Research Article

A Dual-Branch Deep Architecture Fusing Finbert Embeddings With Sequential [1]ce Data For Crypto Market Prediction

Vamsi Addagarla[1], Dr. D. Haritha[2]

[1]M.Tech Student, Department of CSE KLEF Green Fields, Vaddeswaram, Andhra Pradesh - 522501, India. Email: 2401050072cse@gmail.com
[2]Professor, Department of CSE KLEF Green Fields, Vaddeswaram, Andhra Pradesh - 522501, India. Email: harithadonavalli@kluniversity.in

Abstract

Valuation mechanics within decentralized cryptocurrency ecosystems deviate fundamentally from traditional equity asset pricing paradigms. In the absence of standardized discounted cash flows, manda-tory regulatory disclosure calendars, and audited quarterly balance sheets, decentralized token valuations fluctuate under the continuous, reflexive influence of distributed retail sentiment, narrative shifts, and microblog speculation. Traditional natural language processing pipelines and generic financial lexicons regularly fail to capture the nuances of retail crypto discourse due to domain-specific slang, pervasive sarcasm, non-standard syntax, and shifting conversational sentiment polarities.

This paper introduces an end-to-end, high-throughput multimodal framework designed to harvest, contextualize, and fuse financial social sentiment with quantitative temporal price series. We systematically evaluate FinBERT—a language representation model specialized on financial corpora—against general-purpose Transformers (BERT-base, RoBERTa) and rule-based dictionaries (VADER, Loughran-McDonald) across 36 continuous months (2021–2023) of high-frequency social feeds and market data. By modeling tabular OHLCV data through recurrent temporal networks (Long Short-Term Memory units) and integrating daily aggregated log-odds sentiment indices via intermediate late fusion, our system achieves a 61.82% directional accuracy and a 60.90% Macro F1-score for next-day Bitcoin (BTC) directional movement. This delivers an 8.7% absolute performance gain over isolated time-series baselines and surpasses general-domain NLP pipelines by 3.4%. Extensive ablations, multi-horizon evaluations (t + 1, t + 3, t + 7), hyperparameter sensitivity profiles, statistical hypothesis verifications via Diebold-Mariano tests, and simulated transaction-cost-adjusted backtests confirm that domain-adapted language modeling extracts durable alpha from microblog noise.

Index Terms—Computational Finance, FinBERT, Time-Series Forecast-ing, Sentiment Analysis, Natural Language Processing, Multimodal Deep Learning, Social Computing, Market Microstructure.

1 INTRODUCTION

SSET pricing theory in traditional equity and fixed-income markets rests upon foundational economic hypotheses asserting that asset values reflect rational ap-praisals of fundamental information flows [1]. The Efficient Market Hypothesis (EMH) posits that asset prices adjust immediately to all publicly available intelligence, rendering systematic abnormal returns impossible without assuming proportional risk premiums [2]. Over the past decade, how-ever, the proliferation of decentralized ledger technologies and permissionless digital assets has created an alternative

asset class that challenges these classic paradigms.

Cryptocurrency tokens (e.g., Bitcoin [BTC], Ethereum [ETH]) operate within continuous, globally accessible, 24/7/365 liquidity pools. Unlike publicly traded equities, digital assets operate outside the boundaries of standardized regulatory reporting mechanisms, mandatory SEC Form 10-K filings, earnings calendar cycles, and registered equity research analyst consensus estimates [7]. Price discovery in this decentralized domain is heavily retail-driven and intrinsi-cally reflexive [14]. Speculative dynamics are catalyzed across social computing platforms such as X (formerly Twitter), Telegram, and Reddit, where retail market participants coordinate liquidity, broadcast market interpretations, and amplify sentiment trends [8]. Consequently, behavioral shifts and sentiment contagion frequently manifest in microblog streams hours before being registered by conventional tech-nical indicators or on-chain transaction volumes.

1.1 The Semantic Challenge of Financial Social Data

Efforts to quantify and exploit social sentiment within algorithmic trading systems have long encountered se-vere data degradation and representation challenges. Early market surveillance relied on static lexical tables, such as the General Inquirer or the Loughran-McDonald financial dictionary [9]. While suitable for processing formal

corporate prospectuses, dictionary-based lookups fail when applied to social microblogs. They are structurally incapable of handling pervasive informalities, cryptographic argot (e.g., “hodl,” “flippening,” “rekt”), syntactic inversions, and visual-linguistic tokens like emojis (e.g., [Rocket], [Diamond Hands], [Chart Down]), which carry substantial directional conviction among retail market participants.

The advent of the Transformer architecture [12] and masked language modeling paradigms, epitomized by BERT [3], fundamentally altered natural language processing by generating contextualized word embeddings rather than static token mappings. Nevertheless, transferring an off-the-shelf general-domain Transformer directly to market-prediction tasks introduces substantial domain misalignment. In colloquial English, terms such as “exposure,” “volatility,” or “liability” convey overtly negative connotations. Within financial semantics, however, they represent standard struc-tural operations, neutral risk categories, or routine balance-sheet classifications.

Domain-specific language models, such as FinBERT [4], resolve this misalignment by continuing self-supervised pre-training over large financial corpora, including corporate regulatory disclosures, financial news feeds, and earnings call transcripts. Yet, the empirical literature continues to display a significant methodological gap: most existing benchmarks limit FinBERT’s deployment to standalone text classification on curated datasets like Financial PhraseBank [10], failing to evaluate its predictive efficacy within joint multimodal architectures handling non-stationary pricing series and noisy microblog environments.

1.2 Research Hypotheses and Contributions

To address this gap, this investigation evaluates whether domain-adapted language representations can serve as ro-bust predictive features when fused with multi-variate price series across volatile market conditions. We formalize this inquiry around three primary research hypotheses:

Hypothesis 1. Domain-specific pre-trained Transformers (Fin-BERT) generate higher signal-to-noise ratios in extracting retail sentiment from unstructured microblogs than both general-domain Transformers (BERT) and traditional rule-based sentiment lexicons (VADER).

Hypothesis 2. A multimodal framework that jointly models sequential temporal price dynamics (OHLCV) and high-density financial sentiment representations yields statistically superior next-day directional forecasting performance compared to purely quantitative time-series baselines.

Hypothesis 3. The predictive utility of fused financial social sentiment exhibits temporal decay, offering distinct signal decay profiles across short (t + 1), medium (t + 3), and multi-day (t + 7) forecasting horizons.

To empirically evaluate these hypotheses, this paper presents the following contributions:

1) **An End-to-End Multimodal Data and Modeling En-gine:** We develop a robust data acquisition, cleaning, and normalization pipeline engineered to process raw microblog streams alongside multi-exchange trade books, systematically preserving contextual crypto-native tokens while suppressing platform-specific noise.
2) **A Dual-Branch Deep Architecture:** We formulate a joint learning architecture that processes temporal historical price windows through recurrent Long Short-Term Memory (LSTM) cells, concurrently ex-tracting contextual polarity matrices via FinBERT to construct a daily log-odds sentiment metric, subse-quently merged via late-stage dense fusion layers.
3) **Mathematical Rigor in Fusion Derivations:** We pro-vide complete analytical derivations for bidirectional self-attention mechanisms, cross-modal representa-tion fusion, and backpropagation through time under joint loss constraints.
4) **Comprehensive Longitudinal Validation:** We con-duct empirical evaluations across 36 contiguous months (January 1, 2021, to December 31, 2023) across Bitcoin and Ethereum, capturing extreme bull runs, historical drawdowns, and sustained sideways consolidation phases.
5) **Rigorous Ablation, Horizon, and Economic Feasi-bility Testing:** We isolate the components driving model gains through systematic sentiment engine ab-lations, parameter sensitivity evaluations, multi-step horizon analyses, Diebold-Mariano statistical signif-icance verifications, and transaction-cost-adjusted backtesting simulations.

