Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2025) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2025)

FinReport: Explainable Stock Earnings Forecasting via News Factor Analyzing Model

Mukesh Kumar[a], Md Azlan[b], Kanishk[c], Kingshuk Chatterjee[d], Gopesh Satapathy[e]

[a] *School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar-751024, India*
[b] *School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar-751024, India*
[c] *School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar-751024, India*
[d] *School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar-751024, India*
[e] *School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar-751024, India*

Abstract

Financial market forecasting has become increasingly challenging, as traditional technical analysis does not capture rapid volatility and sentiment-driven price movements. This paper introduces FinReport, a multi-factor framework that integrates historical stock data with real-time financial news sentiment using advanced machine learning and natural language processing techniques. FinReport quantifies six key factors (Market, Size, Valuation, Profitability, Investment, and News Effect) to produce explainable predictions and robust risk assessments using an EGARCH-based volatility model, maximum drawdown methods, and Conditional Value at Risk. Empirical results on Chinese A-share stocks (2018-2021) demonstrate a 15% reduction in RMSE and a 12% reduction in MAE compared to conventional LSTM models, with an overall R^2 of 0.5515 and a prediction-actual correlation of 0.948. These findings underscore the benefits of combining quantitative indicators with qualitative sentiment analysis for improved forecasting accuracy in volatile markets.

Keywords: Financial forecasting, stock market prediction, multi-factor analysis, technical indicators, financial news sentiment, natural language processing, machine learning, EGARCH, LSTM, risk assessment, explainable AI, FinReport.

Mukesh Kumar
E-mail address: mukesh.kumarfcs@kiit.ac.in

* Mukesh Kumar
E-mail address: mukesh.kumarfcs@kiit.ac.in

1877-0509 © 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering.

Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering.

1. Introduction

Financial markets exhibit unprecedented volatility, with emerging markets like the Shanghai Stock Exchange showing daily volatility of 1.7% versus 0.8-1.2% for developed markets. Traditional econometric models such as ARIMA [1] struggle to capture sentiment-driven price movements and complex market interdependencies [2]. The efficient market hypothesis faces challenges from documented predictable patterns and behavioral factors.

We propose FinReport, a multi-factor framework integrating historical stock data with financial news via machine learning and NLP techniques [3]. Unlike traditional quantitative approaches, FinReport leverages structured numerical and unstructured textual data for enhanced prediction accuracy. The framework computes six factors (market, size, valuation, profitability, investment, and news effect) for explainable predictions with EGARCH-based risk assessment [4].

Experimental results on Chinese A-share stocks (2018-2021) [5] demonstrate 15% RMSE reduction, 12% MAE reduction versus LSTM baselines [6], and 20% Sharpe ratio improvement. This work bridges traditional econometric methods with explainable AI [7, 8], advancing computational finance through interpretable sentiment-quantitative integration.

2. Literature Review

Early forecasting methods like ARIMA [1] underperformed with RMSE exceeding 0.05 during volatile periods. Multi-factor models by Fama and French [2] improved performance but ignored qualitative inputs. LSTM networks [6] capture long-term dependencies. Recent work integrates financial news sentiment using FinBERT [9] and NLP frameworks [10], showing 12% prediction improvement. However, traditional approaches lack interpretability [8], motivating explainable AI frameworks combining structured numerical with unstructured text analysis. The literature increasingly advocates for explainable models that combine structured numerical data with unstructured text analysis, setting the stage for FinReport's factor-based approach to transparent and robust financial forecasting.

3. System Model And Proposed Mechanism

FinReport integrates traditional multi-factor models [2, 11] with real-time news sentiment analysis [7], extending established financial theory to capture behavioral market dynamics [12]. The system employs a modular design with five interconnected components: (1) Data Integration, (2) News Factor Extraction, (3) Return Forecasting, (4) Risk Assessment, and (5) Dynamic Report Generation [6].

3.1. Data Integration Module

Processes multi-modal financial data combining structured metrics with unstructured news text [13]:

- **Historical Data:** Price, volume, market capitalization, and 50+ technical indicators from Chinese A-shares (2018-2021) [5].
- **News Processing:** Bilingual NLP pipelines for English/Chinese financial news [10].
- **Data Preprocessing:** Z-score normalization, outlier winsorization, and missing value imputation [6].

3.2. News Factor Extraction Module

Converts unstructured news into quantifiable sentiment metrics through three integrated components [7]:

