine learning models often exceed edge hardware capability. An intelligent edge–cloud split is thus necessary:

- light inference at the edge during stable markets
- cloud-assisted inference during volatile markets

This selective activation strategy allows computational expense to scale in direct proportion to market complexity.

Taken together, these observations guide the creation of a regime-aware, funding-rate-driven, hybrid-architecture forecasting system intended to balance efficient accuracy with deep robustness.

2 PROPOSED METHOD AND SYSTEM DESIGN

The proposed methodology integrates economic regime detection, conditional model selection, and distributed execution planning into a single coherent pipeline designed for continuous forecasting.

2.1 System Architecture Overview

To generalize across global financial settings, the architecture is composed of three core layers:

- State Discriminator
- Model Routing & Forecasting Module
- Edge–Cloud Execution Layer

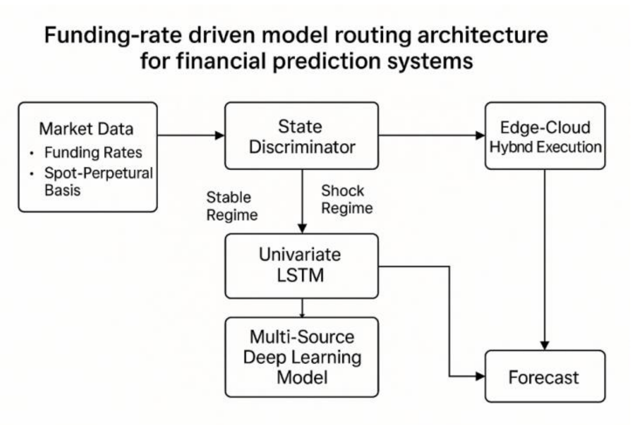

Figure 1. Funding-rate driven model routing architecture for financial prediction systems

2.2 State Discriminator

The purpose of the state discriminator is to evaluate market conditions in real time and classify them into operational regimes. Unlike conventional volatility measures that lag price behavior, this discriminator emphasizes proactive signals. It draws on:

- Funding rate exponential deviation index
- Spot–perpetual basis acceleration
- Short-window realized volatility
- Volume imbalance and liquidity thinning
- Sentiment polarity extracted from textual financial streams

Each indicator is transformed into a standardized score relative to historical norms. A weighted scoring mechanism determines whether the present moment resembles a Stable Regime or a Shock Regime.

Decision rules include:

- Stable Regime:

funding deviation < F_thresh and volatility < sigma_thresh

- Shock Regime:

funding deviation >= F_thresh or volatility >= sigma_thresh

This classification occurs continuously so that shifts in directional pressure are recognized before price volatility fully manifests.

2.3 Model Routing and Forecasting

Once the regime is identified, the model routing component assigns the forecasting task accordingly. The routing logic is:

Table 1. Model Routing Strategy

Market Regime	Forecasting Strategy	Core Objective	Trade-off
Stable Regime	Univariate LSTM	Efficiency	Simplified signal space
Shock Regime	Multi-Source Deep Model	Robustness and context sensitivity	Higher compute cost

Stable Regime Model

Focuses exclusively on price sequences:

- Efficient runtime
- Fast deployment on low-power devices
- Low-risk error profile when markets are calm

Shock Regime Model

Incorporates expanded predictors:

- LSTM component capturing price dynamics

- Attention modules extracting key volume shifts
- Embeddings of funding and basis signals
- Optional transformer sentiment encoder

This model is designed to resist structural breaks — a limitation of purely univariate forecasting [4].

Automated switching prevents human bottlenecks and eliminates latency between detection and response [4].

2.4 Edge–Cloud Hybrid Execution Layer

Latency is a first-order constraint in algorithmic and risk-management environments. Because model complexity should scale with event severity, the system distributes inference tasks according to execution difficulty:

TABLE 2. EDGE–CLOUD EXECUTION STRATEGY FOR REGIME-BASED MODEL ROUTING

Layer	Location	Advantages
LSTM during Stable Regime	Edge device	Fast, resilient, privacy-aware
Multi-source model	Cloud service	High compute availability and model capacity

Routing decisions incorporate network congestion status, device load, and prediction urgency. By minimizing unnecessary cloud offload, operational cost remains proportional to market stress[2].

2.5 System Objectives

The hybrid framework has three central targets:

- Low-risk prediction performance in unstable conditions
- Latency control to support high-frequency decision environments
- Computational efficiency and cost governance

This forms a self-tuning system that continuously aligns method complexity with market requirements.

3 EXPERIMENTAL PLAN

To rigorously validate the framework, this study will conduct a multi-stage evaluation structured around the following components:

3.1 Datasets

TABLE 3. DATASET DESCRIPTION

Domain	Dataset	Purpose
Equity	Apple Inc. (AAPL) historical price + volume	Benchmark univariate prediction
Crypto Derivatives	BTC perpetual futures funding rates & basis	Calibrate shock-state triggers

Where possible, external data sources such as macroeconomic event logs and financial news feeds will enrich contextual evaluation.

3.2 Evaluation Framework

This study uses a mixed methodology that incorporates both statistical accuracy and systems performance metrics.

