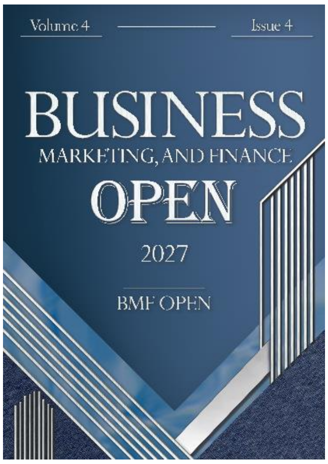
Evaluating the Performance of a Hybrid Model in Improving Key Risk Management Metrics in Financial Markets

Morteza Vahdati [1], Mohammad Reza Karimi [2,*] and Abdolrahim Rahimi [3]

[1] Department of Media Management, ST.C., Islamic Azad University, Tehran, Iran;
[2] Department of Media Management, ST.C., Islamic Azad University, Tehran, Iran;
[3] Department of Cultural Management and Planning, ST.C., Islamic Azad University, Tehran, Iran;

* Correspondence: Mr_karimi@iau.ac.ir



Citation: Vahdati, M., Karimi, M. R., & Rahimi, A. R. (2027). Evaluating the Performance of a Hybrid Model in Improving Key Risk Management Metrics in Financial Markets. *Business, Marketing, and Finance Open,* 4(4), 1-26.

Received: 18 March 2026
Revised: 02 August 2026
Accepted: 10 August 2026
Initial Publication: 15 August 2026
Final Publication: 01 July 2027

Abstract: This study aimed to design and evaluate a hybrid financial–behavioral model integrating technical market indicators, volatility features, and social-media sentiment to improve key risk-management metrics in Bitcoin and Ethereum markets. This quantitative, ex post facto, developmental–applied study used hourly and daily financial data for Bitcoin and Ethereum from January 2019 to December 2025, together with English-language posts from Twitter/X containing relevant cryptocurrency keywords and hashtags. Financial variables included prices, trading volume, returns, volatility measures, and more than 50 technical indicators. Textual data were processed using natural language processing procedures, and sentiment scores were extracted through a finance-specific BERT model. The proposed FinBERT–LSTM model combined financial, technical, and behavioral features. Its performance was compared with naïve persistence, ARIMA, GARCH, random forest, financial LSTM, and sentiment-only models using chronological train–validation–test partitions and walk-forward validation. Predictive accuracy, directional classification, value-at-risk calibration, maximum drawdown, expected shortfall, and risk-adjusted performance were evaluated. The hybrid model significantly outperformed all benchmark models, achieving the lowest mean absolute error, root mean squared error, and mean absolute percentage error, as well as the highest directional accuracy, F1 score, and area under the curve. Diebold–Mariano tests confirmed significantly lower forecasting errors than the financial LSTM and sentiment-only models ($p < 0.001$). The hybrid strategy also produced lower annualized volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall, while yielding higher Sharpe, Sortino, and Calmar ratios. Kupiec and Christoffersen tests indicated adequate value-at-risk coverage and independence at the 95% and 99% confidence levels. Ablation analyses further showed that removing sentiment, technical, engagement, or volatility features significantly weakened predictive and risk-management performance. Integrating technical, temporal, and behavioral information within a hybrid deep-learning architecture improves both forecasting accuracy and the management of downside and tail risk in cryptocurrency markets.

Keywords: Hybrid model, financial risk management, FinBERT, long short-term memory, investor sentiment, cryptocurrency, value at risk, expected shortfall

1. Introduction

Risk management constitutes one of the central functions of modern financial systems because investment decisions are invariably made under conditions of uncertainty, volatility, incomplete information, and exposure to potentially severe losses. Financial institutions, portfolio managers, individual investors, regulators, and corporate decision-makers must continuously estimate the probability and magnitude of adverse outcomes while simultaneously attempting to preserve return-generating opportunities. In this context, risk cannot be understood merely as the possibility of loss; rather, it is a multidimensional construct involving market volatility, downside deviation, liquidity constraints, drawdown exposure, tail events, behavioral reactions, model uncertainty, and the changing relationships among financial assets. The quality of financial decision-making therefore depends on the ability of analytical systems to identify emerging risk patterns, distinguish transient fluctuations from structural changes, and translate forecasts into practical exposure-management strategies [1, 2]. Traditional risk-management frameworks have provided essential measures such as variance, beta, value at risk, expected shortfall, and maximum drawdown, yet their practical performance may deteriorate when markets exhibit nonlinear dynamics, structural breaks, heavy-tailed return distributions, volatility clustering, and rapid shifts in investor expectations.

Conventional financial theory has historically assumed that market participants process information rationally and that prices incorporate available information with sufficient speed and efficiency. Under this view, market behavior can be represented through probabilistic models derived primarily from historical prices, returns, and macroeconomic variables. Nevertheless, the assumptions of rationality, stable preferences, and normally distributed returns are frequently violated in actual markets. Behavioral finance has demonstrated that investor decisions are influenced by overconfidence, loss aversion, herding, anchoring, representativeness, limited attention, and emotionally driven responses to uncertainty. These behavioral forces can produce persistent mispricing, speculative bubbles, panic selling, abnormal trading volume, and contagious reactions across market segments [3, 4]. Investor risk perception is also shaped by financial literacy, subjective knowledge, prior experience, and sentiment, meaning that individuals with access to similar information may form substantially different expectations and assume different levels of risk [5, 6]. Consequently, models based exclusively on historical numerical information may overlook a significant component of the mechanisms through which market risk is generated and transmitted.

The influence of sentiment becomes particularly visible during financial crises and periods of heightened uncertainty. Emotional reactions can magnify fluctuations by causing investors to interpret information asymmetrically, follow collective behavior, or adjust their positions faster than underlying economic conditions justify. Evidence from capital markets indicates that emotionally driven decisions can contribute to crisis formation and intensification, while changes in collective sentiment may affect financing decisions, trading activity, and asset valuation [7, 8]. These findings suggest that market risk is not exclusively embedded in price histories but is also reflected in the language, attention, engagement, and expectations of market participants. Developments in behavioral economics and big-data analytics have further made it possible to study decision biases at a scale that was not feasible through conventional surveys or laboratory studies [9]. Digital traces generated through social-media platforms, online forums, financial news channels, search behavior, and visual media may therefore serve as timely indicators of optimism, fear, uncertainty, disagreement, and speculative intensity.

The emergence of digital financial markets has increased both the importance and complexity of such behavioral information. Cryptocurrency markets, in particular, operate continuously, are globally accessible, and are highly responsive to technological announcements, regulatory decisions, macroeconomic developments, exchange-related events, influential users, and social-media narratives. Unlike traditional exchanges with fixed trading hours, cryptocurrency markets react to information around the clock, creating an environment in which behavioral signals can rapidly translate into price and volume changes. Bitcoin and Ethereum are especially appropriate for examining these dynamics because they combine high liquidity and extensive public attention with substantial volatility, frequent speculative episodes, and pronounced sensitivity to market narratives. Reviews of cryptocurrency trading and price-forecasting research have shown that successful strategies increasingly depend on the integration of technical, statistical, machine-learning, and behavioral information rather than reliance on any single analytical tradition [10]. Research on cryptocurrency investor behavior likewise indicates that sentiment, herding, overconfidence, attention, and perceived technological value play important roles in investment decisions and market fluctuations [4].

Traditional financial forecasting methods, including autoregressive models, moving-average models, volatility models, and technical-analysis systems, remain valuable because they provide interpretable structures for modeling temporal dependence and market variability. Technical analysis attempts to infer future behavior from recurring patterns in prices, returns, trading volume, momentum, trends, and volatility. Indicators such as moving averages, the relative strength index, moving-average convergence divergence, Bollinger Bands, stochastic oscillators, and average true range continue to be widely used to identify changes in market direction and risk exposure [11]. However, the performance of traditional methods is constrained when relationships are nonlinear, variables interact dynamically, or the predictive structure changes across market regimes. Comparative assessments of conventional time-series techniques and machine-learning approaches indicate that no single method consistently dominates across all assets, frequencies, and market conditions, and that model selection must consider data structure, interpretability, computational requirements, and sensitivity to structural change [12].

Artificial intelligence and machine learning have expanded the analytical capacity available for financial forecasting and risk assessment. Algorithms can process large numbers of variables, detect nonlinear interactions, update predictions as new data become available, and identify patterns that may be difficult to specify through conventional econometric models. Machine learning has been applied to algorithmic trading, credit assessment, fraud detection, portfolio construction, volatility prediction, and market-risk management, although its effectiveness depends on data quality, feature engineering, validation procedures, and protection against overfitting [13]. Artificial-intelligence applications in finance have also evolved from narrow forecasting tasks toward broader decision-support systems that combine prediction, risk estimation, scenario evaluation, and strategic recommendation [14]. The integration of artificial intelligence with big-data analytics can improve the speed and breadth of financial decision-making by enabling simultaneous analysis of structured market observations and unstructured textual information [15].

Earlier studies have demonstrated the capacity of intelligent models to address specific financial prediction problems. Artificial intelligence has been used to predict accounting behavior such as income smoothing, illustrating its applicability to financial phenomena involving complex and potentially hidden patterns [16]. Hybrid genetic algorithm and support-vector machine structures have been employed to improve stock-price prediction through optimized feature or parameter selection [17]. Neuro-fuzzy systems combining market indicators have likewise been proposed to capture uncertainty and nonlinear relationships in stock prices [18]. Other hybrid intelligent models have been designed to identify critical turning points in price series, including potential peaks and troughs that are highly relevant to risk control and timing decisions [19]. One-day-ahead forecasting research has also shown that combining multiple modeling components can yield more accurate predictions than isolated methods, particularly when short-horizon market movements depend on both recent dynamics and complex indicator interactions [20].

More recent studies have extended hybridization through ensemble learning, optimization algorithms, deep neural networks, and advanced temporal architectures. Portfolio rebalancing models that integrate ensemble machine learning with genetic algorithms indicate that predictive outputs can be converted into practical allocation decisions, thereby linking statistical accuracy with portfolio performance [21]. Metaheuristic algorithms combined with multivariate adaptive regression splines have been used to improve market-price forecasting by optimizing model structure and capturing nonlinear relationships [22]. Hybrid stock-analysis systems have also been proposed to integrate complementary analytical components and generate more comprehensive market forecasts [23]. Reviews and conceptual examinations of hybrid machine-learning models emphasize that combining algorithms can compensate for the limitations of individual techniques by integrating different learning mechanisms, reducing prediction variance, and improving robustness under heterogeneous market conditions [24].