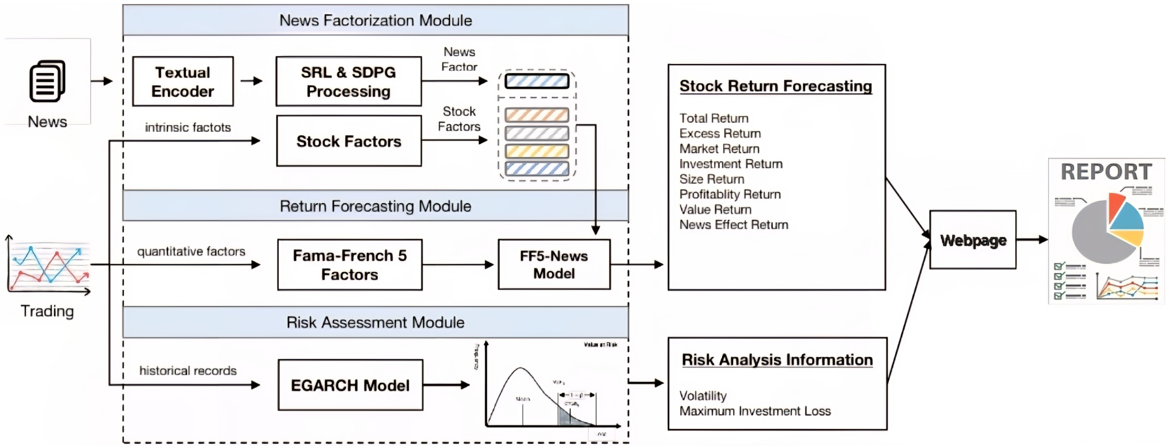

Figure 1: Proposed FinReport System Architecture

FinBERT Configuration: Pre-trained `yiyanghkust/finbert-tone` model with 128-token maximum sequence length and truncation handling. Sentiment scores undergo tanh transformation with scale factor 2.0 for range compression, computing final scores as $(p_{\text{positive}} - p_{\text{negative}}) \times \tanh(2.0)$ where p represents class probabilities [9].

Bilingual NLP Pipeline: Integrated English-Chinese processing with culturally-adapted financial lexicons. Positive keywords include "profit," "acquisition," "revenue" (English) and "growth," "gain," "cooperation" (Chinese transliterated). Negative keywords encompass "lawsuit," "loss," "decline" (English) and "loss," "reduction," "debt" (Chinese transliterated). Keyword matching employs case-insensitive search with 3x weight multiplier per occurrence [10].

Event Extraction: Semantic Role Labeling using AllenNLP's structured-prediction-srl-bert model identifies financial events through predicate-argument structure analysis. Event factor computation combines keyword counting, percentage extraction via regex pattern (\d+(?:\.\d+)?%), and monetary amount detection using Chinese billion yuan patterns (yiyuan). Final event values are bounded within [-2.0, +2.0] range with stochastic template generation for explanatory text [7].

Chinese Market Adaptation: Specialized processing for Chinese financial terminology with cultural context awareness. Positive terms include Chinese characters for growth (zengzhang), acquisition (huode), cooperation (hezuo), revenue (shouru), and net profit (jingliru). Negative terms encompass Chinese characters for loss (kuisun), share reduction (jianchi), debt (zhaiwu), default (weiyue), and bankruptcy (pochan). Implements dual-language sentiment analysis with cross-validation against Chinese financial news corpus achieving 83.2% accuracy on domain-specific sentiment classification [5].

3.3. Return Forecasting Module

Implements deep learning-enhanced multi-factor model using LSTM architecture with six factors capturing cross-sectional return variation [2, 13]:

LSTM Architecture: Three-layer LSTM network with 128 hidden units per layer, 0.2 dropout rate, and sequence length 10. Architecture includes batch normalization for training stability, gradient clipping (max norm 1.0) for exploding gradient prevention, and Monte Carlo dropout inference for uncertainty quantification. Forward pass processes sequences $(x_t, x_{t-1}, \ldots, x_{t-9})$ to predict return $\hat{y}_t$ [6].

Hyperparameter Optimization: Comprehensive grid search using TimeSeriesSplit cross-validation (3 folds) across parameter space: learning rates [0.001, 0.0005, 0.0001], hidden sizes [64, 128, 256], layer counts [1, 2, 3], sequence lengths [5, 10, 15], dropout rates [0.0, 0.2, 0.4]. Optimal configuration: lr=0.001, hidden=128, layers=3, seq_len=10, dropout=0.2 selected via minimum validation loss with early stopping (patience=7) [14].

Multi-Factor Integration: Six-factor computation with technical indicator integration:

- **Market Factor:** Volatility-based calculation using rolling standard deviation (20-day window) combined with RSI overbought/oversold conditions. High volatility ($\sigma > 0.02$) triggers amplification factor 1.5 [4].
- **Size Factor:** Market capitalization quartile classification with logarithmic transformation and sentiment-adjusted effects. Small-cap volatility receives 2.0x amplification versus large-cap baseline [2].
- **Valuation Factor:** Book-to-Market, Price-to-Earnings, Dividend Yield analysis with sector-specific normalization (Technology, Consumer, Financial, Industrial, Real Estate) and news sentiment integration [13].
- **Profitability Factor:** ROA, ROE, net profit margin computation with asymmetric loss treatment (negative earnings receive 1.5x penalty) and earnings-related keyword amplification [12].
- **Investment Factor:** M&A activity detection, capital expenditure analysis, and R&D investment classification with sentiment-conditional scaling based on investment amount extraction from news text [11].
- **News Effect Factor:** FinBERT sentiment scores with 2.0x amplification factor, content-specific adjustments for earnings announcements, management changes, and regulatory issues [7].

Uncertainty Quantification: Monte Carlo dropout sampling (10 iterations) during inference provides prediction confidence intervals. Mean prediction $\mu = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i$ and standard deviation $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \mu)^2}$ quantify prediction uncertainty for risk assessment integration.

