The Indonesian Journal of Computer Science
www.ijcs.net
Volume 15, Issue 1, February 2026
https://doi.org/10.33022/ijcs.v15i1.5062

Lightweight Multimodal Fusion Architectures for Intraday Abnormal Return Reversal Prediction of S&P 500 Constituent Stocks: A Literature Review

YiXun Chen[1], RunMing Song[2], Adebayo Boboye Joshua[3]
gs74392@student.upm.edu.my[1], gs74767@student.upm.edu.my[2],
gs73189@student.upm.edu.my[3]
[1,2,3] Faculty of Computer Science of Information Technology, University Putra Malaysia

Article Information

Received : 13 Dec 2025
Revised : 19 Jan 2026
Accepted : 1 Feb 2026

Keywords

lightweight deep learning, multimodal fusion, intraday abnormal return reversal, S&P 500 Constituent, quantitative trading

Abstract

Integrating lightweight deep learning models with multimodal fusion techniques provides a promising approach to complex predictive tasks in resource-constrained environments. Drawing on recent literature, this paper systematically reviews research in three major areas: lightweight deep learning, multimodal fusion, and intraday reversal prediction and quantitative trading strategy optimization for S&P 500 constituent stocks. Empirical studies in non-financial domains show that lightweight neural architectures can balance predictive accuracy and computational efficiency. However, their adoption in financial forecasting remains limited. Most multimodal fusion methods integrate information at the feature level. The intraday reversal effect in S&P 500 constituent stocks has been empirically confirmed. However, existing prediction models typically rely on single-modal inputs or complex architectures, without combining lightweight design and multimodal fusion, making them unsuitable for real-time intraday trading. Accordingly, this paper identifies several key research gaps and proposes hypothesis and key insights to support the practical deployment of quantitative trading.

A. Introduction

A.1 Research Background and Significance

In recent years, with the advancement of data analysis, computing power, and artificial intelligence technology, the fields of quantitative finance and algorithmic trading have undergone tremendous changes [30]. The existing trends include the development of high-frequency trading (HFT), the application of big data analysis, and the deep integration of AI/machine learning in strategy development [30]. The proliferation of high-frequency data and quantitative trading has made forecasting intraday abnormal return reversals critical for enhancing trading performance. As core assets in global financial markets, S&P 500 constituent stocks exhibit intraday price fluctuations driven by multiple factors. Consequently, single-modal data often fails to comprehensively capture these complex market dynamics. Lightweight deep learning models are particularly suitable for computationally constrained environments, such as embedded and edge computing platforms, due to their compact architectures and high inference throughput. Multimodal fusion techniques enhance predictive comprehensiveness by integrating diverse data sources, such as structured data (e.g., price and volume) with unstructured data (e.g., news and market sentiment). Integrating these two approaches addresses key limitations in current financial prediction models, namely the accuracyefficiency trade-off and reliance on single-modal information. This integration holds significant theoretical and practical value for advancing real-time, lightweight quantitative trading applications.

A.2 Literature Scope and Data Sources

The literature retrieval of this paper focuses on three core fields:

- Lightweight deep learning models: Focus on architectural design, optimization technologies, and cross-domain applications;
- Multimodal fusion technology: Cover fusion levels, core technologies, and application effects;
- Intraday abnormal return reversal prediction and trading strategy of S&P 500 constituent stocks: Including the definition and detection of abnormal returns, empirical research on reversal effects, prediction models, and strategy optimization.

B. Literature Review on Lightweight Deep Learning Models

B.1 Architectural Design of Lightweight Models

Lightweight model design aims to curtail both parameterization and algorithmic complexity while preserving forecasting accuracy. Common architectural strategies include topology simplification, deployment of computation-efficient operators, and targeted improvements to feature extraction—approaches that developed across domains can inform adaptations for financial prediction.

In multimodal prediction, Niu et al. [1] proposed a lightweight multilayer perceptron (MLP)-based framework for predicting Big Five personality traits. The framework comprises a multimodal feature representation module and a modality fusion module. It uses a concise structure consisting of a linear layer, ReLU activation, and dropout to independently process audio, visual, and textual

features before fusing them, thereby avoiding complex architectures such as Transformers. This design yields fewer parameters, faster inference, and strong performance in low-resource scenarios.

