CITE PUBLICATIONS

Journal of Analytical and Applied Computer Sciences
Volume: 2, Issue: 2, 2026 (July - December), Pages: 1-12
ISSN (Online): 3139-3349 | DOI: https://doi.org/10.67914/jaacs/ra/2.2/1-12
Review Article

Machine Learning in Time Series

Sura Mohamed Jamalalden Hussein[1], Fatimah Abdulrazzaq Alsultan[1], Saif Ramzi Ahmed[2], Heyam A.A. Hayawi[1*]

[1]Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq.
[2]Ministry of Planning, Authority of Statistics & Geographic Information Systems, Nineveh Statistics Office, Mosul, Iraq.

[*]**Correspondence**
Heyam A.A. Hayawi
he.hayawi@uomosul.edu.iq

Received: 1 June 2026 / Accepted: 30 August 2026 / Published: 15 September 2026

The field of time series analysis is rapidly developing as a result of the significant expansion of data production and the diversity of its applications in the fields of economics, finance, energy, healthcare, Meteorology, transport and other areas that rely on accurate forecasting to support decision-making. Traditional studies have relied for many years on statistical models, such as autoregressive models, moving averages (ARIMA) and exponential bootstrapping models, but these models face challenges when dealing with data with nonlinear relationships and complex time patterns. In contrast, machine learning techniques have emerged as advanced tools capable of detecting underlying patterns and extracting complex relationships from data, which has contributed to improving prediction accuracy in many applications. This study aims to provide a comprehensive scientific review of machine learning applications in time series analysis and prediction, by reviewing the theoretical foundations of this field, and discussing the most prominent machine learning algorithms used, including decision trees, random forests, supporting vector machines, artificial neural networks, deep learning models such as long-memory neural networks (LSTM), repeating gate modules (GRU), bypass neural networks (CNN), and transformer models (Transformers). The study also deals with data processing methods, feature extraction, performance evaluation metrics, and the most important challenges associated with data instability, missing values, interpretability of models, and high computer requirements. The study also reviews the results of recent research that compared machine learning models with traditional statistical models, showing the advantages of each and the limits of its use according to the nature of the data and forecasting objectives. Recent trends in the development of hybrid models combining statistical methods and machine learning technologies with the aim of improving the accuracy of forecasting and increasing the reliability of results are also discussed. The study concludes by identifying the most prominent research gaps and proposing a number of future directions, such as explainable artificial intelligence, automated machine learning (AutoML), transfer learning, probabilistic forecasting, and foundational models of time series, which contributes to guiding future research and developing more efficient and effective models in time prediction applications.

Keywords: *machine learning, time series, Time Series Prediction, deep learning, artificial neural networks, long-memory neural networks, transformer models, hybrid models, temporal data analysis, scientific review*

Introduction

Time Series are one of the most important types of data used by scientists and practitioners to study time-changing phenomena. They are used in Economics and finance, Meteorology, Energy, Health Care, Transport and other areas where future values need to be forecasted in order to make decision. Models such as autoregressive models, moving averages

(ARIMA) and exponential bootstrapping models have been employed in the past. decades in the analysis of time series and have proven remarkable efficiency when dealing with linear data and relatively stable patterns. But, due to rapid growth in data volume and complexity, the emergence of nonlinear relationships and multiple important variables, these models were not well suited to capture the complex dynamical processes typical of many real-world phenomena. But machine learning has become one of the most popular recent trends in data analysis and forecasting due to its ability to learn from data, as well as identify hidden relationships among variables, without being dependent on statistical assumptions about relations between variables. Decision trees, random forests, support vector machines, artificial neural networks, and deep learning models have demonstrated great efficiency in handling large amounts of data as well as dealing with nonlinear relationships and interaction within time series. As computing power and data resources continue to improve, these models have become a hot topic in forecasting research. In the past few years, more and more studies have been conducted comparing traditional statistical models with machine learning approaches. In many applications, these models outperform traditional models in many cases, especially for nonlinear or multivariate data. In contrast, some experiments have shown that traditional models remain competitive in some applications, where the application is linear and stable. So, to improve prediction accuracy and interpretability, hybrid models have been developed. In light of these developments, the aim of this review is to provide a scientific overview of the literature on machine learning applied to time series with an overview of the theoretical background, the most popular algorithms, their applications, comparison with classical statistics, and the recent trends and challenges. Our goal is to provide an integrated view of the current state of this field, areas for research, and directions for future developments to build more efficient and accurate models for time series prediction.