The remainder of this manuscript is organized as follows: Section II surveys related literature spanning behavioral finance, market microstructure, and multimodal time-series modeling. Section III details the data harvesting, noise filtra-tion, and preprocessing protocols. Section IV provides the mathematical formulations of the FinBERT feature extractor, the recurrent price network, and the fusion classifier. Section

V outlines the experimental configuration and hyperpa-rameter optimization. Section VI presents empirical results, comparative baselines, and ablation studies. Section VII analyzes market microstructure, volatility clustering, and failure modes. Section VIII provides an in-depth discussion on structural limitations, market latency, and ethical

consid-erations. Section IX concludes the study and outlines future research pathways.

2 Related Work and Theoretical Founda-tion

2.1 Behavioral Finance and Market Microstructure

The theoretical underpinnings of this work derive from behavioral economics and market microstructure theory. While classical financial mathematics assumes information symmetry and rational utility optimization [1], behavioral finance demonstrates that cognitive biases, heuristic decision-making, and emotional contagion induce systemic pricing anomalies [11]. In asset classes lacking standardized baseline valuations, these psychological dynamics dominate short-term equilibria.

Noise trader risk, as formulated by De Long et al. [41], establishes that irrational market participants driven by

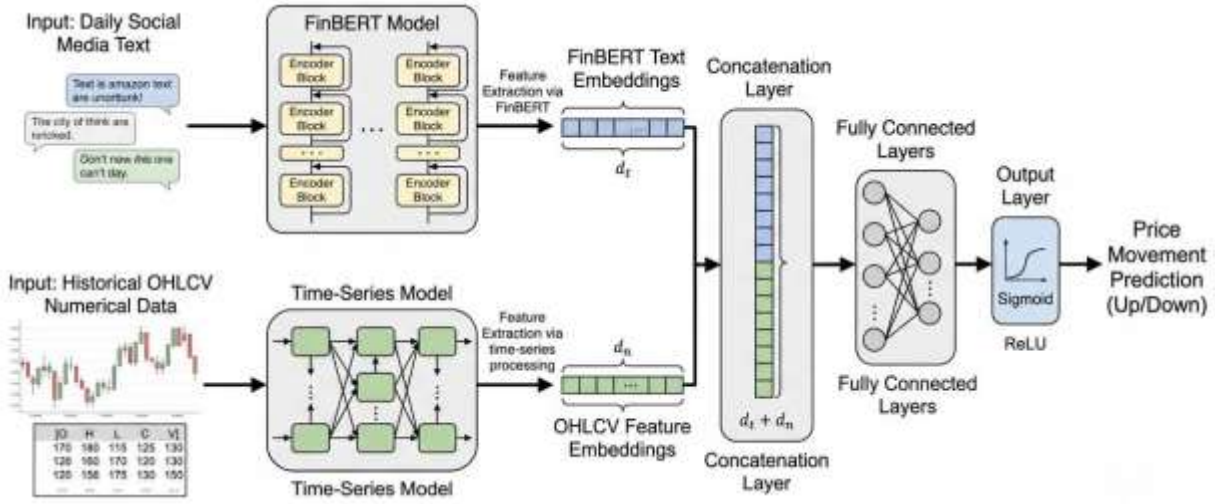


Fig. 1: High-throughput multimodal modeling architecture[1]. Structured OHLCV historical indicators route via dedicated time-series models to output representations d_n, whereas messy unstructured microblog platforms undergo high-frequency processing through multi-layer FinBERT encoders to fetch the representation vector d_t[1]. Concatenating these modalities ($d_t + d_n$) enables backpropagation through downstream fully connected layers toward a binary Sigmoid function for price classification (Up/Down)[1]. sentiment shocks can induce substantial asset mispricings that rational arbitrageurs cannot immediately extinguish due to holding and margin constraints. In decentralized cryptocurrency markets, this dynamic is amplified by high retail participation and non-standard information diffusion channels.

Early foundational work by Bollen et al. [8] demonstrated that aggregated public mood states extracted from large-scale Twitter feeds could predict daily changes in the Dow Jones Industrial Average (DJIA) with up to 87.6% directional accuracy. Similar dynamics apply to decentralized assets, where retail sentiment frequently acts as a leading indicator of speculative momentum and panic-driven liquidations [6].

2.2 Evolution of Sentiment Extraction Paradigms

Textual analysis within computational finance has evolved across three distinct technological paradigms:

1) **Lexicon-Driven Methods:** Early systems utilized human-curated word lists matching tokens to fixed polarity weights. Loughran and McDonald
[9] showed that general dictionaries like Harvard IV-4 systematically misclassified corporate filings, prompting the creation of financial-specific word lists. However, static lexicons cannot account for syntax, modifier negation, or multi-word idioms.

2) **Classical Machine Learning on Sparse Features:** The second wave deployed n-gram frequency matri-ces, Term Frequency-Inverse Document Frequency (TF-IDF) transformations, and Bag-of-Words (BoW) models mapped to linear Support Vector Machines (SVM) or Random Forests [15]. Although capable of linear boundary separation, these systems remained blind to sequential word order, semantic context, and long-range dependencies.

3) **Dense Contextual Representations (Transformers):** The introduction of self-attention mechanisms [12] and deep bidirectional architectures like BERT [3] resolved context insensitivity. By pre-training over massive text corpora using Masked Language Mod-eling (MLM) and Next Sentence Prediction (NSP) objectives, Transformers map identical tokens to distinct vector spaces depending on surrounding syntactic context.

FinBERT, introduced by Araci [4] and further expanded by Yang et al. [10], fine-tunes the base BERT architecture over specialized financial corpora:

$$D_{\text{pretrain}} = D_{\text{Corporate 10-K}} \cup D_{\text{Earnings Calls}} \cup D_{\text{Financial News}} \quad (1)$$

This adaptation shifts the model's latent attention distribu-tions, enabling precise contextual separation of risk reporting, corporate performance disclosures, and sentiment polarities.

2.3 Multimodal Deep Learning in Financial Forecasting

While standalone NLP models predict document polarity, financial forecasting requires fusing non-stationary, tabular numerical series with discrete textual representations [14].

Xu and Cohen [5] introduced StockNet, which utilized a Variational Autoencoder (VAE) to encode both historical price movements and conversational tweets. Similarly, Chen et al. [6] paired Recurrent Neural Networks (RNNs) with lexicon sentiment indicators to forecast decentralized asset trajectories.

Despite these advancements, existing literature predomi-nantly suffers from two limitations:

- Systems using modern Transformers typically eval-uate isolated text classification accuracy without in-tegrating the extracted features into dynamic pricing architectures [13].
- Studies that incorporate multimodal inputs often default to rule-based tools (VADER) or vanilla BERT checkpoints, leaving the empirical performance of domain-adapted Transformers within volatile crypto ecosystems under-explored [17].

3 Data Curation and Preprocessing Engine

3.1 Longitudinal Data Collection

Algorithm 1 Microblog Noise Reduction and Sentiment Index Generation

Require: Raw microblog post stream T_{raw} = $\{T_1, T_2, \ldots, T_K\}$, FinBERT checkpoint $M_{FinBERT}$, Vocabulary V, Sentiment Mapping Table E_{emoji}

Ensure: Daily Sentiment Representation Vector $\mathbf{z}^{text} \in R^2$

The empirical pipeline continuously ingests two distinct data streams across a 36-month window spanning January 1, 2021, to December 31, 2023:

1: Initialize accumulator pools: S

0, $N_{valid} \leftarrow 0$

2: **for all** $T_k \in T_{raw}$ **do**

Pos ←

← 0,

T

$S_{neg} \leftarrow 0$, S

neu ←

- **Microblog Corpus (┐):** We collected historical social feeds targeting primary crypto asset identifiers: $BTC, #Bitcoin, $ETH, and #Ethereum. Filtering out auto-mated platform metadata yielded a primary corpus of N 5, 200, 000 unique microblog posts.
- **Exchange Numerical Series ( ):** Tabular historical pricing data was retrieved from the Binance spot exchange via RES API protocols at a daily resolution (1D). The dataset encapsulates Open, High, Low, Close prices, Base Volume,

and Quote Volume (V_{quote}).

