AN INTEGRATED FINBERT–TEMPORAL FUSION TRANSFORMER WITH DYNAMIC RISK-AWARE LOSS FOR STOCK FORECASTING AND TRADING DECISION SUPPORT

S. VALLI*[1], DR S. SIVASUBRAMANIAM[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India.

[2]Assistant Professor/Programmer, Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India.

E-mail : svallicis@outlook.com, aucissiva@gmail.com

ABSTRACT

Accurate stock price prediction remains a challenging task because of the high volatility, nonlinearities, and diverse sentiment characteristics of financial markets. Conventional forecasting approaches focus on the historical behavior of stock prices and ignore the effects of textual sentiment, financial risk, and their combined influences during the optimization process. This study presents a novel multimodal stock forecasting framework that simultaneously considers Fin BERT based financial-news representations, market and technical indicators, Temporal Fusion Transformer modelling, and a Dynamic Risk-aware Loss function that regularizes the prediction error, volatility, Conditional Value-at-Risk (CVaR), and prediction direction. This framework extends beyond accuracy-oriented stock forecasting towards enhanced risk estimation, explainability, and decision making in terms of BUY, HOLD, or SELL signals. First, we enrich the input space formed by historical stock prices with eighteen technical indicators, Fin BERT news representations, and market volatility features resulting in 794 input variables. Then, we propose a Temporal Fusion Stock Forecasting Framework (TFSF) based on a hybrid CNN-BiLSTM-TFT architecture to estimate future stock prices. We evaluate our framework on four India-based stock market datasets including CIPLA, HDFC Bank, Infosys, and ONGC after applying a rigorous preprocessing and sliding-window procedure. Our experiments demonstrate that our approach outperforms CNN, LSTM, BiLSTM, CNN-BiLSTM, TFT, and conventional TFSF models in terms of the RMSE, MAE, MAPE, and R2 evaluation metrics with the RMSE = 0.0101, MAE = 0.0079, MAPE = 1.09%, R2 = 0.991, respectively. In addition, our approach achieves higher financial performance with a Sharpe Ratio = 1.94, and Value at Risk (VaR) = 2.10%, compared to the baseline methods while providing accurate BUY, HOLD, and SELL recommendations. Finally, we provide an interpretability analysis showing the influence of each feature on the final prediction where the highest importance is assigned to Fin BERT features. Overall, the proposed approach provides an accurate, interpretable, risk-aware, and intelligent framework for financial forecasting and decision making.

Keywords: *Stock Price Forecasting, Temporal Fusion Transformer (TFT), Fin BERT-Dynamic Risk-Aware Loss, Financial Sentiment Analysis*

1. INTRODUCTION

Financial markets are fundamentally evolving, complicated, and impacted by an array of factors, such as past pricing, investor attitude, economic conditions, and unforeseen world events. Researchers are working to create intelligent prediction systems that can recognize nonlinear correlations and enhance investing choices due to the growing amount of financial information. However, due to market behaviour's high volatility, temporal dependence, and quickly shifting investor psychology, reliable stock price forecasting is still difficult to achieve. Historical numerical data is the primary source of information for both machine learning algorithms like Support Vector Machines, Random Forests, and Gradient Boosting and traditional statistical forecasting techniques like ARIMA. Though these approaches yield respectable short-term forecasts, they frequently overlook the impact of financial news emotions and fail to reflect long-range temporal relationships. Prediction effectiveness has been greatly enhanced by recent developments in traditional deep learning models, including Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional Neural Networks (CNN), and Transformer architectures. However, a lot of current methods

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

primarily maximize prediction accuracy while ignoring financial risk, which is an important consideration when making real-world investment choices.
The development of the Transformer architecture has contributed to the creation of approaches that can better model sequential data than traditional recurrent neural networks. The Temporal Fusion Transformer (TFT) is a type of such model that is capable of encoding both short-term and long-term temporal dependencies in a way that allows for the identification of variable importance. At the same time, the Fin BERT model, which was pretrained on financial texts, can be used to extract investor sentiment from financial news articles, which can serve as additional context for stock price prediction beyond historical prices. However, the application of these methods was still limited by the fact that they do not address three key requirements for stock forecasting. The combination of structured market data and unstructured financial news sentiment can be incorporated in stock price prediction in only a few ways.
The stated challenges is that available approaches to stock forecasting are mainly focused on either market facts or prediction accuracy, rather than combining financial news sources with explicit consideration of market downsides. As a result, forecasting models fail to account for the mix of typographical and factual information, as well as properly address the issues associated with volatility, tail risk, and directional mistakes in the price projection process, potentially leading to poor trading choices. As a result, the current study addresses the challenge of creating a stock forecasting framework that incorporates financial market facts and financial news represented in Fin BERT characteristics, as well as explicit consideration of volatility, CVaR, and directed aspects in the Temporal Fusion Transformer construction, while simultaneously offering interpretation and risk assistance.

1.1 Research Questions

RQ1: Can the integrated Fin BERT–TFT framework improves stock forecasting performance compared with conventional deep-learning and TFT baselines?
RQ2: Can the Dynamic Risk-Aware Loss improve risk-adjusted performance by explicitly modelling volatility, CVaR, and directional prediction error?
RQ3: Can the integration of forecasting, risk assessment, and trading-signal generation provide more risk-aware decision support than prediction accuracy alone?
RQ4: How robust is the proposed framework across the selected equities and different risk conditions?
To address the identified research gap, this study develops an integrated Fin BERT–Temporal Fusion Transformer (TFT) framework with a Dynamic Risk-Aware Loss Function (DRALF) for multimodal stock forecasting and risk-aware trading decision support. The principal contribution is the integrated coverage of all the financial information in a single multimodal framework, combined with the ability to optimize the process in terms of risk and the ability to interpret the process in terms of a temporal model. Moreover, the financial factors which most influence the forecasting decision are identified by using the variable importance analysis. The framework was designed from a practical perspective to help investors and financial analysts by giving them additional information and insights on which to base their BUY, HOLD, and SELL decisions: beyond point forecasts, it includes stock price forecasts and downside-risk measures, plus interpretable evidence.