1. Literature Review - Machine Learning in Time Series

These time series analysis methods began to be applied to traditional statistical models that sought to describe the structure of the data and predict future values. These models, such as autoregressive (AR), autoregressive and moving averages (ARMA), integrated autoregressive and moving averages (ARIMA), have formed the theoretical basis for analyzing time series for many decades, because of their ability to represent linear and stable data, but their efficiency was limited when dealing with nonlinear relationships and complex time patterns. In 1997, Hochreiter and Schmidhuber introduced the long-Term memory neural network model (LSTM), which marked a turning point in sequential data processing, addressing the gradient fading problem that was limiting the performance of traditional repetitive neural networks, and later became one of the most widely used models in predicting time series. With the expansion of the use of deep learning, Cho and colleagues (2014) introduced the Gated Recurrent Unit: GRU module, which provided a simpler architecture than LSTM while maintaining a high capacity in representing long-term time dependence, making it a suitable option in many forecasting applications that require speed in training and efficiency in performance. In 2015, the fifth edition of the book by Box, Jenkins, reinsel and Ljung was published, which is the basic reference in the modeling of statistical time series, where it provided an integrated presentation of ARIMA models and methods for determining, diagnosing and predicting the use of the model, and still forms the basis on which modern models are compared in most Applied Studies. In 2016, Goodfellow, Bengio and Courville published their book on deep learning, which contributed to the consolidation of theoretical concepts of modern neural networks and their applications in various types of data, including time series, and paved the way for the development of more complex models capable of extracting nonlinear temporal patterns. With the increasing volume and complexity of data, the subsequent years saw an expansion in the use of neural bypass networks (CNN) and hybrid deep learning models combining CNN and LSTM, as several studies have shown that these models achieve higher levels of accuracy compared to traditional models in Energy, Finance, Meteorology and healthcare applications. In 2021, Hyndman and Athanasopoulos presented the third edition of Forecasting: Principles and Practice, which reviewed recent developments in time series forecasting, explaining the advantages of traditional statistical models and their limitations compared to modern trends in machine learning and deep learning, emphasizing the importance of choosing the appropriate model according to the characteristics of the data.

In 2022, in the third edition of his book Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow discussed practical applications of machine learning and deep learning algorithms, explaining how to build predictive models using decision trees, random forests, supporting vector machines, neural networks, and their use in processing temporal data.
During the period 2023-2026, recent studies have focused on the development of more advanced models based on the attention Mechanism and the structure of transformers, where specialized models such as Informer, autoformer, fedformer and patchst have appeared, which have shown superior performance in predicting long-term time series compared to traditional recurrent networks. The use of hybrid models that combine statistical models and machine learning increased. Explainable artificial intelligence (Explainable AI), transfer Learning (Transfer Learning), automated machine learning (AutoML), foundation models (Foundation Models) were also increasing in interest during this period, with the goal of increasing prediction accuracy, interpretability of models, and adaptability of models to dynamic and always changing data. In general, according to the chronology of the studies, the field of machine learning in time series has moved from relying on traditional statistical models to incorporating machine learning. algorithms, then to deep learning models, up

to modern transformer models and hybrid models, which reflects the continuous development of forecasting techniques and their ability to handle complex temporal data in various applied fields.

2. Classical Foundations and Transition to Machine Learning

Time series analysis studies have relied on traditional statistical models that have characterized the structure of time series data and utilized the correlation of successive observations to predict future outcomes for decades. These models included autoregressive (AR), moving Average (MA), compact (ARMA) models, and then the most widely applied model, the autoregressive model with moving averages (ARIMA). These models have the ability to predict the linear relationship between the past and current values of the time series by considering statistical assumptions of stability (Stability) and linearity of relationships between variables (Box et al., 2015).

Due to the wide variety of practical uses and the vast number of enormous data sets, it has been discovered that many temporal phenomena can no longer be accurately represented by just linear models. For example, financial markets, energy, climate, medicine, and user behavior on digital platforms are all interconnected, structurally diverse, and may have a long-term time dependence, making traditional statistical models in many modern applications less useful. With the rapid pace of computing technology and increasing ability to process Big Data, Machine Learning has evolved as a new way of learning directly from the data rather than building a model based on certain assumptions. Such a technique can detect hidden patterns and relationships without needing to define the function formula of the model at the outset. this made it suitable for dealing with time series with nonlinear or continuously changing behavior (Goodfellow et al., 2016). Machine learning applications in time series began using traditional algorithms such as Decision Trees, support vector machines, and random Forests, before evolving towards the use of artificial neural networks, which have shown a greater ability to represent nonlinear relationships. As deep learning continued to advance, more specialized models of sequential data processing emerged, such as long-memory neural networks (LSTM) and repetitive gateway modules (GRU), which were able to overcome many of the problems that faced traditional neural networks, especially with regard to the representation of long-term temporal dependence (Hochreiter & Schmidhuber, 1997; Cho et al., 2014). In recent years, this field has developed further with the advent of Transformers models that rely on the self-Attention mechanism, as these models have shown a high ability to capture long-term temporal relationships with reduced training time compared to repetitive networks. This has led to the development of specialized models such as Informer, autoformer and patchst, which have achieved advanced results in many long-term forecasting applications (Zhou et al., 2021; Wu et al., 2021; Nie et al., 2023). This move from statistical models based on assumptions to machine learning models based on data, and finally to deep learning models based on complex temporal patterns, is an evolution of statistical models from those based on assumptions towards models based on data. In spite of this progress, the literature on statistics in recent years suggests that traditional statistical models have not gone away completely and are still being used in many applications, even as a component of hybrid models that combine the best of both methods to improve prediction accuracy and predictability (Hyndman & Athanasopoulos, 2021).