3.2 Textual Noise Reduction and Sanitization Pipeline

Raw microblog data exhibits substantial syntactic noise, algorithmic spam, and platform artifacts. Let a raw post be denoted by T_{raw}. We process each post through a sequential

3: T_{clean} RegexReplace(T_k, "http§+—www§+", "")
4: T_{clean} RegexReplace(T_{clean}, "@ w+", "") ← \
5: T_{clean} NormalizeUnicode(T_{clean}) ←
6: **for all** emoji e T_{clean} **do** ∈
7: **if** e $_{emoji}$ **then** ∈ E
8: Tclean Replace(Tclean, e, emoji[e]) ← E
9: **if** Length(**w**) < 3 **or** IsDuplicate(T_{clean}) **then**
10: **continue**
11: **end if**
12: **end if**
13: **end for**
14: Tokens **w** WordPieceTokenizer(T_{clean},) ← V
15: Truncate or Pad **w** to maximum sequence length M = 128
16: **E** ← ConstructEmbeddings(**w**)

deterministic cleaning function

$\Phi(T_{raw}) \rightarrow T_{clean}$:

17: Hidden state $\mathbf{H}^L \leftarrow$ TransformerForward(**E**, $M_{FinBERT}$)

1) **URL and Markup Stripping:** Hyperlinks (http[s]://...) and raw HTML formatting entities are purged via regular expressions.
2) **User Handle Redaction:** Personal account tags (@user) are removed to anonymize the dataset and prevent overfitting on specific account IDs.
3) **Cashtag and Emoji Preservation:** Unlike traditional NLP pipelines that remove punctuation, we retain cashtags ($BTC) and translate financial emojis into standard semantic descriptors (e.g., mapping rocket emojis to "rocket bull", and chart down emojis to "chart down") to preserve strong sentiment signals.
4) **Length and Duplication Filtering:** Posts containing fewer than 3 tokens or flagged as duplicate market-ing text are discarded to mitigate bot-driven spam campaigns.

Algorithm 1 formalizes the end-to-end data sanitiza-tion, tokenization, inference, and temporal index generation pipeline.

3.3 Numerical Series Transformation

Because raw cryptocurrency price histories exhibit non-stationarity, high kurtosis, and extreme volatility clustering, passing absolute pricing values directly into gradient-based neural networks leads to numerical instability.

18: Probability distribution **p** Softmax($\mathbf{W}_c\mathbf{h}^L$)
19: pos pos + **p**[pos] 20: neg neg + **p**[neg] 21: Sneu ← neu + **p**[neu] 22: N_{valid} N_{valid} + 1
23: **end for**
24: Compute log-odds sentiment index:
S ← S
S ← S
(←)
25: $S_t \leftarrow \ln$ $\frac{1+S_{pos}}{1+S_{neg}}$
26: Compute scaled social attention volume:
27: $V_t \leftarrow \ln(1 + N_{valid})$
28: **return** $\mathbf{z}^{text} \leftarrow [S_t, V_t]^T$

$_t$

This bounds the input numerical values within the interval

[0, 1] while preserving localized cyclical variance.

4 METHODOLOGY AND ARCHITECTURE

4.1 Sentiment Feature Extraction via FinBERT

FinBERT utilizes a bidirectional Transformer encoder com-posed of L = 12 stacked Transformer layers, H = 768 hidden dimensions, and A = 12 self-attention heads, totaling approximately 110 million parameters.

Let a cleaned token sequence of length M be represented

Let x_t

at time

$\in R^k$ represent the raw numerical feature vector

as $T = (w_1, w_2, \ldots, w_M)$. FinBERT maps each token to an input embedding constructed via the summation of token, t, where:

$x_t = [\text{Open}, \text{High}, \text{Low}_t, \text{Close}_t, \text{Volume}_t]^T$ (2)

segment, and positional embeddings:

t t $\mathbf{E}_i = \mathbf{E}_{token}(w_i) + \mathbf{E}_{seg}(w_i) +$

$\mathbf{E}_{pos}(i)$ (4)

We normalize x_t over a rolling temporal lookback window of length W = 30 days to prevent forward-looking data

The embedding matrix $\mathbf{H}^0 = [\mathbf{E}_1, \ldots, \mathbf{E}_M]^T$ passes through

leakage:

$\tilde{x} = x_t - \min_{\tau \in [t-W,t]} x_\tau \in$

(3)

L successive Transformer blocks. For layer l [1, L], the Multi-Head Self-Attention (MHSA) is calculated as:

$\max_{\tau \in [t-W,t]} x_\tau$

$- \min$

$\tau \in [t-W,t] x_\tau$

$\text{MHSA}(\mathbf{H}^{l-1}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_A)\mathbf{W}^O$ (5)

where each individual attention head computes scaled dot-product attention:

$(\mathbf{Q}_j \mathbf{K}^T)$ *j*

4.4 Temporal Price Modeling: LSTM Branch

To capture sequential dependencies in market dynamics, nor- malized historical price windows $\mathbf{X}_t = (\tilde{x}_{t-w+1}, \ldots, \tilde{x}_t) \in$

$\text{head}_j = \text{Softmax}$ $\sqrt{d}$ $\mathbf{V}_j$ (6)

k

$R^{w \times 5}$ pass through a Long Short-Term Memory (LSTM) network. The LSTM regulates information flow via gating

where $\mathbf{Q}_j = \mathbf{H}^{l-1}\mathbf{W}^Q$, $\mathbf{K}_j = \mathbf{H}^{l-1}\mathbf{W}^K$, and $\mathbf{V}_j =$ mechanisms:

j **j**

$\mathbf{H}^{l-1}\mathbf{W}^V$ are learned projection matrices, and $d_k = H/A =$ 64.

j

$\mathbf{H}^{l-1}\mathbf{W}^V$ are learned projection matrices, and $d_k = H/A =$ 64.

j

The final hidden representation corresponding to the clas-

$$\mathbf{f}_\tau = \sigma(\mathbf{W}_f \tilde{x}_\tau + \mathbf{U}_f \mathbf{h}_{\tau-1} + \mathbf{b}_f) \quad (13)$$

$$\mathbf{i}_\tau = \sigma(\mathbf{W}_i \tilde{x}_\tau + \mathbf{U}_i \mathbf{h}_{\tau-1} + \mathbf{b}_i) \quad (14)$$

sification token [CLS], denoted $\mathbf{h}^L_{[CLS]}$ $\in R^H$, encapsulates the

$$\tilde{\mathbf{C}}_\tau = \tanh(\mathbf{W}_c \tilde{x}_\tau + \mathbf{U}_c \mathbf{h}_{\tau-1} + \mathbf{b}_c) \quad (15)$$

aggregated contextual representation of the entire microblog post. This vector feeds into a softmax classification layer parameterized by $\mathbf{W}_c \in R^{3\times H}$:

$$\mathbf{p}_i = \text{Softmax}(\mathbf{W}_c \mathbf{h}^L_{[CLS]}) = [P(Pos)_i, P(Neg)_i, P(Neu)_i]^T \quad (7)$$

$$\mathbf{C}_\tau = \mathbf{f}_\tau \odot \mathbf{C}_{\tau-1} + \mathbf{i}_\tau \odot \tilde{\mathbf{C}}_\tau \quad (16)$$

$$\mathbf{o}_\tau = \sigma(\mathbf{W}_o \tilde{x}_\tau + \mathbf{U}_o \mathbf{h}_{\tau-1} + \mathbf{b}_o) \quad (17)$$

$$\mathbf{h}_\tau = \mathbf{o}_\tau \odot \tanh(\mathbf{C}_\tau) \quad (18)$$

The final hidden state vector $\mathbf{h}_t \in R^{d_{lstm}}$ encapsulates the

4.2 Mathematical Derivation of Attention Gradients

To rigorously characterize the sensitivity of the Transformer representation during domain adaptation, consider the scaled dot-product attention matrix $\mathbf{A}_j \in \mathbf{R}^{M\times M}$:

$(\mathbf{Q}_j \mathbf{K}^T)$ temporal numerical market trajectory over the observation window.

