Climate Change Forecasting Using Time Series Techniques: A Comprehensive Review

Hamid Ghous[1], Aqsa Malik[2], Zubair Ahmad[2], Uzma Jabeen[2], Majid Khawar[3]

[1]Department of Computer Science, Australian Scientific and Engineering Solution, Sydney, NSW, Australia

[2]Visions, Linguistic and Machine Intelligence Lab, Multan, Pakistan

[3]Department of Computer Science, NFC Institutes of Engineering and Technology Multan, Pakistan

Abstract: One of the most crucial issues facing the world today is climate change, which substantially impacts ecosystems. It is necessary for stakeholders, policymakers, and academics to make decisions and have efficient adaptation plans to forecast climate change phenomena accurately. Techniques for examining time give smart data for getting trends, patterns, and future environment projections. This review paper provides a detailed overview of 42 research papers constructed on time-based analysis methods to forecast climate change. The picked publications cover a wide range of subjects, procedures, datasets, and modeling approaches utilized in environmental change research. This survey tries to shed light on the ongoing status of environmental change anticipating and suggests potential ways for additional examination by investigating, as far as possible significant findings of these investigations.

Keywords: Time Series Techniques, Climate Change Forecasting, Machine Learning, Deep Learning.

Email: hamidghous@gmail.com

1. Introduction

An ever-increasing issue affecting the whole world, environmental change puts the economies, society, and biological systems in danger. Precise and timely predicting climate change anticipating is turning out to be increasingly more significant as the earth goes through exceptional changes in temperature, rainfall patterns, and extreme weather happenings. It is important to grasp and convert these actions to formulate convincing measures for variations, to impact strategy decisions, and to guarantee the well-being of generations Pardo-Igúzquizaet al [1]. The study of forecasting climate change has changed intensely over the last several years, with rising attention on utilizing advanced time series techniques to model and examine complex climatic data.

Even if they are useful, traditional climate models frequently struggle to capture the complexities of actual climate dynamics. A type of statistical modeling called time series analysis provides a dynamic framework for identifying temporal correlations in climate data, allowing scientists to find patterns, trends, and feasible future scenarios.

The gradual change of the typical weather patterns that have come to define local, regional, and global climates on Earth is known as climate change. Most experts believe that human action, particularly the release of greenhouse gases (GHGs) into the atmosphere mostly from the burning of fossil fuels like coal, oil, and natural gas, is the main cause of the unquestionable melting of the global climate system. Despite Earth's climate having changed significantly during recorded history, ecosystems and civilizations like ours are finding it difficult to adjust to the current warming trend, which is occurring far more promptly than previous natural climate shifts. A few of the numerous effects of climate change include increasing sea levels, changing precipitation patterns, ocean acidification, and an increase in the frequency and intensity of extreme weather events including hurricanes, heat waves, droughts, and floods. Climate change is putting ecological systems and human societies at major risk worldwide.

Climate change represents one of the most pressing global challenges, with far-reaching consequences affecting ecosystems, economies, and societies. As a result of the obvious impacts of climate change, the requirement for accurate and reliable prediction tools increases. Implementing smart decisions and knowing climate change highly depends on accurate forecasting of temperature change. Precise forecasting is now very important for policymakers and researchers to address the socioeconomic effects of climate change, plan sustainable strategies, and efficient allocation of resources. This research explores complications of change in temperature data to deliver evidence-based decision-making by disclosing fundamental trends and patterns and increasing forecasting accuracy.

Time series methods of predicting climate change highly depend on the ability to handle gigantic amounts of climate-related data using high-resolution models of climate. This huge amount and difficulty of the climate data need scalable algorithms and parallel processing procedures. To achieve this task and meet the requirements to forecast climate change, this research highlights the connection of parallel processing, scalable algorithms, and time series techniques, including GAN and LSTM.

1.1 Importance of Forecasting

The prediction of future data based on past and present events known as forecasting is of high importance to people, organizations, and governments. Forecasting is also applied to forecast the results related to computer programs, resource necessities, system gradation and maintenance, and performance improvements. The prediction of accurate forecasting is valuable in various areas like climate forecasting, inventory supervision, sales and marketing, finance, and resource development. Precise forecasting permits organizations to recover control risk, improve decision-making, and maximize capital.

