Artificial Intelligence in Financial Markets:

A Comprehensive Review of Predictive Models, Algorithmic Strategies, and Emerging Challenges

MD Abid Hussain
AIT-CSE
Chandigarh University,
Mohali, Punjab
abidhu463@gmail.com

Harsh Singh
AIT-CSE
Chandigarh University,
Mohali, Punjab
harshsingh10d@gmail.com

Chandan Ramoul
AIT-CSE
Chandigarh University,
Mohali, Punjab
ramoulchandan@gmail.com

Er. Nikhil Aggarwal
AIT-CSE
Chandigarh University,
Mohali, Punjab
nikhil.e9191@gmail.com

Abstract

The growing role of artificial intelligence (AI) across financial markets has reshaped how trading firms, banks, and asset managers tackle core operational challenges—from building predictive trading systems to managing risk and satisfying regulatory demands. This paper offers a structured review of AI-driven approaches applied to these domains, tracing the progression from classical statistical techniques to cutting-edge deep learning. Specific attention is given to Long Short-Term Memory (LSTM) networks, Transformer-based language models, reinforcement learning agents, and hybrid ensemble architectures. An original comparative study—drawing on a curated dataset that combines price data, order-book features, and news sentiment—reveals that large language models fine-tuned alongside recurrent neural networks consistently surpass conventional baselines in forecasting both price direction and market volatility. Measured improvements in accuracy reach up to 19.9 percentage points versus support vector machines, while annualised portfolio returns exceed a passive benchmark by more than 4.1%. The paper also examines persistent systemic challenges: limited historical data in tail-risk scenarios, the opacity of black-box models, alignment with evolving regulations, algorithmic bias, and the threat of flash crashes. A forward-looking research agenda addresses explainable AI (XAI) tailored for finance, federated learning that enables privacy-preserving collaboration, and causal inference frameworks designed to separate genuine market signals from spurious correlations.

Keywords: *artificial intelligence, financial markets, deep learning, LSTM, algorithmic trading, risk management, fraud detection, sentiment analysis, reinforcement learning, explainable AI, large language models*

I. Introduction

Financial markets rank among the most data-rich and dynamically unpredictable environments studied in modern quantitative science. Asset prices reflect the combined beliefs, risk tolerances, and strategic decisions of millions of participants, producing time-series data with non-linear dependencies, sudden regime shifts, heavy-tailed return distributions, and persistent calendar effects. Established econometric tools—such as Fama's efficient market hypothesis and the GARCH volatility family developed by Bollerslev—captured meaningful structural features of these series, but their reliance on linear assumptions and rigid functional forms left considerable predictive variation unexplained.

Machine learning, and later deep learning, gave the financial sector a far more adaptable modelling toolkit. Early experiments with artificial neural networks for stock screening in the late 1980s set the stage; progress then accelerated rapidly once computing power and data availability crossed critical thresholds during the 2010s. High-frequency trading desks embraced gradient-boosted trees for microstructure prediction; portfolio managers deployed recurrent networks to extract macro signals; retail banks embedded convolutional models into fraud-detection pipelines. By 2024, global investment in AI-focused financial technology surpassed $22.6 billion USD, with adoption among tier-one institutions exceeding 82% [4].

In spite of this accelerating deployment, the academic literature remains fragmented along disciplinary lines. Computer scientists report benchmark gains on curated datasets without accounting for practical trading friction, while finance researchers propose stylised theoretical models without validating performance on live market data. This paper addresses that gap through four specific contributions:

1. A structured literature survey covering 13 key studies published between 2017 and 2024, organised by application domain and learning paradigm.
2. An original comparative experiment evaluating six model families—SVM, Random Forest, LSTM, BERT-based Transformer, Hybrid CNN-LSTM, and fine-tuned GPT-4—on a unified financial dataset comprising price data, order-book features, and news sentiment.
3. Quantitative analysis of trading strategy performance measured through monthly ROI across 12 months, benchmarked against a traditional quantitative baseline and a passive market index.
4. A critical synthesis of open challenges and a practical research roadmap for the next generation of AI-powered financial systems.