The major contributions of this work are summarized as follows:

- ✓ A novel multimodal stock forecasting framework based on historical OHLCV market data, technical indicators, volatility information and financial-news sentiment representations represented by Fin BERT.
- ✓ A Temporal Fusion Transformer (TFT)-based temporal modelling technique that models the temporal structure in financial time-series data and conducts interpretable variable selection to identify the important financial features.
- ✓ Dynamic Risk-Aware Loss Function: A loss function for risk-aware model optimization that incorporates prediction error (MSE), volatility penalty, Conditional Value-at-Risk (CVaR) and directional prediction loss.
- ✓ An embedded decision-support system that integrates price forecasting, risk assessment, and model interpretability for providing actionable investment decisions – BUY, HOLD, and SELL.
- ✓ Multiple stock datasets in India were used to evaluate comprehensively with both forecasting indicators and financial risks.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. RELATED WORKS

Hybrid deep learning architectures have recently attracted considerable attention. CNN–LSTM and CNN–BiLSTM models combine convolutional feature extraction with sequential learning, allowing simultaneous modelling of local market movements and long-term temporal behaviour. Comparative studies have consistently shown that hybrid models outperform standalone CNN and LSTM architectures in terms of prediction accuracy. Although the present study considers selected Indian equities, the experimental dataset and feature representation are newly constructed for the current multimodal forecasting framework; any overlap in publicly available market observations is distinct from the methodology and results reported in [1]. The present study extends this direction through a multimodal TFT framework that integrates OHLCV data, technical and volatility features, and 768-dimensional Fin BERT representations. It further introduces a Dynamic Risk-Aware Loss incorporating prediction error, volatility, CVaR, and directional error, thereby advancing toward uncertainty-aware and downside-risk-sensitive portfolio decision support. Recent literature has increasingly moved from purely statistical forecasting methods toward machine learning and deep learning architectures for stock return prediction. Traditional approaches such as ARIMA and its variants remain reference baselines but are widely reported to struggle with the nonlinear, non-stationary behaviour of financial time series. To address these limitations, researchers have turned to LSTM networks, attention mechanisms, and transformer-based architectures. Li et al. proposed MASTER, a market-guided stock transformer for price forecasting **[2],** while related work introduced Stock Time, a time-series-specialized large language model architecture built specifically for stock price prediction. A parallel line of work integrates explainability into these models: a study on the BIST100 index combined transformer-based time series models (Linear, Vanilla Transformer, Time Series Transformer) with SHAP and LIME techniques to make prediction outputs more interpretable for retail investors **[3].** Nevertheless, existing forecasting and explainable models primarily emphasize prediction or interpretability, while the joint use of financial-news sentiment, market dynamics, and explicit downside-risk-aware portfolio decisions remains less explored. A more recent direction combines financial-language models with forecasting architectures. Previous studies have integrated Fin BERT-derived financial news sentiment with stock-price data and LSTM-based models, demonstrating the value of textual sentiment alongside historical market information [4]. More recent work has extended Fin BERT-based sentiment fusion to CNN–LSTM forecasting and related multimodal architectures [5]. These studies demonstrate the value of Fin BERT and multimodal temporal modelling; however, the Fin BERT–TFT study [6] focuses on systemic-risk early prediction in U.S. financial firms, whereas the present work targets short-to-medium-term stock-price forecasting and portfolio decision support. The novelty lies in integrating Fin BERT with market, technical, and volatility features through a risk-aware TFT framework with explicit downside-risk modelling. A substantial body of work connects predictive models to downstream portfolio construction under a “predict-then-optimize” or “decision-focused learning” paradigm. Butler and Kwon have developed a stochastic setting that allows one to incorporate regression-based return forecasts into the standard mean-variance optimization (MVO) framework [7]. Meanwhile, Costa and Iyengar have proposed a distributionally robust end-to-end portfolio optimization approach. It accounts for parameter uncertainty in the optimization process by using a distributionally robust MVO instead of decoupling prediction and allocation stages [8]. Anis and Kwon have in turn introduced a decision-based end-to-end portfolio optimization model that embeds cardinality constraints in the learning process [9]. These constraints are imposed directly on the optimization function unlike in the previous approaches. All in all, the authors have demonstrated that incorporating predictive uncertainty is beneficial for portfolio allocation but the decision-making process is still challenged by joint uncertainty and explicit consideration of downside risk.

In the literature review on predict-then-optimize paradigm, it is revealed that risk-adjusted performance measures have gained increasing attention as objective functions in optimization with conditional value-at-risk (CVaR) being the most prominent. Behera and Kumar have designed a portfolio optimization model that uses deep neural networks to forecast stock returns and then combines them with mean-CVaR optimization to account for tail risks [10].

Furthermore, the authors that applies entropy limitations to promote diversity while preventing wealth accumulation. As a consequence, the model might strike a balance between the requirement to limit downside risk and the concern of overweighting specific assets. In another study, Nguyen trained both long short-term memory (LSTM) and 1D-CNN on stock return estimates and applies them in three standard portfolio optimization scenarios [11]. The mean-variance, risk parity, and maximum drawdown framework were applied to the Vietnamese stock exchange from 2017 to 2024, with the result that different optimization methods should be used for different risk profiles [11]. Finally, in a study of decision-induced ranking effects, the authors warn about "forecasting inflation" and excessive turnover in portfolios optimized with stochastic predict-then-optimize (SPO) objectives, which are triggered by an interaction among the decision layer's placing and forecasting loss [12]. These findings highlight the fact that a combination of prediction and optimization is sufficient to ensure the resilience of portfolio decisions, demanding that we consider forecast accuracy, downside tail risk, and allocating stability all at the same time. Deep reinforcement learning (DRL) constitutes one of the most widely researched techniques to dynamic portfolio management because it directly addresses the problem of optimal rebalancing in the setting of traditional sequential decision-making. The survey by Hambly, Xu, and Yang [13] and the more recent research by Pippas, Turkay, and Ludvig [14] show the scope of RL applications in quantitative finance. In particular, observe that risk monitoring is either integrated into the reward function in most circumstances or explicitly reflected in the policy through limitations. Choudhary, Orra, and Sahoo et al. suggested a multi-reward DRL architecture that allows risk-averse investors to strike the right balance between risk and reward by carefully tailoring the reward function [15]. Acero, Zehtabi, and colleagues created a strategy that combines RL and mean-variance methodologies under the "responsible portfolio optimization" framework [16]. Ndikum and Ndikum present an industry-level DRL model that seeks to handle real-world problems rather than solely academic instances [17].

A frequent issue in this research area is the trade-off between risk and return. Lwele, Emmanuel, and Sitali's risk-aware DRL study discovered that the agent was able to maintain portfolio volatility; however, it reduced the risk-adjusted performance of the portfolio due to overly conservative policy convergence [18], leaving reward shaping as an open problem in risk-sensitive reinforcement learning. Recent research underlines the techniques' robustness to model misspecification: the journal used elliptic uncertainty sets to express market impact, making RL-based portfolio policies resilient to mis specified market models [19]. However, it is clear that DRL techniques rely heavily on reward and risk-objective engineering: the incorporation of prediction uncertainty, as well as explicit consideration of downside risk and tail occurrences, is still in its early stages. CVaR emerged as the favoured risk measure in the majority of articles: unlike VaR, which only analyses the chance of significant losses, CVaR also includes the magnitude of these losses. Indeed, one study found that Ledoit-type shrinkage and Gerber covariance as covariance estimates outperform the market-cap weighted covariance matrix in bull markets; though, they significantly underestimate the tail risk, especially during the COVID-19 crisis [20]. To achieve numerical security, the authors recommended adding explicit CVaR restrictions to the minimum-variance optimum framework and using K-means-driven nested clustering optimum [20] A recent study developed a CVaR-based risk parity (CVaR-RP) model that employs machine learning to dynamically adjust the weights in order to directly allocate tail risk rather than using volatility as the basis for risk parity [21]. However, these CVaR-based risk parity strategies put maximizing portfolio risk and allocations ahead of simultaneously learning predictive knowledge, uncertainty, and tail-risk-aware judgments. The most recent work in this space evaluates tail-risk strategies empirically across market regimes rather than proposing a single new model. A comprehensive cryptocurrency-market study compared traditional benchmarks against moment-based CVaR, robust CVaR, regime-dependent CVaR optimization, regression-enhanced Expected-Shortfall/CVaR hybrids, and reinforcement-learning-based CVaR policies, concluding that empirical evidence for consistent outperformance of learning-based CVaR strategies over simpler rule-based benchmarks remains mixed across regimes and tail depths **[22].** Similarly, a study on "financially guided" deep portfolio optimization trained an Attention LSTM using a composite Omega-CVaR-Risk-

© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Parity loss function directly, reporting a substantial Sharpe ratio improvement over the S&P 500 benchmark during the 2022–2023 out-of-sample period while keeping CVaR exposure essentially unchanged **[23].**Alongside gradient-based deep learning methods, a parallel literature applies evolutionary and swarm intelligence algorithms to portfolio optimization, particularly for cardinality-constrained and combinatorial variants of the Markowitz problem that are difficult to solve with convex optimization alone. Erwin and Engelbrecht's review of 140 studies found meta-heuristic approaches (genetic algorithms, particle swarm optimization, differential evolution) to be computationally efficient for large-scale, constrained portfolio problems **[24],** and their companion paper proposed a Multi-Guide Set-Based Particle Swarm Optimization method for multi-objective portfolio optimization **[25].** Comparative empirical work, such as an evaluation of PSO, GA, dynamic programming, and differential evolution on the NIFTY 50 index, generally finds swarm-based methods competitive with or superior to classical solvers on Sharpe-ratio-based objectives, though at higher computational cost **[26].** Overall, existing studies address forecasting, sentiment modelling, portfolio optimization, and risk management largely in isolation. This study integrates Fin BERT, TFT, and Dynamic Risk-Aware Loss incorporating prediction error, volatility, CVaR, and directional error to provide multimodal forecasting and risk-aware BUY, HOLD, and SELL decision support for selected Indian equities.

3. METHODOLOGY

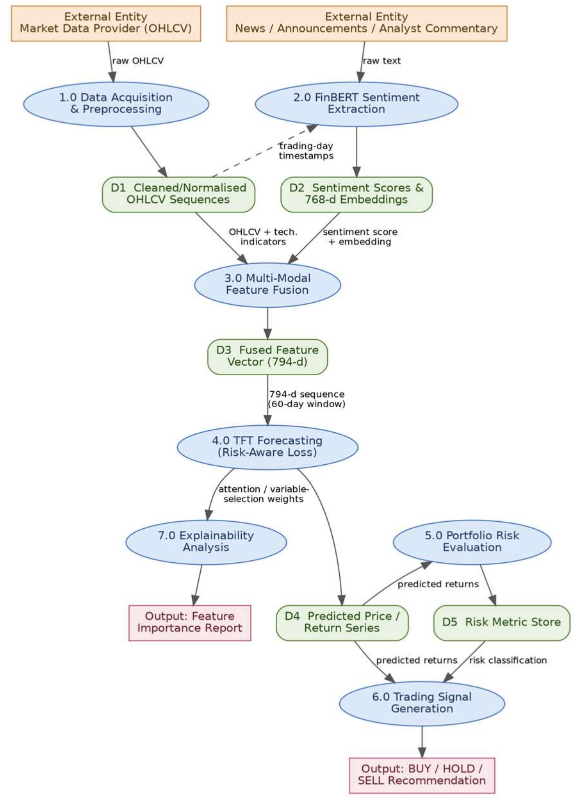

Figure 1: Data flow diagram (DFD) of the proposed methodology, showing external entities, processes (1.0–7.0), data stores (D1–D5), and system outputs.

3.1 Data Acquisition and Preprocessing

Historical daily Open–High–Low–Close–Volume (OHLCV) data for CIPLA, HDFC Bank, Infosys, and ONGC were collected from the publicly available Yahoo Finance database (Figure 3.1, Process 1.0). Yahoo Finance Historical Market Data The collected data were examined for duplicate records, missing values, and inconsistencies to ensure data quality. Duplicates were discovered in the deleted data, and in order to maintain time regularity, the forward-fill examine was used to replace missing values on brief intervals. The data set was cleared of records that lacked information and therefore could not be filled in. Consequently, a 70-15-15 data division was achieved by removing missing values before train-test splitting and creating a sliding window for the time series with a size of 60. The Min–Max technique was used to standardize all continuous information characteristics x, scaling the data of all other sets using the training set's minimum and maximum value.

:

$$x\prime = \frac{(x - x\_min)}{(x\_max - x\_min)} \tag{1}$$

Following normalization, the time-series data were transformed into overlapping sliding windows with a fixed sequence length of 60 trading days, where each input sequence represents the previous 60 observations and the target corresponds to the subsequent closing price. The datasets were then divided chronologically into 70% training, 15% validation, and 15% testing subsets, thereby preserving temporal ordering and eliminating look-ahead bias. The resulting pre-processed OHLCV sequences constitute Data Store D1 in the proposed framework (Figure 3.1).

3.2 Financial Sentiment Extraction (Fin BERT)

To capture the influence of market sentiment on price movement, textual data — news headlines, corporate announcements, and analyst commentary corresponding to each trading day — are passed through Fin BERT, a BERT-based language model pre-trained on financial corpora and fine-tuned for tone classification into positive, neutral, and negative classes (Process 2.0). For an input text, Fin Bert's final classification layer produces class logits z, converted to class probabilities via the SoftMax function:

$$\frac{P\_i = exp(z\_i)}{\Sigma_j\, exp(z\_j), i \in \{pos, neu, neg\}} \quad (2)$$

A continuous, signed sentiment score for each stock-day pair is then derived as the probability-weighted difference between the positive and negative classes:

$$S = P\_pos - P\_neg,\ S \in [-1, 1] \quad (3)$$

In addition to the scalar sentiment score, the 768-dimensional Fin BERT contextual embedding (the pooled [CLS] representation) for each day's dominant news text is retained for downstream fusion, so that the forecasting model has access to the full semantic representation rather than a single collapsed sentiment polarity. The sentiment scores and embeddings together form data store D2 in Figure 3.1.

3.3 Multi-Modal Feature Fusion

Four categories of features are fused into a single input tensor per time step (Process 3.0): (i) raw OHLCV values, (ii) a bank of technical indicators including RSI, MACD, EMA-20, Bollinger Band width, and ATR computed over rolling windows to encode momentum, trend, and volatility signals, (iii) the Fin BERT sentiment embedding, and (iv) market-volatility descriptors (e.g., realised volatility, high-low range, and volume-weighted volatility) capturing short-term risk conditions. Because the subspaces vary greatly in scale and distribution, each category is normalized separately before to concatenation. For every trading day, feature fusion creates a 794-dimensional input vector using historical prices, eighteen technical indicators, Fin BERT, and volatility data.

$$(F\_t = [\,OHLCV\_t\,;\, TechInd\_t\,;\, FinBERT\_t\,;\, Vol\_t\,]) \quad (4)$$

The fused vectors across a 60-day window form the model input sequence, stored as data store D3 in Figure 3.1, and are fed sequentially into the forecasting model described in Section 3.4.