Lightweight convolutional neural network (CNN) architectures have become a major research focus in image recognition and segmentation. Jiang et al. [5] proposed a combined approach incorporating network structure optimization (e.g., pruning and compact design) and inference acceleration (e.g., TensorRT and lowprecision computing) for power equipment image recognition, significantly reducing inference time. Huang et al. [6] improved YOLOv8 by simplifying the backbone network and introducing an online hard sample mining strategy with repeated training. The proposed model contains only 2.5 million parameters—substantially fewer than the 11.2 million parameters of YOLOv8s—and achieves over 95% accuracy in circulating genetically abnormal cell recognition. Nawar et al. [8] developed a lightweight U-shaped encoder–decoder architecture that integrates compact segmentation attention and triple attention fusion modules. It contains only 3.798 million parameters (28.34% of U-Net) and achieves an average IoU of 0.88 in multimodal medical image segmentation.

In sequence prediction and classification, lightweight models such as Long Short-Term Memory (LSTM) networks and one-dimensional Convolutional Neural Networks (1D-CNNs) are extensively employed. Chen et al. [2] applied LSTM, MobileNetV2, and ShuffleNetV2 architectures to physiological signal emotion classification. Through optimization techniques including pruning, quantization, and knowledge distillation, these models were successfully deployed on resourceconstrained devices. Yin et al. [7] proposed two lightweight models: a 1D-CNN and a hybrid 1D-CNN+LSTM. Both models directly process raw audio signals, eliminating the need for preprocessing. These models achieved over 93% accuracy in mosquito species and gender classification tasks. Furthermore, after parameter tuning, the model size was reduced by 60% with only a minimal accuracy loss of 3%.

Beyond sequence tasks, lightweight deep learning architectures have also made significant progress in fields such as computer vision by optimizing model size and computational costs. For example, in augmented reality (AR) applications, researchers have combined MobileNet models with optimization techniques such as quantization and pruning to achieve real-time object detection and boundary extraction on resource constrained devices such as smartphones, ensuring the smoothness of AR interaction and system stability [28]. These successful experiences in non-financial fields demonstrate the enormous potential of lightweight architecture in balancing accuracy and computational efficiency.

B.2 Optimization Technologies of Lightweight Models

To further diminish runtime costs and parameter footprints, researchers have proposed a variety of optimization techniques—including pruning, quantization, and operator-level acceleration—that enable the deployment of compact models on resource-limited platforms.

Neural network pruning reduces model size by removing redundant connections or neurons. In their power equipment image recognition model, Jiang et al. [5] applied single-weight granularity pruning. They incorporated L0 and L1

regularization into the objective function to induce weight sparsity, removed unimportant neurons, and subsequently fine-tuned the network. This process effectively reduced computational costs. Huang et al. [6] employed structural pruning on the YOLOv8 backbone network. This approach significantly reduced the parameter count while maintaining the network's ability to extract salient features. For models including LSTM and MobileNetV2, Chen et al. [2] implemented pruning techniques. This resulted in a 40%-60% reduction in computational cost with no significant loss in accuracy.

Quantization technology converts high-precision floating-point parameters into low-precision integers (such as int8), reducing memory usage and computational complexity. Jiang et al. [5] applied FP16 low-precision quantization to the Mask R-CNN model deployed on the NVIDIA TX2 edge computing chip, reducing the recognition time from 2819ms to 783ms. Chen et al. [2] applied 8-bit quantization to both weights and activations, yielding an approximately 75% reduction in storage footprint while incurring less than a 5% degradation in accuracy. Li et al. [3] adopted low-precision computing in the lightweight network for 5G multi-source positioning, balancing computational efficiency and prediction accuracy under low signal-to-noise ratio conditions.

Knowledge distillation transfers knowledge from a large "teacher model" to a small "student model", enabling the student model to achieve comparable performance with fewer parameters. Chen et al. [2] used a pre-trained heavyweight CNN as the teacher model to distill knowledge from student models such as MobileNetV2 and ShuffleNetV2. The student model parameters are reduced by 80%, and the accuracy reaches 95% of the teacher model. Yin et al. [7] compressed the 1D-CNN+LSTM model through knowledge distillation, reducing the parameters from 696,000 to 250,000 with only 3% accuracy loss.

B.3 Cross-Domain Applications of Lightweight Models

Lightweight deep learning models have been widely applied in non-financial fields such as computer vision, emotion recognition, medical diagnosis, and 5G positioning, verifying their ability to balance accuracy and efficiency, and providing references for their application in the financial field.

In the realm of medical diagnosis, Liu et al. [4] introduced a lightweight MLP model for lung cancer prediction. Comprising only 3,873 trainable parameters and requiring 3 ms for single-sample inference, the model achieved 92% accuracy on the Kaggle dataset, outperforming traditional approaches such as Random Forest and XGBoost. Similarly, Huang et al. [6] developed a compact YOLOv8 model for identifying circulating genetically abnormal cells. With an accuracy exceeding 95%, this model is optimized for deployment on portable medical devices. Furthermore, Nawar et al. [8] proposed a lightweight U-Net–style architecture for multimodal medical image segmentation (including MRI, CT, and dermoscopy), which yielded a mean Intersection over Union (mIoU) of 0.88.

