Effective Prediction on Time Series Data Using Deep Learning: An Incisive Review

[*1]Rupa Rajakumari R Peter, [2]Dr. Ujwal Ambadas Lanjewar

[a]Research Scholar, Department of Electronics and computer Science, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur.
[b]Assistant Professor, Department of Computer Science, Hislop College, Nagpur.

Professor and Principal, Binzani Mahila Mahavidyalai, Nagpur.

[*1]Corresponding author mail ID: : ruparajakumari@gmail.com

Abstract

Time-series evaluation explores the way in which data differs over time and optimal prediction could determine the pathway wherein the data seems to be changing. Prediction of time-series might possess significant value for business development when one possess the access to existing information with time-constituent. Concurrently, DL (Deep Learning) algorithms are capable of offering promising solution to predict time-series due to their advantages in automatic temporal learning. With extensive usage of DL based models in varied areas like science, one might be tangled as to determine the suitable model to resolve the existing issues. Moreover, these methods might be used by existing works in various forms like hybrid. Hence, this study is instigated by an intention to afford a precise review about the recent DL based models to predict time-series. Hence, this study reviews the recent research works (2018-2023) related to the use of DL based models for effective time-series prediction in various areas. Further, a comparatively assessment is undertaken by considering the conventional DL models and varied applications of widely employed DL models for efficient prediction of time-series. Lastly, the research gaps are emphasized from conventional studies with future recommendations. This would support future researches in resolving the prevailing challenges in this area, thereby attempt to bring innovations for effective time-series prediction.

Keywords: Time-Series Prediction, Deep Learning, Hybrid Model, Comparative Assessment

1. Introduction

Prognosticating the future using Time-Series Prediction is a crucial research area in the contemporary era in varied fields like energy consumption, anomaly detection, retail sales, traffic detection etc. Unique features of time-series data wherein observations possess chronological order make their evaluation as a challenging one. Due to this complexity, time-series prediction is a sector gaining paramount significance. The time-series prediction models require to consider varied issues like seasonal and trend alterations of series and correlation amongst the observed values which area nearer in time. Hence, over last decades, investigators have attempted to develop special frameworks which could gain underlying time-series sequences, by which, it could be inferred to future efficiently [1]. Recently, employment of DL (Deep Learning) based methods have gained significance in time-series prediction. Unlike traditional statistical based frameworks that could model the linear associations in data, the deep NNs have exposed the potential for mapping complex and non-linear feature associations. Contemporary neural systems have their success in its deep structure, dense connected neurons and various stacking layers. Enhancement in computing ability during last few years have permitted to develop deep models that possess significant enhancement in their learning ability in comparison to the shallow networks. Hence, DL based algorithms could be comprehended as the large scale optimization process. Provided with the need for accurate classification of time-series data, the researchers have endorsed varied methodologies for solving this task in varied applications.

Accordingly, the study [2] has developed a method that has employed transformer based ML models for forecasting the time-series data. This methodology has worked by leveraging a self-attention method to learn the complex dynamics and patterns from time-series data. Besides, it seems to be a generic model that could be employed to multi-variate and uni-variate time-series data for embedding time-series. Through the use of ILI (Influenza like Illness) prediction as case-study, it has been exposed that, prediction outcomes exposed by the suggested system has been better than conventional system. Following this, the research [3] has endorsed SARIMA-SVR to predict the statistical indicators in aviation sector that could be utilized for planning purpose and capacity management. Initially, time-series has been assessed by SARIMA. Subsequently, Gaussian-white noise has been reversely computed. Subsequently, 4 hybrid frameworks have been suggested and employed to determine the future indictors of aviation sector. Outcomes of experimentation has recommended that, suggested system namely SARIMA-SVR could accomplish satisfactory accuracy. It has also been confirmed that, including Gaussian-white noise possess the ability to enhance the prediction rate. Similarly, the study [4] has suggested ISSA (Improved Singular Spectrum Analysis) that has modified determination methodology of sub-series consideration in reconstructing SSA. Integrating ISSA with LSTM, a hybrid model has been developed. In addition, to assess the superiority and robustness of the suggested model, 4 historical datasets positioned in 3 climatic zones have been gathered as training datasets and testing datasets. Comparison has been undertaken with 5 models. Outcomes have exposed that, the suggested SSA possess the ability to retrieve negative values while, time-series seems to be near to zero and contributing SSA in enhancing the LSTM has been insignificant. Contrarily, ISSA has averted from generating negative-values. It has also minimized MARNE (Mean Absolute Range Normalized Error) of LSTM with range as 0.86 to 11.86%. From the considered models, the suggested ISSA-LSTM has accomplished ideal performance. Outcomes have exposed that, MARNE corresponding to tropical city seems to be high than others, that is instigated by complex consumption pattern of natural gas. However, it is important to note that the SARIMA-SVR model's potential drawbacks, such as its high computational complexity, sensitivity to parameter tuning, and limitations in handling non-stationary data, have not been adequately addressed. These challenges could significantly impact the model's scalability, robustness, and applicability in real-world, dynamic environments. Without a more thorough exploration and mitigation of these limitations, the practical use of the SARIMA-SVR model in various sectors may be restricted.

In addition, the article [5] has used a DL model integrating multiple nested LSTM to accurately predict AQI emphasized with federated learning. Performance of suggested model has been comparatively assessed with traditional ML models, hybrid and DL models. Empirical outcomes have revealed that, performance of suggested system has been better than conventional system. Similarly, conventional studies have used different DL based models for time-series prediction. However, a comprehensive view about these conventional works that have used different DL based approaches (in varied forms like hybrid) for extensive applications of time-series prediction are needed. Hence, this study intends to solve this gap by affording an incisive evaluation of all successful DL applications for time-series prediction. This literature survey comprise of works from diverse time-series prediction domains that has taken into account all the renowned DL models. The overall study is undertaken in accordance with the below objectives,

- To review the recent research works (2018-2023) related to the use of DL based models for predicting time-series in various areas.
- To comparatively assess the conventional DL models and different applications of extensively utilized DL models for efficient prediction of time-series
- To highlight the research gaps from conventional studies and afford future recommendations that would aid future researches in resolving the existing issues in this area.

1.1. Paper Organization

The paper is organized in the following way. Initially, an overview of time-series prediction is discussed in Section.2. Following this, various methodologies to predict-time series are explored in Section.3, following this, DL for efficient prognostication of time-series data are presented in Section.4. Subsequently, the applications of extensively utilized DL models for time-series prediction are described in Section.5. Then, advantages of DL for predicting time-series are given in Section.6, the research gaps and future recommendations are discussed in Section.7, while, the entire study is concluded in Section.8.

2. Time-Series Prediction: An Overview

Time-series is claimed as a data point sequence that rely on certain time. This indicates that, individual data point possess certain time-stamp allocated to it. Such data point computation is performed in a chronological manner at regular intervals. Analyzing time series is identical to EDA (Exploratory Data Analysis) in data science flow as it evaluates the association among outliers and variables. During the analysis of time-series, certain data limitations have to be regarded. General issues include generalization from single source of data, accurate identification of suitable models for data representation and complexity in attaining suitable computations. Initial phase lies in evaluating the time series to outline a model, thereby find and comprehend the underlying data sequences over time. Typically, such underlying patterns are classified into 4 elements,

- Trend: long term steady alteration in series. It involves the simple trend pattern that reveals long-term decline or growth.
- Seasonality: short term patterns which occur in a single time unit and iterate indefinitely.

➢ Cyclic element: long term data swings which might consume several years to come into reality. Such swings occur in predictable way and often leads to external economic circumstances.
➢ Noise: random alterations due to abandoned situations.

The time-series prediction happens when scientific predictions are undertaken in accordance with historical and time-stamped data. In addition, it includes the building frameworks by historical evaluation. These are utilized for making observations and determine strategic decisions. Through the analysis of data managed in past, informed decisions could be taken that could assist in business strategy. This would also assist in comprehending future trends. Vital phase while considering the prediction of time series lies in comprehending the data model, thereby understanding the business queries that have to be answered through this data. Through the leap into problem domain, developer could easily differentiate the random variations from constant and stable historical data trends. This will arrive with convenience while tuning the model for generating the ideal forecasts and taking into account the ideal forecasting methods to be used.