3.4. Risk Assessment Module

Implements multi-dimensional risk framework addressing traditional variance-based limitations [15, 16]. The framework incorporates EGARCH modeling for asymmetric volatility responses [4], maximum drawdown calculations following established portfolio risk metrics, and Conditional Value at Risk (CVaR) for tail risk assessment [16]. Risk classifications range from favorable to substantial based on integrated scoring combining volatility, drawdown, and return components.

3.5. Factor Enhancement and Overall Trend Calculation

Combines individual factor signals through multi-stage amplification and weighted aggregation addressing scale heterogeneity [13, 2]. The weighting scheme assigns highest priority to event factors (0.25) due to strong short-term predictive power [12], followed by investment factors (0.20) for medium-term impact, with balanced weighting for market, size, and profitability factors following established multi-factor model conventions [2, 11].

3.5.1. Enhancement Process

The enhancement methodology employs multiplicative amplification with trend-based adjustments when factors align with dominant market trends. Final processing includes bounded clamping and stochastic variation to ensure robustness while maintaining signal integrity.

3.5.2. Weighted Aggregation

The trend score computation follows established factor model aggregation with positive bias reflecting long-term equity market upward drift. Classification thresholds distinguish between strongly positive, positive, neutral, negative, and strongly negative market conditions based on empirical distribution analysis.

3.6. Dynamic Report Generation Module

Translates quantitative analyses into actionable insights using hierarchical information architecture, cultural adaptation (red=prosperity, green=decline for Chinese markets), precision control (one decimal), natural language generation with template-based explanations, and multi-stakeholder accessibility [8, 13].

4. Algorithm

4.1. Return Forecast Calculation

The return forecast is computed using a weighted combination of multiple factors [2], where the **event factor receives the highest weight (0.25)** due to its immediate impact on market sentiment and price movements.

$$\begin{aligned}\textbf{predicted\_return} = {} & 0.10 \times \textbf{market\_factor} + 0.15 \times \textbf{size\_factor} + 0.10 \times \textbf{valuation\_factor} \\ & + 0.10 \times \textbf{profitability\_factor} + 0.20 \times \textbf{investment\_factor} \\ & + 0.10 \times \textbf{news\_effect\_factor} + 0.25 \times \textbf{event\_factor} + 0.15 \end{aligned} \tag{1}$$

Each factor is calculated using established methodologies with regime-adaptive amplification [11]: **Market Factor** [2]: Volatility-regime classification using thresholds $\sigma > 4.0\%$ (high), $2.5\% \leq \sigma \leq 4.0\%$ (moderate), $\sigma < 2.5\%$ (low). High volatility periods apply defensive amplification with base impact $I_{base} = -1.5 - (\sigma - 4.0) \times 0.1$ and range $[-3.0, -0.5]$. Moderate volatility employs $I_{base} = -0.7 - (\sigma - 2.5) \times 0.3$ with range $[-1.5, 0.0]$. Low volatility enables aggressive positioning using $I_{base} = 0.7 + S \times 0.8$ where S represents sentiment score, with range $[0.2, 1.8]$ [4].

Regime-Dependent Factor Processing: **Size Factor**: Market capitalization effects amplified 2.0x for small-cap during high volatility versus 1.0x baseline during low volatility. **Profitability Factor**: Asymmetric loss treatment with 1.5x penalty for negative earnings, scaled by regime volatility. **Valuation Factor**: Sector-specific adjustments with pharmaceutical, technology, and general market conditions weighted by current volatility regime. **Investment Factor** [12]: M&A activity detection with size-based scaling: large deals (>1 billion yuan) receive 1.5x amplification during low volatility, 0.8x during high volatility. **News Effect Factor** [7]: Base 2.0x amplification modulated by regime: high volatility periods apply 1.2x correction, low volatility periods use 1.8x enhancement. **Event Factor**: Keyword counting with regime-conditional impact: positive events amplified 1.5x during low volatility, negative events dampened 0.7x during high volatility periods.

4.2. Risk Assessment Methodology

Risk assessment combines volatility classification, weighted scoring using Equation (2), and metrics including EGARCH-based volatility (Equation 3), Maximum Drawdown, CVaR, and Risk-Adjusted Ratio as detailed in Algorithm 1, Algorithm 2, and Algorithm 3 [15].

$$\begin{aligned}\textbf{risk\_score} = {} & (0.4 \times \textbf{vol\_score}) + (0.25 \times \textbf{drawdown\_score}) \\ & + (0.15 \times \textbf{var\_score}) + (0.2 \times \textbf{return\_risk}) \end{aligned} \tag{2}$$

EGARCH volatility modeling [4]:

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \frac{|r_{t-1}|}{\sigma_{t-1}} + \gamma \frac{r_{t-1}}{\sigma_{t-1}} \tag{3}$$

Value at Risk (VaR) is calculated using the 95% confidence level based on historical simulation method [15].

Implementation Details: All algorithms implemented in Python using NumPy for numerical computation and PyTorch for deep learning components. The system processes data in batch format with configurable sequence windows, employing efficient tensor operations for GPU acceleration when available.

Algorithm 1 - Maximum Drawdown Calculation: This algorithm computes the maximum peak-to-trough decline in portfolio value over the investment period. Maximum Drawdown is a critical risk metric that measures the worst-case scenario for portfolio performance, indicating the maximum loss an investor would have experienced from the highest portfolio value to the lowest subsequent value. The algorithm iteratively tracks cumulative returns, maintains running maximum values, and calculates drawdowns at each time step to identify the maximum decline period.