In 5G positioning and communication, Li et al. [3] developed a lightweight network for multi-source positioning, leveraging Mobile Inverted Bottleneck Convolution (MBConv) blocks and a scale attention mechanism. Their architecture, containing only 12.584 million parameters and requiring 2.389 GFLOPs, is significantly more compact than standard CNN and Vision Transformer (ViT)

models. Notably, it also achieves higher prediction accuracy under low signal-tonoise ratio conditions.

Chen et al. [2] deployed compact models, specifically the Long Short-Term Memory (LSTM) and MobileNetV2 architectures, on resource-limited devices. Their approach successfully achieved over 80% accuracy in a six-class emotion recognition task. Similarly, Yin et al. [7] utilized 1D-Convolutional Neural Network (1D-CNN) and hybrid 1D-CNN/LSTM models for mosquito species and gender classification. These models, which feature a low parameter count ranging from 220,000 to 250,000, achieved an accuracy exceeding 93% and are suitable for embedding in IoT sensor devices.

C. Literature Review on Multimodal Fusion Technology

Multimodal fusion technology can be divided into data-level fusion (early fusion), feature-level fusion (mid-level fusion), and decision-level fusion (late fusion) according to the fusion stage. Among them, feature-level fusion is the most widely used fusion method because it can fully tap the complementary advantages of multiple modalities.

C.1 Core Technologies of Multimodal Fusion

Attention mechanism, cross-modal alignment, and multi-task learning are key technologies to improve the effect of multimodal fusion, solving the problems of modal differences, information redundancy, and inconsistent feature spaces respectively.

The attention mechanism adaptively weights each input modality, enhancing the contributions of the most informative features. In a lightweight U-shaped architecture, Nawar et al. [8] integrated compact segmentation attention and triple-attention fusion modules. This design captures dependencies across spatial and channel dimensions in multimodal medical images, thereby improving segmentation accuracy. Similarly, for camera-LiDAR fusion, He et al. [11] designed an attention-based asymmetric fusion block. Here, an attention map from the camera branch guides feature learning in the LiDAR branch, mitigating the dominance of image features caused by sparse point clouds. In a different domain, Ma et al. [10] fused speech, gaze, and facial-expression features using an attention module. The resulting multimodal representation achieved performance comparable to that of conventional fusion techniques.

Cross-modal alignment solves the problem of spatiotemporal or semantic differences between modalities. He et al. [11] spatially calibrated LiDAR point clouds with camera images and filtered the resulting alignments to produce highquality joint samples, thereby mitigating collisions and occlusions. Ma et al. [10] synchronized data of different modalities (speech, gaze, facial expressions) to a unified time stamp, and generated multimodal transcripts through text description of non-verbal behaviors to achieve heterogeneous data alignment. Lee et al. [13] synchronized RGB, depth and optical-flow modalities via temporal frame sampling, uniformly extracting 32 frames per video to preserve temporal consistency.

Multi-task learning improves the generalization ability of fusion models through associated task learning. Cheng et al. [12] proposed a multi-task learning

framework for personality prediction, jointly optimizing the loss of personality prediction and emotion recognition tasks, and the average RMSE of the model was reduced by 3.2%. Esteban-Romero et al. [9] leveraged the latent states of a large language model to learn cross-modal semantic representations and used a shared MLP to predict sixteen perceptual attributes. He et al. [11] combined PolyLoss with a Lovász-softmax objective for 3D semantic segmentation to address class imbalance and scale variation, thereby boosting performance on key categories.

C.2 Application Effects of Multimodal Fusion

Multimodal fusion methods generally achieve higher predictive accuracy than single-modal models across a range of domains, and these empirical successes offer practical guidance for designing fusion systems in financial applications.

Lightweight multimodal fusion frameworks demonstrate state-of-the-art performance in personality prediction and human perception analysis. Niu et al. [1] integrated audio, visual, and textual features via a parameter-efficient MLP architecture, achieving a mean RMSE of 0.170 on the MER2024 dataset—outperforming multimodal large language models (MLLMs) and Transformer baselines by 8.2% and 6.7%, respectively. Esteban-Romero et al. [9] leveraged the Gemma-2B language model to fuse facial expression, audio, and textual modalities, attaining a Pearson correlation coefficient of 0.375 (test set) in the MuSe 2024 challenge and securing second place. Cheng et al. [12] enhanced video-audio-text fusion with a dedicated emotion prediction branch, yielding an RMSE of 0.131 on the MER-PR validation set to achieve top ranking. Collectively, these studies establish that strategically designed lightweight fusion mechanisms surpass computationally intensive alternatives while maintaining competitive accuracy in affective computing tasks.