The availability of difficulty and non-linear dynamics in climate change data makes it difficult for traditional modeling techniques of climate change to completely represent. This research utilizes deep learning approaches and advanced time series methods such as Convolutional GAN (C GAN), Long Short-Term Memory (LSTM), Bidirectional LSTM, and Generative Adversarial Network (GAN) in reply. These methodologies are best suited for predicting climate change since they are very useful in capturing non-linear patterns, managing time-based dependencies, and simulating realistic situations.

1.2 Time Series Analysis in Climate Research

Time series analysis is a key method to predict climate change in climate research. This systematic procedure offers researchers, leasers, and societies a valuable data which enables up-to-date policymaking formation of effective justification and adaptation strategies. Using deep learning techniques and time series analysis approaches, scholars can accurately predict the effects of global climate variations.

1.3 Background

The area of predicting climate variations is a multi-disciplinary field of research that is rapidly evolving to improve the effectiveness and accuracy of climate change forecasting. This field of study incorporates climate science research and computer science with the use of algorithms, computational procedures, and data analysis tools. The ultimate objective of this research is to enhance our information on global climate change, assess the effects of climate variations, and form real-world strategies and adaptation plans.

Forecasting climate change is assessed and evaluated with various techniques. Some of them are machine learning methods, time series analysis, and regression analysis. Regression analysis techniques deal with assuming the relationship of variables, for example how a variable imposes effects on another, while time series analysis is concerned with data over time to explore trends and patterns. The patterns in data and prediction of results are done using artificial intelligence with the help of machine learning algorithms.

1.3.1 Increased Intensity and Frequency of Weather

The variations in climate result in regular and extreme happenings of extreme weather events for example, heat waves, storms, heavy and irregular rains, forest fires, and droughts. These happenings can significantly affect the ecosystem, societies, and infrastructure.

1.3.2 Changes in Species Distribution and Ecosystem

The variations in climate affect the ecosystem and consequently employ power to change the behavior of animal species and plants. These changes may become a challenge for specific species to adjust to these varying conditions which result deviations in biodiversity and eventually leads to the ecosystem destruction.

1.3.3 Phonological Alteration

Phenology defines the timeline of biological happenings, for example, flowering, migration, and hibernation. Unfortunately, global warming can disrupt seasonal patterns, resulting in mismatches between species' lifecycles and crucial ecological interactions.

1.3.4 Forecasting

Statistical modeling plays a vital role in predicting climate change, allowing us to examine past data on climate, discover trends and patterns, and make statistical projections in forecasting climate change. Robust statistical models, developed using computational techniques, assist the estimation of model parameters and carry out ambiguity analysis. Furthermore, the integration of deep learning and artificial intelligence algorithms with climate data is gaining importance. These algorithms help in identifying complex relationships, capturing nonlinear dynamics, and enhancing foretelling accuracy.

1.3.5 Forecasting Through Simulation

Simulation-based predictions of climate change involve the development of computer models that simulate real-world scenarios to predict future outcomes. These models utilize input parameters, statistical distributions, and mathematical formulas to simulate complicated methods and forecast their actions under various situations. Simulation-based forecasting includes weather and traffic flow simulations and financial models.

1.3.6 Ensemble Forecasting

Ensemble forecasting incorporates the outcomes of various methods or algorithms to enhance the reliability and accuracy of predictions. Ensemble predicts diminish bias by connecting forecasts from various methods, taking a broader range of possible outcomes, and providing more robust predictions.

1.3.7 Probabilistic Forecasting

Probabilistic forecasting allocates independent breaks and probabilities to various possible outcomes, coming up with the measure of ambiguity related to forecast. This kind of forecasting is generally applicable in climate predictions, risk analysis, and economic modeling to comprehend various potential future situations and their likelihoods.

1.3.8 Time Series Analysis

A statistical method called time series analysis, objects to examine and evaluate data varying with time. This includes investigating a set of data points or observations found at even intervals to identify patterns trends and other various features of data. Time series analysis can be useful in numerous areas like engineering, economics, social and environmental science. Its effectiveness lies in efficiently forecasting the upcoming events of parameters based on previous observations.

