Proceedings of CONF-CDS 2026 Symposium: Data-Centric AI Security: Securing Models, Learning Agents, and Autonomous Systems
DOI: 10.54254/2755-2721/2026.CH36590

Review of The Evolution, Challenges, and Future Directions of Forecasting Methods

Zhiting Chen

International Department, The Affiliated High School of South China Normal University, Guangzhou, China
203273849@qq.com

Abstract. Forecasting serves as a critical cornerstone for strategic planning, operational efficiency, and risk mitigation across modern civilization. By converting historical data into actionable forward-looking insights, it enables organizations and governments to anticipate market shifts, optimize resource distribution, and safeguard against systemic uncertainties. Predictive modeling is a fundamental task in many fields, such as finance, economics, engineering, and artificial intelligence. The aim of this paper is to summarize the methods of statistics and machine learning, outline their inherent challenges, and project future research directions. This paper mainly discusses traditional statistical methods (including Autoregressive Integrated Moving Average [ARIMA] and regression analysis), machine learning approaches (such as Random Forest and Support Vector Machines [SVM]), and deep learning models (such as Long Short-Term Memory [LSTM] networks and hybrid time series-ML models). Nowadays, as these interconnected fields become increasingly complicated, practitioners face severe challenges regarding data quality, computational complexity, and mathematical interpretability. This paper comprehensively reviews these methodologies, establishes a comparative taxonomy, and delineates the evolutionary trajectory of future forecasting applications.

Keywords: Forecasting methods, Machine learning, Deep learning, LSTM, Statistical modeling

1. Introduction

Forecasting plays a significant role in modern society. It is extensively utilized across a broad spectrum of domains, including stock markets, weather forecasting, transportation planning, business management, energy grid regulation, and supply chain logistics. Accurate predictions help allocate resources efficiently, optimize macroeconomic policies, and systematically manage operational risks. In an increasingly volatile global environment, the capacity to generate precise, robust, and timely forecasts differentiates successful enterprises from failing ones and underpins resilient public infrastructure.

Historically, predictive paradigms were dominated by classical statistical models that relied on structural assumptions such as linearity and stationarity. According to Akbar Siami Namin, the ARIMA model has demonstrated outstanding performance in precision and accuracy when

DOI: 10.54254/2755-2721/2026.CH36590

predicting the next lags of time series data [1]. These classical models provided a rigorous mathematical foundation for univariate data patterns. However, as the volume, velocity, and variety of data exploded in the digital era, traditional methods began struggling with highly non-linear, multi-dimensional interactions.

With recent advancements in the computational power of computers—and, more importantly, the development of more advanced machine learning algorithms and approaches such as deep learning—new algorithms have emerged to forecast complex time series data. These paradigms shift away from rigid structural assumptions toward data-driven pattern recognition. For instance, the average reduction in error rates obtained by LSTM is between 84 and 87 percent when compared to ARIMA, clearly indicating the technical superiority of LSTM over ARIMA under complex sequential conditions [1].

This paper reviews and compares these three major categories of forecasting methods based on existing literature. It aims to reveal the specific advantages and structural weaknesses unique to each category. Besides evaluating individual models, this paper also discusses current operational challenges and future directions to provide a useful reference for researchers and practitioners in this rapidly evolving field.

2. Traditional statistical methods

Traditional statistical methods include ARIMA and linear regression. These models represent the foundational bedrock of time series analysis, deriving their strength from mathematical transparency, statistical rigor, and minimal data requirements.

2.1. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is a widely used statistical method for time series analysis and forecasting. It systematically combines autoregressive (AR) components, differencing operators (I), and moving average (MA) components to model underlying patterns in historical data and generate future predictions. The AR component captures the linear dependency between an observation and a specified number of lagged observations. The integrated (I) component represents the differencing steps required to eliminate non-stationarity, such as trends or seasonality, thereby stabilizing the mean of the time series. The MA component models the relationship between an observation and a residual error derived from a moving average model applied to lagged observations.