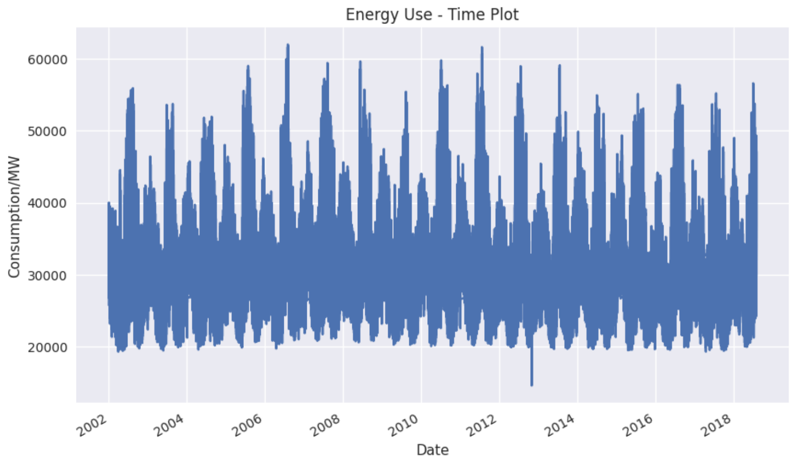

Figure.1 Energy Consumption

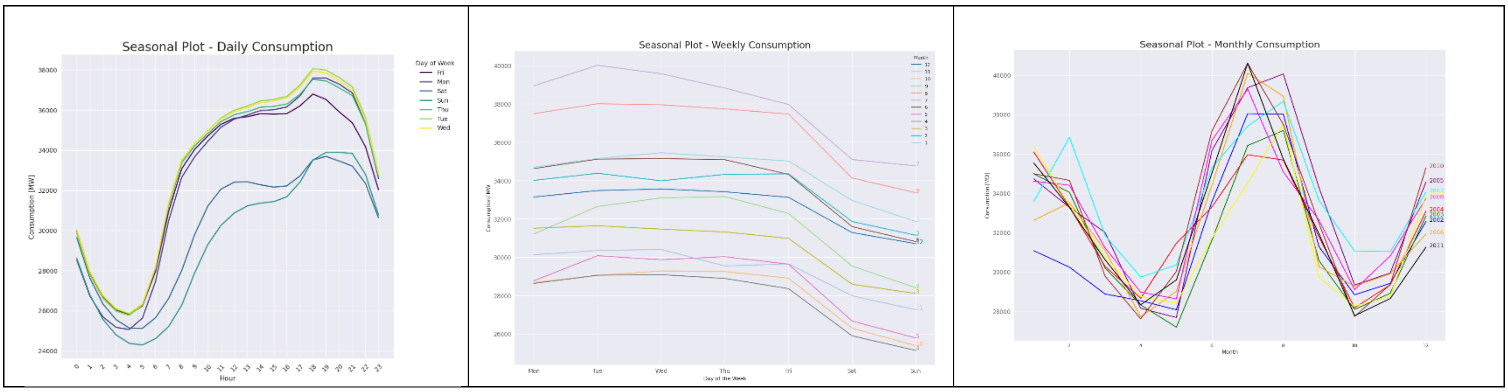

Figure.2 Seasonal Plot- Daily, Weekly and Monthly Comparison

Figure.2 depicts the electricity consumption throughout the day across different days of the week. Fridays exhibit the highest consumption, particularly in the late afternoon, while Sundays tend to have lower consumption, especially in the early morning and evening hours, electricity consumption patterns across various years, with peaks typically occurring in summer months (June to August) and a dip during winter months (December to February). The year 2010 shows higher consumption, especially in the latter months of the year. The last graph shows variations in electricity consumption during the week across different months. Monday, Tuesday, and Wednesday show higher consumption rates, while weekends (Saturday and Sunday) consistently experience lower electricity usage, especially in the mornings.

3. Different Methodologies to Predict Time-Series

Several classical methodologies exists that are trained and then tested that remain as default for several time-series prediction instances and are extensively utilized. They could be effectively armoured with strengths of ML (Machine Learning) for accomplishing ideal outcomes. On contrary, topical methodologies concentrate on certain goals and conditions that fit certain prediction cases. Similar to cases of functioning with ML, prediction could be unsupervised or supervised. Techniques listed below could pertain to both natures in interchangeable format,

ARIMA (Auto Regressive Integrated Moving Average)

ARIMA applies observations gathered during prior time stages as input for taking regression equations which assist in predicting values to be retrieved at subsequent time phase. It integrates the principles of moving average and auto regression, making the predictions correspond to the linear integrations of conventional variables and prediction errors being the renowned methods as a result of it.However, the impact of absent data on model performance is significant and can lead to inaccurate predictions in ARIMA-based models. When some data points are missing, the ARIMA model cannot effectively identify these patterns and may produce biased or suboptimal results. However, integrating ARIMA with machine learning techniques such as Support Vector Regression (SVR) and neural networks has shown great promise in enhancing the accuracy of time-series predictions, as these models can capture both linear and non-linear relationships in the data.

Considering ARIMA, conventional works have undertaken different processes for time-series prediction. Accordingly, the study [6] has used a hybrid method that searched suitable operation for integrating the predictions of non-linear and linear models. The study has combined the ARIMA which is linear and MLP and SVR which is nonlinear for their effective capture of linear trends and the autocorrelations in the time series data which are well-trained with effective computations. SVR and MLP excel at modeling the complex non-linear relationships between the residual errors left by the ARIMA model, thereby improving overall predictive accuracy by introducing flexibility in error-correction mechanisms. The MLP and SVRs which are enabled for modelling the difficult non-linear relationships among the errors which are left over by the ARIMA model and enhances the accuracy by generating the flexibility. Hence, the suggested system has performed linear time-series modelling, non-linear error-series modelling and data oriented integration which searches for suitable operations amongst non-linear and linear formalisms. Two hybrid versions have been assessed. In both these versions, ARIMA model has been utilized. Additionally, SVR (Support Vector Regression) and MLP (Multi-Layer Perceptron) have been utilized. Empirical simulations with 6 real-time series have been assessed with renowned performance metrics such as MSE, MAE and MAPE. Outcomes have exposed the better performance of the suggested method in comparison to hybrid and single models. Further, BHT-ARIMA has been employed, which has considered the distinct advantages of tensorization based MDT, tensor ARIMA, low rank tucker-decomposition in specified framework. The study has combined the MDT (Multi-way Delay Embedding Transform) with the tucker decomposition and tensor ARIMA which has effectively decreased the dimensions of numerous time series by manipulating the inter-connections among the series and displays the compacted low rank format and has reduced the computing costs and enhanced the prediction accuracy. Through the low rank tensor within in the embedded space with MDT hankelization based temporal mode, compressed tensors have been attained through tucker decomposition. Core tensors arrest intrinsic correlations and have been explicitly utilized for training the tensor-ARIMA. It has been claimed that, suggested system could enhance the prediction rate and computational-speed, particularly for multiple TS. Robustness of the recommended model to several parameters have also been empirically researched by comparison with orthogonality version. Experimentations undertaken on 5 real-time TS datasets have exposed the outperformance of BHT-ARIMA in terms of accuracy and computational speed [7]. Moreover, auto ARIMA has been used. Dual customized ARIMA has been used for attaining suitable stock prediction model through the use of Netflix stock-historical information for 5 years. Amongst the considered models, ARIMA (1, 1, and 33) has explored better outcomes in computing holdout testing and MAPE. This has revealed the actual potential of utilizing ARIMA to perform suitable stock prediction [8].

KNN

KNN (K-Nearest Neighbour) is a renowned non-parametric methodology utilized to accomplish classification. It has also been used in regression due to its simplicity and effectiveness in handling various data types and is a valuable tool for many applications. It's particularly useful for time-series forecasting due to its ability to capture patterns and relationships without making strong assumptions about data distribution. It is a supervised ML method which encompasses of features, target components and instances. Selection of features and neighbours is a tedious process. To make this process easy, KNN algorithm finds applicability in several research sectors like image interpolation, visual recognition and financial modelling. In accordance with this, the study [9] has attempted to create an effective, automatic and effective technique to apply KNN on time-series prediction in automating the process of parameter selection (k and input variables), addressing a key challenge in applying KNN to time series forecasting. For accomplishing this, experimentations have been undertaken with several pre-processing and modelling methods through the use of NN3-competition dataset. With regard to experimentation, modelling has been significant than pre-processing. Choosing input variables have been a main factor in attaining effective prediction rate. Deseasonalizing methods have seemed to be redundant as KNN seems to possess the ability for handling seasonal patterns. In accordance with the altering distribution sequences, non-parametric method like quantile regression has been challenging to estimate the prediction error distribution. Prediction integration concept that affords efficient probabilistic prediction outcomes have been used to determine PV generation. Usage of sensible and complementary point predictors featuring several facets of complicated PV production sequences in KNN has been the endorsed model. Accuracy of the suggested system has been evaluated with real-time PV retrieval from USA for several instants. Probabilistic prediction of the suggested quantile KNN model has been assessed against conventional quantile regression, quantile KNN and averaging model with regard to error measures and accuracy. Suggested method has exposed better performance than conventional system by exposing minimum error rate [10].

CART (Classification and Regression Trees)

CART is intended to afford rules which prognosticate the outcome value variable from the explanatory variables. It is capable of exploring ideal accuracy rate. However, it could not detect and acclimatize to concept drift or any alterations. This method is robust with regard to unsupervised practices. However, it seems to lack with regard to maturity. It has to be claimed that, less focus has been given to robust ML methods to design the ensemble CART in forecasting time series [11]. To accomplish statistical deposit processing, the study [12] has modelled uni-variate time series through the use of flexible and powerful ML method of CART bagging and ensembles. Purpose has been to evaluate the accessible data on currency deposits of the Bulgarian citizens. Study has been undertaken in accordance with fundamental schema as selecting predictors from the lagged elements and probable uni-variate time series trend through the use of reference box ARIMA model. Moreover, construction, evaluation and comparison of suggested models have been performed to detect the time series values of deposited data. Further, variable significance of the predictors have been determined. Variable significance in the CART model is determined by assessing the reduction in impurity (e.g., Gini index or entropy) when a feature is used for splitting the data. The importance of each feature is calculated based on its contribution to improving prediction accuracy. Additionally, ideal models have been selected and error diagnosis have been accomplished. Suggested system has been applied to attain short term prediction with three months ahead. Research has been undertaken with the use of SPM (Salford Predictive modeller) and IBM SPSS. Various models with ideal performance has been constructed and assessed. Ideal model has been found to be ensemble of forty CART trees that has afforded determination coefficient (R^2=94%), RMSE=29.95. However, when dealing with large-scale time-series data, there is a notable lack of insight into the considerable time, computational resources, and technical expertise required to effectively train and deploy SPM models. This gap in understanding can hinder the scalability and efficiency of these models, especially when processing vast amounts of data. Without clear strategies for managing these resource demands, organizations may face significant challenges in maintaining performance and accuracy as datasets grow in size and complexity.