Deep-learning architectures have become especially prominent because financial observations are sequential and often contain long- and short-term temporal dependencies. Recurrent neural networks and long short-term memory models are designed to preserve relevant information across time while mitigating the vanishing-gradient limitations of conventional recurrent structures. Their ability to model complex sequences has made them suitable for forecasting prices, returns, volatility, and market direction. Empirical research using long short-term memory architectures has reported improved stock-market predictive performance, particularly when models are trained on multivariate sequences and validated under realistic temporal partitions [25]. Artificial neural machine-learning models have also been applied across BRICS stock markets, demonstrating the broader relevance of nonlinear learning systems for economies with diverse institutional structures and volatility patterns [26]. Models that integrate bidirectional gated recurrent units, long-memory considerations, and optimized parameter-search procedures have further shown that explicitly modeling long-range dependence can improve stock-index prediction [27].

Despite these advances, financial forecasting should not be equated with risk-management effectiveness. A model may generate relatively accurate price predictions while still producing undesirable portfolio outcomes if its errors are concentrated during market crashes or if its signals lead to excessive turnover, unstable exposure, or large tail losses. Risk management requires an evaluation framework that extends beyond mean prediction error to include realized volatility, downside deviation, maximum drawdown, value at risk, expected shortfall, and risk-adjusted performance ratios. Research on hybrid volatility prediction has shown that combining models can strengthen financial-market risk analysis by improving the representation of nonlinear and time-varying volatility structures [28]. Hybrid fuzzy-logic and machine-learning systems have similarly been proposed for financial risk assessment because fuzzy reasoning can represent ambiguity while machine learning detects complex empirical patterns [29]. These developments indicate that an effective model should not only predict market behavior but also support decisions concerning position size, capital preservation, exposure limits, and responses to changing levels of uncertainty.

The practical use of algorithmic systems is also influenced by investor acceptance, trust, perceived usefulness, and perceived risk. Retail investors may recognize the potential benefits of algorithmic trading while remaining concerned about complexity, transparency, reliability, and loss of control. Research on intentions to use algorithmic trading platforms has shown that adoption is shaped through interconnected psychological and technological mechanisms rather than technical performance alone [30]. This issue is relevant because risk-management models must ultimately generate outputs that decision-makers can understand and apply. Artificial-intelligence tools may also help reduce emotional bias by offering structured analysis that counterbalances impulsive reactions and selective interpretation of information. Evidence concerning GPT-4-assisted financial analysis suggests that generative artificial intelligence may reduce certain emotional biases in investment decisions by encouraging more systematic evaluation of evidence [31]. Nevertheless, automated advice does not eliminate model risk, and decision-makers must remain attentive to data limitations, unstable relationships, algorithmic opacity, and the possibility that models may reproduce biases embedded in their training information.

The growing availability of unstructured data provides an opportunity to strengthen financial models through natural language processing and multimodal analysis. Advances in artificial intelligence, semantic technologies, and machine-readable knowledge structures enable systems to extract meaning from large volumes of text and connect linguistic information with quantitative observations [32]. Transformer-based language models are particularly valuable because they identify contextual meaning more effectively than simple word-frequency or dictionary-based sentiment methods. Finance-specific transformer models can distinguish between positive, negative, and neutral meanings in statements containing technical terminology, negation, conditional expectations, and domain-specific expressions. Behavioral information can then be aggregated into indicators such as sentiment intensity, disagreement, discussion volume, engagement-weighted sentiment, and social fear and greed. Recent research has expanded sentiment measurement beyond conventional text by extracting investor sentiment from images through few-shot learning, illustrating the increasing breadth of data sources available for behavioral financial analysis [33].

Sentiment should also be examined as a connected and systemic phenomenon rather than solely as an isolated predictor for a single asset. Changes in sentiment may spread across assets, sectors, and investor communities, creating channels through which fear or optimism becomes correlated. Portfolio-optimization research based on sentiment connectedness suggests that measures of behavioral interdependence can be used to reduce exposure to highly connected sources of emotional contagion and improve portfolio construction [34]. Behavioral analytics has similarly been integrated into credit-risk and financial-stability modeling, demonstrating that attitudes, responses, and behavioral patterns may complement conventional financial indicators in the identification of emerging risk [35]. These perspectives are important for cryptocurrency markets because Bitcoin and Ethereum are connected through investor attention, portfolio allocation, common macroeconomic drivers, and shared narratives concerning the digital-asset ecosystem.

At the same time, the usefulness of advanced models varies across institutional and market contexts. Risk may arise from the structure of asset listings, regulatory environments, geographical exposure, or differences in information availability. Research on overseas listings, for example, demonstrates that market structure and listing status can create distinctive forms of risk that cannot be inferred from price history alone [36]. Financial risk-management frameworks also differ across countries and levels of market development, requiring models to account for institutional variation and differences in market depth, regulatory capacity, and investor composition [2]. Even when the present study focuses on global cryptocurrency markets, these broader findings reinforce the need to evaluate model performance across different regimes rather than assuming that a single average result represents all conditions.

An additional challenge concerns the distinction between predictive improvement and genuine economic value. Financial models are often assessed using statistical indicators such as mean absolute error or root mean squared error, but these measures do not directly reveal whether a model protects capital or improves risk-adjusted returns. A small reduction in average forecasting error may have limited practical importance if the model fails during rare but severe losses. Conversely, a model with modest average accuracy may be highly valuable if it correctly identifies periods of extreme downside risk. Therefore, model evaluation should include out-of-sample testing, walk-forward validation, transaction costs, regime-specific analysis, tail-risk backtesting, and comparisons with both passive and active benchmarks. Comprehensive reviews of cryptocurrency forecasting and trading strategies emphasize that methodological rigor, realistic backtesting, and attention to profitability after costs are essential for distinguishing practically useful models from statistically attractive but fragile results [10].

The literature collectively suggests that the integration of technical, financial, and behavioral information offers a promising direction, yet several gaps remain. Many studies focus on price or return prediction without assessing whether the resulting signals improve multiple risk-management metrics. Other investigations use only structured numerical variables or treat sentiment as a separate explanatory factor rather than embedding it in a unified temporal architecture. Some hybrid models combine algorithms but do not integrate fundamentally different data modalities. In addition, a substantial proportion of the available evidence is based on stock markets, relatively short observation periods, single market regimes, or evaluation criteria that do not sufficiently capture extreme losses. These limitations reduce the extent to which existing findings can be generalized to continuously traded and sentiment-sensitive cryptocurrency markets.

The need for an integrated approach is further supported by the complementary strengths of technical and behavioral variables. Price, volume, trend, momentum, and volatility indicators summarize realized market behavior, while social-media sentiment and engagement may reveal expectations that have not yet been fully reflected in prices. A hybrid model can use sequential neural networks to learn temporal patterns from financial data and transformer-based language models to quantify behavioral information. The resulting representation may improve the detection of volatility changes, downside risk, and market turning points. Hybridization may also reduce dependence on any single data source: when technical indicators become less informative during a sudden narrative-driven event, sentiment variables may provide early signals, whereas price and volatility indicators may correct exaggerated or noisy behavioral reactions. This logic is consistent with the broader movement toward artificial-intelligence-supported financial decision-making, in which diverse data sources are integrated to improve prediction, interpretation, and strategic response [14, 15].

The proposed analytical framework also responds to earlier evidence that intelligent and hybrid systems can improve forecasting through complementary learning mechanisms. Studies using genetic algorithms, support-vector machines, neuro-fuzzy systems, ensemble methods, recurrent networks, and hybrid volatility models have each demonstrated partial advantages under particular conditions [17, 18, 21, 28]. However, the practical contribution of a hybrid model should be judged through its ability to reduce realized volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall while improving the Sharpe, Sortino, and Calmar ratios. Such an evaluation connects computational modeling with the central purpose of financial risk management: obtaining acceptable returns without exposing capital to disproportionate or poorly anticipated losses.

Accordingly, this study aims to design, implement, and evaluate a hybrid financial–behavioral model integrating technical market indicators, volatility features, and social-media sentiment to determine its effectiveness in improving key risk-management metrics in the Bitcoin and Ethereum markets.

2. Methodology

This study was conducted within a quantitative research paradigm using an ex post facto analytical design. The research was intended to identify causal, predictive, and correlational relationships among financial, technical, and behavioral variables in the complex and highly dynamic environment of financial markets. In terms of purpose, the study was developmental and applied because it sought to design, implement, and validate an innovative hybrid computational model capable of addressing a practical risk management problem. The research process was based on computational modeling and involved the systematic integration of historical market data, technical indicators, social-media sentiment, deep-learning algorithms, and quantitative risk measures. Because the study relied exclusively on secondary financial and textual data, it did not involve human participants in the conventional experimental or survey-based sense. The units of analysis consisted of time-indexed observations related to cryptocurrency prices, trading activity, technical market conditions, and social-media behavior.

The study involved two distinct but temporally aligned statistical populations: financial market data and textual social-media data. The population of financial data theoretically comprised the price and trading records of assets listed in global financial markets. Using purposive sampling, Bitcoin and Ethereum were selected as the focal assets. These cryptocurrencies were chosen because they consistently represented highly traded and widely discussed digital assets, exhibited substantial price volatility, and demonstrated pronounced sensitivity to market expectations, investor sentiment, macroeconomic news, regulatory developments, and social-media activity. These characteristics made them appropriate empirical settings for evaluating whether a hybrid model could improve the prediction and management of financial risk under bullish, bearish, and relatively stable market conditions. The observation period extended from January 1, 2019, to December 31, 2025. This seven-year period was selected to capture heterogeneous market regimes, including rapid expansion, severe corrections, liquidity shocks, speculative cycles, periods of elevated uncertainty, and relatively neutral market phases.

The financial dataset consisted of daily and hourly observations for Bitcoin and Ethereum. Each observation included the opening price, closing price, highest price, lowest price, and trading volume. Where available, supplementary variables such as the number of trades, quote-asset volume, taker-buy volume, bid–ask information, and market capitalization were also retained to enrich the representation of market liquidity and trading pressure. Hourly data were used to capture short-term fluctuations and rapid sentiment transmission, whereas daily data were used to evaluate medium-term risk patterns and portfolio-level implications. The use of multiple temporal resolutions enabled the robustness of the hybrid model to be examined under different forecasting horizons and reduced the likelihood that the findings would be limited to a single market frequency.

