International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Systematic Review of Machine Learning Approaches for Medium-Horizon Stock Direction Prediction in Emerging Markets

Pramod T Talole [1], Pavan Tour[2], Nitesh Chavan[3] Nilesh Ambekar[4] Pranit Bodade[5]

[1]*Department of Information Technology, Anuradha College of Engineering and Technology, Chikhli, India*

[2]*Department of Information Technology, Anuradha College of Engineering and Technology, Chikhli, India*

[3]*Department of Information Technology, Anuradha College of Engineering and Technology, Chikhli, India*

[4]*Department of Information Technology, Anuradha College of Engineering and Technology, Chikhli, India*

[5]*Department of Information Technology, Anuradha College of Engineering and Technology, Chikhli, India*

Abstract

Machine learning (ML) methods are widely applied to stock market prediction, and many studies report that ML models can capture nonlinear patterns and sometimes outperform simple benchmarks. However, most of this work focuses on very short horizons such as intraday or next-day movements, while medium-horizon prediction over one to six months remains comparatively underexplored, particularly in emerging markets. Existing studies show that both classical ML models and deep learning architectures can extract useful patterns from prices, volumes, fundamentals, and macro indicators, but reported performance is highly sensitive to the prediction horizon, market type, and evaluation protocol. At the same time, recent reviews and methodological analyses highlight common problems such as unrealistic random train–test splits, data leakage, and lack of reproducibility, which make it difficult to judge how reliable these models truly are.

This review provides a concise overview of ML-based methods for stock direction prediction with an emphasis on medium-term horizons and emerging markets. It clarifies the prediction problem and horizons, summarizes typical data sources and feature families, outlines the main model types used in the literature, and discusses evaluation practices and pitfalls. The review then identifies open challenges related to non-stationarity, regime shifts, and real-world deployability, and outlines directions for more standardized and reproducible research on medium-horizon stock direction prediction in emerging markets.

Keywords: Machine Learning, Stock Prediction, Finance ML, Stock Features Engineering.

1. INTRODUCTION

Survey and empirical studies show that ML models can capture nonlinear patterns in financial time series and, in some settings, outperform simple statistical benchmarks for stock prediction [1][2]. Most of these studies, however, focus on short horizons—typically intraday to a few days ahead—where large amounts of data are available and trading signals are driven mainly by microstructure effects, technical patterns, and short-lived news [1][2]. Medium-horizon prediction over one to six months, which is important for portfolio rebalancing, tactical asset allocation, and risk management, is less frequently studied.

This gap is more evident in emerging markets such as the BRICS economies. Research on stock trend prediction in emerging markets reports that model performance and stability differ from those in developed markets, with stronger influence from macroeconomic conditions, policy changes, and country risk [3][4]. Studies on BRICS indices in particular find that macroeconomic and fundamental factors play a larger role at monthly and quarterly horizons, and that ML models must cope with higher volatility and structural breaks [3][4].

Existing reviews cover a broad range of ML methods for stock prediction but usually mix different horizons and market types, which can hide important differences in data requirements, feature design, and evaluation practice [1]. In parallel, several works emphasize that many reported results rely on random train–test splits, do not respect the time order of data, or suffer from lookahead bias and other forms of data leakage [1][2][7]. As a result, nominally high accuracies or returns may not be reproducible in realistic trading setups.

This paper focuses specifically on ML-based stock **direction** prediction over medium horizons in emerging markets. Its contributions are: (i) a clear problem formulation that distinguishes direction classification from price regression and separates short-, medium-, and long-term horizons; (ii) a compact taxonomy of data sources, features, and model families, with attention to studies on emerging markets [1]–[6]; and (iii) a discussion of evaluation protocols and common pitfalls such as data leakage and overfitting in time-series settings [1][2][7].

