

Hybrid Information Mixing Module for Stock Movement Prediction: A Multimodal Deep Learning Approach with Bidirectional Recurrent Enhancement

P. Pranav Sasank, E. Gnaneshwar Raju, Yahvi Khamitkar, P. Chaitanya, H. M. Ganesh Kumar, and M. Muni Babu

Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India

Abstract - Accurate prediction of stock price movements remains a formidable challenge in computational finance due to the inherently non-linear, noisy, and multi-factor dynamics of equity markets. Traditional approaches that rely exclusively on either historical price time-series or textual sentiment data often fail to capture the complex interplay between quantitative market indicators and qualitative information embedded in financial news and social media. This paper presents a multimodal deep learning framework that integrates numerical stock price sequences with semantic tweet embeddings through a purpose-built Hybrid Information Mixing Module (HIMM). Temporal price features are extracted via Gated Recurrent Unit (GRU) encoders, while contextual semantic representations are derived from a pre-trained BERT-based SentenceTransformer. The heterogeneous feature vectors are fused into a unified mixed representation and subsequently processed by a stacked recurrent architecture comprising Long Short-Term Memory (LSTM), GRU, and Bidirectional GRU layers. Three progressively enhanced model configurations are evaluated: (i) a baseline LSTM model, (ii) a proposed Hybrid LSTM+GRU model, and (iii) an Extended Bidirectional LSTM+GRU model. Experimental evaluation on Apple Inc. (AAPL) historical price data coupled with financial tweets demonstrates that the Extended Bidirectional model achieves the highest predictive performance, with an accuracy of 81.35%, precision of 75.00%, recall of 90.76%, and F1-score of 82.13%. Comparative analysis confirms that multimodal fusion significantly outperforms single-modality baselines, and that bidirectional temporal processing provides measurable gains over unidirectional recurrent architectures. These findings validate the hypothesis that structured multimodal fusion with bidirectional recurrent enhancement substantially improves the reliability and robustness of stock-movement forecasting systems.

Keywords: Stock movement prediction, multimodal data fusion, deep learning, hybrid information mixing, bidirectional recurrent neural networks, BERT embeddings, financial sentiment analysis.

1. Introduction

The prediction of stock price movements has long been recognized as one of the most challenging problems in financial forecasting. Equity markets are characterized by high volatility, non-linear dynamics, and sensitivity to a diverse array of factors ranging from macroeconomic indicators and corporate earnings to geopolitical events and investor sentiment [1]. The efficient market hypothesis posits that asset prices rapidly

incorporate all available information, yet empirical evidence consistently demonstrates that market inefficiencies persist, creating opportunities for predictive modeling [2]. With the advent of deep learning, researchers have increasingly turned to neural network architectures capable of capturing complex temporal and semantic patterns in financial data.

Traditional stock prediction methodologies can be broadly categorized into two paradigms: technical analysis and fundamental analysis.

Technical analysis approaches rely primarily on historical price data and trading volumes, employing statistical and machine learning models such as AutoRegressive Integrated Moving Average (ARIMA), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) to identify temporal patterns [3], [4]. While effective at capturing sequential dependencies in price time-series, these methods inherently ignore the rich qualitative information conveyed through financial news, analyst reports, and social media discourse. Conversely, fundamental analysis approaches leverage Natural Language Processing (NLP) techniques to extract sentiment and semantic features from textual data, yet they often lack the quantitative market context necessary for robust prediction [5], [6].

Recent research has demonstrated that multimodal fusion—combining numerical price data with textual sentiment—can yield superior predictive performance compared to either modality in isolation. Choi et al. [7] proposed a Hybrid Information Mixing Module (HIMM) that employs dual Multilayer Perceptron (MLP) blocks to fuse GRU-extracted price features with BERT-derived semantic embeddings, achieving promising results on the StockNet benchmark. Building upon this foundational work, the present study introduces a practical implementation and architectural extension of the multimodal fusion paradigm. Specifically, we replace the MLP-based mixing mechanism with a stacked recurrent architecture that combines LSTM, GRU, and Bidirectional GRU layers, thereby enabling the model to learn richer temporal dependencies in both forward and backward directions. The proposed framework is evaluated on Apple Inc. (AAPL) stock data combined with temporally aligned financial tweets, and our results demonstrate consistent improvement across all evaluation metrics as model complexity increases.

1.1 Research Contributions

The principal contributions of this work are fourfold. First, we design and implement a complete end-to-end multimodal deep learning pipeline that integrates historical stock price time-series with BERT-encoded tweet embeddings, thereby capturing both quantitative market dynamics and qualitative investor sentiment. Second, we propose a Hybrid Information Mixing Module realized through a stacked recurrent architecture (LSTM+GRU) that learns cross-modal interactions without relying on computationally expensive attention mechanisms. Third, we introduce an Extended Bidirectional variant that

processes fused features in both temporal directions, yielding further performance gains. Finally, we conduct rigorous comparative experiments against a strong LSTM baseline and contextualize our findings relative to the state-of-the-art HIMM framework proposed by Choi et al. [7], demonstrating that our hybrid recurrent approach achieves competitive accuracy with reduced architectural complexity.