The population of textual data comprised publicly available English-language content generated by users on Twitter, subsequently renamed X, during the same observation period. Purposive sampling was used to identify posts directly related to Bitcoin and Ethereum. The textual sample included English-language posts containing predefined keywords, cashtags, and hashtags, including “Bitcoin,” “BTC,” “$BTC,” “#Bitcoin,” “Ethereum,” “ETH,” “$ETH,” and “#Ethereum.” Posts were retained when their publication timestamps fell within the financial data period and could be matched to the hourly or daily market intervals used in the analysis. Duplicate posts, advertisements, irrelevant content, and posts lacking sufficient textual information were excluded during preprocessing. Retweets or reposts were either removed to prevent duplication or retained as engagement signals, depending on their analytical function. When reposts were retained, the original textual content was counted once, while repost frequency was incorporated as an indicator of message diffusion and public attention.

The sampling unit for the financial component was each valid hourly or daily trading interval, while the sampling unit for the textual component was each eligible social-media post. The two datasets were integrated according to Coordinated Universal Time to prevent inconsistencies caused by exchange location, user location, daylight-saving changes, or differences between data providers. Textual observations were subsequently aggregated into the same temporal intervals as the market data. This temporal synchronization ensured that sentiment and discussion indicators represented information available within each corresponding market period.

To limit look-ahead bias, textual information published after the end of a given forecasting interval was not used to predict outcomes within that interval.

Financial and textual data were collected automatically through programming scripts developed in Python. Automated data acquisition was selected because the study required the extraction, management, and synchronization of a large volume of high-frequency observations across an extended period. The Python environment was also used to implement preprocessing procedures, feature engineering, model training, validation, and statistical evaluation, thereby ensuring consistency across the different stages of the research. Data acquisition scripts incorporated error-handling procedures, timestamp verification, duplicate detection, rate-limit management, and periodic data-integrity checks. Raw data were stored separately from processed data to preserve the original observations and permit the reproducibility of all preprocessing and analytical procedures.

Historical financial data were collected through official cryptocurrency-exchange application programming interfaces and standardized exchange-access libraries such as CCXT. Binance was treated as the principal data source because of its high liquidity and extensive historical trading records for Bitcoin and Ethereum. Where necessary, data from an additional high-volume exchange or recognized financial data provider were used to check the consistency of timestamps, prices, and trading volumes. The scripts retrieved open, high, low, close, and volume data for the selected hourly and daily intervals. The raw observations were stored in a time-series database or structured data repository in which each record was indexed by asset, trading interval, exchange, and timestamp. Data-quality procedures were applied to identify missing intervals, duplicated records, impossible price values, non-positive trading volumes, and abrupt discontinuities resulting from technical interruptions rather than genuine market movements.

Textual data were collected through authorized academic, commercial, or archival access to the Twitter/X application programming interface. Python-based libraries such as Tweepy were used to formulate search queries, manage authentication, retrieve eligible posts, and store the returned metadata. For every collected post, the dataset retained the text, exact publication time, language, number of likes, number of replies, number of reposts, and, where access conditions permitted, non-sensitive public profile indicators relevant to user influence. Such indicators included follower count, account age, verification status, and posting frequency. These variables were not used to identify individual users; instead, they were transformed into aggregated measures of message visibility, engagement, and potential influence. Usernames and direct personal identifiers were removed or pseudonymized before analysis, and the study focused on aggregate behavioral patterns rather than individual-level profiling.

Financial feature construction was carried out using Python libraries such as pandas, NumPy, pandas-ta, and scikit-learn. More than 50 technical indicators were initially generated to represent trend, momentum, volatility, liquidity, and volume conditions. Trend-related variables included simple moving averages, exponential moving averages, moving-average convergence divergence, directional movement indicators, and the average directional index. Momentum indicators included the relative strength index, stochastic oscillator, rate of change, commodity channel index, and Williams percentage range. Volatility measures included Bollinger Bands, average true range, rolling standard deviation, historical volatility, and high–low price ranges. Volume-related features included on-balance volume, volume-weighted average price, money flow indicators, accumulation–distribution measures, and changes in trading volume. Additional market variables, such as lagged returns, logarithmic returns, downside returns, rolling skewness, rolling kurtosis, drawdown, and rolling correlations, were calculated to provide a more comprehensive representation of changing risk conditions.

Textual preprocessing and natural language processing were implemented using Python libraries for text manipulation and transformer-based modeling. The collected posts were converted to a standardized encoding and cleaned by removing hyperlinks, unnecessary punctuation, HTML elements, repeated characters, irrelevant symbols, and user mentions. Cashtags and hashtags were preserved when they carried meaningful information about the focal assets, while the hashtag symbol itself could be removed before tokenization. English stop words were removed when appropriate, although stop-word removal was not applied mechanically when transformer-based models required the preservation of the original sentence structure. Tokenization, lemmatization, stemming, spelling normalization, and lowercasing were applied according to the requirements of each analytical model. Emojis and common cryptocurrency expressions were either translated into textual sentiment tokens or retained as separate features because they could communicate optimism, fear, sarcasm, urgency, or speculative enthusiasm.

Sentiment extraction was performed using a transformer-based language model derived from Bidirectional Encoder Representations from Transformers and fine-tuned on financial texts. A finance-specific model was selected because conventional sentiment dictionaries may incorrectly interpret technical financial expressions, informal market terminology, and context-dependent statements. For each post, the model generated probabilities or scores corresponding to positive, negative, and neutral sentiment. These outputs were transformed into continuous sentiment measures and aggregated within hourly and daily intervals. Aggregated behavioral indicators included average sentiment, median sentiment, sentiment dispersion, the proportion of positive and negative posts, discussion volume, engagement-weighted sentiment, user-influence-weighted sentiment, and volume-weighted sentiment. A social fear-and-greed indicator was also constructed by combining the balance of positive and negative sentiment, posting intensity, engagement growth, and abnormal changes in discussion volume. These behavioral features were aligned with the technical and financial variables to form the final multimodal dataset.

Data analysis was performed in two interconnected phases. The first phase involved exploratory analysis, preprocessing, feature engineering, and dataset integration. The second phase involved hybrid-model development, comparative evaluation, and assessment of the model's effects on key risk management metrics. Exploratory analysis was initially conducted to examine the distributions, temporal patterns, stationarity characteristics, volatility clustering, outliers, missing observations, and cross-correlations of the financial and behavioral variables. Descriptive statistics, rolling statistics, return distributions, autocorrelation functions, and correlation matrices were examined to identify irregularities and potentially redundant features. The financial prices were transformed into logarithmic returns when necessary because raw prices were non-stationary and could produce misleading statistical relationships. Stationarity was assessed using appropriate unit-root procedures, and differencing or return transformation was applied when variables exhibited persistent stochastic trends.

Missing financial observations were investigated to determine whether they reflected temporary data-access interruptions, exchange outages, or genuine periods without trading. Since cryptocurrency markets operate continuously, short gaps were imputed using time-based interpolation, forward interpolation, or local estimates when the missingness was limited and non-systematic. Extended gaps were not imputed mechanically and were either reconstructed from a verified alternative data source or excluded from the analysis. Missing textual intervals were distinguished from technical data loss because the absence of eligible posts could itself indicate low public attention. Accordingly, intervals with no relevant posts were assigned a discussion volume of zero, while sentiment-related variables were treated using predetermined neutral or missing-value rules. Extreme financial observations were examined through robust statistical procedures and market-event verification. Genuine market crashes, rallies, and liquidity shocks were retained because they represented critical risk-management conditions, whereas obvious data-entry or extraction errors were corrected or removed.

Financial features were normalized to improve the numerical stability and convergence of machine-learning and deep-learning models. Min–max normalization was applied to variables used as inputs to neural networks, with scaling parameters estimated exclusively from the training data and subsequently applied to the validation and test sets. This procedure prevented information from future observations from influencing the preprocessing of earlier data. Variables with highly skewed distributions, such as trading volume and discussion intensity, were transformed using logarithmic or robust scaling procedures when required. The initial technical and behavioral feature set was screened for multicollinearity, excessive missingness, low variance, and instability. Feature-selection methods, including correlation filtering, recursive feature elimination, regularization, permutation importance, and model-based importance measures, were applied to reduce dimensionality while preserving economically meaningful information.

The hybrid model was developed by integrating two complementary analytical streams. The financial stream processed historical prices, returns, trading volumes, and technical indicators, whereas the behavioral stream processed social-media sentiment, discussion volume, engagement, and user-influence measures. Sequential deep-learning architectures, such as long short-term memory or gated recurrent unit networks, were used to model temporal dependencies in the financial data. The textual component relied on the outputs of the fine-tuned financial BERT model and the aggregated behavioral features. The two feature representations were combined through feature-level or decision-level fusion and passed through fully connected layers to generate the final predictions. Depending on the specific analytical task, the model produced forecasts of future returns, volatility, downside risk, extreme-loss probability, or trading-position signals. Hyperparameters, including the number of recurrent layers, hidden units, learning rate, batch size, dropout rate, sequence length, and optimization algorithm, were determined using the validation dataset and a structured search procedure.

The data were divided chronologically rather than randomly to preserve their time-series structure. The earliest observations were allocated to the training set, the subsequent observations to the validation set, and the most recent observations to the out-of-sample test set. Chronological separation was necessary to prevent future information from leaking into model development. In addition to the primary holdout analysis, rolling-window or walk-forward validation was used to examine the model's stability across changing market regimes. Under this procedure, the model was repeatedly trained using information available up to a given date and evaluated on the immediately subsequent period. This approach reproduced the conditions under which the model would have been used in practice and enabled its performance to be assessed during bullish, bearish, high-volatility, and low-volatility periods.

The hybrid model was compared with several benchmark approaches to determine whether combining technical and behavioral information produced incremental value. The benchmarks included a historical or naïve forecasting model, conventional econometric models, machine-learning models based only on technical indicators, deep-learning models based only on financial time series, and sentiment-based models without technical variables. An ablation analysis was also performed by successively removing sentiment variables, engagement measures, technical indicators, and specific model components. This procedure made it possible to determine whether improvements were attributable to the integrated architecture rather than to model complexity alone.

Predictive accuracy was evaluated using measures appropriate to the dependent variable. Continuous forecasts were assessed through mean absolute error, root mean squared error, mean absolute percentage error, and the coefficient of determination. Directional forecasting performance was assessed using directional accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve when the outcome was expressed as an upward or downward movement. Volatility and tail-risk forecasts were evaluated using loss functions appropriate to variance, value at risk, and expected shortfall. Statistical comparisons of forecasting errors were conducted to determine whether differences between the hybrid model and benchmark models were systematic rather than attributable to random variation.