SVR (Support Vector Regression)

SVR is a ML approach that could be utilized for regression and classification. Capability of SVR to resolve the non-linear regression issues make SVM to be successful in predicting time-series. Considering these advantages, the study [13] has used SVR to predict COVID-19 in India. SVR has exposed better performance in prediction in accordance with linear, logistic regression and polynomial models. Variation in the considered dataset has been addressed through the suggested system. Model has explored accuracy rate of 97% in determining deaths, cumulative confirmed and recovered cases. Contrarily, 87% accuracy has been attained while determining daily cases. Furthermore, a SVR methodology has been recommended in the study [14]. With the use of several distortions in sample data or high dimension feature space represented by vector area, ideal feature space has been exposed wherein high non-linearity amongst the outputs and inputs have been linearly approximated. Thus, the suggested system has ensured better accuracy, robustness and generalization to construct energy consumption prediction. Huge office constructions in china has been utilized as a case-study. Its summer hour based cooling data load has been utilized as energy consuming data. Outcomes have represented that, suggested system has accomplished ideal performance in comparison to generally utilized methodologies in accordance with robustness and generalization. Likewise, the article [15] has used a financial time-series prediction model through the usage of SVM (Support Vector Machine). Suggested system has been employed to determine the closing rate of EUR/TRY and USD/TRY exchange rates. CCI (Commodity Channel Index) and CPV (Closing Price Values) have been utilized as suitable inputs to predict the time-series The CCI values are typically calculated using the following formula: CCI = (Price - MA) / (0.015 * Mean Deviation), where Price is the typical price (average of high, low, and close), MA is the moving average of the typical price over a given period, and the mean deviation is the average of the absolute differences between the typical price and the moving average.Varied models have been attained with different kernel values and model that has estimated better financial time-series has been recommended. Overall performance has been evaluated in accordance with statistical indicators like RMSE (Root Mean Square Error), MSE (Mean Square Error) and MAE (Mean Absolute Error). It has been revealed that, prediction ability of suggested system has seemed to be better than conventional systems. SVR can be adapted to handle non-stationary time-series data by incorporating methods such as differencing and detrending. Differencing involves subtracting previous observations from current ones to remove trends and make the data stationary, while detrending removes long-term trends to isolate short-term fluctuations. These techniques enable SVR to better capture non-linear relationships in non-stationary time-series data.

XGBoost

XGBoost is an employed ML approach that functions with structured and tabular data. This algorithm requires a transformation of time-series data into supervised-learning issues. However, the outcomes have to be valuable. Taking the advantages into account, the study [16] has introduced multi-step prediction approach that relies on XGBoost for HFRS (Hemorrhagic Fever with Renal Syndrome). Moreover, the prediction and fitting rate of the suggested model has been comparatively assessed with ARIMA model in accordance with several evaluation metrics. Outcomes have confirmed that, XGBoost has worked better in determining non-linear and complicated information like HFRS. In addition, multi-step prediction frameworks have been practical than single step forecasting frameworks in predicting infectious diseases. Correspondingly, the article [17] has utilized ISOA (Improved Seagull Optimization Algorithm) to attain better performance and to enhance optimization processes by balancing exploration and exploitation effectively. ISOA has been applied to optimize ML models, such as XGBoost, by refining their parameters and improving their ability to capture complex patterns within data. By simulating the collective search behaviour of seagulls, ISOA aids in achieving better convergence rates and enhanced model performance, making it particularly useful in tasks like time-series forecasting, where the optimization of hyperparameter is crucial for achieving high accuracy.

In addition, XGBoost has been used to predict time-series of regular COVID-19 positive cases. The empirical outcomes have confirmed the better performance of suggested system. Following this, the article [18] has endorsed a multiple model integrating XGBoost. Recommended methodology has comprised of new time-series as features by fitting and predicting the outcomes. Subsequently, XGBoost model has been used for prediction. High precision rate has been attained through the suggested system. For evaluating the prediction performance, 1095 data epochs in up-co-ordinates of sixteen GNSS stations have been chosen to perform test prediction. In comparison to CNN-LSTM, empirical outcomes of the suggested technique has exposed reduction in MAE value. Prediction outcomes have maximum accuracy rate and it possess high correlation to actual time series that could better predict vertical motion of GNSS stations. Hence, prediction methodology could be employed to up-element of GNSS co-ordination of time-series.

AdaBoost

AdaBoost is a kind of prediction approach that has been estimated as an ideal out-of-box approach. This represents that, it has been ideally utilized at elaborating the classification of data in association with other effective approaches. Example: while using AdaBoost with DTs (Decision Trees), it automatically learns to summarize the hard to classify data. Consequently, the study [19] has considered AdaBoost as one of the ensemble model to prognosticate the time-series of stock market. In this case, various algorithm namely bagging, ADA of ET (AdaBoost of Extra Trees), ADA of BAG (AdaBoost of Bagging) and AdaBoost of RF (AdaBoost of Random Forest) have been used. The stock data has been randomly gathered from 3 varied stock exchanges. In this case, 40 technical variables have been computed and utilized as features. Dataset has been split into testset and training set. Model performance has been assessed with testset in accordance with metrics namely AUC (Area under Curve), F1-score, precision, recall, specificity and accuracy. Kendall W-concordance test has been utilized for performance ranking of varied models. Empirical outcomes have represented that, ADA of BAG has exposed better performance amongst the other tree based models. On contrary, bagging ensemble frameworks have enhanced the performance. Further, to evaluate the inner passenger flow features as well as to make accurate detection with minimum training time, an integration of SSA (Singular Spectrum Analysis) and AWEKM (AdaBoost Weighted eXtreme Learning Machine) has been suggested [20]. In this case, SSA has been developed for decomposing actual data into varied elements of periodicity, residue and trend. In addition, AWELM has been developed for predicting individual elements. Three prognosticated outcomes have been concluded as final results. From experiments, dataset has been gathered from AFC (Automatic Fare Collection)-hangzhou metro system in China. Three week passenger flow has been extracted to undertake multiple stage prediction tests. Outcomes have represented the better performance of suggested model with regard to training time and prediction errors. Particularly, in comparison to existing DL framework, LSTM (Long Short Term Memory) has minimized 22%

of testing errors, thereby saved 84% of time. Table-1 shows the comparative analysis of several methods involved in time series prediction for different applications.

Table-1. Methods Involved in Time Series Prediction in Various Applications

References	Techniques	Applications	Performance Metrics	Findings
[21]	ARIMA and KNN	Climate change forecasting	MAPE, MAE and RMSE	The prediction of 'k' values has been done and based on the 'k' value, the validity of the model has been identified using error measures. It has been found that model has been good at predicting rainfall.
[22]	ARIMA	Estimation of Ultra-long time series in a distributed system	MASE, MSIS and ACD	Performance has been analysed in both simulations and real data application and has predicted that the model outperform in uncertainty estimation and point forecasting.
[23]	ARIMA and Least Square SVM	COVID-19	MAE, MSE, RMSE and R^2	The values of forecasting errors in the validation part shows improved accuracy of the model. The prediction results has aided in estimating the future cases of COVID-19.
[24]	ARIMA	Time series analysis of Emission and performance characteristics of biogas fuelled compression ignition engine.	RMSE, R^2 and BIC (Bayesian Information Criterion)	The results has shown the potential of biogas-diesel dual fuel combustion mode at varying engine operating conditions.
[25]	FireFly and AdaBoost Algorithm	Industrial Load Forecasting Model	MAE, RMSE, MAPE and time	Prediction accuracy of the model has been high and easy to implement due to its simplicity
[26]	XGBoost, ARIMA and PSO optimiser	Thermoelectric Power Plants Diesel and Heavy Fuel Oil Engine Fuel Consumption	MAE, RMSE, R^2	The models have been better at capturing disturbances correlated with sudden fluctuations steps shut-down and fuel consumption ramp.

4. Deep Learning for Effective Prediction of Time Series Data

Prior to the advent of DM (Data Mining) methodologies, conventional techniques utilized box Jenkins technique like ARIMA and statistical models like exponential smoothing. Such models depend on constructing linear operations from current past observations for affording future prognostications and is seemed to be widely employed to predicting tasks upon last decades. Nevertheless, these statistical methodologies generally fail while they have been directly employed without the consideration of theoretical needs of ergodicity, stationarity and pre-processing. For example, in ARIMA, time-series must be transferred into stationary form by varied differencing transformations. Example: in ARIMA, time-series must be transformed into motionless form by diverse variance transformations. Another challenge is that, employing these models have been restricted to circumstances, wherein enough historical data with comprehensive outline has been accessible. When data accessibility for single series have been scarce, such models often deteriorate for effective extraction of underlying patterns and features. Further, as these methodologies develop a model for distinct time-series, it seems to be not probable for sharing learning across identical instances. Hence, it seems to be not feasible for usage of these methods to handle with huge data amounts as this would remain computationally exorbitant. Thus, AI (Artificial Intelligence) methods which permit in the construction of global models for predicting several associated series initiated to earn popularity. In the meantime, linear methodologies have initiated to be considered as base-lines to perform comparison. Some of the widely used DL methods are discussed below,