2. Problem Formulation

2.1 Direction Prediction vs. Price Regression

In stock prediction, ML tasks are commonly framed either as **regression**, where the model predicts a continuous future price or return, or as **direction classification**, where the model predicts whether the future return will be positive or negative over a given horizon [1][2][3][6]. Many empirical studies use a binary label defined by the sign of the cumulative return over the prediction window and train classifiers such as support vector machines (SVM), random forests, gradient boosting, or neural networks on these labels [2][3][6].

Direction prediction is attractive because it maps naturally to buy/sell decisions and allows the use of standard ML metrics such as accuracy or F1-score, while avoiding some of the noise sensitivity of raw price regression [1][2][3]. At the same time, financial returns are noisy and often close to zero, which can lead to imbalanced or unstable class labels and makes it possible to achieve seemingly good accuracy with only modest economic value [2][3]. This underlines the importance of linking prediction targets and metrics to realistic trading or portfolio objectives [2][3].

2.2 Prediction Horizons

The prediction **horizon** strongly affects model design and the type of information that is useful. The literature typically distinguishes:

Table 1: Prediction Horizons

Horizon	Time Range	Key Features	Main Challenges
Short-term	Intraday–few days	OHLCV, technical indicators	Noise, microstructure effects
Medium-term	1–6 months	Fundamentals, macro, technical	Small sample size, regime shifts
Long-term	6+ months	Macro, structural factors	Very limited samples, cycles

Studies comparing different models and horizons report that architectures which perform well for one-day-ahead prediction do not necessarily retain their edge at multi-week or multi-month horizons [2][3]. In emerging markets, evidence suggests that combining technical indicators with fundamental and macro features is more beneficial at medium horizons than relying on price-based signals alone [3][4][5].

2.3 Why Medium-Horizon Prediction Is Different

Medium-horizon direction prediction sits between short-term trading and long-term investing. Several aspects make it distinct from the short-term case:

a. **Signal composition**: Over months, returns reflect not only microstructure noise and short-term technical patterns but also earnings, valuation adjustments, macro shocks, and changes in investor sentiment, which are more pronounced in emerging markets [3][4].

b. **Feature dynamics**: Many very short-window technical indicators lose relevance, while slower-moving variables such as valuation ratios, fundamentals, and macro indicators gain importance [2][3][4][5].

c. **Sample size and overfitting**: Because each label aggregates several weeks or months, the effective sample size is smaller, making complex models more prone to overfitting [1][2][3][5].

d. **Evaluation complexity**: With fewer non-overlapping periods, robust time-series cross-validation and realistic backtesting become crucial, and naive random splits can produce misleading results [1][2][7].

These differences suggest that simply extending short-term models to longer horizons is not sufficient. Instead, medium-horizon prediction in emerging markets requires careful feature design, model selection, and evaluation protocols that respect the temporal structure and economic context [2][3][4].

3. Data Sources In Financial Machine Learning

3.1 Market Data

Most ML studies on stock prediction start from market data such as daily or intraday open, high, low, close prices and trading volume (OHLCV) [1][2][3]. These series are used directly as inputs to classical models or as the basis for constructing technical indicators and sequences for deep learning models. Experiments across multiple indices show that OHLCV data remain the core input for both developed and emerging markets, including BRICS economies [3][4]. However, emerging markets often exhibit higher volatility and lower liquidity, which can affect both model training and the implementation of trading strategies [3][4].

3.2 Fundamental and Macroeconomic Data

To go beyond pure price information, many works incorporate **fundamental** variables (such as earnings, book value, leverage, and valuation ratios) and **macroeconomic** indicators (such as interest rates, inflation, exchange rates, or commodity prices) [2][3][4][5]. Studies on feature learning for stock prediction find that combining technical and fundamental inputs can improve generalization compared to technical indicators alone [5]. Research on emerging and BRICS markets similarly reports that macroeconomic and country-specific risk factors are important drivers of medium-horizon index behavior [3][4]. For ML modeling, these variables are typically aligned with monthly or quarterly returns, which requires respecting publication lags to avoid using future information [1][2][7].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