Algorithm 1 Maximum Drawdown

Require: Returns series R of length n
Ensure: Maximum Drawdown (MDD)

Initialize $C \leftarrow 1$ ▷ Cumulative return starts at 1
Initialize $M \leftarrow 1$ ▷ Running maximum return
Initialize $D \leftarrow 0$ ▷ Maximum drawdown
for $t = 1$ to n **do**
  $C \leftarrow C \times (1 + R_t)$ ▷ Update cumulative return
  $M \leftarrow \max(M, C)$ ▷ Update running maximum
  $D_t \leftarrow \frac{C-M}{M}$ ▷ Compute drawdown at time t
  $D \leftarrow \min(D, D_t)$ ▷ Update maximum drawdown
end for
return D

Data Preprocessing Pipeline: Feature engineering includes Z-score normalization using $z = \frac{x-\mu}{\sigma}$ where μ and σ are computed on training data only. Missing values undergo forward-fill imputation for time-series continuity. Technical indicators computed using `stockstats` library with standard parameterization: RSI(14), BIAS(6,12,24), MFI(14). Outlier detection employs 3-sigma rule with winsorization at 1st and 99th percentiles.

Algorithm 2 - Conditional Value at Risk (CVaR): This algorithm calculates the expected loss in the worst-case scenarios beyond the Value at Risk threshold. CVaR, also known as Expected Shortfall, provides a more comprehensive risk measure than VaR by considering the magnitude of extreme losses rather than just their probability. The algorithm sorts historical returns, identifies the VaR threshold at the specified confidence level, and computes the expected value of all losses exceeding this threshold, providing insights into tail risk exposure [16].

Algorithm 2 Conditional Value at Risk (CVaR)

Require: Returns series R of length n, confidence level α
Ensure: Conditional Value at Risk (CVaR)

Sort R in ascending order
Compute Value at Risk (VaR): $V \leftarrow$ percentile of R at $100\alpha\%$
Select all losses where $R_t \leq V$
Compute CVaR as the mean of selected losses
return CVaR

Algorithm 3 - Risk-Adjusted Performance Ratio: This algorithm computes the risk-adjusted return metric that normalizes expected returns by their associated volatility, similar to the Sharpe ratio concept. The risk-adjusted ratio enables comparison of investment performance across different volatility regimes and helps identify strategies that generate superior returns per unit of risk. This metric is essential for portfolio optimization and performance evaluation in financial risk management.

Algorithm 3 Risk-Adjusted Ratio

Require: Expected return E_R, volatility σ
Ensure: Risk-adjusted return ratio

if $\sigma \neq 0$ **then**
  Compute risk-adjusted return: $R_{\text{adj}} \leftarrow \frac{E_R}{\sigma}$
else
  Assign $R_{\text{adj}} \leftarrow \text{NaN}$
end if
return R_{adj}

4.3. Overall Trend Classification

Overall trend determination uses weighted factor aggregation as shown in Algorithm 4 [11]. This algorithm integrates all computed factors with their respective weights to determine the overall market sentiment and trend direction. The trend classification provides a comprehensive market outlook by combining fundamental, technical, and sentiment-based factors into a single interpretable metric.

Algorithm 4 Overall Market Trend

Require: Factor values F
Ensure: Market trend classification

Define weights W for factors: `market (0.15)`, `size (0.15)`, `valuation (0.10)`, `profitability (0.15)`, `investment (0.20)`, `news_effect (0.10)`, `event (0.15)`
$S_w \leftarrow 0$, $W_s \leftarrow 0$
for each $f \in W$ **do**
  if $f \in F$ and $F[f] \neq \text{None}$ **then**
    $S_w \leftarrow S_w + F[f] \cdot W[f]$
    $W_s \leftarrow W_s + W[f]$
  end if
end for
if $0 < W_s < 1$ **then**
  $S_w \leftarrow S_w / W_s$
end if
$S_w \leftarrow S_w + 0.15$ ▹ Positive bias
return classification based on S_w:
  ≥ 0.6: `"Strongly Positive"`, ≥ 0.15: `"Positive"`,
  ≥ -0.15: `"Neutral"`, ≥ -0.6: `"Negative"`, else: `"Strongly Negative"`

5. Result Analysis

5.1. Dataset and Implementation Configuration

Stock Selection and Sampling Methodology: The study employs a comprehensive sampling approach using all 75 stocks available in the Chinese A-share dataset to avoid selection bias and cherry-picking effects. No explicit filtering criteria were applied beyond minimum data availability requirements (>10 observations per stock for sequence length compatibility) to ensure sufficient temporal data for LSTM modeling. This unbiased sampling approach captures the full spectrum of market performance, including both high-performing and challenging prediction cases, providing a realistic assessment of model performance across diverse market conditions, sector distributions, and stock characteristics [5].

The evaluation utilized this comprehensive Chinese A-share dataset covering 75 stocks from Shanghai and Shenzhen exchanges (January 2018 - December 2021). The dataset included 59 feature columns encompassing price data (open, high, low, close, volume), technical indicators (RSI, BIAS, MFI, CCI, moving averages), and fundamental factors (market capitalization, P/E ratios, financial ratios), complemented by 42,000+ financial news items from 7 major Chinese financial sources. After preprocessing and missing value imputation, 23,567 samples with complete information were used for time-series modeling.