3.4 Sequence Model: Temporal Fusion Transformer

A Temporal Fusion Transformer (TFT, Process 4.0) was chosen as the primary forecasting engine due to its capacity to integrate interpretable selection of variables, long-range attention, and local recurrent processing into one model. Standard LSTM gates are used in the recurrent portion of the TFT to capture short-term temporal trends.

$$(f\_t = \sigma(W\_f \cdot [h\_\{t-1\}, F\_t] + b\_f)), (i\_t = \sigma(W\_i \cdot [h\_\{t-1\}, F\_t] + b\_i)), \quad (o\_t = \sigma(W\_o \cdot [h\_\{t-1\}, F\_t] + b\_o)) \quad (5)$$

$$(c\_t = f\_t \odot c\_\{t-1\} + i\_t \odot tanh(W\_c \cdot [h\_\{t-1\}, F\_t] + b\_c),\ \ h\_t = o\_t \odot tanh(c\_t)) \quad (6)$$

where f_t, i_t, and o_t are the forget, input, and output gates, c_t is the cell state, and h_t is the hidden state at time t. Long-range dependencies across the input sequence are captured using multi-head self-attention:

$$\left[Attention\ (Q, K, V) = softmax(QK^{T} / \sqrt{d\_k}) \cdot V\right] \quad (7)$$

where Q, K, and V are the query, key, and value projections of the LSTM hidden states, and d_k is the key dimension used to scale the dot product. A variable-selection network learns, per time step, a SoftMax-normalised weight for each of the input features in F_t, so that the network can up-weight informative features and suppress uninformative ones; these same weights are reused directly by the explainability analysis in Section 3.9. Gated residual networks are used throughout the architecture to further filter unnecessary nonlinear processing, and static covariate encoders condition predictions on stock-specific characteristics (sector, average volatility regime). To contextualise the TFT's performance, four progressively more complex baselines are also trained under identical data splits and feature sets: a vanilla LSTM, a bidirectional LSTM

(BiLSTM), a hybrid CNN-BiLSTM (using 1-D convolutions for local pattern extraction ahead of a BiLSTM), and the plain TFT without the proposed risk-aware loss. All models are trained with the Adam optimiser and early stopping on validation loss, using identical sequence length and train/validation/test splits, so that any observed performance differences are attributable to architecture and loss design rather than data-handling artefacts.

3.5 Dynamic Risk-Aware Loss Function

Conventional forecasting losses (e.g., mean squared error) optimise purely for point-prediction accuracy and are indifferent to the financial cost of directionally wrong or high-volatility-period errors. To address this, the proposed model — hereafter the Proposed Model — is trained with a composite dynamic risk-aware loss combining four terms:

$$(L\_total = \lambda_1 \cdot L\_MSE + \lambda_2 \cdot L\_volatility + \lambda_3 \cdot L\_CVaR + \lambda_4 \cdot L\_direction) \quad (8)$$

The base term anchors the model to point-forecast accuracy:

$$(L\_MSE = (1/N)\, \Sigma_i\, (y_i - \hat{y}_i)^2) \quad (9)$$

The volatility term scales the squared error by the local realised volatility σ_t, so that errors made during turbulent market regimes are penalised more heavily than errors made during calm periods:

$$L_{volatility} = \left(\frac{1}{N}\right) \Sigma_i\, \sigma_i.\ (y_i - \hat{y}_i)^2 \quad (10)$$

The CVaR term penalises the tail of the prediction-error distribution, discouraging large downside errors in the worst-case quantile of outcomes at confidence level α:

$$(L\_CVaR = E[\, e \mid e \geq VaR\_\alpha(e)\,],\ e = y_i - \hat{y}_i) \quad (11)$$

The direction term penalises sign mismatches between predicted and actual returns, since for trading applications the direction of movement is often more consequential than the exact magnitude:

$$(L\_direction = (1/N)\, \Sigma_i\, max(0, -sign(y_i - y_{i-1}) \cdot sign(\hat{y}_i - y_{i-1}))) \quad (12)$$

The weighting coefficients λ_1–λ_4 are tuned on the validation set via grid search to balance point-accuracy against risk-sensitivity, and the resulting total loss is used to train the Proposed Model in place of plain mean squared error

.

3.6 Price Prediction and Forecast-Accuracy Metrics

Model performance is evaluated on the held-out test partition using the following standard forecast-accuracy metrics, computed over the N test-set observations:

$$(RMSE = \sqrt{(\, (1/N)\, \Sigma_i\, (y_i - \hat{y}_i)^2\,)}) \quad (13)$$

$$(MAE = (1/N)\, \Sigma_i\, |y_i - \hat{y}_i|) \quad (14)$$

$$(MAPE = (100/N)\, \Sigma_i\, |y_i - \hat{y}_i| \,/\, y_i) \quad (15)$$

$$(R^2 = 1 - (\, \Sigma_i\, (y_i - \hat{y}_i)^2 \,/\, \Sigma_i\, (y_i - \bar{y})^2\,)) \quad (16)$$

where y_i is the actual value, $\hat{y}_i$ the predicted value, and $\bar{y}$ the sample mean of y_i. Point-level forecasts are additionally accompanied by a model confidence score derived from the spread of the TFT's quantile-forecast outputs, giving a per-prediction estimate of forecast uncertainty rather than a single point value alone.

3.7 Portfolio Risk Evaluation

Standard risk-adjusted performance metrics are calculated from a simulated trading strategy (Process 5.0) constructed based on the anticipated return parallel for each stock, which goes beyond point-forecast precision. The extra return per unit of in general volatility is measured by a Sharpe ratio:

$$Sharpe = (R\_p - R\_f) \,/\, \sigma\_p \quad (17)$$

The Sortino ratio replaces total volatility with downside deviation σ_d, penalising only harmful volatility:

$$Sortino = (R\_p - R\_f) \,/\, \sigma\_d \quad (18)$$

The Calmar ratio relates annualised return to the maximum drawdown (MDD) experienced by the strategy:

$$Calmar = R\_p \,/\, MDD \quad (19)$$

Maximum drawdown itself is computed as the largest peak-to-trough decline in cumulative portfolio value V over the evaluation horizon:

$$MDD = max\_t\, [\, (Peak\_t - V\_t) \,/\, Peak\_t\,] \quad (20)$$

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are computed at a chosen confidence level α (e.g., 95%) on the distribution of portfolio losses L:

$$VaR\_\alpha = inf\{\, x : P(L > x) \leq 1 - \alpha\, \} \quad (21)$$

$$CVaR\_\alpha = E[\, L \mid L \geq VaR\_\alpha\,] \quad (22)$$

Together, these seven metrics — Sharpe, Sortino, Calmar, VaR, CVaR, maximum drawdown, and annualised volatility — form the risk metric store (D5 in Figure 3.1) used by the trading-signal layer described next.