3. Machine Learning Models for Time series Forecasting

Many time series models are becoming increasingly advanced in the field of timeseries forecasting. Machine Learning models are learning from data to derive patterns and relationships that are hard to make out using standard statistical models. While traditional time series models build predictions on certain assumptions about the nature of the data, machine learning models are focused primarily on the data itself to derive predictions. This makes them more adaptable to nonlinear, variable, and high-dimensional time series (Goodfellow et al., 2016; Géron, 2022). Machine learning models for predicting time series can be classified on three main axes: traditional machine learning algorithms, neural network models, and modern deep learning models.

Traditional machine learning algorithms represent the first generation of models that have been used to predict time series, most notably decision Trees, random forests, support vector machines, nearest neighbor algorithms (K-Nearest Neighbors), and Gradient boosting algorithms such as XGBoost and LightGBM. These models have demonstrated good efficiency in processing multivariate data, are characterized by training speed and the possibility of dealing with non-linear relationships, but their performance largely depends on data quality and feature engineering processes, because they do not directly benefit from the sequential temporal structure of data (Breiman, 2001; Chen & Guestrin, 2016). With the development of artificial intelligence technologies, artificial neural networks have become one of the most widely used models for predicting time series, due to their ability to approximate complex nonlinear relationships between variables. These networks have been used in many applications of economics, finance, energy and medicine, but traditional networks have had difficulties retaining information over long periods of time as a result of the gradient fading problem, which limited their efficiency in long-term time series (Goodfellow et al., 2016). To address this problem, I developed recurrent neural networks (Recurrent Neural Networks: RNN), which are designed to process sequential data by transferring information between successive time steps. Although these networks improved the representation of temporal dependence, they continued to suffer from long-term learning difficulties, which led to the development of the long-term memory neural network model (LSTM), introduced by Hochreiter and Schmidhuber in 1997, which relies on a set of gates that allow the model to retain important information for long periods of time, ignoring unnecessary information. The LSTM model has proven highly efficient in financial forecasting, energy consumption, demand, climate data, and meteorological applications. In 2014, Cho and colleagues introduced the Gated Recurrent Unit: GRU, an LSTM-like simplified version with fewer gates and parameters in order to reduce training time and increase accuracy in many applications. Previous work has shown that GRU performs similarly to LSTM on many issues related to forecasting, although it is more efficient in some cases (Cho et al., 2014).

Convolutional Neural networks (CNN) also show strong ability to extract local structure and short-term temporal characteristics. These networks were developed primarily for image processing but have shown good performance in time series, especially when used together with LSTM or GRU networks that take advantage of CNN's ability to extract feature sets, and of the recurrent networks' ability to represent time dependence. Recent years have witnessed a remarkable development with the advent of transformer models, based on the self-Attention mechanism, which enabled models to learn long-range temporal relationships without relying on the traditional repetitive structure. These models can run parallel computations which reduces training time and enables them to perform better on big data. In particular, time series models like Informer, autoformer, fedformer and patchst show good performance in recent tests, particularly when studying long-term forecasts (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022; Nie et al., 2023). In addition, recent studies have tended to develop hybrid models that combine traditional statistical models with machine learning or deep learning algorithms, with the aim of taking advantage of the ability of statistical models to represent linear components, while using machine learning to model nonlinear patterns. These models have shown a marked improvement in prediction accuracy compared to using either method alone, especially in economic, financial and energy chain applications. Despite the successes of machine learning models, the choice of the optimal model depends on the characteristics of the data, their size, prediction horizon, available computer resources, as well as the need to interpret the results. Therefore, modern literature recommends not relying on one model in all applications, but making systematic comparisons between different models and using techniques of optimal parameter tuning, taking advantage of hybrid models and automated machine learning (AutoML) methods to obtain the best predictive performance.

3.1 Shallow Learning Methods

Traditional machine learning methods (Shallow Learning Methods) represent the first generation of machine learning algorithms used in predicting time series, and they appeared with the aim of overcoming some of the limitations facing traditional statistical models, in particular in dealing with non-linear relationships between variables. These methods rely on learning patterns from data using one or a limited number of layers, without the need for deep architectures that characterize deep learning models (Goodfellow et al., 2016). The most notable models belonging to this category are artificial neural networks (Artificial Neural Networks: ANN), support Vector machines (Support Vector Machines: SVM), the nearest neighbor algorithm (K-Nearest Neighbors: K-NN), as well as decision Trees and random Forests. These models have proven good efficiency in many time series forecasting applications, thanks to their ability to represent nonlinear relationships and improve prediction accuracy compared to traditional statistical models in some applications (Vapnik, 1995; Breiman, 2001). Support vector Regression (SVR) prediction machines have received wide attention in the literature, due to their strong mathematical foundation and their ability to achieve a balance between prediction accuracy and generalizability, reducing the likelihood of overfitting, especially when dealing with small or medium-sized data sets (Vapnik, 1995). In contrast, random forests and decision tree algorithms have shown good ability to process multivariate data and handle complex relationships, in addition to their relative resistance to noise and anomalous values (Breiman, 2001). Despite the advantages offered by these models, they rely to a large extent on feature engineering processes and the selection of appropriate variables before building the model, and their ability to represent long-term temporal dependence remains limited compared to deep learning models. For this reason, recent research has turned to the development of more advanced models that can directly learn from temporal data without the need for significant intervention in the design of characteristics, which paved the way for the emergence of deep learning models currently used in predicting time series (Goodfellow et al., 2016; Géron, 2022).