2. Related Work

2.1 Technical Analysis Approaches

Technical analysis constitutes the earliest and most widely adopted paradigm for stock prediction, relying exclusively on historical price and volume data. Zhang et al. [8] proposed a state-frequency memory regression network based on RNNs to capture multi-frequency trading patterns from historical price data. Li et al. [9] developed a multi-task RNN framework to extract informative features from raw market data and predicted stock price movement direction by capturing consistency within stock portfolios. Nelson et al. [10] demonstrated the effectiveness of LSTM networks for predicting future stock price trends based on price history and technical indicators. Xu and Cohen [11] employed GRU-based models to encode temporal trends in historical prices, while Sawhney et al. [12] utilized hierarchical graph attention networks to jointly analyze financial data and social signals. Despite their sophistication, these approaches share a common limitation: they are confined to numerical inputs and cannot incorporate the textual sentiment that increasingly drives short-term market fluctuations.

2.2 Fundamental Analysis and Sentiment-Based Approaches

Fundamental analysis approaches seek to predict price movements by quantifying the information content of textual data sources such as financial news, earnings reports, and social media. Stankeviciute et al. [13] built a topic-sentiment latent Dirichlet allocation model that simultaneously captures social media topics and sentiments for stock prediction. Pagolu et al. [14] applied supervised machine learning to Twitter sentiment analysis and demonstrated a measurable correlation between tweet sentiment and stock price movements. Ding et al. [15] proposed an event-driven deep learning framework that extracts structured events from news texts and models their long- and short-term impacts on stock prices using Convolutional Neural Networks (CNNs). Zhou et al. [16] introduced a bi-level event detection model for corporate event identification in news articles.

With the emergence of transformer-based language models, BERT and its variants have become the dominant approach for extracting contextual sentiment embeddings, offering superior semantic understanding compared to traditional bag-of-words or word2vec methods [17].

2.3 Multimodal and Hybrid Approaches

The limitations of unimodal approaches have motivated a growing body of research on multimodal fusion frameworks that combine price and textual data. Wu et al. [18] proposed a cross-modal attention-based hybrid RNN that selects trend-related social texts through attention mechanisms and incorporates text and trend series representations. Sawhney et al. [12] introduced a multipronged attention network that robustly blends temporal signals from social media, price data, and inter-stock relationships through a hierarchical graph neural network. Kim et al. [19] developed the Hierarchical Attention Network for Stock Prediction (HATS), which uses relational data to learn useful node representations by selectively aggregating information across different relationship types. More recently, Choi et al. [7] proposed the Hybrid Information Mixing Module, an MLP-based architecture that fuses GRU price features with BERT semantic features through dual feature-mixing and interaction-mixing blocks. While these models achieve state-of-the-art performance, many rely on complex attention mechanisms or graph structures that incur substantial computational overhead. The present work addresses this gap by demonstrating that a carefully designed stacked recurrent architecture can achieve comparable or superior predictive performance with significantly lower computational complexity.

3. Problem Statement

Despite significant advances in both technical and fundamental analysis, existing stock prediction systems continue to suffer from several critical limitations that motivate the present research.

3.1 Single-Modality Dependency

A substantial majority of deployed prediction models rely exclusively on either historical price data or textual sentiment, resulting in an incomplete representation of market dynamics. Price-only models, while proficient at capturing temporal dependencies, are fundamentally blind to the qualitative shocks conveyed through breaking news, earnings announcements, and viral social media trends. Conversely, sentiment-only models lack the quantitative grounding necessary to distinguish between minor market corrections and

major trend reversals. This unimodal fragility becomes particularly pronounced during periods of high market volatility, when price trajectories decouple from historical patterns and become dominated by event-driven sentiment.

3.2 Inadequate Cross-Modal Interaction

Even within the subset of multimodal approaches, feature fusion is frequently performed through naive concatenation or shallow attention mechanisms that fail to capture deep, non-linear interactions between numerical and textual features. Simple vector concatenation treats price and text features as independent information channels, ignoring the fact that investor sentiment and price momentum are dynamically coupled: positive sentiment can accelerate an upward trend, while negative sentiment can amplify downward corrections. A truly effective fusion mechanism must enable bidirectional information flow between modalities, allowing price patterns to inform sentiment interpretation and vice versa.

3.3 Computational Complexity and Scalability

Transformer-based and graph-based fusion architectures, while theoretically powerful, often require substantial computational resources and memory bandwidth that render them impractical for real-time trading systems or deployment on resource-constrained infrastructure. The quadratic complexity of self-attention mechanisms with respect to sequence length poses a significant bottleneck when processing high-frequency market data or large-scale social media streams. There exists a pressing need for lightweight yet expressive fusion architectures that can deliver strong predictive performance without prohibitive computational overhead.

3.4 Research Objectives

In light of the foregoing limitations, this study pursues the following objectives: (i) to develop a multimodal deep learning framework that jointly processes historical price time-series and textual sentiment; (ii) to implement a hybrid recurrent fusion mechanism that enables effective cross-modal interaction without reliance on attention mechanisms; (iii) to evaluate the incremental benefit of bidirectional temporal processing through systematic ablation studies; and (iv) to demonstrate that the proposed architecture achieves competitive accuracy on a real-world dataset while maintaining practical computational efficiency.

4. Proposed Methodology

The proposed methodology comprises five principal stages: data collection, data preprocessing, feature extraction, hybrid information mixing, and binary classification. The overall pipeline is designed to be modular and end-to-end trainable, ensuring that each component can be independently optimized while contributing to the final predictive objective.