The main evaluation focused on the model's ability to improve key risk management metrics. Predictions generated by each model were converted into comparable trading or asset-allocation signals under identical transaction-cost and exposure constraints. The resulting strategies were assessed using realized volatility, downside deviation, maximum drawdown, value at risk, conditional value at risk or expected shortfall, Sharpe ratio, Sortino ratio, Calmar ratio, cumulative return, and risk-adjusted return. Value-at-risk estimates were examined at conventional confidence levels, while conditional value at risk was calculated to quantify the average loss beyond the value-at-risk threshold. Backtesting procedures were used to assess whether the observed frequency of violations was consistent with the expected tail probability. Maximum drawdown was calculated from the cumulative portfolio-value series to determine the largest peak-to-trough loss produced by each model.

The effectiveness of the hybrid model was defined as a statistically and economically meaningful reduction in volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall, accompanied by an improvement in the Sharpe, Sortino, and Calmar ratios relative to the benchmark models. Transaction costs, slippage, and trading frequency were incorporated into the performance calculations to avoid overstating the practical value of the proposed model. Robustness analyses were conducted separately for Bitcoin and Ethereum, for hourly and daily horizons, and for different market regimes. Sensitivity analyses examined alternative sequence lengths, sentiment-aggregation windows, risk thresholds, transaction-cost assumptions, and training periods. The analytical procedures were implemented in Python using appropriate statistical, machine-learning, deep-learning, and natural language processing libraries. Random seeds, model configurations, preprocessing parameters, and evaluation rules were documented to support computational reproducibility.

3. Findings and Results

Because the present research was based on secondary financial and social-media data rather than individual human participants, conventional demographic characteristics such as age, sex, education, marital status, and occupational status were not applicable. Accordingly, the initial description of the research sample focused on the distributional and structural characteristics of the analyzed observations. The final dataset covered the period from January 1, 2019, to December 31, 2025, and included Bitcoin and Ethereum observations at both hourly and daily frequencies. After the removal of duplicate records, correction of timestamp inconsistencies, management of short missing intervals, and exclusion of invalid observations, 122,736 hourly financial observations were retained, consisting of 61,368 observations for Bitcoin and 61,368 observations for Ethereum. In addition, 5,114 daily financial observations were available, including 2,557 observations for each cryptocurrency. The textual dataset initially contained 18,746,392 English-language posts retrieved through keywords, hashtags, and cashtags related to Bitcoin and Ethereum. Following the removal of duplicates, promotional spam, irrelevant content, posts with insufficient linguistic information, and records with incomplete timestamps, 14,283,617 posts remained for analysis. Of these, 8,264,951 posts, representing 57.86% of the textual sample, were related primarily to Bitcoin, while 6,018,666 posts, representing 42.14%, were related primarily to Ethereum.

The temporal distribution of the data indicated that all major cryptocurrency market conditions were represented. Approximately 31.42% of the observations occurred during bullish market periods, 27.83% occurred during bearish periods, 24.61% occurred during relatively neutral or range-bound conditions, and 16.14% occurred during high-volatility transitional periods that could not be classified exclusively as bullish or bearish. The mean number of eligible social-media posts per hour was 116.38 for Bitcoin and 84.74 for Ethereum, although the distributions were positively skewed because discussion volume increased sharply during market crashes, major rallies, regulatory announcements, exchange failures, and other high-attention events. The average proportion of positive, neutral, and negative posts across the entire study period was 39.72%, 34.85%, and 25.43%, respectively. However, the proportion of negative posts exceeded 40% during severe bearish intervals, whereas positive sentiment frequently exceeded 50% during sustained market rallies. These descriptive patterns supported the inclusion of sentiment intensity, sentiment dispersion, discussion volume, and engagement-weighted sentiment as potential predictors of changing market risk.

Table 1 presents the descriptive statistics for the principal variables entered into the analytical models. To facilitate interpretation across assets with substantially different price levels, the primary market variables were expressed as logarithmic returns, standardized trading-volume changes, realized volatility, and normalized behavioral indicators. The sentiment score ranged from −1 to +1, with higher values representing more positive market sentiment. Discussion volume and engagement-weighted sentiment were standardized before model estimation. Maximum drawdown, value at risk, and expected shortfall were expressed as percentages of portfolio value.

Table 1. Descriptive Statistics of the Main Study Variables

Variable	Asset	Number of Observations	Mean	Standard Deviation	Minimum	Maximum	Skewness	Kurtosis
Hourly logarithmic return (%)	Bitcoin	61,368	0.006	0.842	−16.374	14.826	−0.18	18.64
Hourly logarithmic return (%)	Ethereum	61,368	0.008	1.037	−19.682	17.945	−0.27	16.92
Daily logarithmic return (%)	Bitcoin	2,557	0.091	3.617	−46.473	18.746	−1.12	20.37
Daily logarithmic return (%)	Ethereum	2,557	0.104	4.691	−55.071	25.183	−1.34	18.95
Realized volatility (%)	Bitcoin	61,368	1.742	1.286	0.112	14.936	2.48	10.83
Realized volatility (%)	Ethereum	61,368	2.164	1.573	0.148	18.472	2.29	9.94
Standardized trading-volume change	Bitcoin	61,368	0.000	1.000	−3.216	11.472	3.07	18.46
Standardized trading-volume change	Ethereum	61,368	0.000	1.000	−3.084	10.936	2.91	16.73
Aggregate sentiment score	Bitcoin	61,368	0.146	0.284	−0.893	0.927	−0.31	0.84
Aggregate sentiment score	Ethereum	61,368	0.121	0.301	−0.918	0.941	−0.22	0.71
Sentiment dispersion	Bitcoin	61,368	0.363	0.147	0.041	0.912	0.82	1.16
Sentiment dispersion	Ethereum	61,368	0.381	0.155	0.038	0.926	0.74	0.93
Standardized discussion volume	Bitcoin	61,368	0.000	1.000	−1.384	9.762	3.41	19.82
Standardized discussion volume	Ethereum	61,368	0.000	1.000	−1.297	8.945	3.19	17.66

Engagement-weighted sentiment	Bitcoin	61,368	0.173	0.338	−0.951	0.968	−0.41	1.09
Engagement-weighted sentiment	Ethereum	61,368	0.149	0.351	−0.963	0.974	−0.34	0.97
Social fear-and-greed index	Bitcoin	61,368	52.84	18.73	4.16	96.42	−0.09	−0.61
Social fear-and-greed index	Ethereum	61,368	51.37	19.46	3.82	97.15	−0.04	−0.68
One-day 95% historical value at risk (%)	Bitcoin	2,557	5.814	2.372	1.246	17.638	1.29	2.11
One-day 95% historical value at risk (%)	Ethereum	2,557	7.241	3.016	1.534	22.476	1.41	2.68
One-day 95% expected shortfall (%)	Bitcoin	2,557	8.147	3.689	1.786	28.942	1.62	3.96
One-day 95% expected shortfall (%)	Ethereum	2,557	10.386	4.716	2.164	35.871	1.74	4.38

As shown in Table 1, both cryptocurrencies generated small positive mean returns over the complete study period, but the relatively large standard deviations and extreme minimum and maximum values demonstrated that average return alone provided an incomplete description of market behavior. Ethereum exhibited greater variability than Bitcoin at both hourly and daily frequencies. The standard deviation of hourly Ethereum returns was 1.037%, compared with 0.842% for Bitcoin, while the standard deviation of daily Ethereum returns was 4.691%, compared with 3.617% for Bitcoin. The return distributions were characterized by substantial excess kurtosis, indicating a greater frequency of extreme observations than would be expected under a normal distribution. The negative skewness of the daily return distributions also indicated that severe downside movements were more pronounced than comparable positive movements, particularly for Ethereum. These characteristics supported the use of risk measures sensitive to tail losses rather than reliance on variance-based measures alone.

The realized-volatility indicators further confirmed that Ethereum was the riskier of the two assets. Mean hourly realized volatility was 2.164% for Ethereum and 1.742% for Bitcoin. Both volatility distributions were strongly positively skewed, demonstrating that relatively stable intervals were periodically interrupted by abrupt volatility spikes. Trading-volume changes and discussion-volume measures also exhibited high positive skewness and kurtosis, indicating that market participation and social attention were concentrated around exceptional events. Mean aggregate sentiment was positive for both assets, although Bitcoin had a slightly higher average sentiment score than Ethereum. Sentiment dispersion was nevertheless substantial, indicating disagreement among social-media users even when average sentiment appeared moderately positive. The mean social fear-and-greed index remained close to the neutral midpoint, but its broad observed range showed that the study period included episodes of extreme fear and extreme optimism. Finally, the average value at risk and expected shortfall were higher for Ethereum than Bitcoin, demonstrating that Ethereum was exposed to larger potential losses during ordinary and extreme adverse market conditions.

Table 2 compares the out-of-sample predictive performance of the proposed hybrid model with the selected benchmark models. The comparison included a naïve persistence model, an autoregressive integrated moving-average model, a generalized autoregressive conditional heteroskedasticity model, a random forest model based on technical variables, a long short-term memory model based exclusively on financial variables, a FinBERT sentiment model based exclusively on behavioral features, and the proposed hybrid FinBERT–LSTM model combining financial, technical, and behavioral inputs. The results were averaged across the Bitcoin and Ethereum out-of-sample test periods. Lower values of mean absolute error, root mean squared error, and mean absolute percentage error indicated superior continuous forecasting performance, whereas higher directional accuracy, F1 score, and area under the receiver operating characteristic curve indicated better classification of upward and downward market movements.