4.1. MLP (Multi-Layer Perceptron)

ANN concept was initially inspired in brain functioning with neurons functioning in parallel to process the data. MLP (Multi-Layer Perceptron) is the fundamental kind of FF-ANN (Feed Forward Artificial Neural Network). Their model has encompassed of 3-block outline namely: input, hidden and output layer. One or several hidden layers could exists that would determine the network depth. It includes the increase in depth, inclusion of several hidden layers that make MLP network to be a DL model. Individual layer encompasses of stated neuron sets which has connections related with training parameters. NN algorithm updates connection weights for mapping the output or input relationship. The MLP networks only possess forward connections amongst neurons, irrespective of other models which possess feedback loops. In addition, the study [27] has claimed a NN method to prediction of time-series is BPNN (Back Propagation NN) with direct input to output associations. 8 varied datasets have been utilized for verifying the assessment of the suggested model in prognosticating time-series. BPNN has been enhanced to 4 variants in accordance with the absence or presence of result layer bias. Empirical outcomes have exposed that, the suggested system has revealed better prediction rate in comparison to traditional BPNN while, resultant layer bias possess no significant impact. Hence, input to output associations could significantly enhance prediction capability of time-series. Reinforcement learning (RL) is applied to time-series prediction in dynamic settings, which adapts the on real-time feedback, also the RL model learns by interacting with the environment and receiving rewards or penalties for its predictions. This feedback mechanism allows the model to improve its predictions over time by adjusting its parameters to optimize future performance. In dynamic settings, where data is constantly changing, RL enables the model to make decisions based on the most current information, leading to more accurate and responsive time-series predictions. Further, in the research [28], alternative MLP has been used for uni-variate time-series categorization for single, dual and thrice layers. Input data don't need several mathematical processing cycles and hyper-parameters have been dynamically altered in accordance with the training set size and computations in time-series. Though MLP seems to not exist among 3 ideal performers, it is significant for emphasizing the merits of utilizing the suggested model. The Synthetic Minority Over-sampling Technique (SMOTE) has been modified for time-series data to account for temporal dependencies between observations. Variants like SMOTE-TS, TS-SMOTE, and SMOTE-ENC adapt SMOTE by ensuring that synthetic samples maintain the sequential relationships inherent in time-series data. SMOTE-TS considers the temporal ordering between samples, while TS-SMOTE uses a sliding window approach to preserve trends. SMOTE-ENC encodes the temporal features to respect periodicity before over-sampling.

4.2. Bi-LSTM (Bi-Directional Long Short Term Memory)

Bi-LSTM involves the process of constructing a NN to possess information sequences in forward as well as backward directions. In Bi-directional NN, flow of input occurs in dual ways, making Bi-LSTM diverse from typical LSTM. Accordingly, the study [29] has used DAFA-Bi-LSTM (Deep Auto-regression Feature Augmented-Bi-directional LSTM) to prognosticate time-series. In the initial phase, input vectors have been fed into VA (Vector Autoregression) module for indicating time-delayed non-linear and linear functions of input in unsupervised manner. Subsequently, learned non-linear integration VA vectors have been gradually fed into varied Bi-LSTM layers and results of prior Bi-LSTM module has been concatenated with time delaying linear VA vectors as augmented feature for forming additional signals for subsequent Bi-LSTM layer. Wide ranging real-time applications of time-series are resolved to reveal the robustness and superiority of endorsed DAFA-Bi-LSTM. Empirical outcomes have explored the better performance of suggested system in accordance with robustness. Likewise, the study [30] has endorsed a Bi-LSTM framework relying on multiple variable time series data that regards the features of the fields. Suggested model has differed from conventional model in that, input is partitioned into areas for learning the features. Additionally, time-series trends have been considered by concurrently learning data value and its difference. The suggested model has been applied on forecasting the trading area for gathering data purchase and SNS-data upon restaurants for progressive learning. Performance assessment has been undertaken with regard to RMSE (Root Mean Square Error) that has relied on if learning has been performed by individual field and if there has been alteration in input. Empirical outcomes have exposed the better performance of the suggested system.

4.3. GAN (Generative Adversarial Networks)

Generative framework is an unsupervised-learning in ML which includes automatic exploration and learning of patterns or regularities in input in a manner which the model could be utilized for generating new instances. This could probably have been attained from actual dataset. Moreover, GANs exists as a clear approach of training by outlining an issue as a supervised-learning with dual sub frameworks. This includes generator and discriminator model wherein generator model has been trained for generating new instances and discriminator model has been attempted to classify instances as fake (generated) or real (from domain). Two models have been trained in zero sum game, until discriminator model has been fooled half period denoting that, generator model has been generating plausible instances. Moreover, GANs are rapidly altering area, affording generative frameworks with an ability for generating realistic instances throughout wide problem domains, notably in translation of image to image like translation of summer – to – winter photo and is retrieving realistic images of scenes, objects and individuals that human could not interpret as fake. For enhancing the visual impact based on possibility and assure the prediction rate of movement trends and shape, GAN has been utilized. From comparative analysis, optimal integration of Huber-loss function and mish-activation function has been chosen. An enhanced GAN has been suggested for predicting time-series with regard to three aspects. Additionally, multi-neighbourhood gradient variation loss has been adopted for enhancing the prediction nephogram sharpness, multi-scale structure has been utilized for improvising performance in long-term. Further, Wasserstein distance has been introduced for replacing JS-divergence for assuring usual updation of the generator parameters. Empirical outcomes have exposed better performance. For instance, for enhanced model with 4 neighbourhood gradient variation loss, accuracy enhancement in SSIM (Structural Similarity) has been 35.57% and PSNR (Peak Signal to Noise Ratio) has been 6.42% [31].

4.4. CNN (Convolutional Neural Network)

CNNs is claimed as a group of DL models that have been outlined for computer-vision based tasks. These are regarded conventional for several classification tasks like speech recognition, NLP (Natural Language Processing) and object recognition. It could automatically extract the features from data of high dimension with grid topology like image pixels without requiring any feature engineering. Model learns for extracting valuable features from raw-data through the use of convolutional function that is sliding-filter which develops feature-maps. It has also intended to retrieve repetitive sequences at varied data regions. Such feature extraction process has afforded CNNs with significant features termed as distortion invariance that represents that, features have been retrieved in spite of where they seem to be in data. This ability to capture temporal and spatial dependencies within data sequences makes CNNs particularly suitable for time-series analysis, as they can effectively extract features from sequences of data with grid-like structures, such as time-series data points. Such features make CNNs appropriate for handling with data of one dimension like time-series. Data sequences could be viewed as a 1-dimensional image from convolutional function that could extract the features. For time-series data, CNNs utilize their convolutional layers to identify patterns or trends over time, allowing the model to capture both short-term and long-term dependencies, which is essential for tasks like forecasting and anomaly detection in time-series datasets. Accordingly, the study [32] has considered multiple CNN for multi-variate prediction of time-series. Suggested multi-CNNs have been modelled from global patterns and could efficiently retrieve short and long period of sequential information. Residual networks (ResNets) enhance CNNs by introducing residual connections that allow the model to skip layers and directly pass information from earlier to deeper layers. This mitigates the vanishing gradient problem, enabling the training of deeper networks without performance loss. In time-series prediction, these connections help capture both short-term and long-term dependencies, improving the model's ability to make accurate predictions. Concurrently, residual network has been integrated with autoregressive element and further, this has enhanced the model performance. Through the experimentation with real-time datasets and comparison with varied conventional methodologies, the recommended system has exposed ideal performance in diverse aspects. For minimizing the computational-costs, a TL (Transfer Learning) system has been suggested to NN based flood framework. The need for TL is due to the high computational cost of training DNN (Deep Neural Networks) from the scratch and permits in enhancing the knowledge gained from training on larger datasets for enhancing the training effectiveness and accuracy in the targeted domain with restricted data which is crucial in the flood prediction where the data acquisition is costlier and time consumption. For the TL concept, after model has been pre-trained in source domain, it could be reutilized in other targeted domains. Following the retraining model parts with target domain data, training time could be minimized due to re-usage. CNN has also been employed with TL that possess several successful applications for classifying 2D image. Nevertheless, the suggested model has endorsed a time-series component. CNN with TL needs conversion framework from the time-series to image-datasets during pre-processing. Initially, classification performance of CNN has been assessed in source area with lesser than 10% of errors for water-level variation. Following this, CNN with TL in target domain has effectively minimized training time at a rate of 1/5 with mean error variation as 15% with those attained by CNN without TL [33].

4.5. GRU (Gated Recurrent Unit)

GRUs have been introduced to vanishing and exploding-gradient issue. Nevertheless, it could be viewed as a simplified form of LSTM. In GRU, the input and forget unit have been integrated into an update gate. Such modification has minimized trainable parameters affording GRU with ideal performance with regard to computation while accomplishing identical outcomes. Considering these merits, the study [34] has explored RNN (Recurrent Neural Network) to prognosticate water quality by relying on seq2seq (sequence to sequence) model. GRU has been utilized as encoder as well as decoder. In addition, FM (Factorization Machine) has been included into model for solving high dimensioned feature association in data and sparsity issues. Besides, due to prolonged time and timespan, dual attention approach has been integrated to seq2seq model for addressing the failures of a model in DL (Deep Learning). Experimentations have been undertaken and outcomes have exposed the better performance of the suggested system than other water quality prognostication methodologies. GRUs expose application in learning the temporal information to detect time-series. This has explored the efficacy of the suggested system. Nevertheless, from statistical viewpoint, these models have been ergodic with limited memory which instigates learned temporal evidence to be forgotten quickly. Meanwhile, such conventional strategies have entirely ignored temporal reliance amongst gradient flow of optimization approach. This has limited the accuracy rate. For resolving such problems, a GRU model has been suggested that has integrated dual methods namely Adaptive Mixed Gradient and selective state updation for enhancing prediction rate. Further, tensor-discriminator has been utilized for determining if the information of hidden state has to be rationalized at individual time-step to learn extreme fluctuating information of the suggested system. Efficacy of the recommended system has been assessed on 5 varied datasets. Empirical outcomes have exposed the better performance of the suggested system [35].