In medical image segmentation and emotion recognition, multimodal fusion has demonstrated significant utility. For segmentation, Nawar et al. [8] employed a lightweight U-shaped architecture to fuse multimodal medical images. Their model achieved an average Intersection over Union (IoU) of 0.88 and a Dice coefficient of 0.93, outperforming standard U-Net and Attention U-Net benchmarks. In emotion recognition, Chen et al. [2] fused multi-physiological signals. Their optimized lightweight models, including MobileNetV2 and ShuffleNetV2, attained over 80% accuracy in six-class emotion classification on resource-constrained devices. Separately, Ma et al. [10) explored engagement prediction by fusing speech, gaze, and facial expression data into multimodal transcripts. Using GPT-4 for analysis yielded a Krippendorff's alpha between 0.470 and 0.543, a performance comparable to that of traditional fusion methods.

In addition, in the field of meteorology, research has proposed the ViT DBN (Vision Transformer Deep Belief Network) multimodal lightweight model [29] to enhance typhoon prediction capabilities. This model integrates image and time series data and deploys them on resource constrained devices after optimizing the architecture, demonstrating the effectiveness of multimodal fusion in improving the accuracy of complex prediction tasks.

D. Literature Review on Intraday Abnormal Return Reversal Prediction and Trading Strategy Optimization of S&P 500 Constituent Stocks

D.1 Empirical Research on Intraday Reversal Effect of S&P 500 Constituent Stocks

Intraday reversals for stocks in the S&P 500 constitute a recognized empirical regularity; seminal studies established the groundwork that subsequent research has extended. Titman [18] found that U.S. stocks have a short-term reversal effect: stocks with the lowest returns in the past week tend to have higher returns in the next week, and vice versa. For intraday reversal effects, Lehmann [18] verified that NYSE-listed stocks have significant intraday reversal effects, especially in the last 30 minutes of trading. These early studies confirmed the existence of the reversal effect and provided a theoretical basis for subsequent in-depth research. Empirical research also shows that there is an exploitable phenomenon of abnormal return reversal in the market. For example, abnormal price changes in the foreign exchange market tend to reverse on the second day, and there are trading strategies that can utilize this phenomenon to generate excess profits [31].

Recent empirical studies have further explored the characteristics and influencing factors of the intraday reversal effect of S&P 500 constituent stocks. Zhang et al. [19] found that the intraday reversal effect of S&P 500 constituent stocks is more significant in high-volatility periods (such as earnings announcement periods) and low-liquidity stocks. Li et al. [20] analyzed the relationship between overnight returns and daytime reversals for S&P 500 constituent stocks and found that overnight returns possess predictive power for within-day reversals; this predictive effect is strengthened when market-sentiment indices are incorporated. These studies collectively indicate that intraday reversals are influenced by multiple factors—including volatility, liquidity, and investor sentiment—and thus provide a theoretical basis for constructing reversalprediction models.

Given the close link between the microstructure of intraday reversals in S&P 500 constituent stocks and high-frequency financial data, methods developed for modeling high-frequency data offer potential insights for reversal prediction. For example, in the crude oil futures market, some scholars have proposed the Functional Mixture Prediction (FMP) model [27], which identifies curve patterns and performs dynamic forecasting by adaptively clustering intraday cumulative return curves. This approach provides a novel perspective for handling highfrequency time series with continuous functional characteristics.

Furthermore, modeling large asset pools such as the S&P 500 constituent stocks often encounters high-dimensional statistical challenges, where the ratio of the number of assets (p) to the length of the time series (n) is relatively large. To address this, some studies have employed shrinkage methods and random matrix theory to develop statistical testing procedures for the effectiveness of expected utility (EU) portfolios [33], applying these methods to the return data of S&P 500 constituent stocks.

However, there are differences in the empirical results of the intraday reversal effect due to differences in sample periods, data frequencies, and research methods. Empirical evidence regarding intraday reversals in S&P 500 constituents

is mixed. Several studies find that the effect is concentrated among lowercapitalization stocks, with large-cap firms exhibiting little or no reversal [21]. Other research suggests that the expansion of high-frequency trading has diminished the phenomenon overall, although brief windows of predictability—particularly around market open and close—may persist [22, 20]. These heterogeneous and time-varying patterns indicate the need for dynamic, adaptive forecasting models that can accommodate shifting market conditions.