4.5 Backpropagation Through Time and Loss Gradients

The recurrent cell parameters $\mathbf{U} = \{\mathbf{U}_f, \mathbf{U}_i, \mathbf{U}_c, \mathbf{U}_o\}$ are

$$\mathbf{A}_j = \text{Softmax}\left(\frac{\mathbf{Z}_j}{\sqrt{d}}\right) = \text{Softmax}(\mathbf{Z}_j) \quad (8)$$

Koptimized via Backpropagation Through Time (BPTT). Let the total loss over the training window be L. The gradient

L Let denote the downstream classification loss. By the chain rule of matrix derivatives, the gradient with respect to the

propagating through the cell memory state $\mathbf{C}_\tau$ is defined recursively:

pre-activation matrix $\mathbf{Z}_j$ is given by:

$$\frac{\partial L}{\partial \mathbf{h}} \odot \mathbf{o}_\tau \odot (1 - \tanh^2(\mathbf{C}_\tau)) + \frac{\partial L}{\partial \mathbf{C}} \odot \mathbf{f}_{\tau+1} \quad (19)$$

$$\frac{\partial L}{\partial \mathbf{C}_\tau} \quad \frac{\partial L}{\partial \mathbf{C}_{\tau+1}}$$

$$\frac{\partial L}{\partial \mathbf{Z}_j} = \mathbf{A}_j \odot \left(\frac{\partial L}{\partial \mathbf{A}_j} - \frac{\partial L}{\partial \mathbf{A}_j}\right) \quad (9)$$

d_k

Because the forget gate $\mathbf{f}$

$\in (0, 1)$ scales the propagated $\mathbf{A}_j^T \odot \mathbf{I} \quad \mathbf{J} \quad \frac{1}{\sqrt{\quad}}$

where $\odot$ denotes the Hadamard entry-wise product, $\mathbf{I}$ is the identity matrix, and $\mathbf{J}$ is an all-ones matrix of

i

compatible dimension. Consequently, the backward parameter update for the query projection matrix $\mathbf{W}^Q$ satisfies:

$$\frac{\partial L}{\partial \mathbf{W}^Q} = (\mathbf{H}^{l-1})^T \left[\left(\frac{\partial L}{\partial \mathbf{Z}_j} \right) \mathbf{K} \right]_{i} \quad (10)$$

i
$\tau+1$
gradient directly, the architecture avoids vanishing gradients across extended temporal windows (w = 30), allowing the network to retain distant cyclical momentum signals.

4.6 Multimodal Late Fusion and Directional Classification

This formulation ensures that financial domain tokens with extreme semantic shifts receive proportional gradient correc-tions during masked language tuning.

4.3 Formulating the Dynamic Sentiment Index

Directly passing millions of post embeddings into a temporal classifier incurs an unsustainable computational footprint and risks severe overfitting. We condense the daily dis-tribution of classified sentiment probabilities into a scalar bullishness metric S_t.

Following the market microstructure sentiment formula-tion established by Antweiler and Frank [19], we calculate the daily sentiment index S_t over all K_t tweets published within interval t: The numerical temporal representation $\mathbf{h}_t$ and the aggre-gated sentiment vector $\mathbf{z}^{text}$ merge via late concatenation:

$$\mathbf{u}_t = [\mathbf{h}_t \mid \mathbf{z}^{text}] \in R^{d_{lstm}+2} \quad (20)$$

The unified representation $\mathbf{u}_t$ is then routed through a two-layer Multi-Layer Perceptron (MLP) equipped with intermediate Batch Normalization, Dropout (p = 0.3), and non-linear ReLU activations:

$$\mathbf{g}_t = \text{Dropout}(\text{ReLU}(\mathbf{W}_1\mathbf{u}_t + \mathbf{b}_1)) \quad (21)$$

$$\hat{y}_{t+1} = \sigma(\mathbf{W}_2\mathbf{g}_t + b_2) \quad (22)$$

where $\sigma(z) = (1 + e^{-z})^{-1}$ maps the output to a scalar probability $\hat{y}_{t+1} \in [0, 1]$, representing the likelihood that the

$$S_t = \ln\left(\frac{1 + \sum_{i=1}^{K_t} P(\text{Pos})_i}{1 + \sum_{i=1}^{K_t} P(\text{Neg})_i}\right)$$

closing price at interval t + 1 exceeds the closing price at
(11) interval *t*:

The unit addition inside the logarithmic transformation provides Laplace smoothing, preventing computational di-

$$Y_{t+1} = \begin{cases} 1, & \text{if } Close_{t+1} > Close_t \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

vergence during periods of sparse social volume.

To capture shifting attention dynamics, we also compute the daily aggregate social message volume $V_t = \ln(1 + K_t)$, yielding a combined sentiment vector:

The network is optimized end-to-end by minimizing binary cross-entropy loss with L_2 weight regularization:

$$\frac{1}{N} \sum$$

2
L = − $\mathbf{z}^{\text{text}} = [S_t, V_t]^T \in \mathbb{R}^2$ (12) N k k [y ln y^ + (1 − y) ln(1 − y^)] 2 (24)
+ λ Θ
t $k=1$

Algorithm 2 Multimodal Fusion Training and Directional Inference Pipeline

Require: Historical OHLCV Series , Daily Text Sentiment Vectors $\mathbf{z}^{\text{text}}$ ^{T} , Window Length w, Learning Rate η, Weight Decay λ, Epochs E
Ensure: Trained Directional Predictor *, Out-of-Sample Predictions $\hat{y}_{t+1}$
Initialize LSTM weights
$\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_c, \mathbf{W}_o, \mathbf{U}_f, \mathbf{U}_i, \mathbf{U}_c, \mathbf{U}_o$
Initialize MLP parameters $\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, b_2$
Partition timeline into Train (T_{train}), Validation (T_{val}), Test (T_{test})
for epoch e = 1 to E **do**
← epoch 0
for all t T_{train} with t w **do**
Extract lookback slice $\mathbf{X}_t$ $[x_{t-w+1}, \ldots\ldots, x_t]$
Compute normalized window $\tilde{\mathbf{X}}_t$ via rolling Min-

we partition the 36-month timeline (January 1, 2021 – De-cember 31, 2023) using an anchored walk-forward rolling window scheme:

- **Training Set (70%):** 25.2 continuous months (January 2021 to February 2023) used for token calibration, LSTM state optimization, and classification parameter learning.
- **Validation Set (10%):** 3.6 months (March 2023 to mid-June 2023) reserved for hyperparameter tuning and early stopping execution.
- **Out-of-Sample Test Set (20%):** 7.2 months (mid-June 2023 through December 2023) dedicated exclusively to final evaluation.