4.6. TCN (Temporal Convolutional Network)

Recently, studies have considered a special form of CNN namely TCN. This model has found its initiation through the motive of Wavenet autoregressive model. This has been designed to resolve the issues related with audio generation. Considering such advantages, the study [36] has predicted multi-variate time-series by using the suggested method for comparing with conventional models. In the suggested system, dilated network has been applied as meta-network. Additionally, asymmetric residual frameworks have been built. The system has significantly enhanced the time-series prediction outcomes on the datasets of ISO-NE and Beijing PM2.5. Study has concentrated on trade-off amongst the prediction accuracy and implementation complexity. From empirical outcomes, it has been exposed that, suggested system has worked better. TCN is a specialized model which possess merits over RNN (Recurrent Neural Networks) to perform task prediction. It also possess the ability for extracting patterns of utilizing residual blocks and dilated casual-convolutions. It could also be effective with regard to computational time. Further, the study [37] has suggested a TCN based DL model for enhancing the prediction rate in determining energy demand. Two energy oriented time-series with information from Spain has been researched. This has included power demand of charging stations and national electric-demand. Extensive experimentations have been undertaken involving near 1900+ models with distinct parameterizations and models. From the results, it have been exposed that, TCN model has worked better than LSTM model. Various models have been comparatively evaluated and the corresponding details are presented in table-2.

Table-2. Comparative Assessment of Conventional DL Models for Efficient Prediction of Time-Series

References	Methodology	Aim	Outcomes
[28]	MLP	Classifying Uni-variate time-series for single, dual and thrice layers	Though MLP has not been among 3 ideal performers, it has been significant to emphasize the merits of utilizing the suggested model.
[38]	Parallelized MLPs	Electric-demand prediction	Issues have been partitioned into varied prediction sub-issues for determining several samples concurrently.
[39]	LSTM+CNN	Financial data forecast	LSTM has learned the significant features thereby minimized dimensionality.
[40]	Encoder and decoder TCN	Retail-sales prognostication	Integrated with the representation learning, the TCN could learn complex sequences like seasonality. It

			has been effective due to convolutional parallelization.
[41]	Bagged-ESN	Predicting the Consumption of energy	ESN, differential evolution and bagging have been used for minimizing the prediction error, thereby, enhancing generalization ability.
[42]	LSTM+ clustering	NN5 competitions and CIF2016	A model has been built for individual cluster of associated time-series. Suggested system has enhanced the forecasting rate, while minimizing training time.
[43]	CNN	Load construction	Suggested system with direct methodology has afforded optimal outcomes, enhancing the prediction rate of seasonal-ARIMAX.
[37]	TCN	Forecasting energy demand	Results have exposed high ability for handling long input sequences. Model has been sensitive to parameterization in comparison to LSTM.

5. Applications of Extensively Utilized DL Models to Predict Time-Series

For exposing the significance of predicting the issues of time-series, conventional studies have been undertaken by partitioning DL based studies by diverse applications like healthcare, energy, weather and finance prediction in accordance with extensively utilized network models. This section discusses the summary of studies for different application areas, emphasizing the intentions accomplished for individual field and method,

Healthcare

Recently, deep sequential frameworks have been frequently utilized. The prediction abilities of deep learning with its automated ability in automated feature identification making it a convincing tool to diagnose the disease. Numerous researches exists in healthcare area. In accordance with this, the study [44] has suggested a DL method to prognosticate the infection cases of Covid-19. Further, a multi-variate LSTM (Long Short Term Memory) has been trained on varied samples of time-series concurrently, inclusive of mobility-series. Outcomes have revealed that, integrating mobility as variable with the use of multiple samples for network training has enhanced the prediction performance with regard to variance and bias. It has also been exposed that, prediction outcomes have identical accuracy rate with ensemble model. Suggested model has been fascinating in varied aspects inclusive of fine geographic predictions and better prediction performance. Further, computational intensity and data requirements have been minimized by substituting a model for multiple models. Similarly, the research [45] has intended to develop the computational model for determining the weakening health status of patients in a manner that is probable for initiating suitable treatment as early as possible. Model has been developed in accordance with DL method namely RNN (Recurrent Neural Network) and LSTM (Long Short Term Memory) for determining the vital signs of patients. Experimentations have exposed that, it seems to be probable for identifying crucial signs with optimal accuracy rate. The main contribution of suggested system has relied in developing a methodology to prognosticate vital signs by relying on historical information with minimum MSE (Mean Square Error) with its subsequent application in computing predictive indexes with efficacy. Differentiation exposed by the endorsed system stems from a fact that, certain works prognosticate vital signs.

Environmental Pollution

Globally, climate change with its related consequences have developed varied challenges for farmers in managing lands. This pollution also impacts the air quality, rural-development, health and water resources. Hence, its mitigation needs correct prediction. Considering this, the study [46] has employed ANN to prognosticate the river flow by incorporating CSA (Co-operation Search Algorithm) into ANN’s learning process. On contrary, ANN’s computation parameters have been iteratively optimized with CSA in feasible state-space. The study has combined the ANN (Artificial Neural Network) and CSA (Cooperation Search Algorithm) which are flexible and model the difficult non-linear relationships present in the hydrological data which has been more relatable for river flow prediction. The CSA algorithms has enhanced the ANN with the increase in the speed of convergence and transfers the risk of getting stuck in the local minima in the training period with an enhanced accuracy. Suggested system has been applied to the data gathered from real-time hydrological stations. Varied quantitative indexes have been selected for comparing the developed model's performance, while, comprehensive evaluation amongst observed and simulated data flow have been undertaken. Experimentations have revealed that, in varied cases, hybrid CSA-ANN has performed better. Suggested system has exposed 5.42% as coefficient-correlation and 11.10% as nash Sutcliffe efficacy. Findings have exposed the better performance of AI models in forecasting the time-series of river flow. On contrary, the research [47] has employed EWT (Empirical Wavelet Transform) integrated with enhanced ensemble models to remove redundant noise with decomposition of actual series into varied elements as inputs. The empirical construction of wavelets in EWT includes adapting the wavelet function to the characteristics of the data. It uses data-driven approaches to automatically decompose the time series into different frequency components. This process helps in identifying and extracting relevant features while removing noise from the signal, making the data more suitable for further analysis or modeling. Following this, GRU (Gated Recurrent Unit) has been employed for deep learning the association amongst historical stream-flow series. Lastly, overfitting and underfitting have been balanced with Improved GWO to expose the suitable parameter integration. Case study has been undertaken by streamflow data on a monthly basis. When compared with LSSVM (Least Square Support Vector Machine) and ELM (Extreme Learning Machine), validation outcomes have exposed that, GRU could expose satisfactory results in forecasting monthly stream-flow. Suggested work has been better than single GRU model that has exposed the superior performance of the considered method.

Weather Forecasting

Weather predictions have been developed by gathering data on atmospheric status at specific time. It involves the process in finding the atmospheric conditions at specific time through the use of varied weather features. As farmers rely on comprehensive short to medium range of prediction with maximum resolution of regional prediction models, the study [48] has intended to resolve this by assessing and developing light-weight weather prediction system. This has comprised of single or numerous local weather regions and conventional methods to predict weather in accordance with time-series data. By this end, system has exposed conventional TCN and LSTM. Experimentations have revealed the better performance of the suggested system with TCN than LSTM. Recommended model could be utilized as an effective localized tool to forecast weather. In addition, the research [49] has explored AEEMD-ANN (Adaptive Ensemble Empirical Mode Decomposition-Artificial Neural Network) to detect ISMR. Overall performance of suggested system has been compared with conventional works. Suggested methodology has accomplished prediction skill of 0.91 and 0.78 to prognosticate the rainfall for India and Kerala. Suggested work has performed better in attaining hydrologic extremes in comparison to conventional systems.

Finance

Financial prognostication in a conventional and complex area that has gained the attention of scientists and economists. Researchers are utilizing ML based deep sequential frameworks to prognosticate stock market. In accordance with this, the study [50] has intended to identify the energy cost consumed in buildings through ML. To alleviate the dimensions of input, dual methods have been undertaken comprising of wrapper based and filter based system. Subsequently, three ML methodologies have been utilized for predictions including CART, RF (Random Forest), DT (Decision Trees) and ANNs. It has explored the hybrid method which has combined the variable selection by utilizing the Boruta and Random Forest with the prediction models. The hybrid RF-Boruta has been utilized for feature selection which followed by the selection of features as input for the ANN, CART or RF prediction models. Outcomes have explored the ideal prediction with the hybrid ML and boruta, specifically, while, RF has been utilized for prediction purposes and variable reduction. Three ML methodologies: CART, RF and ANN have exposed enhancement in accuracy rate while, the integrated scheme has been utilized to accomplish variable selection. In addition, the research [51] has intended to find the stock price with ANN, thereby, train them with certain meta-heuristic approaches like BA (Bat Algorithm) and SSO (Social Spider Optimization). Certain technical indicators have been utilized as input-variables. Following this, GA (Genetic Algorithm) has been employed to select optimal and relevant features. Few loss functions have also been utilized comprising of MAE. On contrary, certain time-series prediction models have been utilized including ARIMA and ARMA to determine the price of stock. Comparison has been undertaken with BA and SSO. Results have been better than conventional studies.

Stock market forecasting

DL models are constructed by utilizing past stock price data from a reputable company listed on the National Stock Exchange (NSE) of India. Stock prices are documented every five minutes between December 31, 2012 and January 9, 2015. Utilizing the characteristics in the highly detailed stock price data, experts construct four

CNN and five LSTM-based deep learning models, verify their efficiency, and assess their performance [52]. Current techniques include CNN, RNN, LSTM, CNN-RNN, and CNN-LSTM. The outcome indicates that CNN has the lowest performance, LSTM performs better than CNN-LSTM, CNN-RNN outperforms CNN-LSTM, CNN-RNN performs better than LSTM, and recommended single Layer RNN model surpasses all other models. The suggested single Layer RNN model shows enhancements of 2.2%, 0.4%, 0.3%, 0.2%, and 0.1%. The effectiveness of the proposed model has been confirmed by the experimental results, which will help investors boost their profits through wise decision-making [53].