3.2 Ensemble and Hybrid Models

Combinatorial (Ensemble Models) and hybrid (Hybrid Models) appeared with the aim of improving the performance of time series forecasting models and overcoming the limitations associated with the use of an individual model, by taking advantage of the strengths of different models (Dietterich, 2000). Combinatorial models rely on combining the outputs of a number of basic learners to produce a more stable and accurate model, which contributes to reducing variability, reducing the risk of overfitting, and improving generalizability when dealing with complex data (Breiman, 2001). The most prominent examples of such models are random Forest, which relies on the integration of a large number of decision trees, and Gradient boosting algorithms, such as XGBoost and LightGBM, which have proven highly efficient in many forecasting applications due to their ability to gradually improve performance by correcting the errors of previous models (Chen & Guestrin, 2016; Ke et al., 2017). Hybrid models combine traditional statistical models with machine learning or deep learning algorithms to combine the benefits of both methods. For example, the model ARIMA uses statistics to represent linear components of the time series and machine learning algorithms or neural networks are used to model nonlinear relationships and residual patterns that the statistical model cannot explain (Zhang, 2003). Zhang's ARIMA–ANN model is one of the first and most influential hybrid models, proving that combining linear and nonlinear models can lead to a significant improvement in prediction accuracy compared to using either model alone. In recent years, advanced hybrid models based on deep neural networks like LSTM and CNN, optimization algorithms, and signal analysis techniques like Wavelet Transform have begun to be studied to improve the ability of the models to model complex temporal behavior and handle data with high variability and noise. Numerous recent studies indicate that combinatorial and hybrid models perform better in applications such as time series forecasts for financial markets, energy consumption, demand, Meteorology, and health. Combinatorial and hybrid models have become one of the fastest growing areas of research in the field of time series forecasting (Lim & Zohren, 2021; Géron, 2022).

3.3 Deep learning Approaches

The arrival of deep learning marked a new era in time series forecasting because it has made it possible to create models that can learn complex characteristics and patterns from data without any feature engineering. Deep learning utilizes multilayer neural networks capable of learning nonlinear relationships and extracting complex temporal patterns, which provides greater accuracy than the many other machine learning algorithms, especially for large, multidimensional data (LeCun et al., 2015; Goodfellow et al., 2016). One of the early models used in this field was recurrent neural networks (RNNs), which are based on a sequential data set processed by passing information between consecutive time steps, and thus better represent time dependence than conventional neural networks. But, recurrent neural networks struggled with vanishing Gradient, which limited their ability to learn a long-term relationship between time and space (Goodfellow et al., 2016). To overcome this problem, Hochreiter and Schmidhuber (1997) introduced the model of long-term memory neural networks (Long Short-Term Memory: LSTM), which is based on a set of structured portals for the flow of information within the network, allowing to retain important information for long periods of time and discard unnecessary information. LSTM models have proven highly efficient in time series forecasting applications, such as financial markets, electrical loads, climate data, and healthcare (Hochreiter & Schmidhuber, 1997). As part of the development of iterative models, Cho and colleagues (2014) proposed the Gated Recurrent unit: GRU module, which is a simplified version of the LSTM model, relying on fewer parameters and Gates, which contributes to reducing training time while maintaining a high level of prediction accuracy in many temporal applications (Cho et al., 2014).

Convolutional Neural networks (CNN) have also emerged as an effective tool for extracting local characteristics and short-term patterns from time series, although they were developed mainly for image processing. Their combination with LSTM and GRU models has led to the development of hybrid models capable of taking advantage of the advantages of each model, where CNN networks extract characteristics, while iterative networks represent time dependence, which was reflected in improved prediction performance in many practical applications (LeCun et al., 2015; Lim & Zohren, 2021). Recent literature indicates that deep learning models have become the basis from which new generations of forecasting models have been launched, in particular, models based on the structure of transformers, which were able to overcome many of the limitations associated with repetitive networks by relying on the mechanism of self-Attention, paving the way for the development of more efficient models in predicting long-term time series (Lim & Zohren, 2021).