3.3 Data Quality, Survivorship Bias, and Leakage

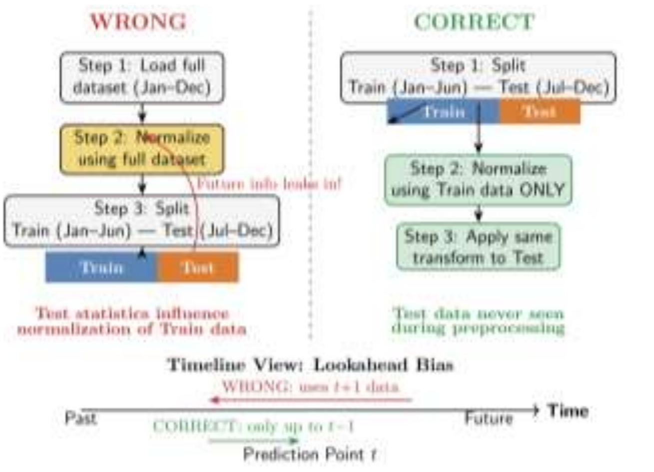

Figure 1: Data Leakage Illustration

Data quality issues such as missing corporate actions, ignoring delisted stocks, or using only current index constituents can introduce survivorship bias and overstate historical performance [1][2][4]. Survey articles stress that many ML studies do not clearly document how they handle splits, dividends, mergers, or index rebalancing, which makes replication difficult and may lead to overly optimistic results [1][2]. This is especially relevant in emerging markets with frequent structural changes and firm turnover [3][4].

A closely related problem is **data leakage**, where information that would not have been available at prediction time accidentally enters the training set. Recent work on lookahead bias shows that such leakage can dramatically inflate reported accuracy or trading performance in financial forecasting tasks [7]. Common sources include misaligned accounting data, rolling windows that include future observations, and random rather than time-ordered train–test splits [1][2][7]. Robust experimental design therefore requires careful alignment of features and targets in time and strict use of chronological splits or walk-forward validation, points that will be revisited in the discussion of evaluation protocols [1][2][7].

4. Feature Engineering Approaches

Table 2: Types of Features

Feature Type	Examples	Horizon	Key Issue
OHLCV	Close, volume	All	Noise
Technical	RSI, MACD	Short	Leakage risk
Fundamental	P/E, earnings	Med/Long	Pub lag
Sentiment	News scores	Short/Med	Quality
Macro	Rates, FX	Med/Long	Low freq
Feature engineering pitfalls	Rolling stats, normalization	All	Leakage

4.1 Raw OHLCV Features

Most ML approaches for stock prediction start from raw market data: open, high, low, close prices, and traded volume (OHLCV) sampled at daily or intraday frequency [1][2][3]. These variables are used directly as model inputs or as the base from which more complex features are derived. Empirical studies across different indices show that even relatively simple models built on lagged prices and volumes can capture some predictability, especially at short horizons [2][3]. In emerging markets, OHLCV data remain the most accessible and consistent source, although higher volatility and occasional liquidity gaps require careful handling of missing values and outliers [3][4].

4.2 Technical Indicators

Technical indicators such as moving averages, momentum, relative strength index (RSI), and moving average convergence–divergence (MACD) are widely used to summarize past price and volume behavior [1][2][3]. Many studies construct dozens of such indicators and feed them into classifiers like SVM, random forests, or gradient boosting to predict short-term or medium-term direction [2][3][6]. Survey work notes that technical-indicator-based feature sets dominate much of the ML literature on stock prediction, including for emerging markets [1][2]. However, indicators derived from overlapping windows can be highly correlated and may not add much independent information for medium horizons, where slower-moving factors become more relevant [2][3][4].