Traffic and transportation

A study on China Railway Express compared CNN-BiLSTM model's accuracy in predicting train travel time to other models like Holt-Winters, RF, SVR, LSTM, BiLSTM, and LSTM with attention mechanism, CLSTM, CNN_LSTM, CNN_GRU1, TCN, and CNN_GRU2. Additionally, the CNN-BiLSTM model yielded values of 4.647 for MSE, 2.156 for RMSE, 2.643 for MAPE, and 1.769 for MAE [54]. In the previous research, scientists utilized the RNN-LSTM model with hyperbolic tangent in hidden layers and rectified linear unit at the final layer to forecast upcoming values based on the time series data. The findings of the study indicate that our system can forecast vibrations with an accuracy exceeding 99% in a sequence of upcoming values [55].

Various applications of DL models in time-series prediction have been comparatively assessed and the respective details are tabulated in table-3.

Table-3. Comparative Evaluation of Different Applications of Extensively Utilized DL models in Time-Series Prediction

References	Methodology	Applications	Results	Values
[56]	Fuzzy C-means and Pi-Sigma ANNs	Stock exchange and temperature prediction	Ideal predictions have been generated with suggested method for test data corresponding to individual dataset in accordance with MAE and RMSE	RMSE – 100.208, MAE-846.00
[57]	LSTM-DP-ARIMA and ARIMA-LSTM	Well-production	Hybrid DEEP-LSTM-ARIMA model has been reliable in accordance with MAPE, RMSE and MAE	RMSE- 11.5840, MAPE-0.1441, MAE-6.5376
[58]	BRKGA-NN	Computer-science	Accurate prediction has been attained. Moreover, better performance has been attained with regard to MAPE, RMSE and MAE	MAE-6.47, RMSE-6.51
[59]	Pi-sigma based ANN	Computer-science	Suggested system has enhanced the prediction rate in accordance with RMSE and MAPE	RMSE-65.30, MAPe-0.731
[60]	TCNN	Computer-science	It has been recommended that, a method to decide scalability has to be developed in accordance with MSE	MSE-0.0000974, MAE-0.00377
[40]	Deep TCNN	Traffic, electricity and spare fragments of the car	Framework has been better than conventional systems in accordance with MASE, SMAPE and NRMSE	NRMSE-0.895, SMAPE-0.268, MASE-0.708
[61]	ES+LSTM	M4-competition	Forecasting the forex price, the LSTM model has permitted for cross learning and non-linear trends from several associated series in accordance with MAPE	-

[52]	CNN and LSTM	Stock market forecasting	Characteristics in the highly detailed stock price data	-
[53]	CNN, RNN, LSTM, CNN-RNN, and CNN-LSTM	Forecasting the stock market	Investors boost their profits through wise decision-making	-
[54]	Holt-Winters, RF, SVR, LSTM, BiLSTM, and LSTM with attention mechanism, CLSTM, CNN_LSTM, CNN_GRU1, TCN, and CNN_GRU2.	Traffic and transportation	Predicting train travel time	-
[55]	RNN-LSTM	Transportation and traffic	Forecast vibrations	-

Table.4 Hyper-parameters Values [62]

Parameters	**Values**
Convolution Layer	1
Convolution Layer Filters	32
Convolution Layer Kernel Size	1
Convolution Layer Activation Function	tanh
Pooling Layer	1
Pooling Method	Max pooling
Pooling Size	2
Strides	1
Flatten Layer	1
LSTM Layer	1
LSTM Layer Units	1
LSTM Layer Activation Function	tanh
RNN Layer	1
RNN Layer Units	64
RNN Layer Activation Function	tanh
Optimizer	Adam
Loss Function	MAE
Time step	10
Batch Size	64
Epochs	100

Table.4 depicts the hyper –parameters values of various layers.

Multivariate Time-Series Analysis

Multivariate time-series analysis involves examining multiple time-dependent variables to understand their relationships and dependencies over time. It is widely used in fields like finance, healthcare, and economics to model interactions between variables and uncover patterns. Key objectives include forecasting future values, detecting anomalies, reducing dimensionality, and selecting relevant features. Techniques like vector autoregressive models (VAR), multivariate ARIMA, and deep learning methods such as RNNs and LSTMs are commonly applied to capture complex relationships.

Anomaly detection [63] has been applied in various domains, with deep learning techniques proposed for multivariate time series analysis. This study classifies anomalies into abnormal time points, time intervals, and time series, reviewing deep learning methods for each. LSTMs and auto encoders are commonly used for detecting time point and interval anomalies, while dynamic graphs have been implemented to explore relational features. However, anomaly detection still faces challenges, particularly in explaining the causes of anomalies.

Similarly, the rise of digitalization [64] has led to the use of multi-sensor systems that generate large amounts of unlabeled multivariate time series data. Multivariate Time Series Anomaly Detection (MTSAD) is crucial for identifying operational states but is challenging due to the need for analyzing temporal and spatial dependencies. Labeling large datasets is often impractical, making unsupervised MTSAD methods essential.

Correspondingly, Time Series Classification (TSC) [65] has advanced with new algorithms, mainly for univariate cases, but multivariate TSC (MTSC) has received less attention. The UCR archive has been valuable for univariate TSC, while the 2018 UEA archive, offering 30 MTSC problems, has made algorithm comparison easier. This review covers recent MTSC methods based on deep learning, shapelets, and bag of words approaches. For cases where algorithms cannot handle multivariate data, univariate classifiers have been adapted by applying them to each dimension. Our comparison of bespoke algorithms and dimension-independent approaches on 26 MTSC problems shows that four classifiers outperform the dynamic time warping benchmark, with ROCKET achieving significant accuracy improvements in much less time.

6. Advantages of DL to Predict Time-Series

Conventionally, time-series prediction has been subjugated by linear methodologies as they are comprehended and efficient on varied forecasting issues. In this case, DL based neural networks have gained attention as it possess the ability for automatic learning of arbitrary complex associations from input-output and assist in several input-output. In addition, the DL based models are optimal in exposing intricate structures in big datasets. They have also exposed to learn the discriminating sequences from several time-series information. Though the model builders have main concern with prediction, typically, end-users make use of prognostications for assisting their forthcoming actions. Instances include a physician making the usage of clinical predictions for assisting in prioritizing tests for formulating diagnosis, thereby, determine, treatment course. While, time-series prediction is a vital stage, better comprehension of motivations behind the forecast of a model and temporal dynamics could assist the users to optimize their operations. Concurrently, with the advantageous nature of DL, this section exposes dual ways in which NNs have been enhanced for enabling decision making with time-series information. This focusses on techniques in causal inference and interpretability.

Causal Inference and Counter-factual Predictions over Time

DL possess the ability to comprehend the associations learnt by networks. It could also assist in facilitating decision assistance by affording predictions external to the counterfactual forecasts or observational datasets. The counterfactual prognostications are valuable for scenario evaluation permitting users to assess the ways in which diverse action sets could influence the target trajectories. It could be useful from historical viewpoint that is, determining the reality when various situations have happened. On contrary, it seems to be valuable from forecasting viewpoint that is, determining the actions for optimizing future results. While big class of DL methodologies exists to find the causal impacts in statistical settings [66], main issues in time-series datasets lie in the existence of time-reliant confounding impacts. This has arouse due to spherical reliance when actions which could impact target are conditional on the target observations. Without any adjustment in time reliant confounders, straightforward estimating methodologies could lead to biased outcomes. In recent time, various methodologies have evolved to train the Deep NNs while altering for time-reliant confounding in accordance with the enhancement of statistical methods with the outline of loss functions. Through the use of statistical methodologies, the study [67] has extended IPTW (Inverse Probability of Treatment Weighting) method of marginal-structural frameworks in epidemiology through the use of network sets to determine the treatment application possibilities and seq2seq model for learning the unbiased predictions.

Interpretation with Time-series data

With NN deployment, there seems to be an enhancing demand for comprehending the why and how a framework accomplishes certain prediction [68]. Besides, end-users could possess little knowledge in accordance with relationships prevailing in the data. Provided with the black box nature of typical NN model, researches have evolved in techniques to interpret DL models as discussed below.

Recurrent Convolutional Networks (RCN) is a hybrid DL model that integrates CNNs for feature extraction and RNNs (or LSTMs) for modeling temporal dependencies, making it effective for capturing both spatial and temporal aspects of time-series data. The Temporal Fusion Transformers (TFT) is a DL model for multi-horizon time-series forecasting, combining transformers and RNNs. It uses attention mechanisms to capture temporal dependencies and complex patterns across time steps, along with gating mechanisms for variable selection and interpretable forecasting.

Inherent interpretation with the attention-weights

An alternative method relies in designing the models directly with explainable elements in a form of strategic positioned attention layers. With the attention weights afforded as results from Softmax, weights are controlled to sum to one. For time-series frameworks, the results could be inferred as weighted average upon temporal features through the weights afforded by attention-layer at individual phase. Then, attention weights could be utilized for understanding relative feature significance at individual time-step. Moreover, instance wise works have been undertaken in [69], wherein, specific instances have been utilized for exposing the way in which magnitudes could represent the actual time points with high significance to accomplish predictions. With the evaluation of attention vector distributions across time, it has exposed the ways in which attention mechanism could be utilized for finding consistent temporal associations like seasonal patterns.