3.4 Transformer-Based Models

The field of time series forecasting has undergone significant development with the advent of models based on the structure of Transformers, proposed by Vaswani and colleagues (2017), which rely on the mechanism of self-Attention to represent the relationships between the elements of the time series without the need for sequential processing adopted by repetitive neural networks. This approach allowed the model to capture long-term time dependence very efficiently, with the possibility of performing calculations in parallel, which reduced training time and improved performance when working with large data sets (Vaswani et al., 2017). In light of the increasing interest in the use of transducer structure in time series, several models have been developed to address the problems of long-term forecasting. The Informer model is one of the first models that used the ProbSparse Self-Attention mechanism to reduce the computational complexity and still keep accurate predictions for a long time series (Zhou et al., 2021). Wu and co-workers (2021) proposed an Autoformer model based on the analysis of time series components and the self-correlation (Auto-Correlation) to capture cyclical trends and improve the performance of long-range forecasting. He also developed the FEDformer model by integrating frequency analysis into the converter structure. It exploited the transformation of data in the frequency domain to extract temporal patterns at lower computational complexity and improve prediction accuracy in many standard datasets (Zhou et al., 2022). Nie and colleagues later developed the FEDformer model where frequency analysis is integrated with the structure of converters. They utilized the transform of data in the frequency domain to improve the recovery of temporal trends and minimize computation complexity. (2023) The patchst model (distributing the data in Patches into separate time series) was developed using the transformer model, which helped in better representing time and in learning more efficiently especially in long-term forecasts. Comparative tests have shown that transformer-based models perform better than many traditional deep learning models (e.g. LSTM and GRU) in the prediction of long-term time series such as energy consumption, financial market, climate data, and intelligent transportation because they are better able to model complex time relationships and take advantage of parallel computing (Lim & Zohren, 2021).

In recent years, research has focused on the design of framework models for time series trained on huge quantities of time data and then re-used in various applications. One of these is TimeGPT, which is one of the most prominent models in this area. Time GPT provides a generic model that can predict many time series without requiring complete retraining for each application, offering an exciting approach to the future of artificial intelligence forecasting (Garza & Mergenthaler-Canseco, 2023).

3.5 Comparison of machine learning models

The machine learning models used to predict time series differ in their mathematical structure, learning methodology, and their ability to represent time patterns. Therefore, the right model will depend on the characteristics of the data, the nature of the problem, and the goals of prediction (Hyndman & Athanasopoulos, 2021). Classical machine learning models such as support vector Machines (SVM), random forest, and gradient boosting models such as XGBoost and Light GBM are easy to train and require less computation than deep models. They are better at handling small or medium size data sets. But, these models are highly dependent on feature Engineering and have limited ability to represent long-term changes in time (Breiman, 2001; Chen & Guestrin, 2016; Géron, 2022). Deep learning models are not easy to contrast. They can learn from data, such as recurrent neural networks (RNN), long-memory neural networks (LSTM), or recurrent gateway modules (GRU), and they can draw complex relationships over time without any property design. These models can also represent nonlinear structures and time dependence over time. It is possible to use these models for applications that require high accuracy for prediction, especially when big data are available. But, they consume more computing resources, take longer training periods, and their mechanism is still difficult to explain in detail than traditional models (Hochreiter & Schmidhuber, 1997; Cho et al., 2014; Goodfellow et al., 2016). As for transformer-Based models, they represent a new generation of time series forecasting models, as they rely on the self-Attention mechanism, which enables them to capture long-term time relationships very efficiently while taking advantage of parallel computing, which contributes to reducing training time and improving performance when dealing with big data (Vaswani et al., 2017). Models such as Informer, autoformer, fedformer and patchst have shown superior results in many long-term forecasting applications compared to LSTM and GRU models, but these models usually require large training data and high computing capabilities to achieve the best performance (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022; Nie et al., 2023). On the other hand, combinatorial and hybrid models represent a trend that combines the advantages of more than one model, taking advantage of the ability of statistical models to represent linear components, and the flexibility of machine learning or deep learning algorithms in modeling non-linear relationships. Studies have shown that the integration of more than one model contributes to improving the accuracy of forecasting and increasing the stability of results, especially in time series characterized by the presence of linear and non-linear patterns at the same time (Zhang, 2003; Lim & Zohren, 2021).

In general, the modern literature confirms that there is no model that can be considered the best in all time series forecasting applications, as the performance of models is influenced by multiple factors, the most important of which are the volume of data, the characteristics of the time series, the forecast horizon and the available computer resources. Study authors recommend comparing the different models using standard methods to evaluate their performance, such as the Mean Absolute Error (MAE), the square Root of the Mean Square of the Error (RMSE), and the Mean Absolute Percentage Error (MAPE). They then recommend choosing the model that achieves the best balance between prediction accuracy, computational efficiency, and interpretability (Makridakis et al., 2022; Hyndman & Athanasopoulos, 2021).