The fundamental steps in time series analysis involve data cleaning and preprocessing, examining data analysis, method recognition, method approximation, method choosing, and predicting. Time series analysis employs various models, like autoregressive integrated moving average (ARIMA), moving average (MA), and autoregressive (AR) models.

2. Literature Review

Despite the growing awareness of how urgent it is to solve climate change, scientists from all over the world are working hard to develop novel approaches that will improve our comprehension of the complex dynamics affecting global temperatures. Building on this background, the literature study that follows explores the range of information currently available for climate change forecasting. This review attempts to establish the framework for the sophisticated methods and models suggested in the following methodology by looking at previous research projects and identifying gaps in the current approaches. The synthesis of earlier efforts becomes essential for understanding and directing new contributions in forecasting techniques as the conversation about climate change heats up.

Our goal is to investigate the topic of climate change prediction, concentrating on time series methods. The objective is to provide an in-depth analysis of the most innovative methods now in use, together with an overview of their benefits and drawbacks for climate change forecasting. The approaches are divided into two categories climate change forecasting and analysis techniques.

2.1 Forecasting and Time Series Techniques

Southern Corners Water Authorities (SCWA) used the data set from 2001 to 2016 is specifically taken into consideration. It has minimal spatial heterogeneity and a slope varying between 1.75 and 2.24, with an average of 0.04. Mechanism analysis, correlation, regression, geographical overlay analysis, and scenario assumptions are some of the preprocessing approaches employed. The technique used to categorize various trends of geographical heterogeneity is trend analysis. The distinctiveness of this study is that the information was acquired from a location with a temperate continental climate and an arid and semi-arid transition zone. The application of the aforementioned approaches and techniques led to the conclusion that while SCWI showed a major downward tendency, SCWA fluctuated noticeably but not significantly Haghi Kashani, et al [2] forecasted how climate change will affect the Himalayan region using data on monthly temperature and rainfall between 1796 and 2013. Based on meteorological and international standards, the data was separated into four seasons. To find trends in the data, the study performed trend analysis and correlation analysis between variables. The study also developed and enhanced an SDLSTM model, which outperformed past studies and the imported FB-Prophet model utilizing the PWL algorithm in terms of prediction accuracy. The Himalayan states' mean annual temperature increased significantly between 1796 and 2013, according to the findings. The study recommended enhancing the computational efficiency of the CDLSTM model and focusing future research on reducing the memory bandwidth needed because of linear layers Haq, M et al [3] presented For CNN, a novel deep learning (DL) technique that leverages recent advancements in deep CNN for pattern recognition and image analysis to forecast univariate time series About 235,000 samples were incorporated in the dataset, which undergo time series preprocessing to generate image time series data. The structure of For CNN model includes constituents like encoder, generalizing CNN, and regress or. The analysis found that For CNN-SD model performed better than time forecasting techniques like Deep AR and N-Beats in terms of accuracy and speed. Deploying visual time series structures and representations from computer visualizations could enhance forecasting and accuracy.

Future studies should investigate the performance of the planned method with high-frequency data, the application of colored plots and recurrence, and the identification of time series patterns using advanced DL models from computer vision.