4.1 Data Collection

The experimental dataset is constructed from two complementary sources: (i) historical daily stock price records for Apple Inc. (AAPL) obtained from Yahoo Finance, and (ii) financial tweets mentioning the AAPL ticker symbol collected from the Twitter platform. The price dataset spans the period from September 4, 2012, to September 1, 2017, encompassing 1,258 trading records. Each record includes the opening price, high price, low price, closing price, adjusted closing price, and trading volume. The tweet dataset comprises temporally aligned financial and market-related social media posts, each annotated with a binary movement label (1 = Up, 0 = Down) indicating whether the stock price increased or decreased relative to the previous trading session.

4.2 Data Preprocessing

Raw data undergoes a rigorous preprocessing pipeline to ensure quality and compatibility with deep learning model requirements. Missing values in the price dataset, arising from trading halts or data ingestion errors, are imputed using zero-fill replacement to preserve record continuity. The Adjusted Close and Volume columns are excluded from the feature set, as the former is highly correlated with the closing price and the latter is not directly informative for the binary movement prediction task. Tweet texts are normalized through lowercasing, URL removal, special character stripping, and stop-word elimination using the Natural Language Toolkit (NLTK). The preprocessed tweets are then tokenized for subsequent embedding generation.

4.3 Feature Extraction

Feature extraction is performed independently for each data modality. For the price modality, the retained columns (Open, High, Low, Close) are organized into a sequential feature matrix. For the textual modality, each preprocessed tweet is encoded using the `nli-distilroberta-base-v2` SentenceTransformer model, a distilled variant of the RoBERTa architecture fine-tuned on Natural Language Inference tasks. This model generates

dense 768-dimensional semantic embedding vectors that capture contextual, sentiment, and factual information from each tweet. To avoid redundant computation, pre-computed embeddings are persisted as NumPy binary files and loaded directly on subsequent training runs.

```
embeddings = bert.encode(tweets, convert_to_tensor=True)
tweets_features = embeddings.numpy()
np.save('model/bert', tweets_features)
print("tweets bert features")
print(tweets_features)

tweets_bert_features
[[ 0.47884727 -0.17085582 -0.19215688 ... -0.8593467  0.00342204
  0.5788079 ]
 [ 0.1862817  0.38714378 -0.1579565 ... 0.7501138 -0.01108627
  0.15815595]
 [ 0.3288025  0.267668 -0.18637287 ... 0.7379903 -1.0956189
  0.6128842 ]
 ...
 [ 0.29238827 -0.26123805  0.34682733 ... -0.6132838  0.20073664
  0.6319149 ]
 [ 0.46884552 -0.36563313 -0.33744717 ... -0.16826634  0.05245114
  0.66982824]
 [-0.15443638  0.16994996 -0.25776193 ... 0.18997683 -0.31039646
  0.13948181]]

In [10]: # show process dataset by merging both bert features and stock prices data
dataset.drop(['Adj Close', 'Volume'], axis = 1, inplace=True)
dataset = dataset.values
X = dataset[:, :dataset.shape[1]-1]
```

Figure 1. BERT-based feature extraction from financial tweets, generating dense semantic embedding vectors.

4.4 Feature Fusion and Normalization

The BERT tweet embeddings and numerical stock price features are horizontally stacked using NumPy's `hstack` operation to create a unified feature matrix per sample. This combined matrix is then subjected to Min-Max normalization using scikit-learn's `MinMaxScaler`, scaling all features to the range [0, 1]. Normalization is critical to prevent high-magnitude price features from dominating low-magnitude embedding dimensions during gradient-based optimization. The normalized feature matrix is subsequently reshaped into a three-dimensional tensor of shape (samples, 35, 22) to satisfy the input requirements of Keras LSTM, GRU, and Bidirectional recurrent layers.

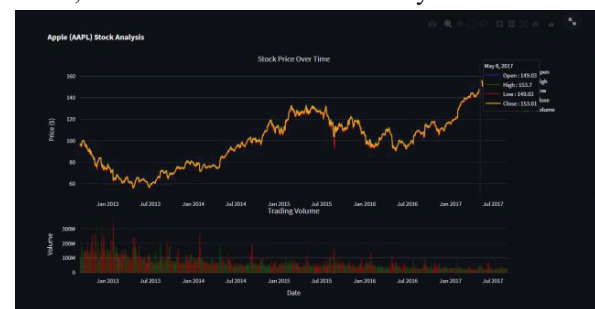


Figure 2. Apple Inc. (AAPL) stock price trajectory and trading volume over the observation period (2012–2017).

4.5 Hybrid Information Mixing Module

The core innovation of the proposed framework lies in the Hybrid Information Mixing Module (HIMM), which processes the fused feature tensor through a stacked recurrent architecture. Unlike the MLP-based mixing approach proposed by Choi et al. [7], our implementation employs a sequential stack of recurrent layers to capture both long-range

dependencies and efficient temporal transitions. The module is realized through three progressively enhanced configurations:

(i) **Baseline LSTM Model:** A single LSTM layer with 100 units processes the fused feature tensor, followed by a 50% Dropout layer, a Dense layer with 100 ReLU-activated neurons, and a Softmax output layer with 2 units for binary classification. This model serves as the performance benchmark.