4.3 Fundamental Features

Fundamental features capture information about firms' financial health and valuation, including earnings, book value, leverage, profitability ratios, and price-to-earnings or price-to-book multiples [2][3][5]. Research on feature learning for stock prediction shows that models which combine technical and fundamental signals often generalize better than those relying on technical indicators alone, especially when the prediction horizon extends beyond a few days [5]. In emerging markets, studies report that valuation and balance-sheet variables are important drivers of index-level behavior at monthly and quarterly frequencies, supporting their use as inputs for medium-horizon direction prediction [3][4]. Proper alignment of these variables with returns is crucial, since financial statements are released with delays and revisions [1][2][7].

4.4 News and Sentiment Features

Some ML studies enrich price- and fundamental-based features with text-derived signals from news, reports, or social media, typically in the form of sentiment scores or topic indicators [1][3]. Survey articles note that such sentiment features are most common in short-term prediction but can also affect medium-horizon returns when they capture persistent changes in investor expectations [1][3]. In emerging markets, where idiosyncratic political or policy news can have outsized effects on prices, incorporating structured sentiment or event-based variables may be particularly valuable, although high-quality text data are less standardized than price data [3][4].

4.5 Macroeconomic and Market-Wide Indicators

Macroeconomic and market-wide indicators aggregate information at the country or global level. Typical examples include interest rates, inflation, exchange rates, commodity prices, volatility indices, and broad credit or risk-premium measures [2][3][4]. Empirical work on emerging and BRICS economies finds that such variables are significant drivers of medium-term stock index movements and can help explain cross-market differences in ML model performance [3][4]. From a feature-engineering perspective, these indicators can be included directly, transformed into growth rates or surprises relative to expectations, or summarized using dimensionality-reduction techniques when many correlated series are available [2][3][4].

4.6 Pitfalls of Precomputed Indicators and Feature Leakage

While engineered features can improve predictive power, they also introduce risks. Reviews highlight that many studies compute technical indicators, rolling statistics, or normalization parameters using windows that inadvertently include future observations, leading to subtle forms of data leakage [1][2][7]. Similarly, precomputed datasets that already contain indicators or labels may hide how those variables were aligned in time, making it difficult to verify whether lookahead bias is present [1][2][7]. These issues are especially problematic for medium-horizon prediction, where each label spans weeks or months and overlapping windows can blur the boundary between past and future. Careful implementation is required to ensure that all features are constructed using only information available at the prediction time and that any scaling or feature selection is performed within a proper time-series cross-validation framework [1][2][7].

5. Machine Learning Models Reviewed

5.1 Classical Machine Learning Models

Classical ML models such as logistic regression, SVM, k-nearest neighbors, decision trees, random forests, and gradient boosting (including XGBoost) are widely used for stock direction prediction [1][2][3][6]. Comparative studies show that tree-based ensembles and SVMs often outperform simpler linear models on technical-indicator feature sets, particularly for short-term horizons [2][3][6]. One study that forecasts trade profitability using random forests, SVM, and XGBoost finds that XGBoost frequently achieves the best classification performance among the tested algorithms [6]. Reviews also note that these models remain competitive baselines for many applications, including in emerging markets, because they handle nonlinearities and interactions without requiring very large datasets [1][2][3][4][6].

5.2 Deep Learning Models

Deep learning models extend beyond handcrafted features by learning hierarchical representations directly from raw or minimally processed inputs. In stock prediction, common architectures include feed-forward neural networks, recurrent networks such as LSTM, and convolutional neural networks applied to temporal sequences or transformed price images [2][3][5]. Empirical work reports that LSTM-based models can capture temporal dependencies in financial time series and sometimes outperform classical baselines on stock trend prediction tasks [2][3]. Feature-learning studies similarly find that neural networks can learn joint representations of technical and fundamental variables that improve prediction accuracy compared to manually designed features [5]. However, deep models require more data, are more prone to overfitting, and their gains are less consistent at longer horizons, especially when the effective sample size is limited [1][2][3][5].