Table 2. Out-of-Sample Predictive Performance of the Hybrid and Benchmark Models

Model	Mean Absolute Error	Root Mean Squared Error	Mean Absolute Percentage Error (%)	Directional Accuracy (%)	Precision	Recall	F1 Score	Area Under the Curve
Naïve persistence model	0.0318	0.0476	12.84	50.73	0.512	0.501	0.506	0.508
ARIMA model	0.0297	0.0441	11.92	53.46	0.538	0.529	0.533	0.552
GARCH model	0.0289	0.0427	11.37	54.82	0.551	0.543	0.547	0.568
Random forest with technical features	0.0254	0.0382	9.76	59.41	0.603	0.586	0.594	0.632
Financial LSTM model	0.0231	0.0356	8.81	62.73	0.631	0.619	0.625	0.671
FinBERT behavioral model	0.0259	0.0391	9.93	58.86	0.597	0.578	0.587	0.624
Proposed hybrid FinBERT–LSTM model	0.0184	0.0297	6.94	69.58	0.704	0.687	0.695	0.756

The results presented in Table 2 demonstrate that the proposed hybrid model achieved the strongest predictive performance across all reported criteria. Its mean absolute error was 0.0184, representing a reduction of 42.14% relative to the naïve model, 38.05% relative to the ARIMA model, 36.33% relative to the GARCH model, and 20.35% relative to the financial LSTM model. Its root mean squared error was 0.0297, which was substantially lower than the corresponding errors of all benchmark models. The reduction in root mean squared error was particularly important because this measure assigns greater weight to large forecasting errors, which are especially consequential in financial risk management. The mean absolute percentage error of the hybrid model was 6.94%, compared with 8.81% for the financial LSTM model and 9.93% for the sentiment-only model. This finding indicated that the integration of behavioral information with market and technical information generated greater predictive accuracy than either data stream used independently.

The proposed hybrid model also correctly predicted the direction of market movement in 69.58% of the out-of-sample observations. This rate was 6.85 percentage points higher than that of the financial LSTM model and 10.72 percentage points higher than that of the FinBERT behavioral model. The model achieved a precision of 0.704, recall of 0.687, and F1 score of 0.695, demonstrating that its performance did not result from disproportionately favoring one directional class. The area under the receiver operating characteristic curve was 0.756, indicating an acceptable-to-strong ability to discriminate between positive and negative future market movements. By contrast, the naïve, ARIMA, and GARCH models produced areas under the curve closer to random classification. The Diebold–Mariano comparisons of out-of-sample forecasting errors showed that the error reduction produced by the hybrid model was statistically significant relative to each benchmark model. The difference was significant compared with the financial LSTM model, with a test statistic of 4.86 and a probability value below 0.001, and compared with the FinBERT behavioral model, with a test statistic of 6.14 and a probability value below 0.001. Thus, the improvement obtained by the hybrid model reflected a systematic predictive advantage rather than random variation in the test sample.

Table 3 presents the portfolio-level risk and performance measures generated when the predictions of each model were converted into comparable trading and exposure-management signals. All strategies were evaluated under identical initial capital, maximum position size, transaction-cost, and slippage assumptions. The buy-and-hold strategy was included as a passive reference condition. Lower realized volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall indicated improved risk control, whereas higher Sharpe, Sortino, and Calmar ratios indicated stronger risk-adjusted performance.

Table 3. Comparison of Portfolio Risk and Risk-Adjusted Performance Measures

Strategy or Model	Annualized Return (%)	Annualized Volatility (%)	Downside Deviation (%)	Maximum Drawdown (%)	95% Value at Risk (%)	95% Expected Shortfall (%)	Sharpe Ratio	Sortino Ratio	Calmar Ratio
Buy-and-hold strategy	31.84	67.26	48.91	−72.48	6.93	10.41	0.43	0.59	0.44
ARIMA-based strategy	24.73	53.18	38.74	−49.36	5.62	8.31	0.42	0.58	0.50
GARCH-based strategy	26.58	48.62	34.81	−43.75	5.08	7.42	0.50	0.68	0.61
Random-forest strategy	32.46	43.79	30.16	−35.82	4.43	6.31	0.68	0.91	0.91
Financial LSTM strategy	36.92	39.84	26.47	−29.64	3.86	5.48	0.85	1.18	1.25
FinBERT behavioral strategy	30.17	45.13	31.92	−38.71	4.61	6.72	0.61	0.82	0.78
Proposed hybrid FinBERT–LSTM strategy	42.76	31.58	19.86	−19.47	2.94	4.07	1.26	1.78	2.20

As indicated in Table 3, the hybrid model produced the most favorable combination of return generation and risk reduction. The annualized return of the hybrid strategy was 42.76%, exceeding the annualized return of the buy-and-hold strategy by 10.92 percentage points and that of the financial LSTM strategy by 5.84 percentage points. More importantly, this increase in return was accompanied by a substantial reduction in risk. Annualized volatility declined from 67.26% under the passive buy-and-hold strategy to 31.58% under the hybrid strategy, corresponding to a relative reduction of 53.05%. Compared with the financial LSTM strategy, the hybrid model reduced annualized volatility by 20.73%. The hybrid strategy also generated the lowest downside deviation, at 19.86%, showing that the model was particularly effective in limiting harmful negative fluctuations rather than merely reducing total variability.

The maximum drawdown of the hybrid strategy was −19.47%, compared with −72.48% for the buy-and-hold strategy, −35.82% for the random-forest strategy, and −29.64% for the financial LSTM strategy. Thus, the hybrid model reduced the magnitude of the largest peak-to-trough portfolio decline by 73.14% relative to the passive strategy and by 34.31% relative to the strongest single-stream deep-learning benchmark. The one-period 95% value at risk was reduced to 2.94%, while the corresponding expected shortfall was reduced to 4.07%. These results indicated that the hybrid model improved both ordinary tail-risk control and the management of losses occurring beyond the value-at-risk threshold. The expected-shortfall reduction was especially important because expected shortfall measures the mean loss under the most adverse market outcomes and is therefore more informative than value at risk when distributions are characterized by heavy tails.

The superiority of the hybrid strategy was also reflected in all risk-adjusted performance ratios. Its Sharpe ratio was 1.26, compared with 0.43 for buy-and-hold and 0.85 for the financial LSTM strategy. Its Sortino ratio reached 1.78, demonstrating that the return achieved by the model was high relative to harmful downside volatility. The Calmar ratio was 2.20, substantially exceeding the values reported for all benchmark strategies. The difference between the Sharpe ratio and the more pronounced improvement in the Sortino ratio indicated that the model's primary benefit arose from its ability to reduce downside fluctuations and severe adverse movements. Overall, the portfolio analysis showed that the hybrid model did not merely improve statistical forecasting accuracy; it translated predictive information into economically meaningful improvements in capital preservation, tail-risk reduction, and risk-adjusted return.

Table 4 reports the results of value-at-risk backtesting for the 95% and 99% confidence levels. The expected violation rate was 5% for the 95% value-at-risk model and 1% for the 99% value-at-risk model. A well-calibrated risk model should generate an observed violation rate close to the expected rate, should not systematically underestimate losses, and should not produce clusters of consecutive violations. The Kupiec unconditional-coverage test evaluated whether the total number of violations was statistically consistent with the expected probability, while the Christoffersen independence test evaluated whether violations occurred independently over time. Probability values above 0.05 indicated that the null hypothesis of adequate coverage or independence could not be rejected.

Table 4. Value-at-Risk Backtesting Results for the Competing Models

Model	Confidence Level	Expected Violations (%)	Observed Violations (%)	Kupiec Test Statistic	Kupiec p Value	Christoffersen Test Statistic	Christoffersen p Value	Backtesting Decision
Historical simulation	95%	5.00	7.36	14.82	<0.001	9.47	0.002	Rejected
GARCH	95%	5.00	6.18	4.96	0.026	5.21	0.022	Rejected
Random forest	95%	5.00	5.72	2.14	0.143	4.38	0.036	Partially rejected
Financial LSTM	95%	5.00	5.39	0.73	0.393	2.81	0.094	Accepted
Proposed hybrid model	95%	5.00	5.08	0.03	0.862	0.91	0.340	Accepted
Historical simulation	99%	1.00	2.14	17.26	<0.001	8.72	0.003	Rejected
GARCH	99%	1.00	1.67	6.43	0.011	4.97	0.026	Rejected
Random forest	99%	1.00	1.43	2.94	0.086	4.19	0.041	Partially rejected
Financial LSTM	99%	1.00	1.21	0.79	0.374	2.36	0.124	Accepted
Proposed hybrid model	99%	1.00	1.06	0.07	0.791	0.84	0.359	Accepted

The backtesting results in Table 4 show that the proposed hybrid model generated the most accurately calibrated value-at-risk forecasts. At the 95% confidence level, the hybrid model produced an observed violation rate of 5.08%, which was almost identical to the expected rate of 5%. The Kupiec test statistic was 0.03 with a probability value of 0.862, indicating that the hypothesis of correct unconditional coverage could not be rejected. The Christoffersen test statistic was 0.91 with a probability value of 0.340, showing that the violations were not significantly clustered over time. The financial LSTM model also passed both tests at the 95% confidence level, although its observed violation rate of 5.39% was farther from the expected level than that of the hybrid model. In contrast, historical simulation and GARCH significantly underestimated risk because their observed violation rates exceeded the expected rate and both coverage and independence tests were rejected.

At the more demanding 99% confidence level, the hybrid model again produced the most accurate result. Its observed violation rate was 1.06%, compared with the expected rate of 1%. Neither the Kupiec test nor the Christoffersen test was statistically significant, confirming that the model adequately estimated both the frequency and temporal distribution of extreme losses. The financial LSTM model also passed the backtesting requirements but produced a higher observed violation rate of 1.21%. The random-forest model passed the unconditional-coverage test at both confidence levels but failed the independence test, indicating that its total number of violations was approximately acceptable while the violations tended to occur in clusters. Such clustering is undesirable in practical risk management because it suggests that the model may perform adequately under ordinary conditions but systematically fail during periods of persistent market stress. The historical-simulation and GARCH models failed at both confidence levels, particularly during abrupt regime changes and severe market declines. Therefore, the evidence supported the conclusion that the integration of market, technical, and sentiment information improved not only point prediction but also the calibration and temporal reliability of tail-risk forecasts.

Table 5 presents the ablation and robustness analysis of the proposed model. In the ablation component, selected feature groups were removed from the full architecture to determine their incremental contribution. In the robustness component, the full hybrid model was evaluated separately for Bitcoin, Ethereum, hourly forecasts, daily forecasts, bullish periods, bearish periods, neutral periods, and high-volatility conditions. Predictive performance was represented by directional accuracy and root mean squared error, while economic performance was represented by the Sharpe ratio and maximum drawdown.