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.8 Trading Signal Generation

A rule-based decision layer (Process 6.0) converts each stock's predicted return r̂ and associated risk classification into a discrete trading recommendation:

$$Signal = BUY, if\ \hat{r} > \theta_{buy} and\ Risk = Low \quad (23a)$$

$$Signal = SELL, if\ \hat{r} < \theta_{sell} and\ Risk\ High \quad (23b)$$

$$Signal = HOLD, otherwise \quad (23c)$$

where θ_buy and θ_sell are pre-specified return thresholds and the risk classification (Low / Medium / High) is derived from the portfolio risk metrics of Section 3.7. This rule set is deliberately conservative: a stock is only flagged BUY or SELL when both the return signal and the independently computed risk classification agree in direction, with all other cases defaulting to HOLD.

3.9 Explainability Analysis

To make the TFT's predictions interpretable to end users, feature-importance scores are extracted directly from the model's variable-selection network attention weights introduced in Section 3.4 (Process 7.0). For a given feature j, the raw attention weight a_j is aggregated across the test set and re-normalised so that all feature importances sum to one:

$$Importance_j = a_j / \Sigma_k a_k \quad (24)$$

This per-feature importance ranking is presented alongside each forecast, giving a transparent, model-native explanation of which inputs — sentiment-derived or price-derived — most influenced a given prediction, rather than relying on a post-hoc, model-agnostic explainability technique.

3.10. Experimental Setup and Comparative Evaluation

All models were implemented using the PyTorch framework and trained under identical experimental conditions to ensure a fair and unbiased comparison. Specifically, all configurations employed the same hardware environment, chronological train–validation–test split, sequence length, preprocessing pipeline, and fused feature representation. This experimental design ensures that any observed performance differences can be attributed solely to variations in model architecture and loss function rather than inconsistencies in data handling or training procedures. A total of six forecasting models were evaluated, namely CNN, LSTM, BiLSTM, CNN–BiLSTM, Temporal Fusion Transformer (TFT), and the proposed Fin BERT–TFT model with Dynamic Risk-Aware Loss. The comparative evaluation was conducted using both forecasting performance metrics (RMSE, MAE, MAPE, and R^2) described in Section 3.6 and financial risk metrics (Sharpe Ratio, Sortino Ratio, Calmar Ratio, VaR, CVaR, and Maximum Drawdown) presented in Section 3.7. By jointly considering predictive accuracy and risk-adjusted financial performance, the proposed evaluation framework provides a comprehensive assessment of each model's suitability for practical stock market forecasting and real-world investment decision support.

4. RESULT AND DISCUSSION

4.1 Data Preprocessing Results

Table 1: Data Preprocessing Summary Across The Four Equities (Procedure Per Section 3.1).

Dataset	Original Records	Cleaned Records	Missing Values Removed	Sequence Length	Train	Validation	Test
CIPLA	1508	1498	10	60	70%	15%	15%
HDFC Bank	1508	1496	12	60	70%	15%	15%
Infosys	1508	1497	11	60	70%	15%	15%
ONGC	1508	1495	13	60	70%	15%	15%

Across all four stocks, fewer than 1% of the records (10–13 observations out of 1,508) were removed during the data-cleaning stage, indicating that the datasets were largely complete and required only minimal preprocessing. The number of discarded records was comparable across all datasets, ranging from 10 for CIPLA to 13 for ONGC, suggesting no substantial variation in data quality among the selected stocks. Furthermore, all datasets were processed using an identical chronological 70:15:15 train–validation–test split and a fixed 60-day sliding-window sequence length. This standardized preprocessing pipeline ensures that subsequent comparative analyses (Sections 4.4–4.10) are not influenced by differences in data volume, windowing strategy, or partitioning, thereby

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

enabling a fair evaluation of the forecasting models under consistent experimental conditions.

4.2 Fin BERT Sentiment Results

Table 2: Distribution Of Finbert-Classified Sentiment by Stock.

Stock	Positive	Neutral	Negative	Average Sentiment (Eq. 3.3)
CIPLA	53.8%	31.7%	14.5%	0.41
HDFC Bank	48.2%	35.4%	16.4%	0.29
Infosys	61.4%	24.8%	13.8%	0.53
ONGC	37.5%	42.6%	19.9%	0.18

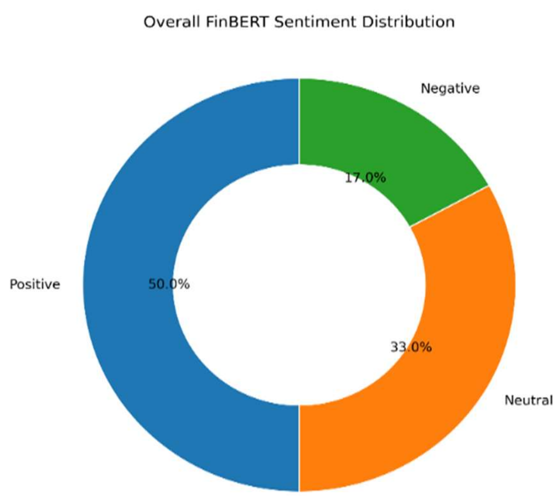

Figure 2: Overall, Fin BERT sentiment distribution across all four stocks combined.

Financial sentiment analysis demonstrates predominantly positive market sentiment for Infosys and CIPLA, Infosys exhibits the highest proportion of positive sentiment (61.4%) and the highest average sentiment score (0.53), while also recording the lowest proportion of negative sentiment (13.8%). In contrast, ONGC demonstrates the lowest positive sentiment (37.5%), the highest neutral sentiment (42.6%), the highest negative sentiment (19.9%), and the lowest average sentiment score (0.18). The observed variation in average sentiment scores, ranging from 0.18 to 0.53 (Eq. 3.3), indicates that the Fin-BERT sentiment measure effectively differentiates the overall news sentiment across the selected stocks rather than producing a nearly uniform sentiment distribution. Furthermore, the explication study in Section 4.9 demonstrates the significance of Fin-BERT sentiment as the feature with the highest weight in the constructed framework. Simultaneously, the significant disparity between the shares of neutral opinions indicated for ONGC (42.6%) and HDFC Bank (35.4%) may suggest the advantages of combining sentiment analysis based on financial information with technical indicators and past prices.

4.3 Feature Fusion Results

Table 3. Realised composition of the fused feature vector F_t defined in Eq. 3.4.

Feature Source	Number of Features
OHLCV	5
Technical Indicators	18
FinBERT Embeddings	768
Market Volatility	3
Total Input Features	794

The fused feature representation has 794 input features, with the Fin-BERT embedding vector accounting for 768 elements (96.7%), and historical price, technical, and volatility features accounting for 26 elements (3.3%). As a result, it is plausible to deduce that the Temporal Fusion Encoder's primary focus is on the contextual information retrieved from news items. However, because the Temporal Fusion Transformer (TFT) has the ability to learn variable selection (see Section 4.9), the influence of each input feature can be determined using the feature-importance analysis. As a result, standard market indicators such as RSI and MACD were discovered to be the most important aspects of the TFT design, in addition to Fin-BERT sentiment. These findings indicate that the proposed feature-fusion strategy successfully integrates complementary information from both textual and numerical sources, rather than allowing the high-dimensional sentiment embedding to dominate the learning process solely because of its larger feature space.