Forecast Accuracy Metrics

- RMSE (error magnitude)
- MAE (average deviation)
- MAPE (normalized relative error)
- Directional Accuracy (trend classification)

System Efficiency Metrics

- End-to-end prediction latency
- Compute load distribution (edge vs cloud)
- Cloud utilization ratio

Shock-Sensitivity Metrics

- Detection delay of volatility onset
- Incorrect switching rate (false-positive regime classification)
- Cost of misclassified volatility windows

A model that is efficient but fails in volatile moments is less valuable than one that can recognize risk escalation and elevate analytical capability accordingly [5].

3.3 Ablation Studies

To justify the necessity of funding-rate guidance, multiple ablations will be conducted:

TABLE 4. ABLATION STUDY DESIGN AND ASSOCIATED HYPOTHESES

Ablation	Hypothesis
Remove funding-rate input	Performance drops significantly in shock windows
Remove basis tracking	Slower shock recognition
Disable routing	Compute cost spikes and volatility error increases
Always LSTM	Inadequate during market turbulence
Always Multi-Source Model	Unnecessary compute overhead; unacceptable latency

3.4 Deployment Simulation

A simulated trading environment will evaluate real-world value.

Key measurements include:

- Profit-loss improvement from enhanced timing
- Risk exposure and maximum drawdown reduction
- Strategic responsiveness to sudden event triggers

Furthermore, the experiment will analyze:

- Robustness under network degradation
- Privacy retention when edge-heavy computation is active
- Cost profile under varying volatility cycles

This robust evaluation ensures the system's economic and operational practicality. During development, we also reviewed recent research and identified an open-source edge–cloud AI agent framework known as SolidGPT [3]. By studying its deployment structure and implementation examples, we were able to accelerate environment configuration, modularize the routing logic, and validate real-time inference behavior more efficiently [6]. The availability of reusable components and tooling support significantly reduced engineering overhead, allowing us to focus on model performance and regime-switching behavior rather than low-level system setup [7]. This practical integration experience further supports the feasibility of deploying our proposed routing architecture in real-world financial systems [8-9].

4 CONCLUSION

Financial forecasting requires models that perform optimally only when their assumptions remain valid. Markets change — and predictive systems must change with them [10]. The funding-rate driven routing architecture proposed here introduces structural adaptability into forecasting workflows by combining:

- Proactive volatility state detection
- Conditional model switching
- Hybrid distributed execution

The system expects the unexpected. By observing leverage imbalance and basis divergence, it can detect hidden instability before it materializes in prices. Model complexity then expands accordingly. Computational resources are preserved when risk is minimal while reinforcement activates only when stakes escalate. This realigns predictive power to serve the trader when it matters most.

The method is scalable across asset classes and market structures because funding-rate logic generalizes to any risk premium that reflects leverage sentiment. The resulting forecasting workflow avoids over-engineering and ensures cost-responsiveness — an increasingly crucial objective for technological operations in finance.

This work demonstrates that a context-aware, regime-adaptive architecture can both guard against shock-induced prediction failure and streamline computation. It is therefore a practical and theoretically justified advancement in the field of financial time-series modeling.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

CHENG, Xueyi

Researcher at Duke University.

REFERENCES

[1] Qi, R., & Hu, L. (2025). Stock Price Prediction of Apple Inc. Based on LSTM Model: An Application of Artificial

Intelligence in Individual Stock Analysis. European Journal of AI, Computing & Informatics, 1(3), 1-9.

[2] Hu, L. (2025). Hybrid Edge-AI Framework for Intelligent Mobile Applications: Leveraging Large Language Models for On-device Contextual Assistance and Code-Aware Automation. Journal of Industrial Engineering and Applied Science, 3(3), 10-22.

[3] Hu, L., Wu, Q., & Qi, R. (2025). Empowering smart app development with SolidGPT: an edge–cloud hybrid AI agent framework. Advances in Engineering Innovation, 16(7), 86-92.

[4] Qi, Ruoyu. "Research on the Application of Robo-Advisory Systems in Personal Wealth Management: A Case Study of Middle-Class Households in the United States." Economics and Management Innovation 2.4 (2025): 46-55.

[5] Qi, Ruoyu. "Loan Default Prediction and Feature Importance Analysis Based on the XGBoost Model." European Journal of Business, Economics & Management 1.2 (2025): 141-149.

[6] Qi, Ruoyu. "An Empirical Study on Credit Risk Assessment Using Machine Learning: Evidence from the Kaggle Credit Card Fraud Detection Dataset." Journal of Computer, Signal, and System Research 2.5 (2025): 48-64.

[7] Galema, R., Koetter, M., & Liesegang, C. (2016). Lend global, fund local? Price and funding cost margins in multinational banking. Review of Finance, 20(5), 1981-2014.

[8] Gross, M. M., & Siebenbrunner, C. (2019). Money creation in fiat and digital currency systems. International Monetary Fund.

[9] Akcay, Ü., & Güngen, A. R. (2019). The making of Turkey's 2018-2019 economic crisis (No. 120/2019). Working Paper.

[10] Jansen, M., Nguyen, H. Q., & Shams, A. (2025). Rise of the machines: The impact of automated underwriting. Management Science, 71(2), 955-975.