5.3 Hybrid and Ensemble Approaches

Hybrid and ensemble approaches aim to combine the strengths of different model types or data sources. Examples include stacking classical models with neural networks, integrating technical, fundamental, and macro features within a single architecture, or ensembling models trained on different markets or sectors [1][2][3][4][5][6]. Survey work notes that hybrid systems are increasingly common in stock prediction, reflecting the view that no single model class dominates across all markets and horizons [1][2]. Empirical studies on emerging and BRICS markets indicate that models which jointly exploit technical signals, fundamentals, and macro factors tend to perform better at medium horizons than models based on any single feature family [3][4].

5.4 Suitability for Medium-Horizon Prediction in Emerging Markets

Evidence from comparative experiments suggests that the relative performance of classical and deep models depends on both the prediction horizon and the richness of available features. For medium-term direction prediction in emerging markets, where datasets are often shorter and structural breaks more frequent, reviews emphasize the importance of starting from strong classical baselines such as regularized logistic regression, SVMs, and tree-based ensembles, then considering deep or hybrid models when sufficient data and careful regularization are available [1][2][3][4][6]. Studies across multiple indices show that more complex architectures do not automatically translate into better out-of-sample performance, particularly when evaluation protocols are made more realistic with walk-forward validation and strict controls for data leakage [1][2][3][7].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

6. Evaluation Protocols In The Literature

6.1 Random Splits vs. Time-Series Splits

Many ML papers on stock prediction report performance using standard random train–test splits that assume independent and identically distributed samples [1][2]. For financial time series, this assumption is unrealistic because data are ordered in time and future observations must not influence model training. Survey and empirical studies emphasize that random splits can leak information from the future into the training set, leading to overstated accuracy and unstable results when models are applied chronologically [1][2][7]. In contrast, time-series-aware splits that train on past data and test on strictly later periods provide a more realistic assessment of predictive power [1][2].

6.2 Walk-Forward Validation

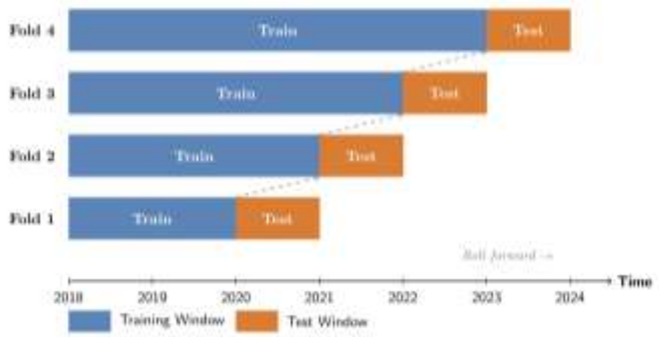

Figure 2: Walk-Forward Validation for Financial Time Series

Walk-forward (rolling) validation is widely recommended for financial applications because it mimics how models are used in practice: train on a historical window, test on the next period, then roll forward and repeat [1][2][3]. Studies that compare random splits with walk-forward evaluation show that reported accuracy and robustness often drop when time ordering is enforced, especially for more complex models [2][3]. For emerging markets, where structural breaks and regime changes are common, walk-forward schemes also allow researchers to observe how performance evolves across different subperiods and crises [3][4].

6.3 Backtesting vs. Pointwise Metrics

A large part of the literature evaluates stock direction models using pointwise metrics such as accuracy, precision, recall, or F1-score on the predicted labels [1][2][3][6]. While these metrics are informative from an ML perspective, they do not directly capture economic usefulness. Empirical work that combines ML predictions with simple trading rules shows that modest directional accuracy can still yield attractive risk-adjusted returns, whereas high accuracy does not guarantee profitable strategies once transaction costs and risk are taken into account [2][3]. For this reason, several studies advocate reporting both statistical metrics and backtest results such as cumulative return, Sharpe ratio, and maximum drawdown [1][2][3].