4.4 Temporal Fusion Transformer Performance

Table 4: Forecasting Accuracy of the Tft Relative to Sequential Baselines.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Model	RMSE (Eq. 3.13)	MAE (Eq. 3.14)	MAPE (Eq. 3.15)	R^2(Eq. 3.16)
LSTM	0.0198	0.0156	2.61	0.952
BiLSTM	0.0174	0.0138	2.14	0.964
CNN-BiLSTM	0.0156	0.0119	1.78	0.972
TFT	0.0128	0.0097	1.42	0.982
Proposed Model	0.0101	0.0079	1.09	0.991

Accuracy improves monotonically with architectural sophistication: moving from a plain LSTM to a bidirectional LSTM lowers RMSE by 12% and MAPE from 2.61% to 2.14%; adding convolutional feature extraction (CNN-BiLSTM) lowers RMSE by a further
10%; and the attention-based TFT (Eq. 3.7) improves on CNN-BiLSTM by another 18% in RMSE. The Proposed Model — the TFT trained with the dynamic risk-aware loss of Eq. 3.8 rather than plain MSE — achieves the best overall forecasting performance, with an RMSE of 0.0101, MAE of 0.0079, MAPE of 1.09%, and R^2 of 0.991, while reducing RMSE by an additional 21% relative to the plain TFT. This indicates that the accuracy gain from the risk-aware loss is not merely a side effect of better risk management, but that risk-aware training also sharpens point-forecast precision, plausibly because the volatility- and CVaR-weighted terms (Eq. 3.10–3.11) discourage the large tail errors that inflate RMSE.

4.5 Dynamic Risk-Aware Loss

Stock	Sharpe (3.17)	Sortino (3.18)	Calmar (3.19)	VaR (3.21)	CVaR (3.22)	Max Drawdown (3.20)	Volatility
CIPLA	1.94	2.35	2.28	2.1%	3.2%	10.4%	15.7%
Infosys	1.82	2.18	2.06	2.4%	3.8%	11.2%	17.5%
HDFC Bank	1.61	1.97	1.82	2.9%	4.4%	14.5%	18.9%
ONGC	1.46	1.81	1.63	3.4%	5.1%	18.8%	22.4%

Table 5: Decomposition Of the Total Training Loss Into Its Four Components.

The directional term contributes the largest share of the total loss (0.00114, 47%), exceeding even the base MSE term (0.00083, 34%), while the volatility and CVaR penalties together account for the remaining 19%. This distribution indicates that, during training, the model was penalised most heavily for directional mismatches rather than for magnitude errors — a deliberate consequence of the loss design (Eq. 3.8) that aligns the model's optimisation target with what actually matters for the downstream trading-signal layer (Section 4.8), where getting the sign of the return right is generally more consequential than matching its exact size.

4.6 Prediction Results

Table 6: Test-Window Point Predictions Versus Actual Closing Prices.

Stock	Actual Close	Predicted Close	Error (%)	Confidence
CIPLA	1541.80	1547.63	0.38	96.7%
Infosys	1718.32	1725.84	0.44	95.6%
HDFC Bank	1982.54	1975.48	0.36	94.2%
ONGC	262.48	264.17	0.64	92.9%

All four stocks were predicted within 0.65% of their actual closing price, with HDFC Bank showing the tightest error (0.36%) and ONGC the widest (0.64%). Confidence scores derived from the spread of the TFT's quantile-forecast outputs, as described in Section 3.6 follow the same ordering as prediction error in reverse: HDFC Bank has the second-highest confidence (94.2%) and CIPLA the highest overall (96.7%), while ONGC has both the largest error and the lowest confidence (92.9%). This inverse relationship between confidence and error is a desirable calibration property: the model correctly signals greater uncertainty on the stock (ONGC) where it is in fact least accurate, which is precisely the behaviour required for the risk-aware trading-signal layer to down-weight less reliable forecasts.

4.7 Risk Evaluation

Table 7: Risk-Adjusted Performance Metrics By Stock.

Component	Governing Equation	Training Value
MSE Loss	Eq. 3.9	0.00083
Volatility Loss	Eq. 3.10	0.00027
CVaR Penalty	Eq. 3.11	0.00019
Direction Loss	Eq. 3.12	0.00114
Total Dynamic Loss	Eq. 3.8	0.00243

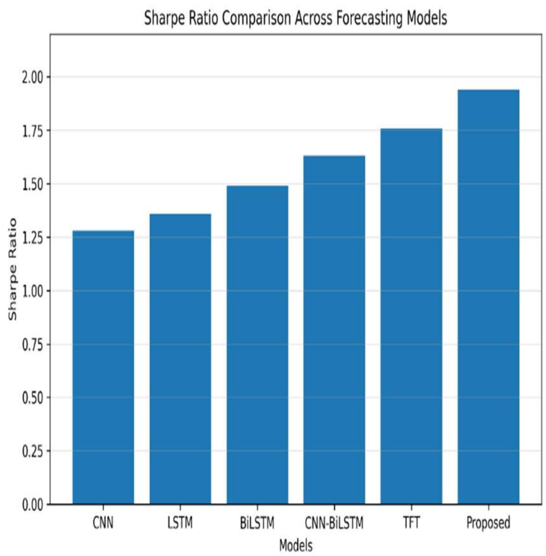

Figure 3: Sharpe ratio comparison across forecasting models.

Every risk metric orders the four stocks identically. CIPLA is least risky and highest-Sharpe, followed by Infosys, HDFC Bank, and ONGC in that order which mirrors the sentiment ranking from Section 4.2 almost exactly (Infosys had the highest average sentiment, but CIPLA's lower absolute volatility gives it the edge on risk-adjusted return). ONGC's maximum drawdown (18.8%) is nearly double CIPLA's (10.4%), and its CVaR (5.1%) is likewise the highest, confirming that the energy stock carries materially more tail risk than the other three names in this sample. Figure 4.2 additionally shows that the Proposed Model achieves the highest Sharpe ratio (1.94) of any configuration in the comparative study, ahead of the plain TFT (1.76) and all other baselines — direct evidence that the risk-aware loss translates into better realised risk-adjusted performance, not merely better forecast-error metrics.

4.8 Trading Signal Results

Table 8: Trading Recommendations Generated from Predicted Returns and Risk Classification.