3.6 Criteria for Selecting an Appropriate model

Choosing the right machine learning model is one of the key steps in building time series forecasting systems, as the literature confirms that there is no single model that can achieve the best performance in all applications, but rather the success of the model depends on its compatibility with the characteristics of the data, the nature of the problem and the prediction goals (Hyndman & Athanasopoulos, 2021). Therefore, the selection process is based on a set of scientific and applied criteria that should be taken into account before building and evaluating the model. The analysis of the characteristics of the time series is one of the most important of these criteria, as it is necessary to determine whether the series is linear or nonlinear, stable or unstable, as well as to study the general trend (Trend), seasonality (Seasonality), cyclic Behavior (Cyclic Behavior), anomalous values (Outliers), the complexity of time patterns, because these characteristics directly affect the relevance of the model and its predictive accuracy (Box et al., 2015; Hyndman & Athanasopoulos, 2021). Data volume and quality are also a decisive factor in model selection, as deep learning models, such as LSTM and transformer-based models (Transformers), require large amounts of data to achieve high performance, while traditional machine learning methods, such as SVM and random Forest, can achieve good results when dealing with small or medium data sets (Goodfellow et al., 2016; Géron, 2022). The efficiency of models varies depending on the target time period, where LSTM models and converters have proven superiority in long-term forecasting thanks to their ability to represent long-term time dependence (Lim & Zohren, 2021). An important criterion is also the available computing resources, as deep models and transformer-based models require higher processing and memory capabilities compared to traditional models, which may limit their use in environments with limited capabilities. The interpretability of the model should also be taken into account, especially in applications that require decision support, such as healthcare and the financial sector, where the possibility of interpreting the results acquires no less importance than the accuracy of the forecast (Géron, 2022).

Moreover, the performance of the models should be evaluated using validated quantitative measures, such as Mean Absolute Error (MAE), Mean Square Error (MSE), square root of mean square error (RMSE), Mean Absolute Percentage Error (MAPE), as well as determination coefficient (R2) when needed, to ensure an objective comparison between different models (Hyndman & Athanasopoulos, 2021). Recent studies also recommend relying on independent training and testing data, applying cross-validation methods for time series (Time Series Cross-Validation), with the use of parameter tuning techniques (Hyperparameter Optimization) to improve model performance and reduce the likelihood of overfitting (Lim & Zohren, 2021; Géron, 2022). Accordingly, the choice of the optimal model depends not only on the level of predictive accuracy, but also represents an integrated process that takes into account the characteristics of the time series, the volume of data, computer resources, interpretability and the results of experimental evaluation, ensuring the construction of an efficient and reliable model that meets the requirements of practical application.

4. Applications Across Domains

Machine Learning has become one of the most widely used techniques for analyzing and predicting time series because of its ability to manipulate complex data and extract temporal patterns that are hard to represent with traditional statistical models. Machine learning and deep learning have been successfully applied to many scientific, economic, and engineering fields because they have shown high levels of accuracy and efficiency in processing temporal data. (Lim & Zohren, 2021). Machine learning models are used in the financial field to predict stock prices, exchange rates, commodity prices, financial market volatility, and risk-management. Models such as LSTM, GRU, and transformers have been shown to better represent time relationships than traditional forecasting techniques, especially in markets with high volatility and uncertainty (Sezer et al., 2020). In the energy sector, machine learning models are important for predicting electrical loads, energy consumption, and the production of renewable energy such as solar and wind energy. They help improve the efficiency of power grids, reduce resources consumed, and aid planning and operations. Deep learning models and converters have shown remarkable results in this domain. (Wang et al., 2019; Lim & Zohren, 2021). Machine learning models are used to predict temperature, rainfall, wind speed, air quality, and extreme weather. Because it can analyze large scale data on climate and predict seasonal trends and time scales, it is possible to improve the accuracy of climate predictions and early warnings. (Rolnick et al., 2022). Health has also seen a significant increase in the use of machine learning models to analyze medical time series such as ECG (ECG) and EEG (EEG), to monitor the vital signs of patients, and to predict the onset of diseases and epidemics. Machine learning models are able to improve diagnostic accuracy and medical decision-making, especially when enacted with deep learning algorithms (Esteva et al., 2019).

Machine learning models for predictive Maintenance in industry enable us to monitor the performance of equipment, detect problems before they happen, and improve the quality of production by mining all of the continuous data collected by industrial sensors. Through this process, we reduce the maintenance costs, keep our industrial equipment reliable, and reduce operating costs. (Zonta et al., 2020). Applications of machine learning have extended to smart transportation, telecommunications, smart agriculture and supply chain management, where forecasting models are used to predict traffic, demand, manage resources, and improve decision making. Recent research suggests that the continued development of deep learning models and converters, in combination with big data, will lead to the use of these models in new areas and to improved accuracy and reliability in time series predictions (Lim & Zohren, 2021).

5. Critical Comparison of Methods

Time series forecasting has been a rapidly growing area over the last few decades and has evolved from simple statistical models to models based on deep learning and models based on transformers. These models are very different in their accuracy, but there seems to be no single model that is better than any other for all cases. Performance is highly dependent on the data itself, its size, its prediction window, and the computer resources available. (Hyndman & Athanasopoulos, 2021). Traditional statistical models, such as ARIMA or wets, offer easy to interpret results and low computational requirements. They are also effective in linear and stable time series. Their ability to represent nonlinear relationships and interactions is limited and may not be efficient in big data, time series with nonlinear behavior, etc. (Box et al., 2015; Hyndman & Athanasopoulos, 2021). In contrast, traditional machine learning models, such as Support Vector Machines (SVM), random Forest and xgboost, have made it possible to model nonlinear relationships and achieve higher accuracy in many applications compared to statistical models. However, these models are highly dependent on feature engineering selection processes, and their ability to represent long-term temporal dependence remains limited, especially in complex time series (Breiman, 2001; Chen & Guestrin, 2016; Géron, 2022).