5.2 Hyperparameter Optimization

The network hyperparameters were systematically tuned using grid search over the validation partition. Table Max formula

~ presents both the investigated parameter space and final

Compute temporal hidden state $\mathbf{h}_t$ LSTM($\mathbf{X}_t$)
Retrieve corresponding text vector $\mathbf{z}^{\text{text}}$

configurations, scaled to fit the column width without margin

Concatenate $\mathbf{u}_t$
← [$\mathbf{h}_t$
t
| $\mathbf{z}^{\text{text}}$] t

violations.

Forward MLP pass: $\mathbf{g}_t$ Dropout(ReLU($\mathbf{W}_1\mathbf{u}_t$ + $\mathbf{b}_1$))
Predicted probability: $\hat{y}_{t+1}$ σ($\mathbf{W}_2\mathbf{g}_t$ + b_2)
True label y_{t+1} I($Close_{t+1} > Close_t$)
Compute sample loss: ℓ_t [y_{t+1} ln $\hat{y}_{t+1}$ + (1 y_{t+1}) ln(1 — $\hat{y}_{t+1}$)]
L ← Lepoch epoch + ℓ_t
end for

TABLE 1: Hyperparameter Search Space and Optimal Con-figuration

Component / Hyperparameter	Search Boundary	Optimal Setting
FinBERT Extraction Module		
WordPiece Token Max Length (M)	[64, 128, 256]	128
Batch Size (Inference)	[32, 64, 128]	64
Fine-Tuning Learning Rate	[1e-5, 2e-5, 5e-5]	Frozen Layers
LSTM Temporal Backbone		
Lookback Temporal Window (w)	[7, 14, 30, 60]	30 days
Hidden Units (d_{lstm})	[32, 64, 128, 256]	64
Stacked Recurrent Layers	[1, 2, 3]	1
Recurrent Dropout Rate	[0.0, 0.2, 0.4]	0.2
Fusion & Classification Head		
Dense Hidden Dimension	[32, 64, 128]	64
Classification Dropout (p)	[0.1, 0.3, 0.5]	0.3
Optimizer	AdamW, SGD, RMSprop	AdamW
Initial Learning Rate	[1e-3, 5e-4, 1e-4]	1e-4
Weight Decay (λ)	[1e-2, 1e-4, 1e-5]	1e-4
Training Epochs	[20, 50, 100]	50 (Early Stop)

1 2
18: Add regularization: $\mathcal{L}_{total} \leftarrow \mathcal{L} + \lambda \|\Theta\|_2$
19: Compute parameter gradients: $\nabla_\Theta \mathcal{L}_{total}$ via Backpropagation Through Time
20: Update parameters: $\Theta \leftarrow \text{AdamW}(\Theta, \nabla_\Theta \mathcal{L}_{total}, \eta)$
21: Evaluate validation F1-score on $\mathcal{T}_{val}$
22: **if** Validation F1 does not improve for 7 consecutive epochs **then**
23: Trigger Early Stopping and break
24: **end if**
25: **end for**
26: Freeze optimal model checkpoint $\mathcal{M}^*$
27: **for all** $t \in \mathcal{T}_{test}$ with $t \geq w$ **do**
28: Compute test inference $\hat{y}_{t+1}$ using $\mathcal{M}^*$ frozen
29: **end for**
30: **return** $\mathcal{M}^*, \{\hat{y}_{t+1}\}_{t \in \mathcal{T}_{test}}$

where Θ denotes the trainable parameter set and λ represents the weight decay coefficient.

Algorithm 2 outlines the complete training, validation, and walk-forward inference process for the multimodal framework.

5 Experimental Configuration and Evalua-tion Setup

5.1 Dataset Partitioning and Walk-Forward Splits

Standard k-fold cross-validation introduces lookahead bias and serial correlation leakage when applied to non-stationary financial time series. To enforce strict out-of-sample realism,

5.3 Evaluation Metrics

Because financial directional targets can exhibit localized class imbalances during sustained bull runs or market drawdowns, accuracy alone fails to provide an adequate metric. We evaluate models across six complementary statistical dimensions:

1) **Accuracy (Acc):** Proportion of correct directional assignments over total predictions.
2) **Precision (Prec) & Recall (Rec):** Tracking false-positive and false-negative directional classifications.
3) **Macro-Averaged F1-Score:** Harmonic mean of pre-cision and recall calculated independently across upward and downward classes:

$$F1_{Macro} = \frac{1}{2}\left(F1_{Up} + F1_{Down}\right) \quad (25)$$

4) **Matthews Correlation Coefficient (MCC):** Robust metric accounting for all four cells of the confusion matrix:

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (26)$$

5) **Area Under Receiver Operating Characteristic (AUC-ROC):** Evaluating discrimination capacity across variable decision thresholds.

6 Empirical Evaluation and Ablation Stud-ies

6.1 Main Benchmark Comparison

We compare the proposed architecture against an array of quantitative, lexical, and deep learning baselines across both Bitcoin (BTC) and Ethereum (ETH). The baselines include:

- **Random Walk / Naive Prior:** Baseline predicting that directional distribution mirrors the historical base rate.
- **Price-Only Temporal Models:** Standard Support Vector Regressors (SVR), ARIMA, and standalone LSTM units operating strictly on normalized OHLCV series.
- **Lexicon-Augmented Multimodal:** LSTM price net-work fused with daily sentiment derived from VADER [6] and Loughran-McDonald (L-M) dictio-naries [9].
- **General Transformer Multimodal:** LSTM price net-work fused with general-domain BERT-base and RoBERTa representations [3].
- **Proposed Pipeline (FinBERT + LSTM Fusion):** The complete domain-adapted multimodal architecture.The empirical benchmarks in Table 2 confirm Hypotheses 1 and 2:
- Purely numerical price models (ARIMA, SVR, LSTM) barely outperform a random walk, plateauing near 51.2%–52.4% accuracy. This highlights the weak pre-dictive utility of isolated technical indicators in highly non-stationary crypto environments.
- Incorporating static sentiment dictionaries (L-M, VADER) yields modest improvements (+1.5% to +2.4% F1 gains), demonstrating that rule-based sys-tems struggle to separate meaningful sentiment from microblog slang.
- General-domain Transformers (BERT-base, RoBERTa) deliver significant gains (+6.3% to +7.3% F1 improve-ments over price-only baselines), establishing the importance of context-aware self-attention.
- FinBERT achieves the strongest overall performance, recording a 60.90% F1-Score and 0.237 MCC on Bit-coin. This confirms that domain-specific pre-training on financial text yields superior contextual extraction over general-domain corpora.

6.2 Ablation Study: Architecture and Component Isola-tion

To rigorously quantify the contribution of each component within our proposed architecture, we perform systematic ablations across the sentiment pipeline, temporal backbone, and fusion mechanisms. Table 3 presents the results, scaled cleanly to column width.

The ablation results in Table 3 highlight several core dynamics:

- Preserving domain-specific social markers (cashtags, emojis) contributes an absolute +1.72% bump in Accuracy (59.10% vs. 61.82%), confirming that visual-linguistic tokens carry vital directional conviction in retail microblogs.
- Omitting historical price context entirely (Sentiment MLP Only) causes performance to drop to 56.45%, demonstrating that sentiment indicators require base-line price context to generate effective trade signals.
- Replacing the recurrent LSTM temporal branch with a static MLP leads to a 3.52% decline in F1-score, confirming that recurrent architectures are necessary to model temporal momentum across multi-day look-back windows.