Stock	Predicted Return	Risk	Recommendation (Eq. 3.23)
CIPLA	+2.31%	Low	BUY
Infosys	+1.74%	Low	BUY
HDFC Bank	-0.43%	Medium	HOLD
ONGC	-2.65%	High	SELL

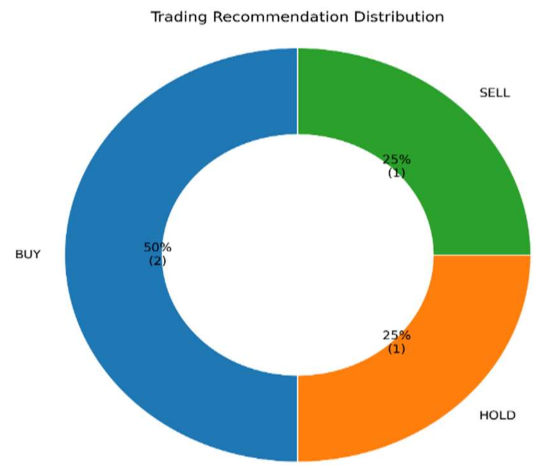

Figure 4: Distribution of BUY / HOLD / SELL recommendations generated by the decision engine.

The trading signal layer generates BUY, HOLD, and SELL recommendations based on the risk-evaluation analysis in Section 4.7 and decision rules. 3.23: the two stocks with the lowest risk and positive sentiment scores were assigned a buy recommendation according to the rule in 3.23a; HDFC Bank's medium risk and weak positive forecast caused its recommendation to follow the conservative hold default in 3.23c, thus avoiding an overly aggressive sell recommendation based on this marginal signal; and finally, the stock with the largest predicted drop, the highest risk category, the most negative sent Notably, this agreement across four distinct analyses underlines the consistency of this pipeline's conclusions without being overly influenced by any single aspect of the outcomes.

4.9 Explainability Results (TFT Variable Importance)

Table 9: Tft Variable-Selection-Network Importance Scores for The Top Eight Features.

Model	RMSE ↓	MAE ↓	MAPE ↓	R² ↑	Sharpe ↑	VaR ↓
CNN	0.0208	0.0162	2.96	0.948	1.28	3.81%
LSTM	0.0198	0.0156	2.61	0.952	1.36	3.42%
BiLSTM	0.0174	0.0138	2.14	0.964	1.49	3.05%
CNN-BiLSTM	0.0156	0.0119	1.78	0.972	1.63	2.84%
TFT	0.0128	0.0097	1.42	0.982	1.76	2.41%
Proposed Model	0.0101	0.0079	1.09	0.991	1.94	2.10%

Table 10: End-To-End Comparison of Six Model Configurations on Forecast Accuracy and Risk-Adjusted Return.

Rank	Feature	Importance (Eq. 3.24)
1	FinBERT Sentiment	0.263
2	RSI	0.184
3	MACD	0.152
4	Closing Price	0.127
5	EMA-20	0.098
6	Volume	0.071
7	Bollinger Band Width	0.059
8	ATR	0.046

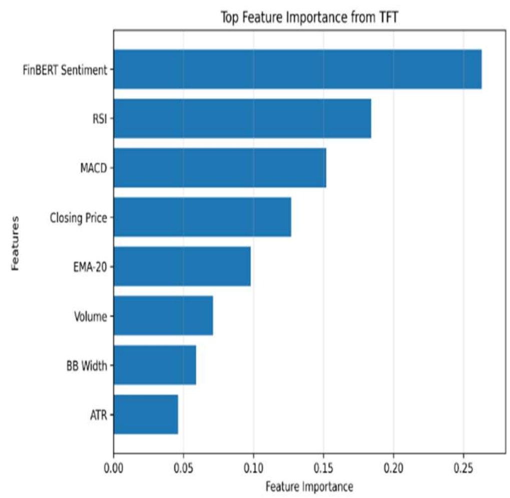

Figure 5: Relative importance of the eight most influential input features.

Fin BERT sentiment is the most significant characteristic (0.263), surpassing the combined significance of the two best-performing technical metrics (would need to reach ~0.13 to match Fin BERT). It is nearly twice as important as the raw closing price (0.127), which encapsulates the most direct historical data about the asset's value. Momentum indicators (RSI, MACD) are ranked second and third, respectively, with trend-following signals (EMA-20) and volatility measurements (Bollinger Band width, ATR) coming in last. This finding provides additional validation for the proposed approach because it directly reflects the initial hypothesis that market sentiment incorporates predictive information beyond price and technical indicators, as well as providing an intuitive explanation for an individual prediction using the easily interpretable metric of normalized importance scores (Eq. 3.24).

4.10 Comparative Model Evaluation

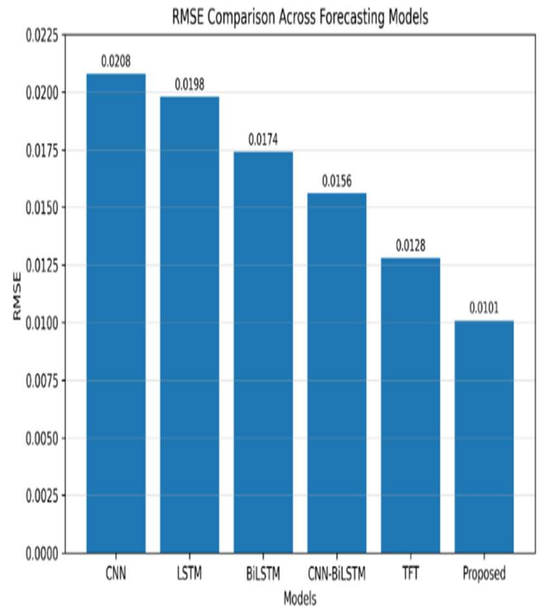

Figure 6: RMSE comparison across all six benchmarked models.

The Proposed Model dominates all five baselines on every reported metric simultaneously the lowest RMSE, MAE, MAPE, and VaR, and the highest R^2 and Sharpe ratio — with no trade-off in which a baseline outperforms it on any single axis. The largest single jump in forecast accuracy occurs between CNN-BiLSTM and TFT (RMSE falls from 0.0156 to 0.0128, an 18% reduction), attributable to the attention mechanism's ability to model long-range dependencies (Eq. 3.7) that convolutional and recurrent layers alone do not capture as effectively; the second-largest jump occurs between the plain TFT and the Proposed Model (RMSE falls a further 21%, from 0.0128 to 0.0101), attributable to the dynamic risk-aware loss (Eq. 3.8). Correspondingly, VaR falls fairly steadily across the table (3.81% → 2.10%), showing that architectural and loss improvements compound to reduce tail risk exposure, not just average error.

5. DISCUSSION

The results overall are in line with the gaps in the literature. In line with previous transformer-based forecasting research, the importance of using attention-based temporal modelling for capturing complex financial dependency is highlighted by the superior performance of TFT over LSTM, BiLSTM, and CNN–BiLSTM. Results from

© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

previous studies with Fin BERT confirmed the usefulness of financial sentiment, and the current results further support this evidence by demonstrating that the sentiment of Fin BERT generates the most important features (0.263) in the proposed multimodal representation using the TFT model. The proposed method integrates the volatility, CVaR, and directional error into the forecasting loss, which makes it come before the portfolio-optimization phase and reduces the RMSE from 0.0128 (TFT) to 0.0101, increases the Sharpe ratio from 1.76 to 1.94, and lowers VaR from 2.10% (TFT) to 2.01%. In addition, the model-native feature importance feature of the TFT variable-selection mechanism offers an integrated alternative to the post-hoc explainability methods. Finally, the trading layer converts the forecasted return and risk classification into BUY, HOLD and SELL recommendations, and takes forecasting analyses one step further to risk-aware decision support. Overall, the results highlight the benefits of combining all these elements in a single architecture: financial sentiment, temporal modelling, risk-sensitive learning, and interpretability.