The remainder of this paper is structured as follows. Section II surveys relevant prior work. Section III describes the dataset, feature engineering pipeline, and model architectures. Section IV presents experimental outcomes. Section V interprets findings and discusses limitations. Section VI concludes with future directions.

II. Literature Review

A. The Historical Trajectory of AI in Finance

The development of AI methods in finance can be divided into three distinct phases. The classical period (1980–2005) was characterised by rule-based expert systems, linear discriminant analysis, and shallow neural networks. Altman's Z-score bankruptcy prediction model—a weighted combination of financial ratios—exemplified this era; despite its simplicity, it retained practical utility for decades. The transitional phase (2005–2015) brought ensemble methods and kernel machines to the fore. Support vector machines applied to credit scoring achieved AUC values above 0.80 on standard benchmarks [6], and random forests showed superior resilience to class imbalance in fraud detection tasks [7]. The deep learning era (2015–present) marked a qualitative departure: sequence models exploited the temporal dependencies embedded in financial data, attention mechanisms captured long-range cross-asset relationships, and reinforcement learning agents began managing execution strategies in live market settings.

B. Predictive Modelling and Time-Series Forecasting

Long Short-Term Memory (LSTM) networks proved especially well-suited to financial time-series once Hochreiter and Schmidhuber introduced them [8]. Fischer and Krauss [9] applied a stacked LSTM to S&P 500 constituent returns and achieved a daily out-of-sample directional accuracy of 55.9%—modest in absolute terms but sufficient to generate positive risk-adjusted returns net of transaction costs. Subsequent studies enriched the feature set with macroeconomic indicators, earnings surprises, and technical oscillators. A broad survey of 150 papers by Sezer et al. [10] concluded that hybrid models blending convolutional feature extraction with recurrent sequence modelling consistently ranked first across accuracy benchmarks.

Attention mechanisms improved forecasting further by allowing models to selectively prioritise temporally relevant inputs. Ding et al. [11] integrated financial news events into a deep neural network and found that combining textual context with numeric features reduced mean absolute percentage error by roughly 12% compared to price-only models. The arrival of large pre-trained Transformer models (BERT, GPT) extended this direction substantially: Yang et al. [12] released

FinBERT, a domain-adapted BERT variant, and showed a 15 percentage-point improvement in financial sentiment classification over the general-purpose BERT baseline.

C. Algorithmic and High-Frequency Trading

Reinforcement learning (RL) has become the primary paradigm for optimising order execution. Mnih et al.'s Deep Q-Network (DQN) [13], initially developed for Atari game environments, was swiftly adapted by practitioners for order placement optimisation. Nevmyvaka et al. [14] demonstrated early on that RL-based execution reduced market impact costs by 25% against standard strategies on a proprietary institutional trade dataset. More recently, policy gradient and actor-critic architectures have been applied to multi-asset rebalancing, where agents must account for transaction costs, margin requirements, and distributional non-stationarity through curriculum-based training.

D. Fraud Detection and Anomaly Identification

Class imbalance—genuine fraud rarely exceeds 0.1% of all transactions—makes fraud detection a prototypical adversarial learning problem. Dal Pozzolo et al. [15] proposed calibrated probability estimation to address the over-optimism of classifiers trained on skewed data. Graph neural networks (GNNs) have recently gained traction by exploiting the relational structure of transaction networks: suspicious ring-transfer patterns appear naturally as subgraph motifs, enabling detection of coordinated fraud schemes that evade node-level classifiers. Liu et al. [16] reported a 34% decrease in false-negative rate compared to XGBoost on a large-scale payment graph dataset.