Table 5. Ablation and Robustness Analysis of the Proposed Hybrid Model

Model Configuration or Subsample	Directional Accuracy (%)	Root Mean Squared Error	Sharpe Ratio	Maximum Drawdown (%)
Full hybrid model	69.58	0.0297	1.26	−19.47
Hybrid model without sentiment features	64.02	0.0348	0.96	−27.31
Hybrid model without discussion-volume and engagement features	66.11	0.0326	1.08	−24.86
Hybrid model without technical indicators	61.74	0.0373	0.84	−31.92
Hybrid model without user-influence weighting	67.43	0.0315	1.13	−23.74
Hybrid model without volatility-related features	62.83	0.0361	0.88	−30.47
Bitcoin subsample	71.26	0.0264	1.34	−17.82
Ethereum subsample	67.91	0.0331	1.17	−21.36
Hourly forecasting horizon	70.43	0.0288	1.29	−18.96
Daily forecasting horizon	67.86	0.0319	1.20	−20.41
Bullish market periods	72.84	0.0256	1.48	−13.72
Bearish market periods	66.37	0.0354	1.09	−18.86
Neutral market periods	68.41	0.0289	1.14	−15.93
High-volatility transitional periods	63.92	0.0416	0.93	−24.71

The ablation results reported in Table 5 demonstrate that every major component of the proposed architecture contributed to its performance. Removing sentiment features reduced directional accuracy from 69.58% to 64.02%, increased root mean squared error from 0.0297 to 0.0348, reduced the Sharpe ratio from 1.26 to 0.96, and increased maximum drawdown from −19.47% to −27.31%. These changes indicated that sentiment information contributed materially to both predictive performance and downside-risk control. Removing discussion-volume and engagement variables also weakened performance, although the decline was smaller than that observed after complete removal of sentiment. This pattern suggested that the emotional polarity of market discussion carried the largest behavioral contribution, while attention and message diffusion provided additional information concerning the strength and likely persistence of sentiment effects.

The removal of technical indicators produced one of the most substantial performance declines. Directional accuracy fell to 61.74%, root mean squared error increased to 0.0373, the Sharpe ratio declined to 0.84, and maximum drawdown deteriorated to −31.92%. Similarly, excluding volatility-related features reduced directional accuracy to 62.83% and increased maximum drawdown to −30.47%. These results demonstrated that behavioral information could not replace market-structure and volatility information. Rather, the highest performance emerged from the complementary interaction of technical, volatility, and sentiment variables. The model without user-influence weighting also underperformed the complete model, indicating that posts generated by highly visible or highly engaged users carried different predictive value from posts receiving limited attention.

The subsample analyses showed that the hybrid model remained effective across assets, frequencies, and market regimes. Performance was stronger for Bitcoin than Ethereum, with directional accuracy of 71.26% and 67.91%, respectively. This difference was consistent with Ethereum's greater volatility, more pronounced tail risk, and greater sensitivity to asset-specific technological developments. The model also performed slightly better at the hourly horizon than the daily horizon, suggesting that social-media signals were transmitted into cryptocurrency prices relatively quickly and that some behavioral information lost predictive strength when aggregated over longer intervals.

Across market regimes, the model achieved its highest directional accuracy and Sharpe ratio during bullish conditions. Nevertheless, it remained effective during bearish periods, producing a directional accuracy of 66.37%, a Sharpe ratio of 1.09, and a maximum drawdown of −18.86%. The ability to retain positive risk-adjusted performance during bearish conditions was particularly important because risk-management models should be evaluated primarily on their ability to control losses during adverse markets. Performance was weakest during high-volatility transitional periods, when directional accuracy declined to 63.92% and root mean squared error rose to 0.0416. Even in these periods, however, the hybrid model maintained a Sharpe ratio of 0.93 and limited maximum drawdown to −24.71%, which remained superior to the corresponding full-period performance of several benchmark strategies. Consequently, the robustness analysis confirmed that the model's overall advantage was not confined to one cryptocurrency, one forecasting horizon, or one favorable market regime.

Taken together, the findings provided consistent support for the central proposition that a hybrid model integrating financial time-series information, technical indicators, volatility measures, social-media sentiment, discussion intensity, engagement, and user-influence information could improve key risk management criteria in cryptocurrency markets. The proposed model outperformed traditional econometric, machine-learning, financial deep-learning, and sentiment-only alternatives in terms of forecasting error, directional classification, value-at-risk calibration, maximum-drawdown control, expected-shortfall reduction, and risk-adjusted return. The improvement was evident not only in aggregate test-period statistics but also across individual assets, alternative forecasting horizons, and different market regimes.

The largest practical benefits of the hybrid model were observed in downside-risk management. Relative to the passive strategy, the model reduced annualized volatility by 53.05%, reduced maximum drawdown by 73.14%, reduced 95% value at risk by 57.58%, and reduced 95% expected shortfall by 60.90%. At the same time, it increased the annualized return from 31.84% to 42.76% and raised the Sharpe ratio from 0.43 to 1.26. The model also produced value-at-risk violations that were close to their expected frequencies and showed no statistically significant clustering. These results indicated that the proposed method was not excessively conservative, because it did not achieve risk reduction merely by eliminating market exposure. Instead, it improved the timing and magnitude of exposure by identifying periods in which technical conditions and collective market behavior jointly indicated elevated or declining risk.

The ablation findings further demonstrated that the performance of the model resulted from multimodal integration rather than from the isolated dominance of one feature group. Technical indicators and volatility measures captured observable price dynamics and market structure, while sentiment and engagement variables captured behavioral information that had not yet been fully incorporated into prices. Their integration therefore provided a more comprehensive representation of market risk. On this basis, the main research hypothesis was supported: the proposed hybrid model significantly and economically improved the principal indicators of financial risk management compared with the selected benchmark approaches.

4. Discussion and Conclusion

The present study evaluated the performance of a hybrid financial–behavioral model that integrated historical market data, technical indicators, volatility features, and social-media sentiment for improving key risk-management metrics in the Bitcoin and Ethereum markets. The findings showed that the proposed FinBERT–LSTM architecture consistently outperformed the naïve persistence, ARIMA, GARCH, random-forest, financial LSTM, and sentiment-only models. The hybrid model achieved the lowest mean absolute error, root mean squared error, and mean absolute percentage error, while producing the highest directional accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve. The model also generated superior economic outcomes by increasing annualized return and reducing annualized volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall. Its value-at-risk forecasts passed both unconditional-coverage and independence tests at the 95% and 99% confidence levels, and the ablation analysis demonstrated that technical indicators, volatility variables, sentiment scores, discussion volume, engagement, and user-influence weighting each made an incremental contribution to performance. Overall, the findings support the view that financial risk in cryptocurrency markets is best represented through an integrated framework capable of simultaneously processing quantitative market behavior and collective investor sentiment.

The superior predictive performance of the hybrid model can first be explained by the nonlinear and temporally dependent nature of cryptocurrency markets. Bitcoin and Ethereum returns exhibited high variance, negative skewness, excess kurtosis, and extreme observations, showing that their statistical behavior could not be adequately represented through assumptions of linearity, normality, and stable parameter relationships. Conventional econometric models such as ARIMA and GARCH captured some temporal dependence and volatility clustering, but their forecasting performance remained weaker than that of the machine-learning and deep-learning alternatives. This result is consistent with comparative evidence indicating that traditional time-series models are useful for parsimonious representation and interpretation but often become less effective when financial relationships are nonlinear, high-dimensional, and subject to structural change [12]. The findings also align with research showing that machine-learning methods can detect complex interactions in market data and improve algorithmic trading and risk-management decisions when they are trained and validated appropriately [13].

The stronger performance of the recurrent component is attributable to the capacity of long short-term memory networks to learn both short- and long-range dependencies in sequential data. Cryptocurrency prices are influenced not only by the immediately preceding observation but also by accumulated momentum, volatility persistence, delayed market responses, and recurring patterns across longer intervals. The financial LSTM therefore outperformed the naïve, ARIMA, GARCH, and random-forest models, suggesting that sequential representation was important for extracting information from the temporal structure of the market. This interpretation is supported by research demonstrating that LSTM-based architectures can improve stock-market prediction by retaining relevant historical information while mitigating the vanishing-gradient limitations of conventional recurrent neural networks [25]. It is also consistent with evidence from neural forecasting across emerging economies, where artificial neural models were able to capture complex and nonlinear stock-market behavior more effectively than simpler approaches [26]. Similarly, the findings correspond with the superior performance reported for architectures combining long-memory perspectives with optimized bidirectional recurrent networks [27].

Nevertheless, the financial LSTM remained inferior to the complete hybrid model, indicating that historical prices, volumes, and technical variables did not contain all the information relevant to future market risk. Incorporating sentiment and behavioral features reduced forecasting errors and improved directional discrimination, which suggests that social-media data contained incremental information that had not yet been fully incorporated into prices. This result is theoretically consistent with behavioral finance, according to which market participants do not always process information rationally and may be influenced by herding, loss aversion, overconfidence, limited attention, and emotionally driven reactions [3, 4]. In highly sentiment-sensitive markets, collective optimism may sustain buying pressure, whereas fear, uncertainty, and negative narratives may accelerate selling and increase volatility. The finding that sentiment improved the model therefore reflects the role of behavioral expectations as a mechanism through which market risk is formed and transmitted.

The result also agrees with evidence that investor sentiment affects financing decisions, trading behavior, and market instability. Emotional decisions have been associated with the intensification of financial crises, particularly when pessimistic reactions become contagious and produce coordinated selling behavior [7]. Investor sentiment has likewise been linked to financial decision-making and financing choices, demonstrating that subjective market expectations affect behavior beyond the information contained in accounting or price variables [8]. The present findings extend this line of research by showing that sentiment is not only associated with financial behavior but can be incorporated into a computational model to improve the prediction and management of downside risk. This extension is important because it moves behavioral information from an explanatory variable in retrospective analysis to an operational component of a forward-looking risk-management system.

The higher accuracy of the finance-specific transformer model relative to conventional sentiment representations may be explained by its contextual processing of financial language. Cryptocurrency posts frequently contain informal terminology, abbreviations, sarcasm, conditional statements, emojis, and expressions whose meanings differ from their general-language interpretations. A domain-specific transformer can distinguish between positive, negative, and neutral meanings more effectively than simple dictionary-based procedures. This interpretation is consistent with developments in artificial intelligence and semantic processing, which have increased the capacity of computational systems to extract contextual meaning from unstructured information [32]. It also accords with the broader development of multimodal sentiment measurement, including the extraction of investor sentiment from visual information through few-shot learning [33]. Together, these advances indicate that investor psychology can be quantified through increasingly diverse digital traces and incorporated into financial decision-support systems.

The improvement in directional accuracy was particularly relevant because correct identification of market direction is central to exposure adjustment and loss avoidance. The hybrid model correctly predicted the direction of movement in approximately seven out of ten out-of-sample observations and achieved balanced precision and recall. This result indicates that the model did not obtain high accuracy by disproportionately predicting one dominant class. Its superior area under the curve further showed that it discriminated effectively between upward and downward movements across different classification thresholds. The result is consistent with previous hybrid forecasting studies in which the combination of multiple methods improved stock-price prediction and turning-point identification [19, 20]. It also supports research showing that the integration of metaheuristic optimization and flexible nonlinear models can improve financial forecasting by refining parameter selection and model structure [22].