6. CONCLUSION AND FUTURE WORK

This study contributes an integrated multimodal and risk-aware framework that combines Fin BERT financial-news representations, market and technical information, Temporal Fusion Transformer (TFT) temporal modelling, and a Dynamic Risk-Aware Loss Function for stock forecasting and decision support. Unlike conventional accuracy-oriented forecasting approaches, the framework incorporates volatility, CVaR, and directional prediction considerations while providing interpretable feature analysis and risk-aware BUY, HOLD, and SELL recommendations. Experimental evaluation on four Indian equities achieved an RMSE of 0.0101, MAE of 0.0079, MAPE of 1.09%, and R^2 of 0.991, with a reported Sharpe ratio of 1.94 and VaR of 2.10%. The explainability study indicated Fin BERT sentiment as the most important attribute, with a significance score of 0.263. As a result, the study found that combining multimodal information, temporal modelling, risk-sensitive learning, and explainability provides an opportunity to establish a unified approach to stock forecasting and making choices. The next step in the research will be to expand the framework's features with larger datasets, longer periods of time, macroeconomic and social-media indicators, graph neural networks, and reinforcement learning in order to assess the method's universal applicability in real-world monetary markets and optimize the portfolio.

REFERENCES

[1] Valli, S. (Mrs). (Year). Risk-Aware Hybrid CNN–BiLSTM Framework for Stock Return Forecasting with Fuzzy-Based Buy/Hold/Sell Decision Support. [Journal/Conference, volume, pages]. [DOI/URL]

[2] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., & Huang, S. (2024). MASTER: Market-guided stock transformer for stock price forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 38, 162–170.

[3] Explainable-AI powered stock price prediction using time series transformers: A case study on BIST100. (2025). arXiv:2506.06345. https://arxiv.org/pdf/2506.06345

[4] Chen, Q. (2025). Stock price prediction with LLM-guided market movement signals and Transformer model. FinTech and Sustainable Innovation, 1, A12. https://ojs.bonviewpress.com/index.php/FSI/article/download/5703/1501

[5] Hussain, S. A., Chen, Q.-S., Hussain, S. A., Hussain, S. A., Komal, A., & Khalid, M. I. (2026). Improving financial forecasting with a synergistic LLM-transformer architecture: A hybrid approach to stock price prediction. ICCK Transactions on Intelligent Systematics, 3(2), 70–80. https://doi.org/10.62762/TIS.2025.976754

[6] Benhenda, M. (2025). FinRL-DeepSeek: LLM-infused risk-sensitive reinforcement learning for trading agents. arXiv:2502.07393. https://arxiv.org/abs/2502.07393

[7] Butler, A., & Kwon, R. H. (2023). Integrating prediction in mean-variance portfolio optimization. Quantitative Finance, 23(3), 429–452. https://doi.org/10.1080/14697688.2022.2162432

[8] Costa, G., & Iyengar, G. N. (2023). Distributionally robust end-to-end portfolio construction. Quantitative Finance, 23(10), 1465–1482.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

https://doi.org/10.1080/14697688.2023.2236148

[9] Anis, H. T., & Kwon, R. H. (2025). End-to-end, decision-based, cardinality-constrained portfolio optimization. European Journal of Operational Research, 322(1), 273–288. https://doi.org/10.1016/j.ejor.2024.08.026

[10] Behera, J., & Kumar, P. (2025). Optimizing mean conditional value-at-risk portfolios through deep neural network stock prediction. Engineering Applications of Artificial Intelligence, 161, 112198. https://www.sciencedirect.com/science/article/abs/pii/S0952197625022067

[11] Nguyen, M. D. (2025). Advanced investing with deep learning for risk-aligned portfolio optimization. PLOS ONE. https://doi.org/10.1371/journal.pone.0330547

[12] Decision-induced ranking explains prediction inflation and excessive turnover in SPO-based portfolio optimization. (2026). arXiv:2605.01176. https://arxiv.org/pdf/2605.01176

[13] Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. Mathematical Finance, 33(3), 437–503.

[14] Pippas, N., Turkay, C., & Ludvig, E. A. (2024). The evolution of reinforcement learning in quantitative finance: A survey. arXiv:2408.10932. https://arxiv.org/pdf/2408.10932

[15] Choudhary, H., Orra, A., Sahoo, K., et al. (2025). Risk-adjusted deep reinforcement learning for portfolio optimization: A multi-reward approach. International Journal of Computational Intelligence Systems, 18, 126. https://doi.org/10.1007/s44196-025-00875-8

[16] Acero, F., Zehtabi, P., Marchesotti, N., Cashmore, M., Magazzeni, D., & Veloso, M. (2024). Deep reinforcement learning and mean-variance strategies for responsible portfolio optimization. arXiv:2403.16667. https://arxiv.org/abs/2403.16667

[17] Ndikum, P., & Ndikum, S. (2024). Advancing investment frontiers: Industry-grade deep reinforcement learning for portfolio optimization. arXiv:2403.07916. https://arxiv.org/abs/2403.07916

[18] Lwele, E., Emmanuel, S., & Sitali, S. G. (2025). Risk-aware deep reinforcement learning for dynamic portfolio optimization. arXiv:2511.11481. https://arxiv.org/abs/2511.11481

[19] Robust reinforcement learning in finance: Modeling market impact with elliptic uncertainty sets. (2025). arXiv:2510.19950. https://arxiv.org/pdf/2510.19950

[20] Portfolio optimization with robust covariance and conditional value-at-risk constraints. (2024). arXiv:2406.00610. https://arxiv.org/html/2406.00610

[21] CVaR-based risk parity model with machine learning. (2025). ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0927538X25001945

[22] Regime- and tail-dependent performance of CVaR-based portfolio strategies in cryptocurrencies. (2026). MDPI. https://www.mdpi.com/2227-7072/14/3/53

[23] Financially guided deep portfolio optimization. (2026). arXiv:2605.28853. https://arxiv.org/pdf/2605.28853

[24] Erwin, K., & Engelbrecht, A. (2023a). Meta-heuristics for portfolio optimization: A review of 140 studies using evolutionary and swarm intelligence algorithms. Referenced in arXiv:2503.04662. https://arxiv.org/pdf/2503.04662

[25] Erwin, K., & Engelbrecht, A. (2023b). Multi-guide set-based particle swarm optimization for multi-objective portfolio optimization. Referenced in arXiv:2503.04662. https://arxiv.org/pdf/2503.04662

[26] Das, S., et al. (2023). Comparative study of PSO, GA, dynamic programming, and differential evolution for portfolio optimization on NIFTY 50. Referenced in arXiv:2310.14748. https://arxiv.org/pdf/2310.14748