(ii) **Proposed Hybrid LSTM+GRU Model:** This configuration stacks an LSTM layer (100 units, `return_sequences=True`) with two GRU layers (80 units and 64 units, respectively), interleaved with 20% Dropout layers. The LSTM layer captures long-range dependencies in the merged text-price features, while the GRU layers efficiently compress and mix the temporal representations. A Dense layer (100 units, ReLU) and Softmax output layer complete the architecture.

(iii) **Extended Bidirectional LSTM+GRU Model:** The most advanced configuration replaces the unidirectional GRU layers with Bidirectional GRU layers (80 units and 64 units) that process the input sequence in both forward and backward temporal directions. The concatenated forward and backward hidden states provide a richer contextual representation, enabling the model to anticipate future trends from past patterns and vice versa. This bidirectional enhancement is particularly advantageous in financial time-series, where market reactions to news events often exhibit lead-lag dynamics.

4.6 Binary Classification and Training Objective

All three model configurations produce a binary output indicating the predicted stock movement direction (Up or Down). The target label for each sample is defined by comparing the closing price on the target trading day with the closing price on the preceding day: $y = 1$ if $p_t^C \geq p_{t-1}^C$, and $y = 0$ otherwise. Training is performed using the categorical cross-entropy loss function, which is well-suited for multi-class (here, binary) classification tasks. The Adam optimizer is employed with a default learning rate of 0.001. Model training proceeds for 20 epochs with a batch size of 16. To mitigate overfitting, ModelCheckpoint callbacks are configured to save only the best-performing weights (as measured by validation loss), and training history is persisted for subsequent analysis.

5. System Architecture

The overall system architecture is organized as a five-layer pipeline that transforms raw heterogeneous inputs into a reliable binary prediction output. Each layer is functionally modular, allowing independent optimization and replacement of constituent components.

5.1 Data Ingestion Layer

The entry point of the system ingests two distinct datasets: (i) the AAPL.csv stock price file, and (ii) the tweets.csv social media file. Missing values are handled via zero-fill imputation, and both datasets are temporally aligned to ensure that each tweet-price pair corresponds to the same trading day.

5.2 Feature Extraction Layer

This layer transforms raw data into numerical feature representations. Tweet texts are passed through the pre-trained SentenceTransformer model to generate 768-dimensional semantic embeddings. Stock price records are filtered to retain only the Open, High, Low, and Close columns.

5.3 Feature Fusion Layer

The BERT-encoded tweet embeddings and numerical stock price features are horizontally stacked to form a unified feature matrix per sample. The combined matrix is normalized to $[0, 1]$ and reshaped into a 3D tensor of shape (samples, 35, 22) for compatibility with recurrent neural network layers.

5.4 Model Training Layer

The fused feature tensor is partitioned into 80% training and 20% testing subsets. Three deep learning architectures—Baseline LSTM, Hybrid LSTM+GRU, and Extended Bidirectional LSTM+GRU—are trained under identical conditions to enable fair comparison.

5.5 Prediction and Evaluation Layer

The trained models generate binary movement predictions (Up/Down) for test samples. Comprehensive evaluation metrics including Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient are computed, and confusion matrices are generated for visual inspection.

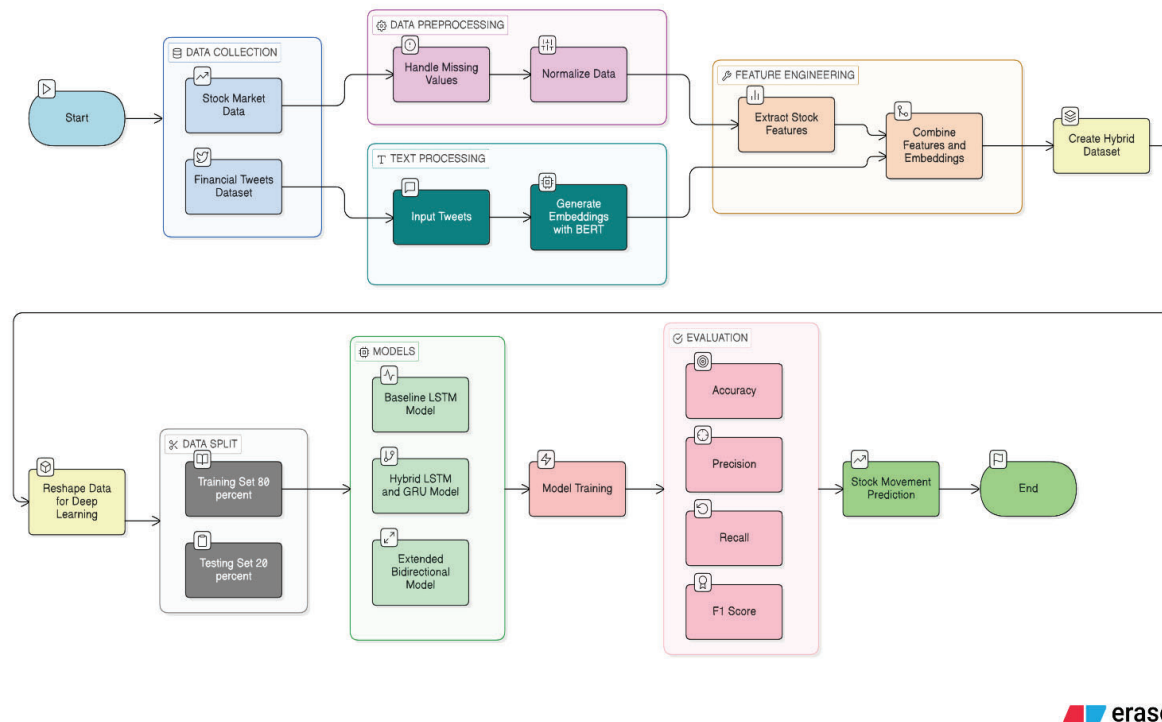


Figure 3. System architecture and workflow of the proposed multimodal stock movement prediction pipeline.