ARIMA relies heavily on past observations to identify internal trends, cyclical behaviors, and serial dependencies within a time series, making it particularly useful for forecasting based entirely on historical behavior. Depending on the unique characteristics of the data, ARIMA can be effectively applied to both non-seasonal and seasonal time series (the latter formalized as SARIMA) by incorporating recurring patterns over fixed chronological intervals. Due to its mathematical effectiveness, minimal computational cost, and relatively simple structure, ARIMA remains one of the most commonly used forecasting techniques in statistical modeling (Joshua Noble). However, it requires strict adherence to assumptions of linearity and stationarity, often struggling when faced with abrupt structural breaks or highly irregular non-linear shocks.

DOI: 10.54254/2755-2721/2026.CH36590

2.2. Linear regression analysis

Linear regression is a foundational statistical technique used to examine, quantify, and predict the relationship between dependent and independent variables. The method estimates the numerical coefficients of a linear equation that best describes how changes in one or more independent variables systematically affect the primary dependent variable. By applying the standard ordinary least squares (OLS) approach, linear regression fits a geometric line (or hyperplane in multivariate contexts) that minimizes the sum of squared differences between observed and predicted values.

Due to its simplicity, computational speed, and exceptional interpretability, it is one of the most widely used methods for predictive modeling and data analysis (IBM). Every coefficient in a linear regression model yields a direct, quantifiable physical interpretation, allowing analysts to understand not just what the prediction is, but why it was made based on individual driver variables. Nevertheless, linear regression possesses inherent limitations: it assumes that explanatory variables are completely independent of one another (absence of multicollinearity), that error terms are uncorrelated and homoscedastic, and that the underlying real-world relationships are fundamentally linear. When these restrictive assumptions are violated, the model's predictive validity degrades significantly.

3. Machine learning approaches

Machine learning (ML) approaches relax the rigid structural assumptions of classical statistics, leaning instead on algorithmic architectures capable of extracting complex, non-linear patterns directly from large datasets.

3.1. Random Forest

Random Forest is a widely applied machine learning technique originally developed by Leo Breiman and Adele Cutler. It operates as an ensemble learning method, constructing an expansive forest of multiple individual decision trees at training time and combining their distinct predictions to generate a stabilized final result. For classification tasks, the output is the mode of the classes; for regression and forecasting tasks, the final prediction is calculated as the mathematical average of the outputs from all individual trees. Random Forest relies heavily on bootstrap aggregating (bagging) and feature randomness, ensuring that each tree develops independently and minimizes correlation with its peers.

Due to its simplicity, adaptability, and consistently strong performance, Random Forest has become immensely popular across many scientific and commercial fields. In addition, it can be seamlessly used for both classification and regression tasks, making it a highly versatile tool for multi-dimensional data analysis.

Random Forest is highly valued for its natural ability to drastically reduce overfitting—a common flaw of single decision trees—while achieving exceptional predictive accuracy and managing high-dimensional datasets without requiring complex feature selection. The algorithm is also robust to missing data and provides useful, intuitive measures of feature importance, which rank the relative predictive power of different input variables. This makes it an effective and versatile tool for complex data analysis, pattern recognition, and multivariate forecasting.

DOI: 10.54254/2755-2721/2026.CH36590

3.2. Support Vector Machines (SVM)

Support Vector Machine (SVM) is a highly robust supervised learning method extensively applied to classification and regression problems. When adapted for continuous numerical prediction and forecasting, it is frequently referred to as Support Vector Regression (SVR). The algorithm fundamentally aims to construct an optimal decision boundary, known as a hyperplane, in a high-dimensional space. In classification, this hyperplane maximizes the physical margin of separation between different categorical classes; in regression, it establishes an ϵ-insensitive tube around the training data, ensuring that errors smaller than a specified threshold are ignored while penalizing deviations outside this envelope.

Its strong performance in complex binary classification and regression problems has made it a popular choice in various predictive modeling applications. One of SVM's most powerful attributes is the "kernel trick," which map-transforms non-linear, inseparable input data into a higher-dimensional space where it becomes linearly separable, utilizing mathematical kernel functions such as Radial Basis Functions (RBF) or polynomials.