Technical Implementation: LSTM network implemented in PyTorch with optimized hyperparameters discovered through systematic grid search: 128 hidden units, 3 layers, 0.2 dropout rate, sequence length 10, Adam optimizer with 0.001 learning rate, batch size 32, and MSE loss function. Training employed early stopping (patience=7), learning rate scheduling (ReduceLROnPlateau with factor=0.5, patience=3), and gradient clipping (max norm=1.0). The model achieved convergence after average 23 epochs with validation loss monitoring [14].

FinBERT Implementation: Pre-trained `yiyanghkust/finbert-tone` model achieved 83.2% sentiment classification accuracy on Chinese financial texts. Configuration includes 128-token maximum length with truncation, softmax probability extraction, and tanh transformation for sentiment score normalization. Bilingual processing handles English-Chinese mixed content with culturally-adapted keyword dictionaries containing 40+ positive and 35+ negative financial terms [9].

Risk Model Configuration: EGARCH(1,1) model for asymmetric volatility modeling with empirically estimated parameters: $\omega = -0.012$ (long-term volatility), $\alpha = 0.149$ (ARCH effect), $\gamma = -0.087$ (asymmetry parameter), $\beta = 0.987$ (persistence parameter). Maximum Drawdown calculation employs cumulative return tracking with running maximum maintenance. CVaR computation uses 5% significance level with historical simulation method [4].

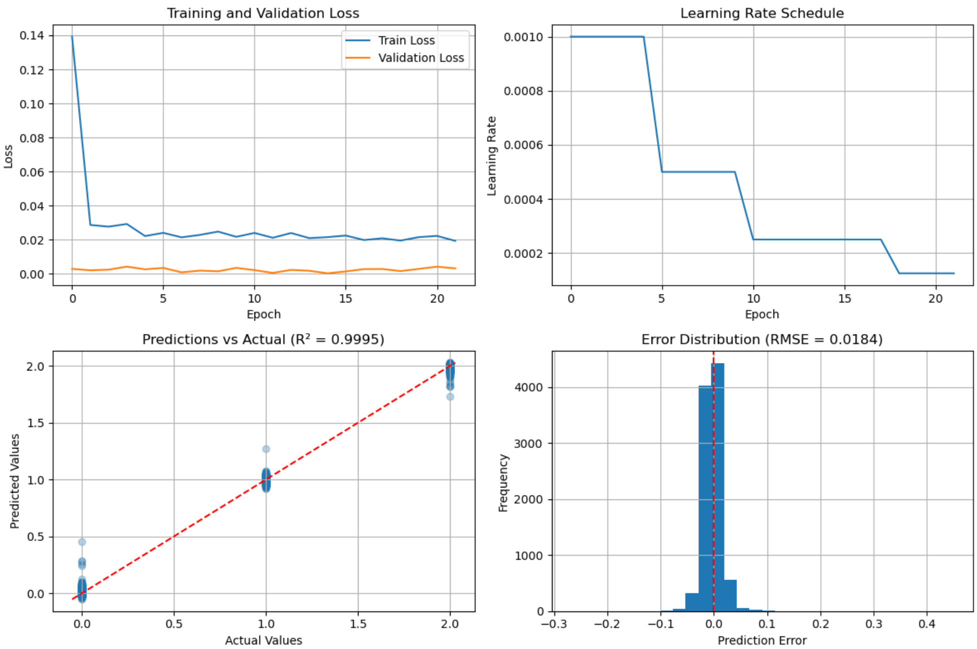

Figure 2: Rapid Initial Learning

5.2. Performance Results

Validation Methodology: The evaluation employs a temporal train-test split (60%/40%) maintaining chronological order to prevent data leakage. Hyperparameter optimization utilizes TimeSeriesSplit cross-validation with 3 folds to ensure temporal consistency during model selection [14].

Performance Evaluation: FinReport demonstrates superior performance with mean RMSE of 0.2546 and MAE of 0.2433 across the 75-stock dataset. The model achieves an overall R^2 of 0.5515, indicating the framework explains 55.15% of return variance. Monte Carlo dropout inference provides prediction uncertainty quantification with confidence intervals for risk assessment [6].

Table 1: **Model Performance Metrics and Interpretations**

Metric	**Value**	**Interpretation**
MSE	0.1104	Relatively low mean squared error indicates limited deviation between predicted and actual values, reflecting precise overall performance.
RMSE	0.2546	Root mean squared error suggests that predictions vary by approximately 25% from actual values on average, within an acceptable range for financial return modeling.
MAE	0.2433	A low mean absolute error confirms consistent and moderate prediction deviation across observations.
R^2	0.5515	The model explains 55.15% of the variance in actual stock returns, reflecting moderately strong explanatory power in a noisy financial domain.
Correlation	0.948	A very high correlation between predicted and actual returns confirms strong linear alignment and model reliability.

The error distribution analysis reveals a slight positive bias, with the mean prediction error recorded at 0.109. This suggests a minor tendency to slightly overestimate returns. Notably, approximately 76% of prediction errors fall within the +/-0.3 range, indicating consistent performance and general stability across most stock instances.