Buras, et al [4] proposed a newly created SIRN (Stationary and Instant Recurrent Network) technique for multivariate and univariate prediction settings in long-term time series forecasting. The dataset had several categories, including wind, exchange, and weather. To absorb the hidden states in SIRN, discover the underlying distribution, and reconstruct the target series directly in a generative manner, the study made use of normalizing flow architecture. Comprehensive empirical research was done on six real-world datasets, and the results verified Conformer's state-of-the-art performance in long-term time series forecasting. The main techniques employed in this study were the multivariate TF setup, encoder-decoder architecture, and input representation block. The implementation specifics in the study also included nine baselines with time-determined lengths Tan, F, et al [5], the hourly meteorological readings of temperature, pressure, humidity, wind direction and speed, percentage of clouds, rain and snow volume, and volume of precipitation in five major UK cities were used to construct the dataset used in this study. They employed a variety of machine learning models to predict rainfall after preprocessing the data, which included filling in missing values and feature engineering. These models included XGBoost, automated machine learning, and LSTM networks. They discovered that a Bidirectional-LSTM network outperformed a Stacked LSTM network with two hidden layers in terms of performance. To enhance the efficacy of the models, future studies will focus on fine-tuning parameters and examining the significance of features and other meteorological phenomena Spiliotis, et al [6] worked on the dataset consisting of 234 locations' worth of graphical and picture data on European beech woods. Climate data from CHELSAcruts were combined with a forecast growth model. A tree-ring network was the methodology used, and it led to the discovery of a common growth rate reduction between 1955 and 2016. Except for sites in Denmark, Norway, and Sweden that are closer to the distribution's northern limit and those that are higher up in mountainous areas, this drop was widespread throughout Europe Linden, et al [7] utilized a dataset consisting of 35 papers from different scholarly databases, including Google Scholar, IEEE, Science Direct, Springer, Taylor & Francis, Wiley, Inderscience, Sage, Emerald, World Scientific, and Hindawi. The paper selection process involved multi-label classification and multi-objective optimization. The study employed Artificial Fish Swarm Algorithm (AFSA) along with non-heuristic and heuristic algorithms. The results of the study showed that AFSA achieved 24.2% accuracy, 16.5% scalability, 12.1% MSE, RMSE, precision, and 5.5% reliability. The study's novelty was its focus on scalability and integration for accurate forecasts. The future work includes analyzing data from before 2014 and after 2020 and developing a data mining algorithm for a Map Reduce solution Ustrnul, et al [8] used several datasets, including data from synoptic process data from 2015 to 2020, Data from air quality monitoring stations from 2016 to 2020, and data from ground meteorological stations from 1981 to 2020. The study used time series analysis and machine learning methods. And performance analysis to examine temperature-related data. The findings demonstrated that PDK values were exceeded for NO2, NO, and HCOH during 80-90% of the heating duration. The study's future work involves conducting numerical analyses and using machine learning and deep learning approaches.

The dataset in Oyedele, et al [9] included a single treatment group and multiple-group compared data. The preprocessing process involved single-group ITSA, multiple-group ITSA, and multiple treatment periods. The study employed ITSA, multiple-group analysis, single-group ITSA, and multiple-group ITSA. The results showed that ITSA could easily estimate complex models by including additional covariates Aggarwal, et al [10] used the dataset made up of Web of Science citations, English-language publications using interrupted time series techniques, and MEDLINE keywords. Preprocessing entailed determining that interrupted time series analysis was being used in more applications. The research used linear regression, ARIMA models, and segmented regression. The results showed that interrupted time series analysis could be applied to health care intervention and methodological and reporting standards development. The study's novelty was its presentation of interrupted time series analysis for ARIMA. The future work includes assessing data variability and appropriateness, identifying relevant ceiling or floor effects and outliers, and examining the direction and magnitude of the results Farmonov et al [11] had a dataset consisting of 500 scientific articles published since 2018, with 250 papers related to numerical and weather prediction and 250 papers related to machine learning and climate? Preprocessing involved using K means clustering. The study employed tidy text and R package and reported results for methods such as RF, XGB, ANN, and DL. The study's novelty was in exploring the intersection of meteorology and synoptic climatology Lu, et al [12] studied a dataset consisting of rainfall measurements from 2020, crop data from Hungary, six months of satellite images of the Sentinel-2, and sunflower-yielding data obtained from Sentinel-2 satellite imagery. Preprocessing involved using image processing classification and selection techniques. The study employed Random forest regression and spatial prediction with TSA. The results reported RMSE values ranging from 121.9 and 284.5 kg/ha for regression. The study's novelty was in the development of a robust RFR model. Future work includes pixel-level identification Iglesias et al [13] used a dataset consisting of power spectral values and time series data obtained from spectral analysis techniques. Preprocessing involved using the Spectral LOMB-Scargleperiodo gram (SLOMBS) permutation test for confidence levels. The study employed multivariate time series analysis, terrain analysis, and spectral analysis methods. The study's novelty was in the development of module-based algorithms, "LR04" stack, and "LR04" unturned. The outcomes demonstrated that the code helps validate and/or calibrate the results obtained from time-frequency approaches, like the wavelet.