The main advantages of SVM include its strong, reliable performance in high-dimensional spaces, its exceptional ability to capture intricate non-linear patterns through these flexible kernel functions, and its notable robustness to outliers due to its reliance on a localized subset of training points called support vectors. Furthermore, it inherently supports both binary and multiclass classification tasks while maintaining highly efficient memory usage during execution. These characteristics collectively make SVM an exceptionally popular and dependable choice for a wide range of complex classification, regression, and time series prediction tasks.

3.3. Comparative framework: machine learning vs. statistics

Although machine learning and statistics are both fundamentally data-driven fields, they serve different primary purposes and rely on distinct analytical approaches. Statistics is historically and primarily concerned with understanding internal relationships, calculating confidence intervals, and drawing rigorous mathematical inferences from data samples. It prioritizes parameters that describe the data-generating mechanism.

Conversely, machine learning focuses primarily on developing empirical models that can autonomously learn intricate patterns from massive training data and maximize out-of-sample predictive accuracy. As a result, the two fields differ substantially in their underlying methods, core objectives, and primary areas of application. Statistics thrives in data-scarce scenarios requiring deep theoretical validation and causality, while machine learning excels in data-rich, complex environments where absolute predictive precision takes precedence over structural simplicity.

4. Deep learning methods

Deep learning represents an advanced evolution of machine learning, relying on multi-layered artificial neural networks to automatically learn hierarchical feature representations from raw data without manual feature engineering.

4.1. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a highly specialized type of Recurrent Neural Network (RNN) specifically engineered to capture long-term dependencies and sequential patterns in complex

DOI: 10.54254/2755-2721/2026.CH36590

temporal data. First proposed by Hochreiter and Schmidhuber in 1997, LSTM has since ascended to become one of the most widely deployed deep learning models globally due to its exceptional performance across a vast variety of sequence-driven applications [2].

Traditional RNNs often struggle to retain critical information over long sequences due to the mathematical phenomenon known as the vanishing or exploding gradient problem; as the sequence lengthens, backpropagated error signals exponentially diminish, blinding the network to past events. Unlike these traditional architectures, LSTM is specifically designed to preserve important information over extended periods, enabling highly effective learning from intricate time series and sequential datasets.

LSTM achieves this memory retention through a sophisticated, gated structural architecture consisting of three primary operational components that systematically regulate the flow of information. First, the forget gate determines what percentage of information from the previous cell state should be discarded or retained by evaluating the new incoming inputs and the prior hidden state. Following this, the input gate evaluates the new contextual information to decide which pieces of data are valuable enough to be written into the evolving cell state. Finally, the output gate controls the updated internal information, regulating the precise amount of data that should be passed out as the hidden state to influence subsequent cells in the sequence [3].

By continuously regulating this internal cell state via these mathematical gates, LSTMs can selectively remember or forget information across hundreds of time steps, making them incredibly powerful for complex time series forecasting, financial market analysis, and natural language processing.

4.2. Hybrid time series-ML models

A notable frontier within modern forecasting is the implementation of hybrid time series-ML models, which seek to combine the unique strengths of different paradigms. Real-world time series data rarely present themselves as purely linear or purely non-linear; instead, they are complex compositions of both structures. For instance, a hybrid ARIMA-LSTM framework utilizes a two-stage modeling approach: first, a linear ARIMA model captures the structured, seasonal, and trend-driven linear components of the data; second, the remaining non-linear residuals are passed into a deep LSTM network to extract complex, volatile interactions [4]. By decomposing the forecasting task into linear and non-linear subproblems, hybrid models have demonstrated significantly lower error rates and enhanced generalization capabilities compared to single-model frameworks across diverse empirical studies [4].