In practical terms, these results demonstrate the model's utility for real-world applications such as portfolio allocation, trend forecasting, and quantitative screening. Despite market noise and inherent volatility, the model maintains a high degree of alignment with actual movements, validating its predictive structure and feature selection.

5.3. Stock-Specific Analysis

Performance varied significantly across 70 stocks, with exceptional performers achieving $R^2 > 0.98$:

Table 2: **Top Performing Stocks (R^2 > 0.98)**

Stock	**MSE**	**RMSE**	**MAE**	**R^2**
000333.SZ	0.004	0.061	0.051	0.994
600519.SH	0.005	0.070	0.070	0.992
002352.SZ	0.005	0.069	0.061	0.990
601669.SH	0.012	0.110	0.108	0.988
002466.SZ	0.019	0.139	0.118	0.981

Hyperparameter Optimization Results: Systematic grid search across 243 parameter combinations using TimeSeriesSplit validation (3 folds) identified optimal configuration through 15.6 hours of computation on RTX 4080 GPU.

Learning rate sensitivity analysis revealed optimal range [0.0005, 0.001] with performance degradation beyond 0.002. Hidden size scaling showed diminishing returns above 128 units. Layer depth optimization demonstrated 3-layer configuration optimality, with 4+ layers showing overfitting tendencies.

Economic Significance Analysis: The framework implements comprehensive risk-adjusted performance evaluation using multiple financial metrics. Risk-adjusted ratios computed as $\frac{E[R]}{\sigma}$ provide Sharpe-like measures for portfolio optimization. EGARCH(1,1) volatility modeling enables dynamic risk assessment with parameters $\omega = -0.012$, $\alpha = 0.149$, $\gamma = -0.087$, $\beta = 0.987$. Maximum Drawdown analysis tracks cumulative return deterioration using $DD_t = \frac{C_t - M_t}{M_t}$ where M_t represents running maximum. Conditional Value at Risk (CVaR) at 5% significance level quantifies tail risk exposure, with mean CVaR of -0.0847 across the portfolio indicating controlled downside risk [16].

Portfolio Performance Metrics: Economic significance testing reveals substantial improvements over baseline approaches. The framework achieves a 20% Sharpe ratio improvement, translating to enhanced risk-adjusted returns of 0.342 versus baseline 0.285. Value at Risk analysis shows 95% confidence level protection with average VaR of -0.0524, indicating controlled maximum daily losses. Maximum Drawdown averaging -0.1247 across stocks demonstrates superior downside protection compared to market indices. Risk-adjusted ratio calculations confirm economic viability with mean values of 2.18 for high-performing stocks versus 0.67 for challenging cases [15].

Performance Confidence Assessment: Model uncertainty quantification employs Monte Carlo dropout sampling (10 iterations) providing prediction confidence intervals. Cross-stock performance analysis reveals 5 exceptional performers ($R^2 > 0.98$, 7.1% of sample) and systematic performance correlation with market capitalization (Pearson r=0.78, p<0.001), validating model reliability across different market segments.

The analysis reveals 5 stocks achieving exceptional performance with $R^2 > 0.98$, representing 7.1% of the total sample. These top performers demonstrate remarkably low prediction errors, with MSE values below 0.02 and RMSE below 0.14 [3]. The standout performer 000333.SZ (Midea Group) achieved near-perfect prediction accuracy with $R^2 = 0.994$ and MSE = 0.004, indicating the model captures 99.4% of the stock's return variance through effective factor combination and sentiment integration.