A central contribution of the study was that predictive improvement translated into economically meaningful risk reduction. The proposed strategy achieved a higher annualized return than all benchmark strategies while simultaneously generating the lowest volatility, downside deviation, maximum drawdown, value at risk, and expected shortfall. This finding demonstrates that the model did not merely fit price movements more accurately; it improved the timing and intensity of market exposure. The reduction in maximum drawdown was especially substantial, suggesting that the model successfully identified periods in which capital exposure should be reduced. This is consistent with the purpose of hybrid risk models, which is to combine complementary information sources in order to respond more effectively to changing market conditions [23, 24]. It also aligns with the literature on portfolio rebalancing, in which machine-learning forecasts and optimization procedures have been combined to translate predictions into improved allocation decisions [21].

The marked reduction in downside deviation and expected shortfall indicates that the hybrid model was particularly effective in controlling harmful negative outcomes. This distinction is important because total volatility treats favorable and unfavorable fluctuations similarly, whereas downside deviation and expected shortfall focus directly on losses. The results therefore suggest that the model recognized warning signals associated with adverse market states and adjusted exposure before or during periods of severe decline. The finding corresponds with research on hybrid volatility prediction showing that the integration of multiple analytical components improves the representation of complex and time-varying financial risk [28]. It is also compatible with hybrid fuzzy-logic and machine-learning approaches that attempt to combine empirical pattern recognition with flexible representations of uncertainty in risk assessment [29].

The higher Sharpe, Sortino, and Calmar ratios further demonstrate that the hybrid strategy's gains were not achieved through uncontrolled risk taking. The Sortino ratio improved more strongly than the Sharpe ratio, indicating that the model's primary benefit arose from reducing harmful downside variability rather than simply lowering all fluctuations. The Calmar ratio also increased substantially because the model generated higher return relative to maximum drawdown. These findings support the argument that financial models should be assessed using economic and risk-adjusted indicators rather than prediction errors alone. A forecasting system may appear statistically accurate but still provide limited practical value if its largest errors occur during market crashes. Reviews of cryptocurrency forecasting similarly emphasize that realistic model evaluation should incorporate profitability, transaction costs, market regimes, and risk exposure rather than relying solely on goodness-of-fit measures [10].

The value-at-risk backtesting results provided additional evidence of the model's reliability. The hybrid model produced violation rates close to the theoretically expected frequencies at both confidence levels, and the violations did not occur in statistically significant clusters. By contrast, historical simulation and GARCH underestimated extreme risk and generated excessive or temporally dependent violations. This suggests that conventional methods responded too slowly to abrupt changes in market conditions or were unable to incorporate behavioral signals preceding episodes of stress. The hybrid model's improved calibration can be attributed to its combination of volatility history, technical conditions, and real-time behavioral information. This finding is consistent with conceptual frameworks emphasizing that financial risk is multidimensional and requires the consideration of market conditions, managerial judgment, behavioral responses, and changing uncertainty [1]. It also supports the broader view that artificial intelligence can improve financial decision-making by integrating heterogeneous information and generating more adaptive risk estimates [14].

The ablation analysis clarified how the model achieved its advantage. Removing technical indicators generated one of the greatest declines in directional accuracy and risk-adjusted performance, demonstrating that trend, momentum, volume, and market-structure variables remained essential. This result supports the continued relevance of technical analysis for identifying recurring patterns in price and trading activity [11]. Previous neuro-fuzzy and hybrid-indicator studies have likewise shown that combinations of technical variables can improve stock-price forecasts by representing nonlinear market relationships [18]. However, the decline observed when sentiment features were removed was also substantial, indicating that technical data alone were insufficient. Thus, the model's performance resulted from complementarity rather than replacement: technical variables represented realized market behavior, while sentiment variables represented expectations, attention, and emotional pressure that could precede price adjustment.

The removal of discussion-volume, engagement, and user-influence measures also weakened performance. These findings imply that the predictive relevance of social-media information depends not only on emotional polarity but also on how widely a message is disseminated and how much attention it receives. A strongly negative statement with minimal engagement may have limited market impact, whereas a moderately negative statement amplified by influential accounts and extensive reposting may produce a broader behavioral response. The importance of behavioral intensity is consistent with big-data approaches that analyze decision bias through large-scale digital information rather than isolated self-reports [9]. It also supports the concept of sentiment connectedness, according to which emotional signals can spread among assets and investors and thereby influence portfolio risk [34].

The robustness analysis showed that the model remained effective for both Bitcoin and Ethereum, although its performance was stronger for Bitcoin. Ethereum displayed higher realized volatility, greater tail risk, and more extreme return fluctuations, making it more difficult to forecast. Ethereum prices may also respond to a broader set of asset-specific factors, including network upgrades, decentralized-finance activity, smart-contract risks, and technological developments. Nevertheless, the hybrid model continued to outperform the benchmarks for Ethereum, indicating that the architecture was not limited to the relatively more established Bitcoin market. Its effectiveness across both assets supports the broader applicability of machine-learning methods in financial markets characterized by heterogeneous structures and risk sources [13].

Performance was also better at the hourly horizon than at the daily horizon. This result suggests that social-media information is rapidly incorporated into cryptocurrency prices and may lose some predictive value when aggregated over longer periods. The continuous operation of digital-asset markets allows narratives and emotional reactions to affect trading almost immediately. Therefore, high-frequency aggregation may preserve the temporal relationship between sentiment and market movement more accurately than daily aggregation. This finding is compatible with the use of big-data analytics and artificial intelligence for rapid financial decision support [15]. It also suggests that behavioral indicators should be aligned carefully with the intended forecasting horizon rather than assuming that sentiment has a uniform effect across time scales.

The model performed most strongly during bullish conditions and least strongly during high-volatility transitional periods. Bullish markets may contain more persistent momentum and relatively coherent positive sentiment, making their direction easier to identify. Transitional periods, by contrast, are characterized by abrupt reversals, conflicting narratives, liquidity shocks, and rapid regime changes. Under these conditions, historical patterns may become less reliable and social-media signals may be noisy or contradictory. The persistence of acceptable risk-adjusted performance during bearish periods is therefore more important than the model's peak performance during bullish markets. It indicates that the model retained practical value when loss control was most necessary. This result also reinforces the view that risk models should be assessed across heterogeneous market regimes and institutional conditions rather than through aggregate averages alone [2, 36].

The findings additionally have implications for the relationship between artificial intelligence and investor decision-making. Algorithmic systems can provide structured analysis that reduces dependence on emotional judgment, although their acceptance depends on perceived usefulness, trust, transparency, and ease of use. Research on retail investors' intention to adopt algorithmic trading platforms indicates that technological and psychological factors jointly determine whether predictive systems are used in practice [30]. The observed performance of the hybrid model may increase perceived usefulness, but decision-makers still require interpretable outputs and clear risk limits. Related evidence suggests that advanced language models can reduce emotional bias by supporting more systematic financial analysis [31]. However, algorithmic assistance should complement rather than eliminate human oversight, particularly in rapidly changing markets where model assumptions may fail.

The present study also extends earlier domestic evidence on artificial intelligence and hybrid forecasting. Artificial-intelligence techniques have been used to predict financial reporting behavior, demonstrating their capacity to identify latent and nonlinear patterns in financial data [16]. Hybrid genetic algorithm–support vector machine models have improved stock-price prediction through feature and parameter optimization [17], while hybrid intelligent approaches have been applied to critical price points and short-horizon forecasting [19, 20]. The present findings advance this literature by integrating structured market variables with transformer-derived sentiment and by evaluating the resulting model through a comprehensive set of predictive and risk-management criteria.

In summary, the proposed hybrid model performed better because it combined three complementary sources of information. The technical component identified trends, momentum, volume dynamics, and volatility patterns; the recurrent component learned temporal dependencies and changing market sequences; and the behavioral component quantified investor expectations, emotional polarity, disagreement, attention, and influence. No single component fully captured the complexity of cryptocurrency risk, but their integration produced more accurate forecasts, better-calibrated tail-risk estimates, lower drawdowns, and higher risk-adjusted returns. The results therefore support the use of multimodal artificial-intelligence systems as decision-support mechanisms for financial risk management, while also emphasizing the need for realistic validation, interpretability, and continuous performance monitoring.

The study had several limitations that should be considered when interpreting its findings. First, the analysis was restricted to Bitcoin and Ethereum, and the results may not be fully generalizable to less liquid cryptocurrencies, traditional equities, foreign-exchange markets, commodities, or fixed-income securities. Second, social-media data were limited to English-language posts from one platform, which may have excluded relevant information from other languages, online forums, news outlets, messaging platforms, and blockchain communities. Third, although spam, duplicates, and irrelevant posts were removed, automated procedures may not have completely identified bots, coordinated promotional activity, sarcasm, or strategically manipulated sentiment. Fourth, transaction costs and slippage were modeled through fixed assumptions, while actual trading costs may vary according to exchange liquidity, order size, volatility, and execution method. Fifth, the historical period included diverse market regimes, but future regulatory, technological, and macroeconomic changes may create patterns not represented in the training data. Finally, the computational complexity and limited interpretability of the hybrid architecture may make implementation more difficult for organizations lacking advanced data infrastructure and machine-learning expertise.

Future studies should evaluate the model across a broader range of digital and traditional assets and examine whether its performance remains stable in markets with different levels of liquidity, regulation, and investor participation. Researchers should incorporate multilingual sentiment, financial news, search trends, blockchain activity, order-book variables, macroeconomic indicators, and visual or audio content to create more comprehensive multimodal systems. Future work should also compare alternative transformer and sequential architectures, including domain-adapted large language models, temporal convolutional networks, attention-based recurrent models, and multimodal transformers. More detailed regime-detection procedures could be used to allow the model to adapt its structure or parameters automatically during bullish, bearish, neutral, and crisis conditions. Researchers should additionally examine explainable-artificial-intelligence methods to identify which technical and behavioral variables drive each risk forecast. Prospective and live-trading evaluations would provide stronger evidence of practical effectiveness than historical backtesting alone, particularly when realistic latency, market impact, variable transaction costs, and execution constraints are incorporated.