6.4 Data Leakage and Lookahead Bias in Evaluation

Recent work explicitly testing for lookahead bias in financial forecasting illustrates how subtle design choices can contaminate evaluation [7]. Examples include constructing rolling windows that inadvertently include future data, normalizing features using information from the full sample, or aligning fundamental and macro variables with returns using publication dates that are not realistic [1][2][7]. Such leakage can lead to very high reported accuracy or returns that disappear when the experiment is corrected. Surveys therefore stress that evaluation protocols must ensure strict temporal separation between training and test data, and that all preprocessing steps, including scaling and feature selection, should be performed within each training window of a time-series cross-validation scheme [1][2][7].

7. Challenges And Open Problems

7.1 Non-Stationarity and Concept Drift

Financial time series are highly non-stationary: statistical relationships can change over time due to evolving market structure, regulation, technology, and investor behavior [1][2][3]. In emerging markets, these shifts can be even stronger because of rapid economic development, policy changes, and varying capital flows [3][4]. Studies that evaluate ML models across different subperiods often find that patterns learned in one regime do not transfer well to another, indicating concept drift and model instability [2][3]. Designing models and training schemes that adapt to such drift remains a central open problem for medium-horizon prediction.

7.2 Overfitting with Limited Data

Medium-horizon direction prediction reduces the effective number of independent samples, since each label aggregates weeks or months of returns. This makes complex models particularly prone to overfitting, especially when many features or deep architectures are used [1][2][3][5]. Comparative experiments show that deep learning models do not always outperform simpler baselines when evaluated with realistic time-series cross-validation, suggesting that their capacity can exceed what the data support [2][3]. Regularization, feature selection, and careful model comparison against strong classical baselines are therefore essential, but guidelines tailored specifically to medium-term horizons in emerging markets are still limited [1][2][3][6].

7.3 Market Regime Shifts and Tail Events

Market crises, policy shocks, and other rare events can have disproportionate impact on returns and can fundamentally alter relationships between predictors and targets [2][3][4]. BRICS-focused studies highlight that emerging markets are especially exposed to global risk sentiment and domestic political developments, which can cause abrupt regime shifts [3][4]. Most ML models are trained on historical averages and may not generalize well to these tail events. Developing techniques that are robust to regime changes—such as regime-switching models, ensemble methods across regimes, or stress-testing models on crisis periods—remains an important open area.

7.4 Lack of Reproducibility and Standardization

Surveys point out that many financial ML studies do not release code, data preprocessing details, or complete hyperparameter settings, which makes replication difficult [1][2]. In addition, evaluation protocols differ widely across papers, with some using random splits, others using different walk-forward schemes, and many omitting transaction costs or realistic constraints in backtests [1][2][3][7]. This lack of standardization hinders fair comparison of methods and slows progress. The problem is particularly acute for emerging markets, where public benchmark datasets and shared evaluation frameworks are rare [1][2][4].

8. Research Gaps And Future Directions

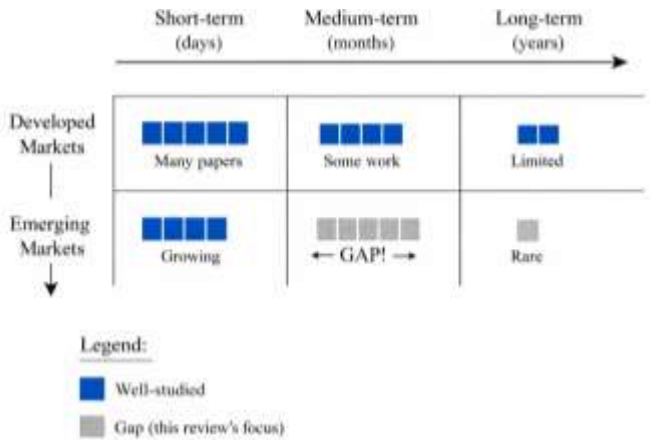

Figure 3: Research Coverage Gap in Stock Prediction Literature

8.1 Standardized Benchmarks for Medium-Horizon Prediction

The literature still lacks widely accepted benchmark datasets and protocols specifically targeting one- to six-month direction prediction, especially for emerging markets. Existing surveys cover many horizons and markets together, which makes it hard to isolate what works best at medium horizons [1][2]. A promising direction is to build open benchmark suites that include multiple emerging-market indices, clearly defined prediction horizons, and standardized walk-forward evaluation with transaction-cost assumptions [2][3][4]. Such benchmarks would enable more meaningful comparisons between classical, deep, and hybrid models.