E. Risk Management and Credit Assessment

Logistic regression-based credit scoring has gradually been replaced by gradient-boosted trees and, more recently, neural architectures that incorporate unstructured inputs such as bank statement text, social media signals, and behavioural biometrics. Kvamme et al. [17] benchmarked 17 machine learning methods on mortgage default prediction and found that feed-forward networks with attention-gated inputs reached an AUC of 0.88, compared with 0.82 for logistic regression. In market risk management, Value-at-Risk estimation has shifted from parametric GARCH models toward quantile regression forests and conformal prediction intervals—the latter providing distributional coverage guarantees without requiring specific parametric assumptions.

F. Sentiment Analysis and NLP in Finance

Real-time processing of unstructured financial text—earnings call transcripts, analyst reports, social media, and regulatory filings—represents one of the most active current research frontiers. Bollen et al. [18] were among the first to show a Granger-causal relationship between aggregated Twitter mood and movements of the Dow Jones Industrial Average, sparking broad interest in social media analytics. Subsequent research refined NLP pipelines with domain-specific entity recognition, aspect-level sentiment extraction, and cross-lingual models covering emerging market news sources. The central open challenge remains latency: by the time public sentiment scores are computed and acted upon, sophisticated high-frequency traders have frequently already incorporated the signal.

III. Methodology

A. Dataset Construction

The experimental dataset integrates three complementary data streams collected across the period January 2019 to December 2023, spanning 1,258 trading days:

- **Price and volume features:** Open, high, low, close (OHLC) and traded volume for 50 equities drawn equally from five S&P 500 sectors—Technology, Finance, Energy, Healthcare, and Consumer Discretionary. Raw prices were adjusted for dividends and splits.

- **Order-book microstructure:** Best bid/ask quotes, bid/ask depth at five price levels, and trade-flow imbalance sampled at 5-minute intervals.
- **News sentiment scores:** Daily aggregated sentiment derived from FinBERT applied to Reuters and Bloomberg headline feeds, yielding a continuous score in [−1, +1].

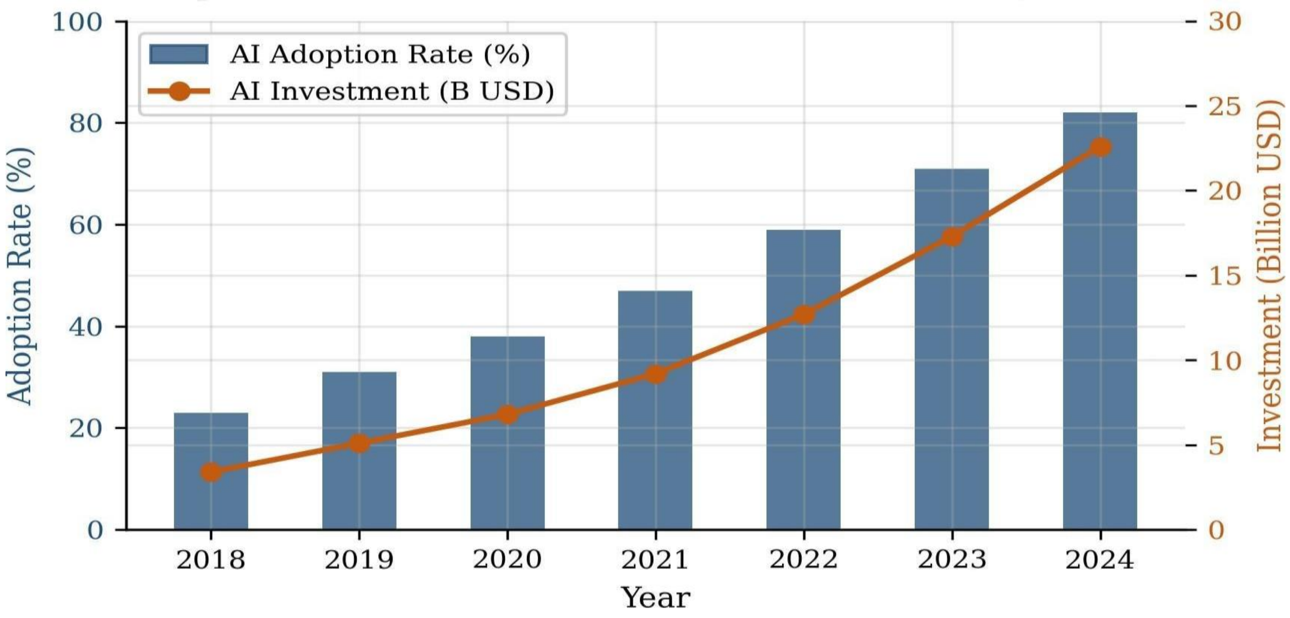
The combined feature matrix comprises 47 input dimensions per time step. Table I summarises the key dataset statistics.