Financial institutions, portfolio managers, exchanges, and professional investors may use hybrid models as supplementary decision-support tools for monitoring volatility, adjusting exposure, setting dynamic risk limits, and identifying periods of elevated downside risk. Practical implementation should include continuous data-quality control, model-drift detection, periodic retraining, stress testing, and independent validation. Trading and risk-management teams should avoid relying on a single prediction and should interpret model outputs alongside liquidity conditions, regulatory developments, macroeconomic events, and organizational risk tolerance. Sentiment indicators should be weighted according to source credibility, engagement quality, and the probability of coordinated manipulation. Institutions should also establish human-review procedures for extreme signals and maintain predefined safeguards, including maximum position sizes, loss limits, capital buffers, and emergency deactivation rules. The model is most appropriately used to strengthen disciplined decision-making rather than to replace professional judgment or established risk-governance processes.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

[1] K. Hrytsiv and J. Kartašova, "Risk in Financial Decision-Making: A Conceptual Framework for Investors and Corporate Managers," 2025, doi: 10.15837/aijes.v19i1.7145.

[2] A. O. Scott, P. Amajuoyi, and K. B. Adeusi, "Theoretical Perspectives on Risk-Management Strategies in Financial Markets: A Comparative Review of African and U.S. Approaches," 2024.

[3] H. Hassanpour, M. Vaezz Barzani, A. Moarefi Mohammadi, and S. M. Emami Jomeh, "A Comparative Analysis of the Epistemological Foundations of Classical and Behavioral Finance," *Islamic Economics Knowledge*, vol. 11, no. 22, pp. 57-74, 2020.

[4] J. Almeida and T. C. Gonçalves, "A Systematic Literature Review of Investor Behavior in Cryptocurrency Markets," *Journal of Behavioral and Experimental Finance*, vol. 37, p. 100785, 2023, doi: 10.1016/j.jbef.2022.100785.

[5] A. Jalilvand and M. Rostami Nowruzabad, "Financial Literacy, Investor Sentiment, Risk Perception, and Investment Intention: Evidence from the Tehran Stock Exchange," *Investment Knowledge*, vol. 7, no. 27, pp. 141-169, 2018.

[6] S. Arabi Monfared, "Examining the Effect of Illusory Financial Literacy on Behavioral Biases among Stock-Market Investors in Shiraz," in *Third International Conference and Fourth National Conference on New Findings in Management, Psychology, and Accounting*, Tehran, Iran, 2023.

[7] G. Gol Arzi and S. Piri, "The Effect of Investors' Emotional Decisions on Crises in the Tehran Stock Exchange," *Financial Management Strategy*, vol. 9, no. 37, pp. 99-116, 2022.

[8] M. Ramesheh, A. Jannati Manzari, and M. Dadpanah, "Investor Sentiment and Financing Decisions in the Tehran Stock Exchange," *Financial Management Perspective*, no. 34, pp. 173-193, 2021.

[9] X. Huang, H. Li, and N. Wei, "Big Data-Driven Behavioral Economics: Analysis of Decision Biases in Financial Markets," 2025, doi: 10.70339/vyet1339.

[10] O. Sattarov and J. Choi, "From Prediction to Profit: A Comprehensive Review of Cryptocurrency Trading Strategies and Price-Forecasting Techniques," *IEEE Access*, vol. 12, pp. 87039-87063, 2024, doi: 10.1109/ACCESS.2024.3417449.

[11] V. Ladyzhets and S. Terenchuk, "Models and Methods of Technical Analysis of Financial Markets," vol. 48, pp. 47-52, 2021, doi: 10.32347/2412-9933.2021.48.47-52.

[12] C. Cappello, A. Congedi, S. De Iaco, and L. Mariella, "Traditional Prediction Techniques and Machine-Learning Approaches for Financial Time-Series Analysis," *Mathematics*, vol. 13, no. 3, p. 537, 2025, doi: 10.3390/math13030537.

[13] W. A. Addy, A. O. Ajayi-Nifise, B. G. Bello, S. T. Tula, O. Odeyemi, and T. Falaiye, "Machine Learning in Financial Markets: A Critical Review of Algorithmic Trading and Risk Management," *International Journal of Science and Research Archive*, vol. 11, no. 1, pp. 1853-1862, 2024, doi: 10.30574/ijsra.2024.11.1.0292.

[14] J. S. Dote-Pardo, M. C. Cordero-Díaz, M. T. Espinosa Jaramillo, and J. Parra-Domínguez, "Leveraging Artificial Intelligence for Enhanced Decision-Making in Finance: Trends and Future Directions," *Journal of Accounting Literature*, 2025, doi: 10.1108/JAL-02-2025-0100.

[15] V. Pillai, "Integrating AI-Driven Techniques in Big-Data Analytics: Enhancing Decision-Making in Financial Markets," 2023, doi: 10.18535/ijecs/v12i07.4745.

[16] F. Fereydouni, R. Darabi, and A. A. Anvar Rostami, "Application of Artificial Intelligence in Predicting Income Smoothing," *Financial Accounting and Auditing Research*, vol. 12, no. 45, pp. 103-134, 2020.

[17] O. M. Ebadati, M. A. Jafari, and N. Davoudifar, "Stock-Price Prediction Using Hybrid GA-SVM Algorithms," *Advances in Finance and Investment,* vol. 2, no. 5, pp. 1-22, 2021.
[18] S. Sarafraz, F. Sefati, and A. Ghiasvand, "Stock-Price Prediction Using Hybrid Market Indicators and a Neuro-Fuzzy Model," in *International Conference on New Ideas in Management, Economics, and Accounting*, Barcelona, Spain, 2016.
[19] M. Mashari, H. Didehkhani, K. Khalili Damghani, and E. Abbasi, "Designing a Hybrid Intelligent Model to Predict Golden Points in Stock Prices," *Investment Knowledge,* vol. 8, no. 29, pp. 45-66, 2019.
[20] V. Vafaei Ghaeni and A. M. Kimiagari, "One-Day-Ahead Stock-Price Prediction Using a Hybrid Model," *Investment Knowledge,* vol. 8, no. 30, pp. 313-328, 2019.
[21] S. Faridi, M. Madanchi Zaj, A. Daneshvar, S. Shahverdiani, and F. R. Roodposhti, "Portfolio Rebalancing Based on a Combined Method of Ensemble Machine Learning and Genetic Algorithm," *Journal of Financial Reporting and Accounting,* vol. 21, no. 1, pp. 105-125, 2023, doi: 10.1108/JFRA-11-2021-0413.
[22] D. Sabancı, S. Kılıçarslan, and K. Adem, "Stock-Market-Price Forecasting Using Metaheuristic Algorithms and Multivariate Adaptive Regression Splines," *International Journal of Intelligent Computing and Cybernetics,* vol. 16, no. 4, pp. 847-866, 2023, doi: 10.1108/IJICC-02-2023-0030.
[23] M. Korablyov, O. Fomichov, D. Antonov, S. Dykyi, I. Ivanisenko, and S. Lutskyy, "Hybrid Stock-Analysis Model for Financial-Market Forecasting," 2023, doi: 10.1109/CSIT61576.2023.10324069.
[24] A. K. Bukhanets, D. A. Dudnik, and O. V. Kosnikova, "Hybrid Machine-Learning Models for Financial-Market Analysis," 2024, doi: 10.47576/2949-1908.2024.1.1.003.
[25] F. Afzal, F. Afzal, M. Kamran, and H. Pan, "Stock-Market Predictive Modeling Using a Recurrent Neural Network with Long Short-Term Memory Architecture," *International Journal of Islamic and Middle Eastern Finance and Management,* 2025, doi: 10.1108/IMEFM-12-2024-0632.
[26] S. Panigrahi, G. Kukreja, and S. Kumaraswamy, "Forecasting Stock-Market Behavior in BRICS Economies Using Artificial Neural Machine-Learning Models," *Journal of Business and Socio-economic Development,* 2025, doi: 10.1108/JBSED-05-2025-0134.
[27] Z. Mao and C. Wu, "Stock-Price-Index Prediction Based on an SSA-BiGRU-GSCV Model from the Perspective of Long Memory," *Kybernetes,* vol. 53, no. 12, pp. 5905-5931, 2024, doi: 10.1108/K-02-2023-0286.
[28] W. Wang and Y. Wu, "Risk Analysis of the Chinese Financial Market Using a Novel Hybrid Volatility-Prediction Model," *Mathematics,* vol. 11, no. 18, p. 3937, 2023, doi: 10.3390/math11183937.
[29] T. Kaur, K. Rani, P. Thakur, and N. S. Talwandi, "Enhanced Decision-Support System for Financial Risk Assessment Using Hybrid Fuzzy Logic and Machine Learning," 2024, doi: 10.1109/ICCICA60014.2024.10584834.
[30] B. Bunkar and K. Ramaiah, "A Serial Mediation Model for Investigating the Intention to Use Algorithmic Trading Platforms among Retail Investors in India," *VILAKSHAN: XIMB Journal of Management,* vol. 21, no. 2, pp. 263-280, 2024, doi: 10.1108/XJM-12-2023-0233.
[31] W. C. Hsu, M. C. Wang, and H. I. Ting, "Reducing Emotional Bias in Investment Decisions: The Role of GPT-4 in Financial Analysis," *Asia-Pacific Journal of Business Administration,* pp. 1-26, 2025, doi: 10.1108/APJBA-03-2025-0181.
[32] L. Bani Fatemeh and S. Ghaffari, "Artificial Intelligence, the Semantic Web, and Their Interaction: A Review," *Scientific Quarterly,* vol. 10, no. 21, pp. 61-85, 2021.
[33] X. Ren, W. Jiang, X. Sun, and S. Wang, "Investor Sentiment from Images: A Few-Shot Learning Investigation," *Journal of Accounting Literature,* pp. 1-30, 2025, doi: 10.1108/JAL-04-2025-0199.
[34] K. Nyakurukwa and Y. Seetharam, "Leveraging Sentiment Connectedness for Portfolio Optimisation: A Minimum-Connectedness Approach," *Review of Behavioral Finance,* pp. 1-10, 2025, doi: 10.1108/RBF-03-2025-0125.
[35] S. Nayak, "Developing Predictive Models for Financial Stability: Integrating Behavioral Analytics into Credit-Risk Management," 2024, doi: 10.47363/JAICC/2024(3)E173.
[36] M. Lobato, J. Rodríguez, and H. Romero-Perez, "Overseas Listing and Risk: The Case of Single-Listed Chinese ADRs," *Managerial Finance,* pp. 1-14, 2025, doi: 10.1108/MF-10-2024-0753.