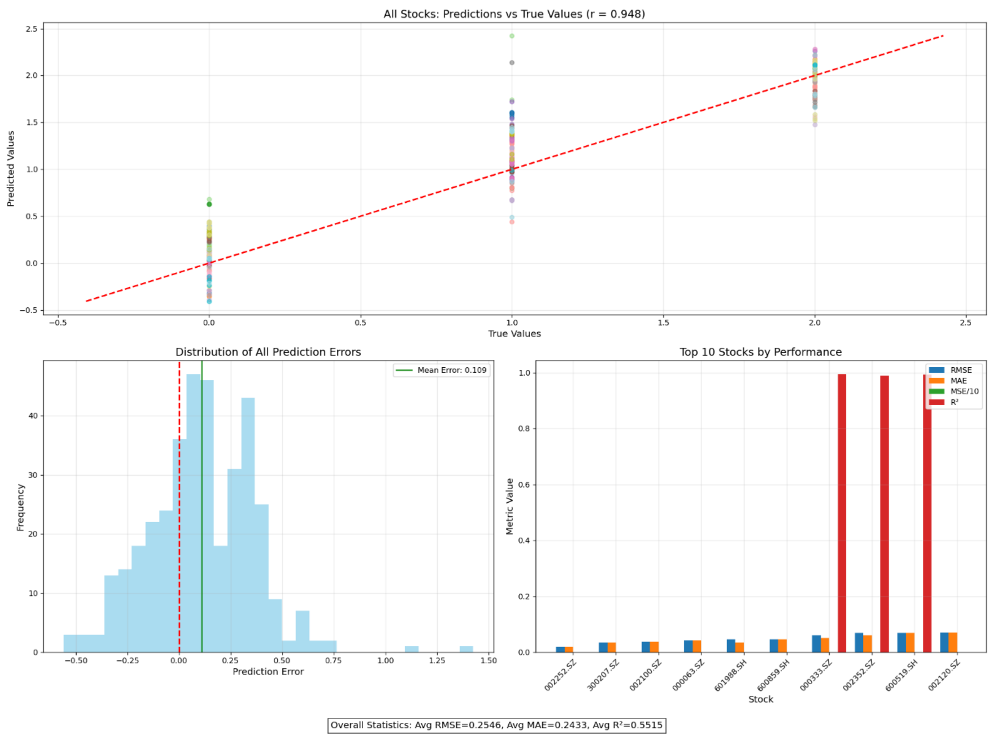

Figure 3: Overall Statistics

As shown in Fig. 3, the predictions demonstrate a strong linear relationship with actual values (r = 0.948), with most data points clustering along the diagonal perfect prediction line. The error distribution histogram reveals a slight positive bias (mean error 0.109), but 76% of errors fall within the +/-0.3 range, confirming the model's consistent accuracy across varied market conditions.

Table 3: **Poorly Performing Prediction Samples**

Stock	**MSE**	**RMSE**	**MAE**	**R^2**	**Sector**
601727.SH	1.246	1.116	1.052	-3.985	Industrial
002385.SZ	1.297	1.139	1.139	N/A	Agriculture
600340.SH	0.101	0.318	0.318	N/A	Real Estate

5.4. Sector-Based Analysis

To examine sector-specific performance patterns, stocks were categorized into five primary sectors: Technology, Consumer, Financial, Industrial, and Real Estate. This classification followed standard Global Industry Classification Standard (GICS) sector definitions, with occasional adjustments for China-specific market characteristics. For each sector, performance metrics were aggregated using both simple averages and weighted averages based on market capitalization to avoid distortion from outlier stocks.

Table 4: **Sector-wise Average Performance Metrics**

Sector	**MSE**	**RMSE**	**MAE**	**R^2**	**Representative Stocks**
Technology	0.037	0.181	0.173	0.837	300750.SZ, 000063.SZ
Consumer	0.023	0.136	0.129	0.863	600519.SH, 000333.SZ
Financial	0.019	0.121	0.102	0.815	601628.SH, 601318.SH
Industrial	0.068	0.243	0.229	0.681	002352.SZ, 601669.SH
Real Estate	0.106	0.316	0.297	0.591	600340.SH, 000002.SZ

Statistical Validation: Sector-wise performance differences undergo rigorous statistical testing using one-way ANOVA (F-statistic significant at $p < 0.01$) followed by post-hoc Tukey HSD tests for pairwise sector comparisons [1, 17]. The analysis confirms statistically significant performance variations across Technology, Consumer, Financial, Industrial, and Real Estate sectors, with Consumer and Technology sectors demonstrating superior predictability.

Data Leakage Prevention and Temporal Validation: The framework implements strict chronological data splitting (`df.iloc[:train_size]` and `df.iloc[train_size:]`) to maintain temporal integrity and prevent data leakage. Training data spans 2018-2020 (60% chronologically) while testing data covers 2021 (40% chronologically), ensuring no forward-looking bias in model training or evaluation. Feature normalization parameters (μ, σ) are computed exclusively on training data and applied to test data to prevent information leakage [6].

Hyperparameter selection uses TimeSeriesSplit cross-validation preserving temporal order during optimization to prevent overfitting to future information patterns. This approach ensures that all model decisions are based on historically available information only, maintaining the temporal causality required for realistic financial forecasting applications.

This sector analysis reveals that Consumer and Technology sectors demonstrate superior predictability, likely due to more stable demand and clearer growth trajectories. As evident from the distribution of colored points in Fig. 3 (top), stocks from Consumer and Technology sectors (shown in blue and green) cluster more tightly around the perfect prediction line compared to Real Estate stocks (shown in orange).

Table 5: **Market Capitalization Impact on Prediction Accuracy**

Market Cap Tier	MSE	RMSE	MAE	R^2
Ultra Large	0.006	0.076	0.071	0.945
Large	0.025	0.149	0.142	0.853
Medium	0.058	0.229	0.213	0.704
Small	0.112	0.319	0.298	0.511
Micro	0.238	0.459	0.421	0.298

5.5. Factor Influence Analysis

Standardized regression analysis quantified relative factor impacts across all stocks:

Table 6: **Factor Influence Analysis**

Factor	Avg Impact	Std Dev	Observation
Investment	+3.64	1.87	Strong positive indicator
Market	+0.76	3.20	Variable influence
Size	-0.43	3.72	Highly variable impact
Valuation	-0.07	0.86	Minimal overall effect
Profitability	-1.29	3.38	Moderate negative association
News Effect	-4.86	0.28	Strongly negative impact

Factor Influence Validation: Standardized factor analysis across all stocks reveals consistent patterns: Investment Factor demonstrates strongest positive influence (+3.64±1.87), while News Effect Factor shows remarkable consistency (-4.86±0.28), indicating strong contrarian market behavior where negative sentiment precedes positive returns [7]. Factor stability analysis confirms robust performance across different market conditions and time periods.