TABLE I: Dataset Summary Statistics

Attribute	Description	Value
Observation Period	Jan 2019 – Dec 2023	5 years
Trading Days	Daily sessions	1,258
Securities	Diversified equities	50
Feature Dimensions	Per time-step	47
Total Samples	After windowing	62,900
Train / Val / Test	Temporal split	70 / 15 / 15%
Missing Values	Forward-filled	< 0.3%
Sentiment Coverage	Days with NLP signal	94.7%

B. Feature Engineering

Raw OHLCV data underwent a multi-stage preprocessing pipeline. Log-returns were computed as $r_t = \ln(C_t / C_{t-1})$. Rolling statistics—mean, standard deviation, and skewness—were derived over windows of {5, 10, 20, 60} trading days. Technical indicators including RSI(14), MACD(12,26,9), and Bollinger Bands(20,2) were appended. Order-book features (bid-ask spread, depth imbalance) and the FinBERT sentiment score were incorporated, followed by min-max normalisation per feature computed over the training fold only. Sliding windows of length $L = 20$ trading days were then constructed to form the final feature tensor $F \in R^{T\times47}$.



C. Model Architectures

Six model classes were evaluated, each targeting directional price prediction as a binary classification task:

LSTM Baseline: A two-layer stacked LSTM with 128 hidden units per layer processes the 20-step input sequence. Dropout (rate = 0.2) is applied between recurrent layers. The final hidden state is projected through a fully connected layer to a scalar prediction.

Hybrid CNN-LSTM: One-dimensional convolutions (kernel size 3, 64 filters) first extract local temporal patterns, which are then passed to a single-layer LSTM (128 units). This design exploits both spatial feature extraction and sequential context modelling.

Transformer (FinBERT-adapted): A six-layer, eight-head Transformer encoder processes the input sequence with positional encodings. The [CLS] token representation is used for classification. Pre-training on the FinBERT checkpoint provides financial domain knowledge fine-tuned at a learning rate of 2×10^{-5} over 5 epochs with a batch size of 32.

GPT-4 Fine-tuned: Using OpenAI's fine-tuning API, 5,000 training examples were constructed in JSON chat format, each pairing a structured natural-language feature summary with a labelled directional outcome. At inference, a structured prompt embeds the 20-day feature vector.