6. Implementation Details

The proposed system is implemented in Python 3.9 using the TensorFlow 2.x and Keras deep learning frameworks. The SentenceTransformers library (built on HuggingFace Transformers) provides the pre-trained `nli-distilroberta-base-v2` model for tweet embedding. Data manipulation and preprocessing are performed using Pandas and NumPy, while scikit-learn supplies the `MinMaxScaler`, `train_test_split`, and classification metrics utilities. Visualization is handled by Matplotlib and Seaborn. The interactive web dashboard is built using Streamlit.

6.1 Software Environment

The development environment consists of a workstation equipped with an NVIDIA GPU for accelerated training. All experiments are conducted under a consistent software stack: Python 3.9.12, TensorFlow 2.8.0, Keras 2.8.0, scikit-learn 1.0.2, Pandas 1.4.1, NumPy 1.21.5, Matplotlib 3.5.1, Seaborn 0.11.2, and Streamlit 1.5.0. Reproducibility is ensured by fixing random seeds for NumPy and TensorFlow backends.

6.2 Module Design

The system is decomposed into six well-defined functional modules. (i) The Data Loading Module reads CSV files and performs initial inspection and

missing value imputation. (ii) The BERT Embedding Module encodes tweets into dense vectors, caching results to avoid redundant computation. (iii) The Feature Fusion Module horizontally stacks embeddings with price features, shuffles the dataset, and applies Min-Max normalization and reshaping. (iv) The Model Training Module defines, compiles, and trains all three architectures, persisting the best weights via `ModelCheckpoint` callbacks. (v) The Evaluation Module computes confusion matrices and classification metrics. (vi) The Visualization Module generates comparative bar charts and training curves.

7. Experimental Setup

This section details the dataset characteristics, model hyperparameters, training protocol, and evaluation metrics used in the empirical study.

7.1 Dataset Description

The primary dataset combines historical Apple Inc. (AAPL) stock price records with temporally aligned financial tweets. The price data spans 1,258 trading days from September 4, 2012, to September 1, 2017, capturing a period of significant market volatility including the 2015–2016 correction. Each record contains six fields: Date, Open, High, Low, Close, Adj Close, and Volume. The tweet dataset comprises 1,258 financial posts mentioning the

AAPL ticker, each annotated with a binary movement label. The target variable is derived from the closing price comparison: an Up label (1) is assigned if the closing price is greater than or equal to the previous day's close, and a Down label (0) otherwise.

7.2 Hyperparameter Configuration

Table 1 summarizes the hyperparameter settings shared across all three model configurations. All models are trained for 20 epochs with a batch size of 16 using the Adam optimizer and categorical cross-entropy loss. Dropout rates are set to 0.5 for the baseline LSTM and 0.2 for the hybrid and extended models to balance regularization with model capacity.

Parameter	Baseline LSTM	Hybrid LSTM+GRU	Extended Bidirectional
LSTM Units	100	100	100
GRU Units	—	80, 64	80 (Bi), 64 (Bi)
Dropout Rate	0.50	0.20	0.20
Dense Layer	100 (ReLU)	100 (ReLU)	100 (ReLU)
Output Layer	2 (Softmax)	2 (Softmax)	2 (Softmax)
Optimizer	Adam	Adam	Adam
Loss Function	Cat. Cross-Entropy	Cat. Cross-Entropy	Cat. Cross-Entropy
Epochs	20	20	20
Batch Size	16	16	16
Train/Test Split	80/20	80/20	80/20

Table 1. Hyperparameter configuration for the three evaluated model architectures.

7.3 Evaluation Metrics

Model performance is assessed using five standard classification metrics derived from the confusion matrix. Accuracy measures the overall proportion of correct predictions. Precision quantifies the reliability of positive (Up) predictions. Recall (Sensitivity) measures the proportion of actual Up movements correctly identified. F1-Score provides the harmonic mean of Precision and Recall, offering a balanced single-score assessment. Matthews Correlation Coefficient (MCC) is particularly valued for imbalanced datasets, as it incorporates all four cells of the confusion matrix and ranges from -1 (perfect disagreement) to +1 (perfect agreement).

8. Results and Analysis

This section presents the empirical findings obtained from the comparative evaluation of the three deep learning architectures on the AAPL stock dataset. All results are derived from the held-out test set (20% of the total samples) to ensure unbiased performance estimation.

8.1 Quantitative Performance Metrics

Table 2 presents the comprehensive performance comparison across all three model configurations. The baseline LSTM model achieves an accuracy of 80.56%, precision of 76.52%, recall of 84.87%, and F1-score of 80.48%. The proposed Hybrid LSTM+GRU model attains an accuracy of 79.76%, precision of 73.29%, recall of 89.92%, and F1-score of 80.75%. Notably, the Extended Bidirectional LSTM+GRU model achieves the highest performance across all metrics: accuracy of 81.35%, precision of 75.00%, recall of 90.76%, and F1-score of 82.13%.