The Post-hoc interpretation models have found its evolution to interpret the trained networks and assist in identifying significant features or instances without altering actual weights. The techniques could be partitioned into two major divisions. Initially, one probable system lies in employing simple interpretable surrogate frameworks amongst input-output of NN. It also depend on approximation model for affording explanations. Instances include LIME (Local Interpretable Model agnostic Explanations) have found relevant features with fitting of instance oriented linear models for distress of input with linear coefficients affording computational significance. In addition, SHAP (Shapley Additive Explanations) have afforded an additional surrogate system which makes use of Shapley values of co-operative game theory for identifying significant features throughout the dataset.

Both SHAP and LIME provide crucial insights into deep learning model predictions by identifying which features have the most influence on a given prediction, enhancing model transparency. SHAP, in particular, offers a global understanding of feature importance across the entire dataset, while LIME provides local interpretability, focusing on specific instances or predictions. Following this, gradient based methodology like influence functions and saliency maps have been suggested that has assessed the network gradients for determining the actual input features possessing high influence upon loss functions [70]. Post-hoc interpretation methodologies could assist with feature attributes, they usually ignore sequential reliance amongst inputs making it complex for employing them in complex time-series data. Attention-based techniques, on the other hand, directly address this limitation by highlighting relevant time steps or regions in sequence data, enabling more meaningful and interpretable predictions for time-series tasks.

7. Research Gap and Future Recommendations

From extensive evaluation of conventional works, this study emphasizes certain significant drawbacks in this section.

- As different design alternatives seem to be attractive, it is better to integrate them [8].
- In future, other detendring methods have to be studied and suitable datasets have to be utilized for evaluating time-series globally [9].
- Future study must enhance by assessing the prediction ability of the stated ML methods through the use of additional features. It also seems to be promising to utilize DL methods for selecting suitable features, thereby make predictions [11].
- In future, hyper-parameters of the model has to be modified for learning to enhance its performance through diverse experimentation for considering the features of time-series [30].

For time-series prognostication, affording explainable outcome is vital to compute its reliability. Hence, improvising the interpretability has turned as future research work attracting intense attention. Though several DL models are introduced to forecast time-series, certain drawbacks prevail. Typically, Deep NNs demand time-series for discretization at periodic intervals, creating it complex for detecting datasets wherein observations could be missing or evolve randomly. While certain preliminary studies on persistent time-models have been undertaken through neural ordinary-differential equations, additional endeavours are needed for enhancing the work for data with complicated inputs, thereby benchmark them in accordance with conventional models. Additionally, as claimed in study [71], time-series possess hierarchical outline with logical categorizations amongst trajectories.

For instance: in retail prediction, wherein, the product sales could be impacted by general trends. Architectural development that categorize account for these hierarchies could be fascinating research way, it probably enhance the forecasting performance upon conventional multi-variate or uni-variate models. It could be concluded that, with DM (Data Mining) repositories becoming frequent, leveraging deep models that could perform automated learning from the annotated data makes DL as an attractive method.

7.1 Critical analysis

Critical analysis is an essential skill that enables individuals to think critically, make informed decisions, and contribute to the ongoing discourse in various fields. It requires a systematic and objective approach to evaluating information, as well as an open mind and a willingness to challenge one's own beliefs.

Table 5 Critical analysis of the time series data by years

Year	Papers
2018	2
2019	16
2020	18
2021	16
2022	7
2023	9

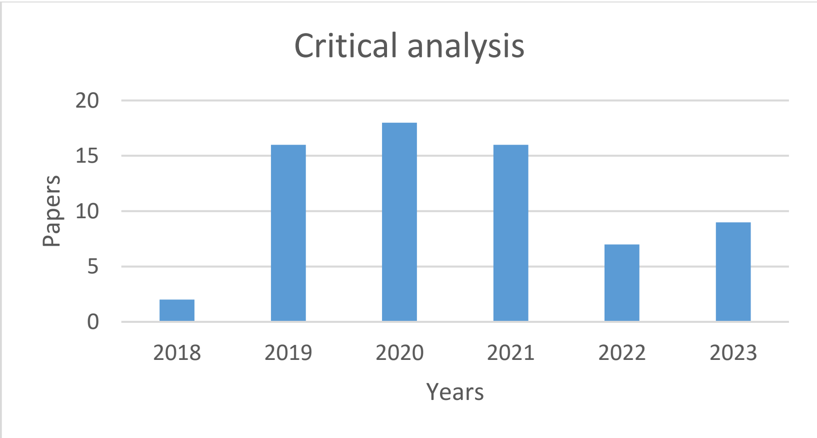

Figure 1 Critical analysis of the time series data by years

The Figure 1 and Table 5 explains about the critical analysis of the models which predicts the time-series data by the years in schematic representation. Figure 1 illustrates the critical analysis of the time data series. In 2018, two papers were published. 2019 brought with it the 16. In 2020, the journal published 18 papers. In 2021, the journal published 16 papers. In 2022, seven articles are released in the journal. In 2023, there are 9 papers that are released in the journal.

8. Conclusion

The study examined recent researches in DL based models for time-series prediction in various applications. To achieve this, the study reviewed about the different methodologies to predict time-series and the use of DL for effective prediction of time-series data. Additionally, a comparative assessment was undertaken in accordance with conventional DL models and different applications of extensively utilized DL models in time-series prediction. From the comparative analysis, it was found that, varied forms of DL models were used in diverse applications. As varied design alternatives exists and seem to be attractive, it is ideal to integrate them and consider as hybrid models. Though DL based methods have found usage in various areas, some common drawbacks prevail which is highlighted in this study so that, the researchers can consider these issues and endeavour to resolve them in future. Further, future recommendations were also afforded through this study. It was exposed that, additional deeds are required for improving the work for data with intricate inputs, thereby, benchmark them with conventional models. Moreover, the future recommendations provided in this study would serve as a roadmap for researches to bring innovations in this area for enhancing the efficacy in time-series prediction.

Declaration

Conflict Of Interest: The author reports that there is no conflict of interest

Funding: This research received no external funding.

Acknowledgement: None

Data Availability: Data sharing not applicable to this article as no datasets were generated.

References

[1] J. Viehweg, K. Worthmann, and P. Mäder, "Parameterizing echo state networks for multi-step time series prediction," *Neurocomputing,* vol. 522, pp. 214-228, 2023.

[2] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case," *arXiv e-prints,* p. arXiv: 2001.08317, 2020.

[3] S. Xu, H. K. Chan, and T. Zhang, "Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach," *Transportation Research Part E: Logistics and Transportation Review,* vol. 122, pp. 169-180, 2019.

[4] N. Wei, C. Li, X. Peng, Y. Li, and F. Zeng, "Daily natural gas consumption forecasting via the application of a novel hybrid model," *Applied Energy,* vol. 250, pp. 358-368, 2019.

[5] N. Jin, Y. Zeng, K. Yan, and Z. Ji, "Multivariate air quality forecasting with nested long short term memory neural network," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 12, pp. 8514-8522, 2021.

[6] S. d. O. Domingos, J. F. de Oliveira, and P. S. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowledge-Based Systems,* vol. 175, pp. 72-86, 2019.

[7] Q. Shi *et al.*, "Block Hankel Tensor ARIMA for Multiple Short Time Series Forecasting," 2020.

[8] S. Khan and H. Alghulaiakh, "ARIMA model for accurate time series stocks forecasting," *International Journal of Advanced Computer Science and Applications,* vol. 11, no. 7, 2020.

[9] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A methodology for applying k-nearest neighbor to time series forecasting," *Artificial Intelligence Review,* vol. 52, no. 3, pp. 2019-2037, 2019.

[10] D. S. Tripathy, B. R. Prusty, and K. Bingi, "A k-nearest neighbor-based averaging model for probabilistic PV generation forecasting," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields,* vol. 35, no. 2, p. e2983, 2022.

[11] A. Azaryan, "Machine learning approaches for financial time series forecasting," 2020.

[12] I. Boyoukliev, S. Gocheva-Ilieva, and H. Kulina, "Time series modeling and forecasting of deposits in foreign currency using CART ensemble and bagging," 2022.

[13] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons & Fractals,* vol. 138, p. 109942, 2020.

[14] H. Zhong, J. Wang, H. Jia, Y. Mu, and S. Lv, "Vector field-based support vector regression for building energy consumption prediction," *Applied Energy,* vol. 242, pp. 403-414, 2019.

[15] A. Altan and S. Karasu, "The effect of kernel values in support vector machine to forecasting performance of financial time series," *The Journal of Cognitive Systems,* vol. 4, no. 1, pp. 17-21, 2019.

[16] C.-X. Lv, S.-Y. An, B.-J. Qiao, and W. Wu, "Time series analysis of hemorrhagic fever with renal syndrome in mainland China by using an XGBoost forecasting model," *BMC infectious diseases,* vol. 21, pp. 1-13, 2021.

[17] S. Xian, K. Chen, and Y. Cheng, "Improved seagull optimization algorithm of partition and XGBoost of prediction for fuzzy time series forecasting of COVID-19 daily confirmed," *Advances in Engineering Software,* vol. 173, p. 103212, 2022.

[18] Z. Li, T. Lu, X. He, J.-P. Montillet, and R. Tao, "An improved cyclic multi model-eXtreme gradient boosting (CMM-XGBoost) forecasting algorithm on the GNSS vertical time series," *Advances in Space Research,* vol. 71, no. 1, pp. 912-935, 2023.

[19] E. K. Ampomah, Z. Qin, G. Nyame, and F. E. Botchey, "Stock market decision support modeling with tree-based AdaBoost ensemble machine learning models," *Informatica,* vol. 44, no. 4, 2021.

[20] W. Zhou, W. Wang, and D. Zhao, "Passenger flow forecasting in metro transfer station based on the combination of singular spectrum analysis and AdaBoost-weighted extreme learning machine," *Sensors,* vol. 20, no. 12, p. 3555, 2020.