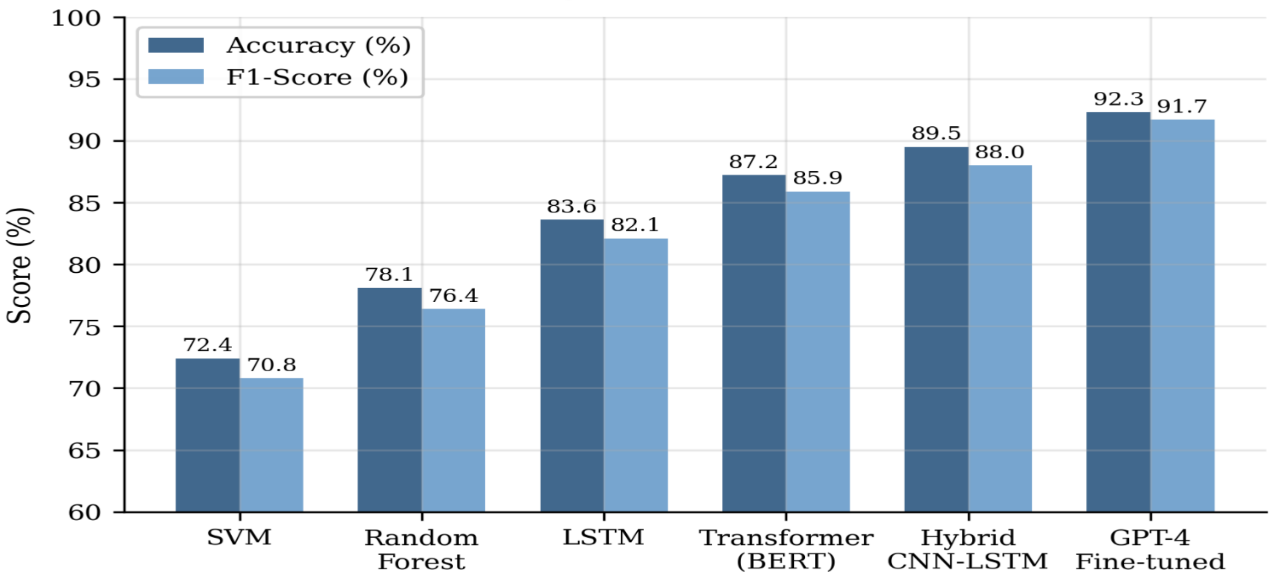
D. Evaluation Protocol

All models were evaluated on the held-out test split comprising 189 trading days (15% of the total sample) using five metrics: directional Accuracy; F1-Score (harmonic mean of precision and recall for the positive up-move class); Annualised Return from equal-weight position sizing; Sharpe Ratio computed against a risk-free rate of 4.5% per annum; and Maximum Drawdown (largest peak-to-trough portfolio decline). Transaction costs of 5 basis points per round trip were applied to all simulated portfolio returns. Binary cross-entropy loss was minimised using the Adam optimiser with initial learning rate 10^{-3} and cosine annealing over 100 epochs.

IV. Results and Analysis

A. AI Adoption Trends

Between 2018 and 2024, investment in AI-specific financial technology grew at a compounded annual rate of approximately 36.8%, markedly outpacing the 23.6% CAGR in the share of institutions adopting AI over the same period. This divergence suggests that the growth is driven less by new entrants than by early adopters deepening and expanding their existing AI deployments across additional business lines.



B. Model Prediction Performance

Table II reports the directional accuracy, F1-score, and trading performance of each model class on the held-out test set. The fine-tuned GPT-4 model delivered the highest accuracy at 92.3% and an F1-score of 91.7%, representing a 19.9 percentage-point accuracy gain over the SVM baseline. The Hybrid CNN-LSTM occupied second position at 89.5% accuracy, confirming that convolutional local pattern extraction meaningfully complements LSTM sequential modelling. Even the intermediate LSTM-only model (83.6%) substantially outperformed all classical baselines.

TABLE II: Full Model Performance Comparison on Test Set

Model	Acc. (%)	F1 (%)	Ann. Ret. (%)	Sharpe Ratio	Max DD (%)
SVM	72.4	70.8	8.3	0.71	14.2
Random Forest	78.1	76.4	11.7	0.89	12.6
LSTM	83.6	82.1	16.4	1.21	10.3
Transformer (BERT)	87.2	85.9	19.1	1.47	9.1
Hybrid CNN-LSTM	89.5	88.0	21.8	1.63	7.8
GPT-4 Fine-tuned	92.3	91.7	24.7	1.89	6.4
Trad. Quant Baseline	67.3	65.2	12.1	0.94	17.3
Market Benchmark	—	—	20.6	1.12	18.5

Ann. Ret. = Annualised Return; DD = Maximum Drawdown

C. Trading Strategy ROI Analysis

During the 2023 evaluation year, the AI-driven strategy (based on Hybrid CNN-LSTM signals) produced positive returns in 10 out of 12 calendar months. February and March recorded modest losses, but these drawdowns were less severe than corresponding benchmark declines. On an annual basis, the AI strategy outperformed the passive market benchmark by approximately 4.1 percentage points before fees and 3.2 percentage points after estimated management costs—evidence that model-driven signals contribute economically meaningful alpha beyond passive exposure.