5. Challenges and future directions

5.1. Inherent limitations and challenges of forecasting methods

Despite their distinct analytical capabilities, traditional statistical, machine learning, and deep learning paradigms all encounter severe operational limitations in real-world applications. Traditional statistical methods rely heavily on the assumptions of linearity and data stationarity, making them structurally incapable of adapting to abrupt external shocks or managing severe outliers. Furthermore, their performance degrades significantly when forecasting for new markets or products due to an absolute scarcity of historical data.

While machine learning approaches successfully relax these rigid statistical assumptions to capture multi-dimensional patterns, they introduces new systemic bottlenecks. Their predictive

DOI: 10.54254/2755-2721/2026.CH36590

accuracy depends completely on continuous access to massive, high-quality training datasets, and they generally lack the capacity for causal inference. Moreover, most machine learning models suffer from low interpretability, operating as non-transparent "black boxes" that are difficult to explain to regulatory stakeholders.

Deep learning methods, particularly specialized sequential architectures like LSTM, offer state-of-the-art accuracy but demand immense computational infrastructure and energy consumption, which drives up implementation costs. Similar to machine learning, deep learning suffers from a complete lack of transparency and a strong propensity for overfitting training noise. Consequently, these models often struggle to generalize when deployed against out-of-distribution scenarios, and their widespread integration remains heavily restricted by lingering concerns regarding data privacy, cybersecurity vulnerabilities, and automated algorithmic bias.

5.2. Future research directions

To overcome these limitations, future research in forecasting is rapidly converging on several transformative paradigms designed to bridge the gap between empirical accuracy and operational viability. A primary focus is the advancement of Explainable Artificial Intelligence (XAI), which involves developing specialized frameworks like SHAP (Shapley Additive exPlanations) and LIME to peer into the "black box" of deep neural networks, thereby providing human-readable explanations for complex forecasts to satisfy strict regulatory requirements. Concurrently, researchers are exploring advanced hybrid frameworks that pursue a deep integration between physics-based models or econometric theory and deep neural networks, effectively forcing data-driven models to respect physical conservation laws or economic equilibria. The domain is also being democratized through Automated Machine Learning (AutoML), which streamlines advanced forecasting by automating the arduous processes of data preprocessing, feature engineering, network architecture selection, and hyperparameter optimization. Finally, the emergence of transfer learning and time series foundation models allows institutions to train massive models on generalized datasets and fine-tune them on data-scarce target domains with minimal training samples, directly addressing the traditional data dependencies of deep learning in niche markets.

6. Conclusion

In conclusion, traditional statistical methods, machine learning approaches, and deep learning methods all play critical, complementary roles in the modern forecasting landscape. Classical statistical methods offer unmatched interpretability and ease of implementation, making them ideal for stable, linear environments, though they struggle significantly when confronted with complex, non-linear real-world data. Machine learning models take a substantial step forward by improving absolute prediction accuracy and capturing intricate, non-linear multi-variable patterns, although they demand high-quality datasets and often lack operational transparency. Deep learning methods, particularly specialized architectures like LSTM, have achieved truly impressive results in managing highly complex forecasting tasks, especially when massive, high-frequency datasets are readily accessible; however, they remain computationally expensive, prone to overfitting, and notoriously difficult to interpret.

Overall, each analytical approach possesses its own unique strengths and structural limitations, and the selection of a specific forecasting methodology must depend heavily on the unique context of the task, the volume and quality of available data, computational constraints, and the level of transparency required by stakeholders. Future developments are likely to focus intensely on

DOI: 10.54254/2755-2721/2026.CH36590

combining these disparate forecasting techniques through robust hybrid models, systematically enhancing model interpretability through Explainable AI, and engineering highly automated, efficient, and universally accurate forecasting systems capable of navigating an increasingly complex world.

References

[1] Namin, A. S., & Namin, A. S. (2018). Forecasting time series data using ARIMA and LSTM deep learning models. *International Journal of Computer Science and Information Security (IJCSIS), 16*(4), 112-119.
[2] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.
[3] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *International Journal of Forecasting, 34*(4), 545-562.
[4] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159-175.
[5] Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. *Business Intelligence: Second European Summer School, eBISS 2012*, 62-77.