[21] P. Khatri, T. Arjariya, and N. Shivhare Mitra, "Climate change forecasting using data mining algorithms," *AQUA-Water Infrastructure, Ecosystems and Society,* 2023.

[22] X. Wang, Y. Kang, R. J. Hyndman, and F. Li, "Distributed ARIMA models for ultra-long time series," *International Journal of Forecasting,* vol. 39, no. 3, pp. 1163-1184, 2023.

[23] S. Singh, K. S. Parmar, S. J. S. Makkhan, J. Kaur, S. Peshoria, and J. Kumar, "Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries," *Chaos, Solitons & Fractals,* vol. 139, p. 110086, 2020.

[24] S. K. Mahla, K. S. Parmar, J. Singh, A. Dhir, S. S. Sandhu, and B. S. Chauhan, "Trend and time series analysis by ARIMA model to predict the emissions and performance characteristics of biogas fueled compression ignition engine," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 45, no. 2, pp. 4293-4304, 2023.

[25] Z. Zhu *et al.*, "A day-ahead industrial load forecasting model using load change rate features and combining FA-ELM and the AdaBoost algorithm," *Energy Reports,* vol. 9, pp. 971-981, 2023.

[26] E. A. Siqueira-Filho, M. F. A. Lira, A. Converti, H. V. Siqueira, and C. J. Bastos-Filho, "Predicting Thermoelectric Power Plants Diesel/Heavy Fuel Oil Engine Fuel Consumption Using Univariate Forecasting and XGBoost Machine Learning Models," *Energies,* vol. 16, no. 7, p. 2942, 2023.

[27] Y. Wang, L. Wang, Q. Chang, and C. Yang, "Effects of direct input–output connections on multilayer perceptron neural networks for time series prediction: L. Wang," *Soft Computing,* vol. 24, pp. 4729-4738, 2020.

[28] F. A. Del Campo, M. C. G. Neri, O. O. V. Villegas, V. G. C. Sánchez, H. d. J. O. Domínguez, and V. G. Jiménez, "Auto-adaptive multilayer perceptron for univariate time series classification," *Expert Systems with Applications,* vol. 181, p. 115147, 2021.

[29] H. Wang, Y. Zhang, J. Liang, and L. Liu, "DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction," *Neural Networks,* vol. 157, pp. 240-256, 2023.

[30] J. Kim and N. Moon, "BiLSTM model based on multivariate time series data in multiple field for forecasting trading area," *Journal of Ambient Intelligence and Humanized Computing,* pp. 1-10, 2019.

[31] R. Wang, D. Teng, W. Yu, and X. Zhang, "Improvement and Application of GAN Models for Time Series Image Prediction—A Case Study of Time Series Satellite Nephograms," 2022.

[32] K. Wang *et al.*, "Multiple convolutional neural networks for multivariate time series prediction," *Neurocomputing,* vol. 360, pp. 107-119, 2019.

[33] N. Kimura, I. Yoshinaga, K. Sekijima, I. Azechi, and D. Baba, "Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions," *Water,* vol. 12, no. 1, p. 96, 2019.

[34] J. Xu, K. Wang, C. Lin, L. Xiao, X. Huang, and Y. Zhang, "FM-GRU: a time series prediction method for water quality based on seq2seq framework," *Water,* vol. 13, no. 8, p. 1031, 2021.

[35] W. Zheng and G. Chen, "An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization," *IEEE Transactions on Cybernetics,* vol. 52, no. 12, pp. 13902-13914, 2021.

[36] R. Wan, S. Mei, J. Wang, M. Liu, and F. Yang, "Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting," *Electronics,* vol. 8, no. 8, p. 876, 2019.

[37] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, "Temporal convolutional networks applied to energy-related time series forecasting," *applied sciences,* vol. 10, no. 7, p. 2322, 2020.
[38] J. F. Torres, A. Galicia, A. Troncoso, and F. Martínez-Álvarez, "A scalable approach based on deep learning for big data time series forecasting," *Integrated Computer-Aided Engineering,* vol. 25, no. 4, pp. 335-348, 2018.
[39] Z. Shen, Y. Zhang, J. Lu, J. Xu, and G. Xiao, "A novel time series forecasting model with deep learning," *Neurocomputing,* vol. 396, pp. 302-313, 2020.
[40] Y. Chen, Y. Kang, Y. Chen, and Z. Wang, "Probabilistic forecasting with temporal convolutional neural network," *Neurocomputing,* vol. 399, pp. 491-501, 2020.
[41] H. Hu, L. Wang, L. Peng, and Y.-R. Zeng, "Effective energy consumption forecasting using enhanced bagged echo state network," *Energy,* vol. 193, p. 116778, 2020.
[42] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach," *Expert systems with applications,* vol. 140, p. 112896, 2020.
[43] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied energy,* vol. 236, pp. 1078-1088, 2019.
[44] B. Nikparvar, M. M. Rahman, F. Hatami, and J.-C. Thill, "Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network," *Scientific reports,* vol. 11, no. 1, p. 21715, 2021.
[45] D. B. da Silva, D. Schmidt, C. A. da Costa, R. da Rosa Righi, and B. Eskofier, "DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration," *Expert Systems with Applications,* vol. 165, p. 113905, 2021.
[46] Z.-k. Feng and W.-j. Niu, "Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions," *Knowledge-Based Systems,* vol. 211, p. 106580, 2021.
[47] X. Zhao, H. Lv, S. Lv, Y. Sang, Y. Wei, and X. Zhu, "Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer," *Journal of Hydrology,* vol. 601, p. 126607, 2021.
[48] P. Hewage *et al.*, "Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station," *Soft Computing,* vol. 24, pp. 16453-16482, 2020.
[49] K. Johny, M. L. Pai, and S. Adarsh, "Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting," *Theoretical and Applied Climatology,* vol. 141, no. 1-2, pp. 1-17, 2020.
[50] M. Zekić-Sušac, A. Has, and M. Knežević, "Predicting energy cost of public buildings by artificial neural networks, CART, and random forest," *Neurocomputing,* vol. 439, pp. 223-233, 2021.
[51] M. Shahvaroughi Farahani and S. H. Razavi Hajiagha, "Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models," *Soft computing,* vol. 25, no. 13, pp. 8483-8513, 2021.
[52] S. Mehtab and J. Sen, "Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models," in *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2021*, 2022, pp. 405-423: Springer.
[53] S. Zaheer *et al.*, "A multi parameter forecasting for stock time series data using LSTM and deep learning model," *Mathematics,* vol. 11, no. 3, p. 590, 2023.
[54] J. Guo, W. Wang, Y. Tang, Y. Zhang, and H. Zhuge, "A CNN-Bi_LSTM parallel network approach for train travel time prediction," *Knowledge-Based Systems,* vol. 256, p. 109796, 2022.
[55] J. Siłka, M. Wieczorek, and M. Woźniak, "Recurrent neural network model for high-speed train vibration prediction from time series," *Neural Computing and Applications,* vol. 34, no. 16, pp. 13305-13318, 2022.

[56] E. Egrioglu, U. Yolcu, and E. Bas, "Intuitionistic high-order fuzzy time series forecasting method based on pi-sigma artificial neural networks trained by artificial bee colony," *Granular Computing,* vol. 4, pp. 639-654, 2019.

[57] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, "Well production forecasting based on ARIMA-LSTM model considering manual operations," *Energy,* vol. 220, p. 119708, 2021.

[58] Z. I. E. Cicek and Z. K. Ozturk, "Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting," *Applied Soft Computing,* vol. 102, p. 107091, 2021.

[59] C. Kocak, A. Z. Dalar, O. Cagcag Yolcu, E. Bas, and E. Egrioglu, "A new fuzzy time series method based on an ARMA-type recurrent Pi-Sigma artificial neural network," *Soft Computing,* vol. 24, pp. 8243-8252, 2020.

[60] E. Golshani and M. Ashtiani, "Proactive auto-scaling for cloud environments using temporal convolutional neural networks," *Journal of Parallel and Distributed Computing,* vol. 154, pp. 119-141, 2021.

[61] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *International Journal of Forecasting,* vol. 36, no. 1, pp. 75-85, 2020.

[62] S. Zaheer *et al.*, "A multi parameter forecasting for stock time series data using LSTM and deep learning model," *Mathematics*

vol. 11, no. 3, p. 590, 2023.

[63] G. Li and J. J. J. I. F. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion,* vol. 91, pp. 93-102, 2023.

[64] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. J. S. Salvo Rossi, "Unsupervised anomaly detection for IoT-based multivariate time series: Existing solutions, performance analysis and future directions," *Sensors*

vol. 23, no. 5, p. 2844, 2023.

[65] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, A. J. D. M. Bagnall, and K. Discovery, "The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining*

Knowledge Discovery

vol. 35, no. 2, pp. 401-449, 2021.

[66] T. Luo, "GANITE: Estimation of Individualized Treatment Effects Using Generative Adversarial Nets," 2019.

[67] B. Lim, "Forecasting treatment responses over time using recurrent marginal structural networks," *Advances in neural information processing systems,* vol. 31, 2018.

[68] R. Moraffah, M. Karami, R. Guo, A. Raglin, and H. Liu, "Causal Interpretability for Machine Learning-Problems, Methods and Evaluation," 2020.

[69] S. Li *et al.*, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," *Advances in neural information processing systems,* vol. 32, 2019.

[70] S. A. Siddiqui, D. Mercier, M. Munir, A. Dengel, and S. Ahmed, "Tsviz: Demystification of deep learning models for time-series analysis," *IEEE Access,* vol. 7, pp. 67027-67040, 2019.

[71] C. Fry and M. Brundage, "The M4 forecasting competition-A practitioner's view," 2019.