Rakholia et al [14] used a forecasted time series dataset on neural architecture where they used various models including AR, MA, ARMA, ACF, PCF plots, and ARIMA for the preprocessing process, and methods used are Independently Recurrent Neural Networks (IndRNN), Long Short Term Memory (LSTM), Recurrent Neural Network (RNN), and ANN, the optimization is done with the development of Time series pipeline optimization (TSPO) which shows that Auto ML can automatically find a reasonable ML model and optimize Auto ML frameworks. Developed models are still limited by the empirical study.

Thapliyal et al [15] the dataset comprised air quality monitoring data from HCMC, AQMN, Meta information such as location and timestamp, statistical summaries, technical specifications, and exploratory data. The preprocessing involved time series data cleaning, measurements of elements, outlier elimination, and collection of raw air pollution data. Methods used included correlation analysis, N-Beats architecture, neural network-based architecture, and random search. The novelty of this study was in the development of a forecasting model for five climatic variables. The results showed that the developed model had an average RMSE value lower than the average standard deviation, with stable accuracy. Health and asthmatic recommendations were also provided. Future work includes forecasting NO2, CO, SO2, and O3 simultaneously with a single model, dispersion model, air pollution forecasting model, and machine learning techniques.

S. M et al [16] used weather datasets from five cities in the UK to predict rainfall volumes using machine learning techniques. Hourly measurements of temperature, pressure, humidity, wind direction and speed, percentage of clouds, rain, and snow volume were all included in the dataset. After preprocessing, including the use of bidirectional-LSTM networks for feature selection, the team applied various machine learning models, including LSTM-Network, stacked-LSTM network, XG Boost model, and TPOT tool, to predict rainfall. The bidirectional-LSTM network outperformed the stacked-LSTM network, with room for improvement to reduce the gap between predicted values and observed rainfall volumes. Future work includes the use of hybrid and ensemble modeling of LSTM-Networks and the Grey Wolf optimizer. Rakholia et al [17] used wheat yield data along with meteorological data and two drought indices, SPEI and SPI, to investigate the impact of drought on wheat cultivation. Pearson's correlation was used to compute the drought indices, and trends in wheat cultivation were analyzed. The study also presented a novel approach for SPEI time-series meteorological representation. The results showed a reduction in dependence on fossil fuels from 80% to 50% to achieve negative carbon emissions. Further work involves the use of indices for quantitative assessment.

Soomro et al [18] analyzed time series meteorological data from 190 countries and 37 other territories from 1961 to 2019. The data were preprocessed by analyzing temperature changes around the world, normalization, and standard deviation. The authors employed several machine learning methods, including Bayesian Ridge, Random Forest, Light Gradient Boosting Machine, Extra Trees, and k-Nearest Neighbors. The Extra Trees algorithm performed the best with an RMSE of 0.3998 C◦ in 47.62 seconds, while Bayesian Ridge had the weakest RMSE of 0.5293 C◦. The study found that the highest error in worldwide temperature change was 3.5 ◦C, while the highest error of the algorithm was 5˚ C.

Li, B et al [19] discussed a dataset consisting of 1200 wind power series data. Preprocessing was done using the Beta model, PSO model, and LSTM model. The study employed the Beta-LSTM model, LSSVM model, Beta IM, and Beta PSO to analyze the data. The novelty of the paper is the combination of the Beta-PSO-LSTM model. The results showed that the Beta-PSO-LSTM model achieved the best results in 4 out of 6 indexes. Future work includes predicting intervals in power systems applications and integrating the model into the optimal scheduling problem of the hydro-thermal-wind system.

Olufemi et al [20] described a study on climate data processing and analysis in Pakistan using monthly precipitation, maximum temperature, relative humidity, and wind speed data from 1990 to 2017. The data was obtained from the Pakistan Meteorological Department and the Climate Data Processing Centre. Breakdown was a phase in the preprocessing procedures. The analysis techniques employed were the Autoregressive Moving Average (ARMA) / Autoregressive Integrated Moving Average (ARIMA), the AR-Model, and the Augmented Dickey-Fuller (ADF) Test. The study was innovative in that it assessed the model's performance using RMSE, MAE, and MAPE. The AR model had the lowest performance indicators, according to the findings. Different studies found that Pakistani climate data may be reliably forecasted by the application of AR models.