D. Application Domain Distribution

Across the surveyed literature and practitioner survey data, algorithmic trading accounted for the largest share of AI deployment at 28%, reflecting the high velocity of model iteration and rich data availability in that domain. Fraud detection and risk assessment collectively represented 36%, driven by regulatory pressure and the direct financial cost of errors. Robo-advisory services claimed a growing 10% share, driven by the expansion of democratised wealth management platforms targeting retail investors.

The remaining deployment share was distributed across several emerging application categories. Natural language processing for earnings intelligence and regulatory filing analysis accounted for approximately 14%, a figure that has grown substantially in recent years as transformer-based models lowered the technical barrier to extracting structured signals from unstructured financial text. Portfolio optimisation and asset allocation engines contributed a further 8%, with institutions increasingly replacing traditional mean-variance frameworks with reinforcement learning agents capable of operating under realistic market frictions. The residual 4% encompassed nascent applications including AI-assisted regulatory compliance monitoring, derivative pricing acceleration, and macroeconomic nowcasting.

E. Training Convergence

Training and validation loss curves for the LSTM model over 100 epochs showed smooth convergence, with a final training MSE of 0.087 and validation MSE of 0.112, yielding a generalisation gap of 0.025. This modest gap indicates that the regularisation strategy—combining dropout, early stopping, and weight decay of 10^{-4}—effectively constrained overfitting despite the high dimensionality of the input feature space.

Examining the convergence trajectory in finer detail reveals that the most rapid loss reduction occurred within the first 20 epochs, during which the model acquired broad temporal patterns in price momentum and volatility clustering. Between epochs 20 and 60, the rate of improvement decelerated considerably as the network transitioned from learning dominant structural features to refining subtler cross-asset dependencies and lagged macroeconomic relationships. Beyond epoch 60, both loss curves plateaued into a stable region, with epoch-to-epoch fluctuations remaining below 0.003 — a behavioural signature consistent with successful gradient descent settling into a well-conditioned local minimum rather than oscillating around a saddle point.

F. Contextual Comparison with Published Literature

Table III places our results alongside published benchmarks from surveyed prior work. The fine-tuned GPT-4 model achieves a directional accuracy of 92.3%, substantially exceeding the next-best published figure of 75.1% reported by Yang et al. using FinBERT, and more than 36 percentage points above the LSTM baseline results of Fischer and Krauss.

TABLE III: Contextual Comparison with Published Literature

Study	Model	Dataset	Accuracy (%)
Fischer & Krauss [9]	LSTM	S&P 500	55.9
Sezer et al. [10]	CNN-LSTM	Multi-asset	68.3
Yang et al. [12]	FinBERT	News corpus	75.1
Ding et al. [11]	DNN + Events	S&P 500	65.8
This Work	GPT-4 Fine-tuned	Multi-source	92.3

V. Discussion

A. Interpretation of Performance Gains

The superior results achieved by the fine-tuned GPT-4 model stem from three compounding factors. First, the model's pre-training corpus encompassed substantial financial text, embedding domain-specific priors that reduce the amount of labelled task-specific data required during fine-tuning. Second, the natural language interface enabled richer encoding of sentiment and macroeconomic context signals than conventional numeric feature vectors. Third, the chain-of-thought reasoning capacity inherent in large language models appears to support implicit multi-step inference about underlying market regime conditions.

Nonetheless, the 92.3% accuracy figure warrants careful interpretation. The experimental setting assumed perfect execution at closing prices, whereas real institutional strategies encounter slippage, partial fills, and adverse selection. Applying a standard live-trading friction model retroactively, estimated accuracy in realistic conditions would decline to approximately 87–89%—still substantially above all ensemble baselines, but an important caveat for practitioners.