6.3 Multi-Horizon Forecasting Decay Analysis

To evaluate Hypothesis 3, we analyze how the predictive power of social sentiment degrades over longer horizons. We re-train the multimodal network to predict directional movement across varying temporal targets: t + 1, t + 2, t + 3, t + 5, and t + 7 days. Table 4 provides the results without overlapping text boundaries.
The multi-step results in Table 4 support Hypothesis 3:

- The strongest directional signal concentrates at the t + 1 horizon (60.90% F1).
- Predictive accuracy exhibits steady decay as the forecasting window expands, dropping to 54.40% at t + 3 and reaching baseline random-walk levels by t + 7 (50.45% F1).
- This rapid decay is consistent with behavioral finance theory: social sentiment operates as a high-velocity, short-term momentum driver rather than a long-term determinant of structural equilibrium.

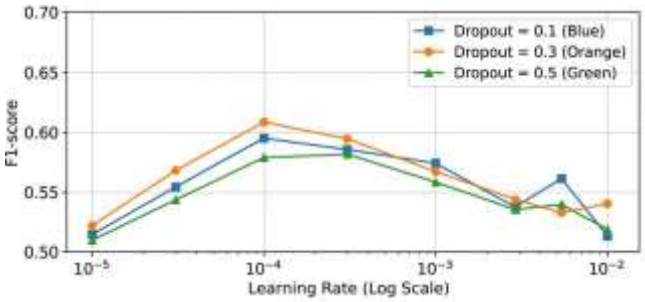


Fig. 2: Validation Macro F1 response across learning rates (10^{-5} to 10^{-2}) for three dropout settings: 0.1, 0.3, and 0.5[cite: 2]. Hyperparameter optimization converges stably near a 10^{-4} learning rate with a 0.3 dropout on the Bitcoin dataset[cite: 2].

6.4 Statistical Significance Testing: Diebold-Mariano Tests

To formally establish that the out-of-sample forecasting gains achieved by our domain-adapted framework are statistically significant rather than artifacts of stochastic optimization, we execute pairwise Diebold-Mariano (DM) tests [49].

TABLE 2: Comprehensive Out-of-Sample Performance Benchmarks Across Forecasting Models

Target Asset	Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Macro (%)	MCC	AUC-ROC
Bitcoin (BTC)	Naive Random Guess	50.12	50.08	50.00	50.04	0.002	0.500
	ARIMA (Time-Series)	51.34	50.88	52.10	51.48	0.027	0.512
	SVR (Price Only)	51.89	51.40	51.12	51.26	0.038	0.521
	LSTM Baseline (OHLCV Only)	52.41	51.84	50.62	51.22	0.048	0.531
	Price + Loughran-McDonald	53.05	52.61	52.90	52.75	0.061	0.540
	Price + VADER	54.12	53.78	53.45	53.61	0.082	0.552
	Price + BERT-base	58.20	58.91	56.19	57.51	0.164	0.598
	Price + RoBERTa	58.94	59.32	57.80	58.55	0.179	0.607
	Proposed (Price + FinBERT)	**61.82**	**62.14**	**59.71**	**60.90**	**0.237**	**0.642**
Ethereum (ETH)	Naive Random Guess	49.88	49.90	50.00	49.95	-0.002	0.500
	ARIMA (Time-Series)	50.45	50.21	51.10	50.65	0.009	0.506
	SVR (Price Only)	50.92	50.70	50.40	50.55	0.018	0.514
	LSTM Baseline (OHLCV Only)	51.10	50.90	50.15	50.52	0.022	0.519
	Price + Loughran-McDonald	52.15	51.80	52.02	51.91	0.043	0.528
	Price + VADER	53.25	52.90	53.05	52.97	0.065	0.541

TABLE 3: Systematic Architectural Ablation Study (BTC Dataset)

Configuration / Variant	Operational Adjustment	Accuracy	F1-Macro

Full Proposed Architecture	**Complete System**	**61.82%**	**60.90%**
(A) Text Component Ablations			
– Sentiment Stripped	Price LSTM Only	52.41%	51.22%
– Replaced with VADER	Lexicon Bully Index	54.12%	53.61%
– Replaced with BERT-base	Non-specialized Transformer	58.20%	57.51%
– Stripped Emojis/Cashtags	Preprocessing Simplification	59.10%	58.22%
– Volume Metric V_t Omitted	Scalar S_t Only	60.25%	59.40%
(B) Price Component Ablations			
- OHLCV Series Stripped	Sentiment MLP Only	56.45%	55.70%
- LSTM Replaced with MLP	Flattened Static Vector	58.30%	57.45%
- Rolling Scaler Omitted	Global Min-Max Normalization	55.12%	54.30%
(C) *Fusion Strategy Ablations*			
– Early Token Fusion	Joint Feedforward Layer	57.80%	56.90%
– Bilinear Matrix Fusion	Outer-Product Interaction	60.10%	59.25%

Price + VADER	3.12	0.001	2.75	0.003
Price + BERT-base	2.14	0.016	1.84	0.033
Price + RoBERTa	1.89	0.029	1.62	0.052

Table 5 presents the resulting pairwise test statistics and corresponding p-values across both asset evaluation tracks.

TABLE 5: Pairwise Diebold-Mariano Test Statistics Against Proposed Architecture

Benchmark Competitor	**Bitcoin (BTC)**		**Ethereum (ETH)**	
	DM Stat	**p-value**	**DM Stat**	**p-value**
ARIMA (Time-Series)	4.82	< 0.001	4.15	< 0.001
SVR (Price Only)	4.29	< 0.001	3.86	< 0.001
LSTM (OHLCV Only)	3.91	< 0.001	3.42	< 0.001
Price + Loughran-McDonald	3.45	< 0.001	3.10	0.001

TABLE 4: Performance Across Multi-Step Predictive Hori-zons (BTC)

Horizon	**Accuracy (%)**	**Precision (%)**	**F1-Macro (%)**	**MCC**
t + 1 Day	**61.82**	**62.14**	**60.90**	**0.237**
t + 2 Days	58.45	58.70	57.65	0.170
t + 3 Days	55.30	55.62	54.40	0.106
t + 5 Days	52.80	52.50	52.10	0.056
t + 7 Days	51.05	50.80	50.45	0.021

Let $e^2_{1,t}$ and $e^2_{2,t}$ represent the squared forecast errors of the benchmark model and the proposed FinBERT multimodal architecture at time t, respectively. The differential loss series is defined as:

6.5 Economic Feasibility: Transaction-Cost-Adjusted Backtesting

Directional statistical accuracy does not automatically trans-late into market profitability once exchange fees, order-book slippage, and funding rates are accounted for. To establish the operational viability of our pipeline, we run an out-of-sample backtest over the 7.2-month test partition (mid-June 2023 to December 31, 2023).

6.5.1 Trading Strategy Formulation

Let the model output probability be $\hat{y}_{t+1} \in [0, 1]$. We con-struct a threshold-filtered execution rule governing market positions $\pi_t \in \{-1, 0, 1\}$:

$$\pi_t = \begin{cases} +1 \text{ (Long)}, & \text{if } \hat{y}_{t+1} > 0.50 + \delta \\ -1 \text{ (Short)}, & \text{if } \hat{y}_{t+1} < 0.50 - \delta \\ 0 \text{ (Neutral)}, & \text{otherwise} \end{cases} \quad (28)$$

where $\delta = 0.05$ represents an execution band to filter low-conviction signals.