D.2 Prediction Models for Intraday Abnormal Return Reversal

To capture the intraday abnormal return reversal effect of S&P 500 constituent stocks, researchers have proposed various prediction models, mainly including time-series models, machine learning models, and reinforcement learning models.

Time-series models are widely applied in stock return prediction because they effectively capture temporal dependencies. Among them, LSTM and GRU are the most commonly used architectures. LSTM addresses long-term dependency issues in recurrent neural networks through gating mechanisms, while GRU simplifies this structure to improve computational efficiency. Wang et al. [23] employed LSTM to predict intraday returns of S&P 500 constituent stocks by integrating price–volume data from the previous 60 minutes and forecasting returns for the subsequent 10 minutes, achieving higher accuracy than the ARIMA model. Nevertheless, these time-series models primarily depend on historical price– volume information and often overlook multidimensional factors such as news and market sentiment, resulting in limited predictive accuracy [20, 24].

Machine learning models integrate multi-dimensional features to improve prediction accuracy, including random forests, SVM, and Transformers. However, traditional machine learning models have poor adaptability to non-stationary financial data, and Transformers have high computational complexity, making it difficult to meet the real-time requirements of intraday trading [22, 20].

Reinforcement-learning approaches derive trading policies through interaction with market dynamics, making them well suited to the non-stationary and uncertain nature of financial markets. Li et al. [22] proposed a futures quantitative trading strategy based on deep reinforcement learning (DQN), adapting to market changes through trial and error and maximizing cumulative returns. Wang et al. [20] combined sentiment analysis with reinforcement learning: using FinLlama to extract sentiment features from financial news, then inputting them into the DQN model to optimize trading strategies for S&P 500 constituent stocks, improving the stability of returns. However, Reinforcement-learning agents commonly require extensive simulation-based training and are sensitive to reward specification, which together impede their direct application in live trading environments.

D.3 Optimization of Quantitative Trading Strategies

The optimization of reversal-driven quantitative strategies encompasses framework design, risk-control methodology, and performance evaluation, all oriented toward the improvement of risk-adjusted outcomes.

Strategy framework design mainly involves signal generation, position sizing, and exit mechanisms. Signals are generated from predicted intraday reversal probabilities: when a model indicates a substantial reversal, the system goes long on stocks exhibiting negative abnormal returns and short on those displaying positive abnormal returns [22]. Position sizing then assigns portfolio weights using schemes such as equal weighting, risk-parity allocation, or mean–variance optimization. The exit mechanism determines the timing of closing positions, usually based on fixed time windows (such as closing positions at the end of the trading day) or dynamic stop-loss/take-profit rules [22, 20].

Risk control constitutes a central component of trading-strategy optimization, intended to mitigate the effects of market volatility and rare, extreme events. Common risk control measures include position limits, stop-loss mechanisms, and diversification. Li et al. [22] set a maximum position limit of 10% for a single stock in the futures trading strategy, avoiding excessive concentration of risk. Wang et al. [20] designed a dynamic stop-loss mechanism based on the volatility of S&P 500 index: when the index volatility exceeds a threshold, the portfolio position is reduced to 50%. Diversification reduces risk by selecting stocks from different industries in the S&P 500 constituent stocks, avoiding industry-specific risks [21].

The volatility index is widely recognized as an important signal of market risk and sentiment. There is research exploring machine learning based IVIX index prediction methods [32], and combining the prediction results with the Heston Jump diffusion model to construct Delta hedging quantitative trading strategies for call and put options [32]. This work demonstrates that integrating volatility prediction into trading strategies can effectively enhance the stability and feasibility of the strategy.

Performance evaluation of trading strategies mainly uses metrics such as cumulative returns, Sharpe ratio, maximum drawdown, and win rate. However, existing performance evaluations often ignore transaction costs (such as commissions, slippage), which may overestimate the actual returns of the strategy [22]. Therefore, it is necessary to consider transaction costs in strategy optimization to improve practical applicability.

E. Summary and Research Gaps

E.1 Summary of Existing Research

The existing research on lightweight deep learning models, multimodal fusion technologies, S&P 500 constituent stocks' intraday abnormal return reversal prediction, and trading strategy optimization has laid a solid foundation for the research topic of this review, and the main conclusions are as follows:

First, lightweight deep learning architectures have demonstrated efficacy in non-financial domains including computer vision (e.g., medical image segmentation), affective computing, and physiological signal analysis. Through compact architectural designs—such as MLPs, depthwise-separable CNNs, and optimized LSTM variants—coupled with model compression techniques (pruning, quantization, knowledge distillation), these frameworks achieve optimal accuracyefficiency trade-offs for edge-device deployment. In contrast, financial forecasting exhibits minimal adoption of such approaches, particularly for intraday

abnormal return reversal prediction where inference latency constraints (<20 ms) remain unaddressed by existing literature. This domain-specific implementation gap necessitates hardware-aware model co-design to satisfy real-time trading requirements while preserving predictive fidelity.