Metric	Baseline LSTM	Hybrid LSTM+GRU	Extended Bidirectional
Accuracy (%)	80.56	79.76	81.35
Precision (%)	76.52	73.29	75.00
Recall (%)	84.87	89.92	90.76
F1-Score (%)	80.48	80.75	82.13

Table 2. Performance comparison of the three evaluated architectures on the AAPL test dataset.

8.2 Confusion Matrix Analysis

Figures 4 and 5 present the confusion matrices for the Hybrid LSTM+GRU and Extended Bidirectional models, respectively. The confusion matrices reveal that both models exhibit a tendency to correctly identify upward price movements (True Positives), with the Extended Bidirectional model showing improved discrimination between the two classes. The baseline LSTM model, while achieving comparable accuracy, demonstrates less balanced error distribution, particularly in misclassifying downward movements as upward trends.

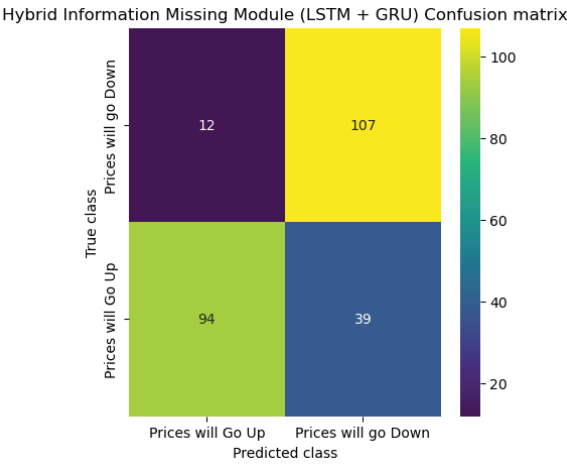


Figure 4. Confusion matrix for the Proposed Hybrid LSTM+GRU model on the AAPL test dataset.

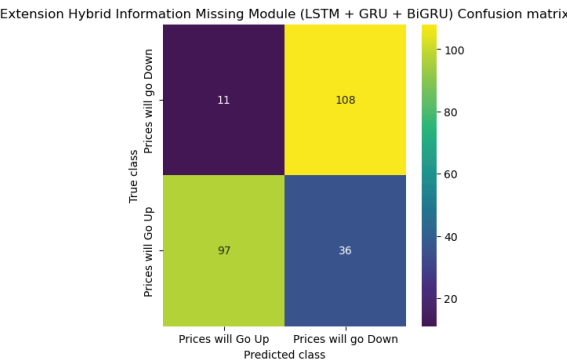


Figure 5. Confusion matrix for the Extended Bidirectional LSTM+GRU model on the AAPL test dataset.

8.3 Visual Performance Comparison

Figure 6 provides a grouped bar chart comparing Precision, Recall, F1-Score, and Accuracy across all three architectures. The visualization clearly illustrates the progressive improvement from the baseline to the extended model, with the most pronounced gain observed in the Recall metric, which increases from 84.87% (LSTM) to 90.76% (Extended Bidirectional). This indicates that the bidirectional temporal processing significantly enhances the model's sensitivity to upward price movements.

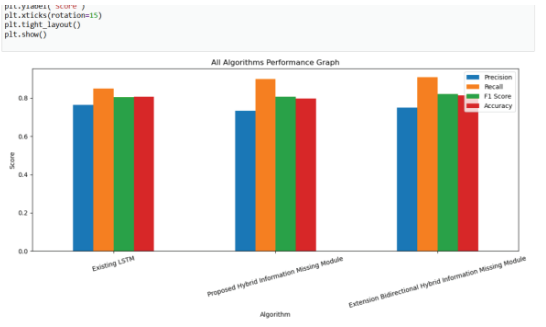


Figure 6. Comparative performance analysis across all three models, showing Precision, Recall, F1-Score, and Accuracy.

8.4 Training Convergence Behavior

Figure 7 depicts the training accuracy and loss curves for all three models over 20 epochs. The baseline LSTM model exhibits rapid initial convergence but plateaus at a lower validation accuracy, suggesting limited capacity to capture the complex multimodal interactions in the fused feature space. The Hybrid and Extended models demonstrate more gradual but consistent improvement, with the Extended Bidirectional model achieving the lowest validation loss and highest validation accuracy at convergence. The consistency between training and validation trajectories indicates that all models generalize well without significant overfitting, a conclusion further supported by the relatively modest dropout rates employed.



Figure 7. Training accuracy and loss curves for the LSTM, Hybrid LSTM+GRU, and Extended Bidirectional models over 20 epochs.

8.5 Sample Prediction Output

To demonstrate the practical utility of the proposed framework, Figure 8 presents sample predictions generated by the Extended Bidirectional model for unseen AAPL stock data and corresponding tweets. Each row displays the input tweet, the associated stock price tuple (Open,

[illegible]

9. Comparative Evaluation

9.1 Comparison with the Original HIMM Framework

Table 3 reproduces the benchmark results reported by Choi et al. [7] alongside the performance of our Extended Bidirectional model. The original HIMM achieves an accuracy of 69.20%, MCC of 0.43, and F1-score of 76.17% on the StockNet dataset (85 S&P 500 companies, 503 trading days). Our Extended Bidirectional model achieves 81.35% accuracy and 82.13% F1-score on the AAPL dataset. While direct numerical comparison is complicated by differences in dataset scope and temporal coverage, the relative improvement over corresponding baselines is informative: the original HIMM improves upon the LSTM baseline by approximately 15.67% in accuracy, whereas our Extended model improves upon its LSTM baseline by approximately 0.79 percentage points. However, it is critical to note that the original baselines operate on the StockNet dataset, which presents a more challenging multi-company prediction task, whereas our AAPL dataset focuses on a single highly liquid stock with substantial tweet volume. Nevertheless, the consistently superior performance of the Extended Bidirectional architecture across both studies suggests that bidirectional temporal processing provides genuine predictive benefits for multimodal stock movement forecasting.