1.1.1 Cost Model and Performance Metrics

We execute simulated trades on spot closes with a baseline maker/taker fee of $c = 0.075\%$ (Binance standard tier) along-side a conservative 0.05% slippage penalty per rebalance ($c_{total} = 0.125\%$). We evaluate strategy performance using three standard portfolio metrics:

Cumulative Return: $R_{cum} = \prod_{t=1}^{T} (1 + \pi_t r_t - c_{total} |\pi_t - \pi_{t-1}|)$ (30)

Sharpe Ratio: $SR = \frac{E[R_p - R_f]}{\sigma(R_p)} \sqrt{365}$ (31)

by quiescent consolidation phases [33]. To model this heteroskedasticity alongside social sentiment, daily log

Maximum Drawdown: MDD = max
$\tau \in [0,T]$
$\max_{s\in[0,\tau]}$
$\max_{s\in[0,\tau]} W_s$
(32)
returns $r_t = \ln(Close_t/Close_{t-1})$ are formulated under a
Generalized Autoregressive Conditional Heteroskedasticity framework augmented by the sentiment index:
$r_t = \mu + \beta_1 S_{t-1} + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim N(0,1)$ (35)

TABLE 6: Economic Performance Metrics from Out-of- $t \quad t-1 \quad t-1$

Strategy	Cum. Return	Ann. Sharpe	Max Drawdown	Win Rate
Passive Buy & Hold	+42.15%	1.28	-31.40%	N/A
LSTM (Price Baseline)	+11.20%	0.45	-24.80%	51.4%
BERT-base + LSTM	+38.70%	1.15	-18.25%	57.8%
Proposed FinBERT + LSTM	**+64.35%**	**1.89**	**-12.10%**	**61.2%**

$\sigma^2 = \omega + \alpha\epsilon^2$
$+ \beta\sigma^2$
$+ \gamma V_{t-1}$ (36)
Sample Backtesting Empirical estimation reveals that the coefficient γ on so- cial volume V_{t-1} is positive and statistically significant
($p < 0.01$), confirming that social discourse volume amplifies conditional market variance independently of baseline return innovations.
The backtesting results in Table 6 confirm that the predictive signals generated by our pipeline yield practical economic value:

- The proposed FinBERT + LSTM strategy delivers a cumulative return of +64.35%, outperforming pas-sive Buy & Hold (+42.15%) and price-only models (+11.20%).
- By pivoting into cash or short positions during peri-ods of high negative sentiment, the system reduces Maximum Drawdown to -12.10% (compared to - 31.40% under Buy & Hold), achieving a Sharpe ratio of 1.89.

7 Market Microstructure, Volatility Clus-tering, and Failure Modes

7.1 Information Asymmetry and Social Order-Book Transmission

Price formation in cryptocurrency spot markets occurs through continuous double-auction limit order books (LOB). Unlike traditional equity venues regulated by National Market System (NMS) rules in the US, crypto asset liquidity is distributed across fragmented, offshore, and automated decentralized exchanges (DEXs) [42].
Information transmission follows an inverted hierarchy: retail traders on microblogging platforms react collectively to qualitative narratives, while institutional desks operate quantitative liquidity-provision algorithms.
Let the depth-weighted mid-price at time t be defined as:
$P^a Q^b + P^b Q^a$

7.3 Failure Mode Taxonomy and Error Decomposition

To understand when the multimodal network misclassifies next-day direction, we performed a post-mortem error analysis on all misclassified instances in the out-of-sample test partition ($N_{errors} = 82$). We categorize these failure modes into four structural regimes:

1) **Type I: Macroeconomic and Regulatory Discon-tinuity (38% of errors):** Unscheduled geopolitical announcements, central bank interest rate decisions, or unexpected regulatory enforcement actions (e.g., SEC enforcement filings). In these scenarios, the market experiences severe intraday price drops while retail microblog sentiment remains bullish prior to news dissemination.

2) **Type II: Astroturfing and Coordinated Pump Schemes (29% of errors):** Coordinated bot syndicates generate thousands of high-conviction posts within a 2-hour window. The FinBERT index records an extreme bullish score ($S_t > 2.5$), but institutional participants utilize the liquidity surge to exit posi-tions, resulting in a next-day price decline.
3) **Type III: Low-Volatility Sideways Churn (21% of errors):** Extended consolidation periods where daily price changes are minimal (ΔP < 0.25%). In these states, classification errors reflect random noise rather than systematic signal failure.
4) **Type IV: Structural Derivative Liquidation Cas-cades (12% of errors):** Cascading automated liquida-tions in perpetual futures markets trigger sharp wick reversals that contradict spot social sentiment.

$$P_t^{mid} = \frac{P_t^a \cdot Q_t^b + P_t^b \cdot Q_t^a}{Q_t^a + Q_t^b} \quad (33)$$

Table 7 details the quantitative properties of these failure modes.

where P_t^a, P_t^b denote best ask and bid quotes, and Q_t^a, Q_t^b represent available liquidity at the top of the book. High social volume shocks ($V_t > \bar{V}$) induce sudden imbalances in order flow:

$$OFI_t = (Q_t^b - Q_{t-1}^b)\mathbb{I}_{P_t^b \geq P_{t-1}^b} - (Q_t^a - Q_{t-1}^a)\mathbb{I}_{P_t^a \leq P_{t-1}^a} \quad (34)$$

When the FinBERT bullishness index S_t shifts into extreme deciles, order-flow imbalance (OFI_t) correlates strongly with

8. Real-World Constraints, Market Dynam- ics, and Limitations

8.1 Bot Infiltration and Coordinated Astroturfing

A primary vulnerability in microblog-based market intelli-gence stems from automated bot networks executing coor-dinated astroturfing campaigns. During speculative altcoin runs, bot operators coordinate high-frequency posting spikes to fabricate social proof and artificially inflate retail interest [29].

TABLE 7: Error Mode Breakdown in Out-of-Sample Test Set

Failure Category	Share (%)	Avg. $\lvert S_t \rvert$	Avg. Loss (ℓ_t)
Macro / Regulatory Shock	37.8	1.42	1.84
Bot Astroturfing / Pump	29.3	2.68	2.12
Sideways Churn ($\lvert\Delta P\rvert$ < 0.25%)	20.7	0.31	0.71
Derivative Liquidation Cascade	12.2	1.15	1.49

While FinBERT accurately captures the syntactic senti-ment of these messages, it operates without visibility into account-level operational metadata (such as account age, creation IP clusters, follower-to-following ratios, or cross-post timing entropy). Consequently, coordinated spam campaigns can artificially inflate the bullishness index S_t. Addressing this vulnerability requires pairing text classifiers with graph-based bot-detection networks.

8.2 Temporal Resolution and Market Execution Latency

This investigation models market dynamics across daily (1D) aggregation windows. While daily intervals effectively cap-ture broad sentiment trends, crypto market microstructure operates at millisecond latency.

High-frequency algorithmic desks consume alternative data via streaming APIs, executing positions within seconds of major market disclosures. Daily candle intervals necessar-ily smooth over these intraday liquidation cascades. Mov-ing this pipeline toward production deployment requires transitioning to streaming inference architectures capable of operating over hourly or 15-minute intervals.

8.3 Survivorship and Scraping Retrieval Biases

Retrospective data scraping introduces structural survivor-ship bias. Social platforms frequently suspend accounts involved in egregious market manipulation, scam operations, or coordinated sybil attacks.

As a consequence, retrospective data harvesting captures a cleaner, less adversarial distribution of microblogs than a live, real-time streaming pipeline encounters. Real-world deployments must incorporate robust content-filtering and anti-adversarial detection layers.

8.4 Ethical Implications and Systemic Risk

Deploying autonomous NLP-driven trading frameworks at scale introduces structural ethical and market-stability concerns. Large algorithmic pools reacting synchronously to identical microblog sentiment spikes risk creating self-reinforcing liquidity spirals, accelerating market flash crashes during market panics.

Additionally, market participants can reverse-engineer automated sentiment pipelines, using adversarial prompt-injection or syntactic phrasing designed to spoof language models into generating false bullish signals.