8.2 Integrating Technical, Fundamental, and Macro Features

Evidence from feature-learning and cross-market studies suggests that combining technical, fundamental, and macroeconomic signals is particularly important at medium horizons [2][3][5]. However, there is no consensus on how best to integrate these heterogeneous inputs. Future work could explore structured architectures that model short-term technical patterns and slower-moving fundamentals or macro variables in separate modules and then fuse them, as well as representation-learning techniques that capture common factors across markets [2][3][5]. Special attention is needed to avoid feature leakage when aligning slow-moving variables with returns [1][2][7].

8.3 Robust and Transparent Evaluation Pipelines

Given the prevalence of data leakage and optimistic reporting, developing robust, transparent evaluation pipelines is a key research need [1][2][7]. Such pipelines should enforce chronological splits, implement walk-forward validation, encapsulate all preprocessing within the training window, and produce both statistical metrics and backtest results under realistic cost assumptions [1][2][3][7]. Making these pipelines open source and reusable across markets would greatly improve reproducibility and comparability, and would be particularly valuable for emerging-market research where proprietary data and code are common [1][2][4].

8.4 Cross-Market Generalization in Emerging Economies

Studies that compare model performance across indices report that methods calibrated on one market do not always transfer well to others, especially from developed to emerging markets [2][3][4]. Future research could investigate domain adaptation, transfer learning, and multi-task learning approaches that share information across related markets while allowing for country-specific effects [2][3]. For medium-horizon prediction, such methods might exploit common macro and risk-factor structures across emerging economies while adapting to local idiosyncrasies [3][4].

9. Conclusion

Machine learning has become a central tool in stock market prediction, yet most published work still focuses on short-term horizons and developed markets. This review has highlighted the distinct challenges and opportunities associated with medium-horizon stock direction prediction in emerging markets. Compared with short-term tasks, medium-term prediction relies more heavily on fundamentals and macroeconomic conditions, faces stronger non-stationarity and regime shifts, and suffers from smaller effective sample sizes. These factors increase the risk of overfitting and make rigorous evaluation protocols essential.

Across the surveyed studies, classical models such as SVMs and tree-based ensembles remain strong baselines, while deep and hybrid models can offer gains when sufficient data and careful regularization are available. However, the reliability of reported results is often undermined by data leakage, ad-hoc evaluation schemes, and limited reproducibility. Addressing these issues will require standardized benchmarks, transparent pipelines, and a stronger focus on economic relevance rather than purely statistical accuracy.

For practitioners and researchers working on emerging markets, the key takeaway is that model choice, feature design, and evaluation must be tailored to the medium-term horizon and local market characteristics. Future work that combines richer feature sets with robust, reproducible evaluation frameworks has the potential to close the gap between promising in-sample results and methods that can be deployed reliably in real-world investment settings.

References

[1] "A Systematic Review of Recent Advances on Stock Markets ...," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 2024.

[2] "How effective is machine learning in stock market predictions?," Heliyon, 2024.

[3] "Applying machine learning algorithms to predict the stock price trend in emerging markets," Humanities and Social Sciences Communications (Nature portfolio), 2024.

[4] "Forecasting stock market behavior in BRICS economies using machine learning methods," Journal of Business and Socio-economic Development (Emerald), 2025.

[5] "Feature Learning for Stock Price Prediction Shows a Significant Role of Technical and Fundamental Indicators," arXiv preprint arXiv:2103.09106, 2021.

[6] "Forecasting profitability in equity trades using random forest, support vector machines, and XGBoost," in Proc. of a National Conference (ConferenceWorld), 2019.

[7] "A Test of Lookahead Bias in LLM Forecasts," arXiv preprint, 2025.