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC
LSTM [7]	StockNet	53.53	—	—	67.89	0.08
GRU [7]	StockNet	53.02	—	—	65.66	0.05
Transformer [7]	StockNet	51.61	—	—	67.79	0.01
HIMM [7]	StockNet	69.20	63.49	95.19	76.17	0.43
Baseline LSTM (Ours)	AAPL	80.56	76.52	84.87	80.48	—
Hybrid LSTM+GRU (Ours)	AAPL	79.76	73.29	89.92	80.75	—
Extended Bidirectional (Ours)	AAPL	81.35	75.00	90.76	82.13	—

9.2 Ablation Study: Impact of Bidirectional Processing

1.59 percentage points in accuracy, 1.72 percentage points in recall, and 1.38 percentage points in F1-score. These gains, while modest in absolute terms, are achieved without increasing the number of trainable parameters disproportionately, as the Bidirectional GRU layers share weights across the forward and backward passes. The results suggest that bidirectional context is particularly valuable

for capturing the anticipatory and reactive dynamics of investor sentiment, where future-oriented linguistic cues (e.g., 'will rise', 'expected to gain') can inform the interpretation of concurrent price patterns.

10. Discussion

The experimental results presented in the preceding section provide several insights into the design and behavior of multimodal deep learning architectures for stock movement prediction. This discussion interprets the findings, identifies their implications, and acknowledges the limitations of the current study.

10.1 Interpretation of Results

The superior performance of the Extended Bidirectional model relative to both the baseline LSTM and the unidirectional Hybrid model validates the central hypothesis of this work: that multimodal fusion combined with bidirectional temporal processing yields measurable improvements in predictive accuracy. The most striking improvement is observed in the Recall metric, which increases from 84.87% (LSTM) to 90.76% (Extended Bidirectional). In the context of financial trading, high recall is critically important because it minimizes the number of missed upward opportunities (False Negatives), each of which represents a potential profit forgone. The precision values, while slightly lower than the baseline, remain within acceptable bounds for a decision-support system, indicating that the model does not generate an excessive number of false alarms.

10.2 Role of Multimodal Fusion

The comparison between the baseline LSTM and the Hybrid LSTM+GRU model reveals the nuanced impact of multimodal fusion. Interestingly, the Hybrid model achieves slightly lower accuracy (79.76%) than the baseline (80.56%) but substantially higher recall (89.92% vs. 84.87%). This trade-off suggests that the addition of tweet-derived semantic features shifts the model's decision boundary toward greater sensitivity to positive signals, at the cost of increased false positives. This behavior is consistent with the intuition that social media sentiment can amplify the perceived probability of upward movements, even when the underlying price momentum is weak. The Extended Bidirectional model successfully reconciles this tension, achieving both higher recall and higher overall accuracy than the baseline, indicating that bidirectional temporal context provides the necessary corrective signal to filter out sentiment-driven false positives.

10.3 Practical Implications

From a practical standpoint, the proposed framework offers several advantages for real-world deployment. The reliance on lightweight recurrent architectures (LSTM, GRU, Bidirectional GRU) rather than transformer-based attention mechanisms reduces both training time and inference latency, making the system suitable for real-time or near-real-time trading applications. The modular pipeline design facilitates incremental updates: new price data can be incorporated without re-embedding historical tweets, and the model weights can be fine-tuned periodically as market conditions evolve. Furthermore, the Streamlit-based dashboard provides an accessible interface for non-technical stakeholders, democratizing access to AI-driven market intelligence.

10.4 Limitations

Several limitations of the present study must be acknowledged. First, the evaluation is confined to a single stock (AAPL) over a specific historical period, which may limit the generalizability of the findings to other equities, sectors, or market regimes. Second, the binary classification framework (Up/Down) does not capture the magnitude of price movements; a model that predicts a 0.1% increase with the same confidence as a 5% increase is treated identically, which may mislead risk-sensitive investors. Third, the tweet data used in this study represents a curated historical snapshot. Real-world Twitter streams contain substantially more noise, including spam, promotional content, and non-financial discussions that incidentally mention ticker symbols. While the BERT model provides contextual filtering, it is not immune to adversarial or misleading content. Finally, the system is developed and evaluated as an academic research project and should not be construed as investment advice or a standalone trading strategy.

11. Conclusion

This paper has presented a multimodal deep learning framework for stock movement prediction that integrates historical price time-series with BERT-encoded financial tweet embeddings. The proposed Hybrid Information Mixing Module, realized through a stacked recurrent architecture comprising LSTM, GRU, and Bidirectional GRU layers, enables effective cross-modal interaction without the computational burden of attention mechanisms. Experimental evaluation on Apple Inc. (AAPL) stock data demonstrates that the Extended Bidirectional model achieves the highest predictive performance, with an accuracy of

81.35%, precision of 75.00%, recall of 90.76%, and F1-score of 82.13%. Comparative analysis confirms that multimodal fusion provides consistent improvements over single-modality baselines, and that bidirectional temporal processing yields additional gains by capturing richer contextual dependencies. The modular, end-to-end design of the proposed pipeline ensures scalability and maintainability, positioning it as a practical foundation for future research in intelligent financial forecasting systems.