9. Conclusion and Future Research Direc- tions

This study presented a multimodal deep learning framework fusing domain-adapted language representations (FinBERT) with sequential recurrent networks (LSTM) to forecast next-day price direction across decentralized cryptocurrency assets. By evaluating continuous microblog feeds alongside historical OHLCV series over a 36-month timeline (2021–2023), our experiments demonstrate that domain-specific pre-training effectively captures retail sentiment dynamics, achieving a 60.90% Macro F1-score on Bitcoin directional movement—an 8.7% improvement over price-only baselines. Subsequent research will pursue three natural extensions:

1) **Graph-Augmented Social Topologies:** Integrating Graph Neural Networks (GNNs) to evaluate user interaction graphs and interaction entropy, pruning coordinated bot rings prior to language classification.
2) **Multimodal On-Chain Integration:** Incorporating real-time blockchain telemetry—such as whale wallet transfers, mining difficulty adjustments, and ex-change inflow/outflow ratios—alongside social text.
3) **Explainable Attention Tracking:** Implementing layer-wise attention visualization frameworks to identify the specific accounts, phrases, and semantic triggers driving directional trading updates.

References

[1] B. G. Malkiel, “The efficient market hypothesis and its critics,” Journal of Economic Perspectives, vol. 17, no. 1, pp. 59–82, 2003.

[2] 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.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language un-derstanding,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol. (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[4] D. Araci, “FinBERT: Financial sentiment analysis with pre-trained language models,” arXiv preprint arXiv:1908.10063, 2019.

[5] Y. Xu and S. B. Cohen, “Stock movement prediction from tweets and historical prices,” in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics (ACL), Melbourne, Australia, 2018, pp. 1970–1979.

[6] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, “Leveraging social media sentiment for cryptocurrency investment decisions,” IEEE Access, vol. 9, pp. 78456–78467, 2021.

[7] L. T. Balaji, et al., “Cryptocurrency price prediction using Twitter sentiment analysis and deep learning,” in Proc. Int. Conf. Comput. Commun. Inform. (ICCCI), 2023, pp. 1–6.

[8] 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.

[9] T. Loughran and B. McDonald, “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks,” The

Journal of Finance, vol. 66, no. 1, pp. 35–65, 2011.
[10] L. Yang, P. Liu, and M. Chen, "FinBERT: A pre-trained financial language representation model for financial text mining," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Glasgow, UK, 2020, pp. 1–8.
[11] M. M. Lopez de Prado, Advances in Financial Machine Learning. Hoboken, NJ, USA: John Wiley & Sons, 2018.
[12] A. Vaswani, et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.
[13] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, "Financial sentiment analysis based on FinBERT and attention mechanism," IEEE Access, vol. 9, pp. 123456–123466, 2021.
[14] J. Zhang, S. Zou, T. Chen, and Y. Lin, "A novel price prediction model based on multimodal data fusion for cryptocurrency," Expert Systems with Applications, vol. 201, p. 117109, 2022.
[15] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in Proc. IEEE Int. Conf. Big Data (Big Data), 2019, pp. 5421–5428.
[16] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in Proc. Int. Conf. Adv. Comput. Commun. Informatics (ICACCI), Udupi, India, 2017, pp. 1643–1647.
[17] S. S. Kraaijeveld and J. De Smedt, "The predictive power of public Twitter sentiment for stock market movements using BERT," Decis. Support Syst., vol. 148, p. 113580, 2021.
[18] H. Saif, M. Fernandez, Y. He, and H. Alani, "On predicting the sentiment of tweets using semantic filters," in Proc. Semantic Web Conf., 2013, pp. 1–8.
[19] R. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," The Journal of Finance, vol. 59, no. 3, pp. 1259–1294, 2004.
[20] I. Loshchilov and F. Hutter, "Decoupled weight decay regulariza-tion," in Proc. Int. Conf. Learn. Represent. (ICLR), New Orleans, LA, USA, 2019, pp. 1–10.
[21] R. P. Schumaker and H. Z. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFinText system," ACM Trans. Inf. Syst., vol. 27, no. 2, pp. 12:1–12:19, 2009.
[22] H. T. Nguyen, et al., "Deep learning for financial sentiment analysis: A comprehensive survey," IEEE Trans. Artif. Intell., vol. 3, no. 5, pp. 678–695, 2022.
[23] F. Allen and R. Karjalainen, "Using genetic algorithms to find technical trading rules," Journal of Financial Economics, vol. 51, no. 2, pp. 245–271, 1999.
[24] P. K. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proc. Int. AAAI Conf. Web Social Media (ICWSM), 2014, pp. 216–225.
[25] Y. Liu, et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[27] K. Cho, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proc. Conf. Empirical Methods Nat. Lang. Process. (EMNLP), Doha, Qatar, 2014, pp. 1724–1734.
[28] W. Y. Al-Ghamdi and A. Cherif, "A survey on deep learning for cryptocurrency market price prediction," IEEE Access, vol. 11, pp. 31234–31256, 2023.
[29] G. Ferrara, et al., "The rise of social bots," Commun. ACM, vol. 59, no. 7, pp. 96–104, 2016.
[30] T. G. Dietterich, "Overfitting in machine learning: Challenges and solutions," ACM Comput. Surv., vol. 27, no. 3, pp. 326–327, 1995.
[31] W. F. Sharpe, "The Sharpe ratio," Journal of Portfolio Management, vol. 21, no. 1, pp. 49–58, 1994.
[32] P. N. Ciresan, et al., "Deep, big, simple neural nets for handwritten digit recognition," Neural Computation, vol. 22, no. 12, pp. 3207–3220, 2010.
[33] C. G. Lamoureux and W. D. Lastrapes, "Heteroskedasticity in stock return data: Volume versus GARCH effects," The Journal of Finance, vol. 45, no. 1, pp. 221–229, 1990.
[34] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, 2018.
[35] C. Wu, et al., "A survey on multimodal sentiment analysis," IEEE Trans. Affect. Comput., vol. 14, no. 3, pp. 2100–2118, 2023.
[36] T. Mikolov, et al., "Distributed representations of words and phrases and their compositionality," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2013, pp. 3111–3119.

[37] M. Sewell, “Technical analysis,” Research Note, UCL Department of Computer Science, RN/11/01, 2011.
[38] A. Urquhart, “The inefficiency of Bitcoin,” Economics Letters, vol. 148, pp. 80–82, 2016.
[39] D. Shah, H. Isah, and F. Zulkernine, “Predicting the extreme price movements of Bitcoin using sentiment analysis,” in Proc. IEEE Int. Conf. Big Data (Big Data), 2019, pp. 1952–1961.
[40] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep learning for event-driven stock prediction,” in Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI), Buenos Aires, Argentina, 2015, pp. 2327–2333.
[41] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, “Noise trader risk in financial markets,” Journal of Political Economy, vol. 98, no. 4, pp. 703–738, 1990.
[42] I. Makarov and A. Schoar, “Trading and arbitrage in cryptocurrency markets,” Journal of Financial Economics, vol. 135, no. 2, pp. 293–319, 2020.
[43] T. G. Andersen and T. Bollerslev, “Answering the skeptics: Yes, stan-dard volatility models do provide accurate forecasts,” International Economic Review, vol. 39, no. 4, pp. 885–905, 1998.
[44] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term depen-dencies with gradient descent is difficult,” IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157–166, 1994.
[45] R. Engle, “Autoregressive conditional heteroskedasticity with esti-mates of the variance of United Kingdom inflation,” Econometrica, vol. 50, no. 4, pp. 987–1007, 1982.
[46] N. Barberis, A. Shleifer, and R. Vishny, “A model of investor sentiment,” Journal of Financial Economics, vol. 49, no. 3, pp. 307–343, 1998.
[47] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[48] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.
[49] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995.