Wang, Y et al [21] Enclosed is a database consisting of regular data on temperature from 2017 to 2018 and wind from 2011 to 2021. The preprocessing techniques used in it were nonlinear mapping, support vector regression (SVR), and primal-dual algorithm. The techniques employed are long short-term memory (LSTM), neural networks (RNN), and random forests (RF). The findings exhibit the usefulness of ARIMA with LSTM and the deployment of SVR for short-range predictions of temperature change.

Utilizing monthly meteorological data, M. H. et al [22] from 2010 to 2020 across four sites within the HelioClim-1 dataset, the Shapiro-Wilks test and ARIMA models were employed for data preprocessing. Comparative statistical analysis was conducted using the MERRA-2 dataset, distinguishing this study through its innovative use of MERRA-2 as a reference data source. The results, with a 98% confidence level, indicated a correlation of 65.59% and an average temperature of 24.29°C.

A study was conducted H. A et al [23] using a sample size of (731 × 2) bi-daily, leveraging forecasting data from the GRAPES-3km model over the period from January 2019 to December 2020. The dataset included station location information; model forecast data, and observational records. Data preprocessing involved interpolation, standardization, and the removal of outliers and missing values. Three methodologies were employed: Fully Convolutional Network (LSTM-FCN), Light GBM, and Linear Regression. The state-of-the-art findings resulted after the usage of the LSTM-FCn model based on GRAPES-3.

The results demonstrated that the accuracy of the LSTM-FCN model was 21.5% at 1.92°C, the RSTM method was 21.4% at 1.95°C, and the Light GBM model was 25.2% at 1.48°C.

Shah et al [24] led to a thorough study of atmospheric temperature, solar radiations, wind track, barometric pressure, and humidity ambiguity data gathered from August 2017 to August 2018. With the authorized dataset, various machine learning methods were applied for analysis comprising of Convolutional Neural Network, Logistic Regression, K-Neart Neighbors, Support Vector Machine, Support Vector, and Random Forest Classifier. The research innovation lies in the assessment of various performance metrics, for example, MBE, MABE, RMSE, and rRMSE. The outcome indicated a MAPE of 0.18 and a range of RMSE from 6.96 to 37.5 MJ/m2.

Shah et al [25] conducted a study of predicting energy demand having electricity load and climate data. The exploratory data analysis did the preprocessing of data. The study employed two approaches named Generalized Regression Neural Network (GRNN) and Support Vector Regression (SVR). According to the study the GRNN model was being particularly innovative. The results demonstrated that the SVR model had a Root Mean Squared Error (RMSE) of 44.40 and a correlation coefficient (CC) of 0.965, while the GRNN model showed a lower RMSE of 28.82 and a CC of 0.956. The study suggests that future research should aim at improving accuracy, exploring deep learning

Paul et al [26] the research utilized CMIP5 and CMIP6 datasets covering the periods 1850 to 2005 and 1850 to 2014, respectively. Data preprocessing was conducted through super ensemble techniques. The methodologies employed include Mediterranean hotspot analysis, projection validation, weighting procedures, and quantification. The novelty of the study lies in the use of High Res MIP for temperature and precipitation projection. The results show a decline ranging from -49% to -16% in CMPI6 and -47% to -22% in CMPI5. Future work involves identifying the weighting method.

Teixeira et al [27] Used MLR, Wavelet-based method, WT, and MRA-based DWT approaches, a dataset of Indian annual rainfall less than 1,000 mm from 2,500 mm portions spanning from 1901 to 2014 was preprocessed. Traditional Multiple Linear Regression (TMLR) models, Wavelet-based Multiple Linear Regression (WMLR) models, and the Pyramid algorithm are among the techniques used. Significant predictor factors and MRA-based MLR for the dataset were found by the investigation. On the original time series, it was discovered that the WMLR method performed better than the conventional TMLR model.

Ayyoub [28] Tree ring data from ten sites, ranging from latitudes 9.15° to 61.30°N, were used in the study. The GLS model was utilized to incorporate the main climatic drivers, which were determined by using the Climwin program. The data was analyzed using the generalized least squares model.