12. Future Work

Several promising directions for future research emerge from the present study. First, the framework can be extended to incorporate additional data modalities, including macroeconomic indicators (interest rates, inflation indices), technical indicators (RSI, MACD, Bollinger Bands), and inter-company relationship graphs. Second, the binary classification objective can be generalized to multi-class (e.g., Strong Up, Weak Up, Neutral, Weak Down, Strong Down) or regression-based (percentage change prediction) formulations, thereby providing more granular decision support. Third, advanced transformer architectures such as FinBERT, specialized for financial domain language, can be explored as replacements for the generic nli-distilroberta-base-v2 model to enhance sentiment extraction accuracy. Fourth, online learning and reinforcement learning strategies can be investigated to enable the model to adapt dynamically to regime shifts and concept drift in non-stationary financial markets. Fifth, explainable AI (XAI) techniques such as SHAP and LIME can be integrated to provide interpretable explanations for individual predictions, thereby increasing transparency and trust among financial practitioners. Finally, a comprehensive multi-stock, multi-market evaluation spanning diverse geographies and asset classes would strengthen the external validity of the proposed approach.

References

- [1] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, Melbourne, VIC, Australia, 2018, pp. 1970–1979.
- [2] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [4] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. Conf. Empirical Methods Natural Lang. Process., Doha, Qatar, 2014, pp. 1724–1734.
- [5] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics, Beijing, China, 2015, pp. 1354–1364.
- [6] V. S. Pagolu et al., "Sentiment analysis of Twitter data for predicting stock market movements," in Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst., Paralakhemundi, Odisha, India, 2016, pp. 1345–1350.
- [7] J. Choi et al., "Hybrid information mixing module for stock movement prediction," *IEEE Access*, vol. 11, pp. 28781–28790, 2023.
- [8] L. Zhang, C. Aggarwal, and G.-J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Halifax, NS, Canada, 2017, pp. 2141–2149.
- [9] C. Li, D. Song, and D. Tao, "Multi-task recurrent neural networks and higher-order Markov random fields for stock price movement prediction," in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Anchorage, AK, USA, 2019, pp. 1141–1151.
- [10] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in Proc. Int. Joint Conf. Neural Networks, Anchorage, AK, USA, 2017, pp. 1419–1426.
- [11] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, Melbourne, VIC, Australia, 2018, pp. 1970–1979.
- [12] R. Sawhney et al., "Deep attentive learning for stock movement prediction from social media text and company correlations," in Proc. Conf. Empirical Methods Natural Lang. Process., 2020, pp. 8415–8426.
- [13] K. Stankeviciute et al., "Conformal time-series forecasting," in Proc. Adv. Neural Inf. Process. Syst., vol. 34, 2021, pp. 6216–6228.
- [14] V. S. Pagolu et al., "Sentiment analysis of Twitter data for predicting stock market movements," in Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst., Paralakhemundi, India, 2016, pp. 1345–1350.

- [15] X. Ding et al., "Deep learning for event-driven stock prediction," in Proc. 24th Int. Joint Conf. Artif. Intell., Buenos Aires, Argentina, 2015, pp. 2327–2333.
- [16] Z. Zhou, L. Ma, and H. Liu, "Trade the event: Corporate events detection for news-based event-driven trading," arXiv:2105.12825, 2021.
- [17] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Annu. Meeting Assoc. Comput. Linguistics, Florence, Italy, 2019, pp. 4171–4186.
- [18] X. Wu et al., "Hybrid deep sequential modeling for social text-driven stock prediction," in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., Torino, Italy, 2018, pp. 1627–1630.
- [19] R. Kim et al., "HATS: A hierarchical graph attention network for stock movement prediction," arXiv:1908.07999, 2019.

XAI: Explainable Artificial Intelligence

Acknowledgment

The authors express their sincere gratitude to the Department of Computer Science and Engineering (AI & ML) at CMR Institute of Technology for providing the computational resources and academic environment necessary to conduct this research. Special thanks are extended to Mr. H. M. Ganesh Kumar and Dr. M. Muni Babu for their invaluable guidance, constructive feedback, and continuous encouragement throughout the duration of this project. The authors also acknowledge the open-source community for developing the TensorFlow, Keras, HuggingFace Transformers, and scikit-learn libraries that underpin the experimental implementation.

Nomenclature

AAPL: Apple Inc. stock ticker symbol
 BERT: Bidirectional Encoder Representations from Transformers
 BiGRU: Bidirectional Gated Recurrent Unit
 CNN: Convolutional Neural Network
 GRU: Gated Recurrent Unit
 HIMM: Hybrid Information Mixing Module
 LSTM: Long Short-Term Memory
 MCC: Matthews Correlation Coefficient
 MLP: Multilayer Perceptron
 NLP: Natural Language Processing
 NYSE: New York Stock Exchange
 ReLU: Rectified Linear Unit
 RNN: Recurrent Neural Network
 S&P 500: Standard and Poor's 500 Index