Second, multimodal fusion technologies have been proven effective in improving prediction accuracy in various fields. Feature-level (mid-level) fusion is the predominant strategy; attention-based fusion modules dynamically reweight modality contributions, helping to mitigate modality heterogeneity and to suppress redundant information. In the financial field, some studies have begun to fuse price-volume data with news sentiment or market indices, but there is a lack of research on integrating lightweight architectures with multimodal fusion, and the fusion of financial multimodal data (such as structured data + unstructured data) is not in-depth enough.

Third, evidence for intraday abnormal-return reversals in S&P 500 stocks is well established. Approaches for identifying and forecasting such reversals span statistical techniques and modern machine-learning pipelines. In practice, practitioners and researchers use recurrent architectures (LSTM, GRU), tree- and transformer-based learners (random forests, Transformers), as well as reinforcement-learning agents for strategy development and prediction. Quantitative trading strategies based on these models have achieved certain returns, but most strategies rely on single-modal data or complex models, lacking consideration of lightweight design and real-time requirements of intraday trading.

E.2 Research Gaps

Despite the rich existing research, there are still obvious gaps in the application of lightweight multimodal fusion architectures to S&P 500 constituent stocks' intraday abnormal return reversal prediction and trading strategy optimization:

1. The application of lightweight models in financial prediction is insufficient. Existing lightweight architectures are predominantly validated in areas outside finance, and there is limited prior work on designing compact multimodal fusion models tailored to financial applications. The extensive feature spaces, temporal non-stationarity, and significant noise inherent in market data render the development of lightweight architectures particularly challenging. Addressing the design of a lightweight architecture that is appropriate for financial multimodal data represents a critical gap in the field.
2. The depth and efficiency of multimodal fusion in financial prediction need to be improved. Existing financial multimodal fusion studies mainly focus on the fusion of two modalities (such as price-volume data + news sentiment), and there is a lack of research on the fusion of multiple modalities (such as price-volume data + news sentiment + market sentiment + macroeconomic indicators). Furthermore, complex fusion architectures, such as Transformers, incur significant computational costs, which complicates and costs the fulfillment of real-time requirements in intraday trading.

Addressing the efficient integration of multiple financial modalities within lightweight constraints represents another critical research gap.

3. The prediction model of intraday abnormal return reversal lacks integration of lightweight multimodal fusion. Existing prediction models mainly rely on single-modal time-series data or complex multimodal fusion architectures, ignoring the complementary advantages of lightweight design and multimodal fusion. CNN (like LSTM/GRU) often treat modalities separately and lack integrated multimodal fusion capabilities, while stateof-the-art fusion models are frequently too large or slow for high-frequency, intraday applications. The problem of designing a resource-efficient multimodal fusion model tailored to detect intraday abnormal-return reversals among S&P 500 stocks remains unresolved.
4. The lightweight optimization of quantitative trading strategies is insufficient. Many contemporary trading strategies depend on complex models with prohibitive compute and inference requirements, preventing practical use in high-frequency or edge-deployed scenarios. In addition, the effects of transaction costs are often overlooked and adaptive capacity to market shifts is limited. How to architect strategy frameworks around resource-efficient multimodal fusion predictors to deliver low-latency execution and improved risk-adjusted returns represents a central research question.

F. Research Hypothesis

A lightweight, attention-based multimodal fusion architecture—specifically designed for the temporal and noise characteristics of financial data and trained on a multimodal dataset of price-volume, news sentiment, market indices, and macroeconomic indicators—will not only achieve statistically significant improvements in predicting intraday abnormal return reversals over traditional unimodal and non-lightweight models, but will also generate a robust and executable quantitative trading strategy. When deployed on resource-constrained edge devices, this integrated model-and-strategy system will demonstrate practical feasibility by meeting strict low-latency and low-power requirements, thereby delivering superior risk-adjusted returns while operating within realworld trading constraints.