Li, D. et al [29] a fuzzy logic model was applied to monthly rainfall data collected from 26 weather stations spanning from 1977 to 2020. The data was preprocessed by cleaning it and removing inconsistencies, and outlier data. The model achieved an accuracy of 100 RMSE. Future work includes dealing with large data volumes.

Zhang, et al [30] the study makes use of actual greenhouse data, including illumination, indoor humidity, and interior temperature. RMSE and matrix factorization are used for preprocessing. The work suggests a novel model, GCP_lstm, and compares the working of the LSTM and GRU methods. The GCP_lstm model demonstrates robustness when dealing with anomalous data. Future research will focus on enhancing forecast accuracy, climate prediction, and sensor anomaly detection.

M. S., et al [31] the study used daily precipitation and temperature data from CPC, along with 14 climate indices, to develop a PRCTOT model. This approach used GCMs and Eta to analyze model uncertainty, as well as initial and boundary conditions and scenario uncertainty, to understand the internal variability of the climate system.

Hashizume et al [32] provides an overview of a study that involved articles published in PubMed. Duplicate articles were eliminated during the process, and the remaining ones were screened according to meteorological conditions. 33 publications in total were chosen, and it's possible that the PRISMA flow was used to guarantee a systematic review. The researchers used predefined lag times, which were frequently backed up by a review of the literature. Interestingly, the immunological population was seldom included in the study model. The results point to possible weaknesses in the treatment of immune-related issues. The synopsis also suggests the potential for additional disease exploration and research. The study also recommends using different techniques for the analysis than Generalized Additive Models (GAMs) and Generalized Linear Models (GLMs).

Russell et al [33] Makes use of the ERA5 Dataset from ECMWF, which is a limited 13-year period of data interpolated onto a spherical grid. Convolutional Neural Networks (CNN), Convolutional Long Short-Term Memory (ConvLSTM), Variational Auto encoders (VAE), SAVP models, Generative Adversarial Networks (GAN), and a hybrid of VAE and GAN are some of the sampling techniques used in the research, which entails building cross-validation experiments. According to the study, an iterative CNN beats forecasts with a 4-hour lead time. Additionally, it shows that, in contrast to other models, the SAVP method more accurately represents global variability as determined by the Structural Similarity Index (SSIM). Local spatial variability, however, only slightly outperforms the ConvLSTM model. The study compares deep neural networks with conventional Numerical Weather Prediction (NWP) models in the forecasting area and concludes that newer NWP models are still significant.

R. P et al [34] a study was primarily concerned with predicting financial and economic statistics about the Brazilian Stock Market. Important international indices are included in the study, including the Nikkei (Japan), DAX (Germany), Hang Seng (Hong Kong), S&P500 (US), and FTSE100 (UK). The Independent and Identically Distributed (IID) pattern of data is followed in the study, which investigates the usage of nonlinear models such as regression trees, random forests, deep neural networks, and shallow neural networks. The research includes approaches like as bagging and complete subset regression, including both linear and nonlinear methodologies. The empirical data of the study specifies that machine learning methods work extraordinarily for economic prediction when connected with large data. It's interesting to note that neural networks (NNs) beat the benchmark in this situation; however, Random Forests (RFs) did not. The research deeply investigates the application of nonlinear factor models, auto encoders, Support Vector Regression, and many other methods to enhance the accuracy and effectiveness of economic forecasting.

To globally measure changes, the study [35] Baggio et al emphasized distinguished indices for example Nikkei (Japan), FTSE 100 (UK), DAX (Germany), Hang Seng (Hong Kong), and S&P 500 (US). Scholars deploy independently and identically distributed (IID) data to evaluate the performance of various methodologies including regression trees, linear and nonlinear models, random forests, and shallow and deep neural networks. Despite the poor performance of Random Forests, empirical data shows that nonlinear machine learning models are quite suitable for financial predictions when connected with big data. Neural networks significantly exceed the standard, and this complex analysis also makes use of other cutting-edge techniques like auto encoders, Support Vector Regressions, and nonlinear component models. Techniques for bagging and complete subset regression are examined, emphasizing the need to combine linear and nonlinear methods to provide more accurate predictions.