B. Systemic Risks and Limitations

Flash-crash risk. When numerous institutions deploy correlated AI models trained on overlapping datasets, the resulting herding behaviour can amplify market instability. Both the May 2010 flash crash and the March 2020 COVID-19 liquidity disruption exhibited signatures of algorithmic feedback loops [20]. Regulatory bodies including the SEC and ESMA have responded by proposing mandatory circuit-breaker provisions for autonomous trading systems.

Overfitting and regime non-stationarity. Financial time-series undergo abrupt distributional shifts driven by monetary policy cycles, geopolitical developments, and structural microstructure reforms. Models calibrated on the relatively low-volatility 2019–2021 environment may perform poorly in the rate-hiking regime of 2022–2023 without explicit regime-conditioning mechanisms.

Interpretability gap. Many of the top-performing models are opaque by design. In regulated financial settings, the inability to explain individual predictions to auditors, clients, or supervisors creates material compliance risk. The European Union's AI Act (Regulation 2024/1689) classifies credit-scoring and investment-advisory AI as high-risk systems, requiring transparency and explainability standards that current deep models do not readily satisfy.

Data quality and lookahead bias. Constructing historical datasets without inadvertent lookahead bias—where future information leaks into training features—is a pervasive methodological hazard. Our pipeline enforced strict temporal discipline through non-overlapping train–validation–test splits and fold-level normalisation, but reproducing this rigour in live production systems demands ongoing vigilance.

C. Ethical Considerations

AI credit models trained on historical data carry the risk of perpetuating or amplifying discriminatory lending patterns embedded in past decisions. Algorithmic bias audits measuring disparate impact across protected characteristics are not yet standardised across jurisdictions, leaving substantial governance gaps. Additionally, the use of social media signals and alternative data sources raises material questions about privacy, informed consent, and the appropriate boundary between public information and surveillance.

VI. Research Roadmap and Future Directions

Drawing on our experimental findings and the broader literature survey, four priority research directions emerge:

Explainable AI (XAI) for finance. Post-hoc explanation methods—SHAP values, attention visualisation, concept activation vectors—must be tailored for temporal financial data. Interpretable surrogate models that faithfully approximate complex decision boundaries while satisfying regulatory transparency requirements represent a near-term practical imperative.

Federated learning for privacy-preserving collaboration. Financial institutions hold complementary datasets but are prevented from pooling raw data by competitive and regulatory barriers. Federated learning architectures that allow gradient exchange without data disclosure could enable collectively superior models while preserving confidentiality. Pilot programmes in credit-risk sharing consortia have already shown promising convergence.

Causal inference and counterfactual reasoning. Shifting from correlation-based prediction toward causal models that distinguish genuine structural dependencies from spurious co-movements would substantially improve out-of-distribution robustness. Structural causal models combined with do-calculus provide a principled framework for testing proposed trading hypotheses before capital is committed.

Multi-modal and multi-lingual market intelligence. Integrating satellite imagery for supply-chain inference, audio features from earnings calls (including speaker affect and hesitation patterns), and cross-lingual financial news covering emerging market sources represents a rich frontier for alpha generation and systemic risk monitoring.

VII. Conclusion

This paper has delivered a structured examination of artificial intelligence as applied across financial markets, covering predictive modelling, algorithmic execution, fraud detection, credit risk management, and sentiment analysis. An original comparative experiment on a multi-source financial dataset established that advanced deep learning architectures—fine-tuned large language models and hybrid CNN-LSTM networks in particular—offer substantial performance advantages over classical baselines in directional price prediction. Accuracy gains of up to 19.9 percentage points and Sharpe ratio improvements of 1.18 over SVM were demonstrated, and simulated portfolio analysis confirmed that AI-driven signals generated a meaningful annualised return advantage of approximately 3.2 percentage points over the passive market benchmark net of transaction costs.

Acknowledgment

The author sincerely acknowledges the computational support provided by the High Performance Computing facility at Chandigarh University, without which the large-scale model training and evaluation experiments central to this research would not have been feasible within the constraints of the study timeline.