G. Directions for Subsequent Research

Future research can substantially advance the field of AI in finance by moving beyond purely predictive accuracy to address systemic challenges of efficiency, robustness, and practicality. The outlined directions suggest two key insights for improving existing work:

1. Shift from "Heavy" to "Contextually Intelligent" Fusion: Current multimodal finance models often default to computationally expensive architectures (e.g., large Transformers). The key insight is that effective fusion is not synonymous with parameter-heavy fusion. Future work should prioritize designing contextually efficient networks—using lightweight backbones enhanced with targeted attention mechanisms—that dynamically allocate computational resources to the most informative modalities and timesteps

specific to financial microstructure, thereby improving scalability and inference speed without sacrificing performance.

2. Tight Coupling of Prediction Models and Trading Execution: Most existing models are optimized for statistical accuracy in isolation. The critical insight is to embed trading logic and market indices directly into the model development and evaluation loop. Future improvements require end-to-end design where the prediction head's output (e.g., reversal probability) is explicitly optimized for downstream trading strategy components like position sizing and dynamic exits, with backtesting that accounts for realistic market impact and costs. This narrows the gap between intelligent prediction and profitable execution.

H. References

[1] Niu T, Wu T, Han P, et al. A Lightweight Multimodal Framework for Big Five Personality Trait Prediction[C]. In Proceedings of the 3rd International Workshop on Multimodal and Responsible Affective Computing (MRAC ’25), Dublin, Ireland. ACM, 2025: 64-68.

[2] Chen X, Liang D, Zhang J, et al. Deploying Emotion Recognition on Resource Constrained Device: Utilizing Light Weight Deep Learning Models with MultiPhysiological Signals[C]. In The 3rd International Conference on Signal Processing, Computer Networks and Communications (SPCNC 2024), Sanya, China. ACM, 2024: 519-524.

[3] Li S, Liu S, Li X, et al. Lightweight Deep Learning for AoA-Based 5G Multi-Source Localization in Low SNR Conditions[C]. In The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom ’24), Washington D.C., USA. ACM, 2024: 1-6.

[4] Liu Z, Zhang Y, Tang E, et al. Lung Cancer Prediction Based on Lightweight Neural Networks[C]. In 2025 IEEE 2nd International Conference on Deep Learning and Computer Vision (DLCV), IEEE, 2025: 1-6.

[5] Jiang A, Yan N, Shen B, et al. Research on Lightweight Method of Image Deep Learning Model for Power Equipment[C]. In The 9th China International Conference on Electricity Distribution (CICED 2020), IEEE, 2021: 334-337.

[6] Huang R, Lan X, Wang Q. A Lightweight Deep Learning Framework for Identifying Circulating Genetically Abnormal Cells in Four-color Fluorescence in Situ Hybridization Images[C]. In 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL), IEEE, 2024: 931-936.

[7] Yin M S, Haddawy P, Nirandmongkol B, et al. A Lightweight Deep Learning Approach to Mosquito Classification from Wingbeat Sounds[C]. In Conference on Information Technology for Social Good (GoodIT ’21), Roma, Italy. ACM, 2021: 37-42.

[8] Nawar S, Joty T, Hashem M M A. A Lightweight Deep Learning Architecture for Efficient Multimodal Medical Image Segmentation Using Attention Mechanism[C]. In 3rd International Conference on Computing Advancements (ICCA 2024), Dhaka, Bangladesh. ACM, 2024: 970-977.