Deep learning models, such as RNN, LSTM and GRU, have represented a quantum leap in the field of time series forecasting, as they have the ability to directly learn from data and extract complex time patterns without the need for manual design of characteristics. These models have proven to be clearly superior in many applications, especially when dealing with big data, but their training requires more time and more computer resources, and their low interpretability is one of the most prominent challenges facing their use in sensitive applications (Hochreiter & Schmidhuber, 1997; Goodfellow et al., 2016; Lim & Zohren, 2021). With the advent of transformer-based models, the applications of time series prediction have undergone further development, as these models have been able to overcome some of the limitations associated with repetitive neural networks by relying on the self-Attention mechanism, which enables them to represent long-term time relationships more efficiently, while taking advantage of parallel computing to reduce training time. Models such as Informer, auto former, fed former and patchst have shown superior results in many standard datasets, but these models usually require large training data and high computing capabilities, and their performance may decrease if the data is limited or the quality of the time series is poor (Vaswani et al., 2017; Zhou et al., 2021; Wu et al., 2021;Nie et al., 2023). On the other hand, the results of numerous studies indicate that hybrid and combinatorial models often achieve better performance than individual ones, since they combine the advantages of statistical models with machine learning or deep learning algorithms, which helps to represent linear and nonlinear components simultaneously. However, this improvement in performance is offset by an increase in the complexity of the model, an increase in the number of parameters to be adjusted, as well as a higher computational cost and difficulty in interpreting the results compared to simpler models (Zhang, 2003; Dietterich, 2000; Lim & Zohren, 2021).

In general, the contemporary literature suggests that the trade-off between models should be made not just on prediction accuracy, but on interpretability, computational efficiency, ease of use, and generalizability of the model when working with new data. The literature also suggests performing an experiment using multiple databases and standard evaluation measures such as MAE, RMSE, and MAPE, and testing for statistical significance of differences between models such as the Diebold–Mariano test, to arrive at an objective and reliable assessment of the prediction performances of different prediction models (Makridakis et al., 2022).

6. Research Gaps and Future Directions

Despite the significant development of machine learning and deep learning models in the field of time series forecasting, the recent literature indicates that there are a number of research gaps that still limit the possibility of generalizing these models in various applications. One of the most prominent of these gaps is the absence of a unified model capable of achieving superior performance in all types of time series, as the performance of models varies depending on data characteristics, such as linearity, seasonality, instability, data volume, and the nature of noise, which makes the choice of the appropriate model largely depends on the nature of the application (Hyndman & Athanasopoulos, 2021; Lim & Zohren, 2021).

Another research gap is the limited interpretability of many deep learning models, in particular deep neural networks and transformer-based models, which are often treated as "black boxes" (Black Box Models). Despite its high predictive accuracy, the difficulty of interpreting the decision-making mechanism limits its use in areas that require transparency, such as healthcare, the financial sector, and public policymaking. Therefore, an increasing research trend has emerged towards the development of explainable Artificial Intelligence (XAI) machine learning models, which contribute to enhancing confidence in prediction results and improving their usability in practical applications (Samek et al., 2021). Studies also show that most modern models rely on the availability of large amounts of high-quality data, which is a challenge in many real-world applications that suffer from limited data, missing data, or high noise. Hence, there is a need to develop models that are more capable of learning from limited data, taking advantage of transfer Learning methods, and self-Supervised learning, allowing to improve the performance of models in environments where it is difficult to obtain sufficient data (Lim & Zohren, 2021).

One of the aspects that still needs further research is also the development of more computationally efficient models, as deep learning models and converters require large computational resources and a long training time, which limits the possibility of using them in real-time systems (Real-Time Systems) or environments with limited capabilities. Therefore, an increasing number of studies are focusing on the design of lightweight models, model Compression techniques, and improving training and inference efficiency without significantly affecting prediction accuracy (Vaswani et al., 2017; Nie et al., 2023). Another promising research direction is to expand the development of hybrid models that combine traditional statistical models, machine learning algorithms, and deep learning techniques, allowing to take advantage of the advantages of each approach in representing linear and nonlinear components of time series. There is also growing interest in the development of foundation models pre-trained on huge and diverse sets of time series, such as TimeGPT, with the aim of providing generic models adaptable to different applications using a limited amount of additional data (Garza & Mergenthaler-Canseco, 2023). Finally, time series forecasting models should be evaluated using a common set of reference data, division procedures should be standardized, and standard statistical evaluation and testing measures should be used, such as MAE, RMSE, MAPE, and Diebold–Mariano to ensure that models are directly comparable. These trends are expected to contribute to the development of models that are more accurate, more interpretable, and more efficient in dealing with the increasing challenges posed by modern time series applications (Makridakis et al., 2022; Hyndman & Athanasopoulos, 2021).