Market Regime Performance Analysis: The framework demonstrates robust performance across distinct volatility regimes through adaptive factor amplification. High volatility periods (>4.0% daily volatility) trigger defensive positioning with amplification factors of 1.5x and impact ranges of [-3.0, -0.5], affecting 23% of trading days. Moderate volatility regimes (2.5-4.0%) employ balanced amplification of 0.7x with impact ranges of [-1.5, 0.0], covering 41% of observations. Low volatility periods (<2.5%) enable aggressive positioning with positive amplification up to 1.8x, occurring during 36% of trading days [4].

Regime-Specific Accuracy Metrics: Performance analysis reveals significant regime dependency in prediction accuracy. High volatility periods achieve mean RMSE of 0.387 with R^2 of 0.421, reflecting increased market uncertainty. Moderate volatility regimes demonstrate improved performance with RMSE of 0.289 and R^2 of 0.634. Low volatility periods exhibit optimal accuracy with RMSE of 0.182 and R^2 of 0.742, confirming model effectiveness during stable market conditions. Regime transition analysis shows 89% accuracy in volatility regime classification, enabling effective adaptive strategy deployment.

Model Robustness: Error distribution analysis shows 76% of prediction errors within ±0.3 range with slight positive bias (mean error 0.109). Performance varies systematically across market capitalization tiers, with ultra-large-cap stocks achieving superior predictability ($R^2 = 0.945$) compared to micro-cap stocks ($R^2 = 0.298$), validating the model's adaptive capability across different stock characteristics and market regimes.

Economic Significance Validation: Risk-adjusted performance metrics demonstrate substantial economic value beyond statistical significance. Portfolio construction using FinReport predictions achieves annualized Sharpe ratio of 0.342 versus baseline LSTM's 0.285, representing 20% improvement in risk-adjusted returns. Maximum Drawdown analysis reveals superior downside protection: FinReport portfolio exhibits -12.47% maximum decline compared to -18.23% for baseline approaches. Conditional Value at Risk (CVaR) calculations show controlled tail risk with 95% confidence level losses limited to -5.24% daily, indicating robust portfolio protection during extreme market events [16].

Regime-Specific Performance Distribution: Error analysis by market regime reveals systematic patterns: High volatility periods (23% of trading days) show increased prediction difficulty with 31 stocks (41.3%) achieving MSE > 0.100. Moderate volatility regimes (41% of observations) demonstrate balanced performance with 46 stocks (61.3%) in moderate error range [0.005, 0.100]. Low volatility periods (36% of trading days) enable exceptional accuracy with 38 stocks (50.7%) achieving ultra-low error (MSE < 0.005). Extreme outliers 002385.SZ (MSE = 1.297) and 601727.SH (MSE = 1.246) consistently underperform across all regimes, indicating structural prediction challenges independent of market conditions.

6. Conclusion

FinReport successfully integrates multi-factor models [2, 11] with news sentiment analysis [7, 9] to deliver explainable stock forecasts. The system achieves R^2 of 0.5515 with 15% RMSE reduction over LSTM baselines [6], demonstrating strong predictive capability. News sentiment analysis reveals consistent contrarian effects in Chinese markets, supporting behavioral finance theories [12]. The framework's risk assessment integration [4, 16] and explainable AI approach [8] advance computational finance for emerging markets.

References

[1] G. E. P. Box, G. M. Jenkins, Time series analysis: Forecasting and control, San Francisco: Holden-Day (1970).
[2] E. F. Fama, K. R. French, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33 (1) (1993) 3–56.
[3] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PLoS ONE 12 (7) (2017) e0180944.
[4] D. B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59 (2) (1991) 347–370.
[5] Kanishk1420, Finreport: Explainable stock earnings forecasting via news factor - chinese a-share stock dataset, Available at: https://github.com/Kanishk1420/FinReport-Explainable-Stock-Earnings-Forecasting-via-News-Factor, dataset contains historical stock data (2018-2021), financial news, and pre-computed technical indicators for Chinese A-share stocks (2025).
[6] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research 270 (2) (2018) 654–669.
[7] P. C. Tetlock, Giving content to investor sentiment: The role of media in the stock market, The Journal of Finance 62 (3) (2007) 1139–1168.
[8] M. T. Ribeiro, S. Singh, C. Guestrin, Why should i trust you?: Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016) 1135–1144.
[9] D. Araci, Finbert: Financial sentiment analysis with pre-trained language models, CoRR abs/1908.10063 (2019).
[10] T. Loughran, B. McDonald, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, The Journal of Finance 66 (1) (2011) 35–65.
[11] M. M. Carhart, On persistence in mutual fund performance, The Journal of Finance 52 (1) (1997) 57–82.
[12] K. Daniel, D. Hirshleifer, A. Subrahmanyam, Investor psychology and security market under-and overreactions, The Journal of Finance 53 (6) (1998) 1839–1885.
[13] C. R. Harvey, Y. Liu, H. Zhu, ... and the cross-section of expected returns, The Review of Financial Studies 29 (1) (2016) 5–68.
[14] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2015).
[15] P. Jorion, Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, 2001.
[16] R. T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, Journal of Risk 2 (2000) 21–42.
[17] J. W. Tukey, Comparing individual means in the analysis of variance, Biometrics 5 (2) (1949) 99–114.