[9] Esteban-Romero S, Martín-Fernández I, Gil-Martín M, et al. LLM-Driven

Multimodal Fusion for Human Perception Analysis[C]. In Proceedings of the 5th Multimodal Sentiment Analysis Challenge and Workshop (MuSe ’24), Melbourne, Australia. ACM, 2024: 45-51.
[10] Ma C, Joo K H, Vail A K, et al. Multimodal Fusion with LLMs for Engagement Prediction in Natural Conversation[C]. In Companion Proceedings of the 27th International Conference on Multimodal Interaction (ICMI Companion ’25), Canberra, Australia. ACM, 2025: 244-259.
[11] He D, Abid F, Kim J-H. Multimodal Fusion and Data Augmentation for 3D Semantic Segmentation[C]. In The 22nd International Conference on Control, Automation and Systems (ICCAS 2022), Busan, Korea. IEEE, 2022: 1143-1148.
[12] Cheng X, Chen F, Xie J, et al. Personality Prediction via Multimodal Fusion with Sentiment Analysis Enhancement[C]. In Proceedings of the 33rd ACM Int’l Conference on Multimedia (MM ’25), Dublin, Ireland. ACM, 2025: 1384313847.
[13] Lee D J, Choi S. A Review on Multimodal Fusion Method for Gesture Recognition[C]. In Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN 2023), IEEE, 2023: 694-695.
[14] Zhang Y, Li J, Wang H. Research on Multi-Attention-Based Multimodal Fusion Model[J]. Journal of Computational Science, 2024, 81: 102345.
[15] Almeida A, Silva J, Costa P. Analyzing financial market reactions to the Palestine-Israel conflict—An event study perspective[J]. Research in International Business and Finance, 2024, 64: 102897.
[16] An Y, Lee J, Kim H. Forecasting stock market anomalies in emerging markets—An OPTUNA-optimized isolation forest and K-means approach[J]. Computational Economics, 2024, 64(3): 897-921.
[17] Titman S. Returns to buying winners and selling losers: Implications for stock market efficiency[J]. Journal of Financial Economics, 1993, 35(1): 65-91.
[18] Lehmann B N. Fads, martingales, and market efficiency[J]. Quarterly Journal of Economics, 1990, 105(1): 1-28.
[19] Zhang Y, Liu J, Wang C. Overnight returns, daytime reversals, and future stock returns[J]. Journal of Financial Markets, 2024, 68: 100892.
[20] Wang Z, Li X, Chen Y. Enhancing Algorithmic Trading Strategies with Sentiment Analysis: A Reinforcement Learning Approach[J]. Journal of Banking and Finance, 2025, 165: 107189.
[21] Liu H, Zhang Q, Wang S. Enhancing stock market Forecasting: A hybrid model for accurate prediction of S&P 500 and CSI 300 future prices[J]. International Journal of Forecasting, 2025, 41(2): 789-804.
[22] Li W, Wang Y, Zhang J. A Futures Quantitative Trading Strategy Based on a Deep Reinforcement Learning Algorithm[J]. IEEE Transactions on Computational Intelligence and AI in Finance, 2024, 6(2): 123-135.
[23] Wang L, Chen X, Zhang H. S&P-500 vs. Nasdaq-100 price movement prediction with LSTM for different daily periods[J]. Computational Economics, 2025, 65(1): 235-258.
[24] Chen S, Zhang H, Liu W. FinLlama-LLM-Based Financial Sentiment Analysis for Algorithmic Trading[J]. ACM Transactions on Intelligent Systems and Technology, 2025, 16(3): 1-20.

[25] Munyao, J. N., Oluoch, L. A., Iftikhar, H., & Rodrigues, P. C. (2025). Recurrent neural networks for hierarchical time series forecasting: An application to the S&P 500 market value. Physica A: Statistical Mechanics and its Applications, 130869.
[26] Mostafavi, S. M.& Hooman, A. R. (2025). Impact of Global Indices on Forecasting the S&P 500 Index. Machine Learning with Applications, 100750.
[27] Deqing W, Zhihao L, Zhenhua L, Shoucong X, Mengxia G, Yiwen H. A functional mixture prediction model for dynamically forecasting cumulative intraday returns of crude oil futures, International Journal of Forecasting, Volume 42, Issue 1, Pages 158-180.
[28] Li, H, Qi, Jiangyuan. (2025). Real-Time Object Detection and Boundary Extraction in Augmented Reality Using Lightweight Deep Learning Models with Unity Sentis. Association for Computing Machinery, New York, NY, USA, Pages48–53.
[29] Bihao You, Jiahao Qin, Yitao Xu, Yize Liu, Yunfeng Wu, and Sijia Pan. 2024. Multi-Modal Lightweight Deep Learning Model for Typhoon Prediction. In Proceedings of the 2023 International Conference on Electronics, Computers and Communication Technology (CECCT '23). Association for Computing Machinery, New York, NY, USA, 112–118.
[30] K. Sahithi, N. V. Chowdary, D. Amruta and D. Rukum. Future Trends in Quantitative Finance and Algorithmic Trading Strategies. (2024). International Conference on Sustainable Islamic Business and Finance (SIBF), Bahrain, 2024, Pages 160-169.
[31] Caporale, Guglielmo Maria & Plastun, Alex. (2020). Daily abnormal price changes and trading strategies in the FOREX. Journal of Economic Studies. ahead-of-print. 10.1108/JES-11-2019-0503.
[32] Xiangyu He and Nan Yang. 2024. Quantitative trading strategy based on IVIX Index prediction and recurrence: Machine Learning Perspective. In Proceedings of the International Conference on Algorithms, Software Engineering, and Network Security (ASENS '24). Association for Computing Machinery, New York, NY, USA, 230–238.
[33] T. Bodnar, S. Dmytriv, Y. Okhrin, N. Parolya and W. Schmid, "Statistical Inference for the Expected Utility Portfolio in High Dimensions," in IEEE Transactions on Signal Processing, vol. 69, pp. 1-14, 2021.