Conclusion

This study reviewed the development of machine learning applications in the field of time series forecasting, starting from traditional statistical models and moving to traditional machine learning methods, then combinatorial and hybrid models, up to deep learning models and transformer-based models. The review showed that this development has significantly contributed to improving the accuracy of forecasting, increasing the ability of models to represent nonlinear relationships and long-term time dependence, which led to the expansion of their use in many fields, such as economics, finance, energy, Meteorology, healthcare and industry. The studies examined in the review showed that the choice of the forecast model does not depend on the novelty or complexity of the model, but is largely related to the characteristics of the time series, the volume of data, their quality, the forecast horizon, and the available computer resources. While traditional statistical models remain suitable for many applications with linear behavior, machine learning and deep learning models provide greater ability to process non-linear data, while transformer-based models have proven highly efficient in predicting long-term time series, especially when large training data are available. The review also highlighted the potential of hybrid or aggregate models, which allow combining the strengths of statistical models with machine learning and deep learning techniques, for greater prediction accuracy and stability in a wide range of applications. But deep models face difficulties due to their complexity, high computation costs, dependence on large amounts of data, and lack of unified method for comparing models equally. Recent results suggest that time series forecasting will evolve towards more efficient models which make use of foundation models, transfer learning, self-supervised learning, and more integration between statistical models and modern artificial intelligence. We should follow consistent evaluation standards and use different reference data sets in order to develop more robust and generalizable models in diverse applied contexts. Overall, this review confirms that machine learning is the major driving force in the development of time series forecasting models. Machine learning will continue to accelerate its progress and improve the forecasting efficiency and decision support in a large number of scientific and economic fields. Research on how to achieve a balanced balance between accuracy, interpretation, and computational efficiency is needed in order to deliver practical solutions that can be used in real world situations.

Recommendations

Based on the contents of this study, following are some suggestions for future research and development in the area of time series forecasting using machine learning.

1.The choice of the model should be based on the nature of the time series, the size of the data, the noise level, and the desired prediction horizon.

2.Consider comparing the models using data and assessments such as MAE, RMSE, and MAPE with appropriate tests of significance, such as the Diebold-Mariano test.

3.Expanding the development of hybrid and aggregative models, which combine traditional statistical models with machine learning and deep learning because it has the potential to improve prediction accuracy and stability of results in many applications.

4.To promote the research in transformer-based models (Transformers) and foundation models (Foundation Models), and study their utility in real life applications, such as multivariate time series and big data.

5.Attention to developing explainable AI models, which will lead to more reliable predictions and usage in transparent areas such as healthcare, finance, and government.

6.Development of computationally efficient models through better training algorithms and reduced memory and energy consumption for use in real time systems and devices with limited resources.

7.Using transfer Learning and self-supervised learning to enhance models' performance in low- or no-data environments.

8.Encouragement of open, diverse, and accessible reference databases for a range of applications, which will help to standardize the evaluation procedures and allow for more objective comparisons between studies.

9.Machine learning models have become increasingly interdisciplinarized in new areas such as smart cities, water resources management, smart agriculture, precision medicine, and environmental risk prediction, which will bring greater use of these technologies.

10.Improve the links between researchers and specialists in Statistics, data science, artificial intelligence and researchers in applied fields to improve the quality of models that are accurate, realistic and practical, as well as research focused on the solving of problems with scientific and economic significance.

Overall, machine learning algorithms are constantly improving, and big data and computing resources are increasing, which will drive the development of more accurate, flexible, and generalizable models. Therefore, focusing on current problems such as interpretability, model efficiency, and standardization of assessment criteria, will be an important step toward developing more reliable and effective forecasting systems in various areas.

Acknowledgments

The authors are very grateful to the College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq, which helped improve this work's quality.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Ethical Approval

Ethical approval was not required for this study as it did not involve human participants, personal data.

AI Tool Usage Declaration: The authors should declare that no AI and associated tools are used for writing scientific content in the article.

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cho, K., Van Merriënboer, B., Gulcehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Garza, A., & Mergenthaler-Canseco, M. (2023). *TimeGPT-1* [Technical report]. Nixtla.

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow* (3rd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems, 30*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379*(2194), 20200209. https://doi.org/10.1098/rsta.2020.0209

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings and conclusions. *International Journal of Forecasting, 38*(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations (ICLR)*.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., Creutzig, F., & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys, 55*(2), 1–96. https://doi.org/10.1145/3485128

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2021). *Explainable AI: Interpreting, explaining and visualizing deep learning* (2nd ed.). Springer.

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review (2005–2019). *Applied Soft Computing, 90*, 106181. https://doi.org/10.1016/j.asoc.2020.106181

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30* (pp. 5998–6008).

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management, 198*, 111799. https://doi.org/10.1016/j.enconman.2019.111799

Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems, 34* (pp. 22419–22430).

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*.

Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering, 150*, 106889. https://doi.org/10.1016/j.cie.2020.106889