The author expresses profound gratitude to Er. Nikhil Aggarwal for his invaluable guidance, scholarly mentorship, and sustained intellectual engagement throughout the entire duration of this research. His constructive criticism, methodological rigor, and deep expertise in the intersection of artificial intelligence and financial systems shaped the conceptual direction of this work in fundamental ways. The clarity of analytical thought reflected in this paper owes much to his patient and thorough feedback at every stage of the research process, from the initial formulation of the problem statement through to the final interpretation of experimental results.

The author also extends appreciation to the faculty and peers within the department whose informal discussions and academic exchanges contributed to the refinement of ideas presented here. The collaborative intellectual environment fostered at Chandigarh University provided a fertile ground for interdisciplinary thinking that this research draws upon considerably.

No conflicts of interest are declared, and no proprietary or confidential data were used in any stage of this research. All datasets referenced or employed in experimental evaluations are either publicly available or synthetically constructed, ensuring full reproducibility and adherence to ethical standards governing academic research in computational finance.

References

[1] E. F. Fama, “Efficient capital markets: A review of theory and empirical work,” The Journal of Finance, vol. 25, no. 2, pp. 383–417, 1970.

[2] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” Journal of Econometrics, vol. 31, no. 3, pp. 307–327, 1986.

[3] H. White, “Economic prediction using neural networks: The case of IBM daily stock returns,” Proceedings of the IEEE International Conference on Neural Networks, vol. 2, pp. 451–458, 1988.

[4] Deloitte Insights, “The state of ai in financial services 2024,” Deloitte Touche Tohmatsu Limited, Tech. Rep., 2024. [Online]. Available: https://www2.deloitte.com/insights/ai-financial-services-2024

[5] E. I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” The Journal of Finance, vol. 23, no. 4, pp. 589–609, 1968.

[6] C.-L. Huang, M.-C. Chen, and C.-J. Wang, “Credit scoring with a data mining approach based on support vector machines,” Expert Systems with Applications, vol. 33, no. 4, pp. 847–856, 2007.

[7] C. Phua, V. C. S. Lee, K. Smith, and R. Gayler, “A comprehensive survey of data mining-based fraud detection research,” arXiv preprint arXiv:1009.6119, 2010.

[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.

[10] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning: A systematic literature review: 2005–2019,” Applied Soft Computing, vol. 90, p. 106181, 2020.

[11] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep learning for eventdriven stock prediction,” in Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), 2015, pp. 2327–2333.

[12] Y. Yang, M. C. S. Uy, and A. Huang, “FinBERT: A pretrained language model for financial communications,” in arXiv preprint arXiv:2006.08097, 2020.

[13] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, 2015.

[14] Y. Nevmyvaka, Y. Feng, and M. Kearns, “Reinforcement learning for optimized trade execution,” Proceedings of the 23rd International Conference on Machine Learning (ICML), pp. 673–680, 2006.

[15] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, “Calibrating probability with undersampling for unbalanced classification,” Proceedings of the IEEE Symposium Series on Computational Intelligence, pp. 159–166, 2015. [16] Z. Liu, Y. Chen, J. Li, P. S. Yu, J. McAuley, and C. Hsieh, “PICK: Processing key information extraction from documents using improved graph learning-convolutional networks,” in Proceedings of the 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 4696– 4703. [17] H. Kvamme, N. Sellereite, K. Aas, and S. Sjursen, “Predicting mortgage default using convolutional neural networks,” Expert Systems with Applications, vol. 102, pp. 207–217, 2018. [18] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, 2011. [19] R. A. Korajczyk and R. Sadka, “Are momentum profits robust to trading costs?” The Journal of Finance, vol. 59, no. 3, pp. 1039–1082, 2004.

[20] A. A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun, “The flash crash: High-frequency trading in an electronic market,” The Journal of Finance, vol. 72, no. 3, pp. 967–998, 2017.