Abbas et al [36] eleven important crops were produced in Pakistan between 2000 and 2019 and their production was studied. Numerous econometric analytical techniques are employed in the study, including the Panel Pooled Mean Group (PMG) approach and panel unit root analysis. Sensitivity analysis is combined with the applications of Fully Modified Ordinary Least Squares (FMOLS) and Dynamic Ordinary Least Squares (DOLS). The study concludes that rising temperatures have a major negative impact on Pakistan's agricultural productivity. It's interesting to note that the research indicates there is no appreciable difference in crop productivity when utilizing higher-quality seeds. The study provides a comprehensive examination of the meteorological factors influencing agricultural production in the region and examines how temperature variations in the Himalayan and Karakoram mountain ranges affect crop yield.

Qasim, et al [37] conducted a thorough analysis of fifty-five papers drawn from a variety of databases WOS, Google Scholar, Scopus Index Journals, Emerald, Elsevier Science Direct, Springer, and Sciverse are summarized in the summary. 40 articles were excluded from the total of 95 using a multistage procedure including broad review, identical removal, and content-based filtering. The study examines sector-specific policies and weather mitigation used in literature, by the application of the International Renewable Energy Agency's (IRENA) model. The significant results include the observation of decay in cereal efficiency, endorsed features like the introduction of new crop yields, seasonal variability’s, and other input concerns. Furthermore, the research explored socioeconomic adaptation policies, offering a thorough analysis of complicated aspects influencing agricultural outcomes relative to mitigation efforts and weather irregularity.

Elnour et al [38] gives a broad summary of research projects related to Building Automation and Management Systems (BAMSs) from January 2015 to February 2022. Data was sourced from well-known research databases including Wiley, IEEE, Scopus, and Elsevier. The paper incorporates various strategies and approaches such as clustering, dimensionality reduction, classification, and regression. The study finds the application of deep neural networks in artificial intelligence, focusing on Bidirectional LSTM (BiLSTM), Long Short-Term Memory (LSTM), and Convolutional LSTM (ConvLSTM). Furthermore, the study incorporates Technology Readiness Levels (TRLs), Decision Readiness Levels (DRLs), and statistical models, and assembles techniques like boosting and bootstrapping. The integration of state-of-the-art technologies such as edge, fog, and hybrid computing broadens the study's span. It identifies three key factors influencing the future development of BAMSs: machine learning (ML), enhanced connectivity capabilities, and the Internet of Things (IoT). The study also finds advancements in equipment standards, and building automation systems and highlights significant issues including market penetration, industry competition barriers, security and privacy protection, regulatory compliance, scalability, and interoperability.

Carvalho et al [39] summarizes the study that incorporates information provided by the International Panel on Climate Change (IPCC) regarding greenhouse emissions. Investigation by Scopus and Web of Science are taken into account, with an importance on getting together based on commonalities. Standardized approaches are used to give a thorough analysis of various methods deployed in the study. The study makes use of content analysis and systematic review methodologies. The findings indicate that there is a significant amount of uncertainty in climate projections, which has an impact on estimates for renewable energy. Consequently, the study recommends enhancing the quality of input data, refining the precision of the climate model, and evaluating the system's overall capacity to generate electricity. Through the resolution of ambiguities and the improvement of climate projection accuracy, this approach aims to clarify the complex relationships among renewable energy sources, greenhouse gas emissions, and climate change.

Meng, J, et al [40] a study that focuses on climatic analyses and uses data from multiple sources, such as CIRI, EIRI, AIRI, and NOAA. Using methods like cross-correlation functions and time-delayed functions, the NCEP-NCAR analysis is used. In particular, the study uses complex network (CN) construction to create dynamical and physical climate networks based on the global near-surface air temperature field. It also builds in-degree time series. The study attains a forecast skill of 0.5, which is on track with the most advanced dynamical and statistical forecast models. Notably, the network's topology has changed and an increase in the in-degree index has been seen. The paper presents the superiority of the ISMR prediction model, demonstrating its superior performance over operational forecast models. The words Network ENSO and ENSO monsoon are introduced as the research examines the connections between climatic phenomena. The methodology also takes the SWAS index into account. All things considered, the work advances our understanding of climate dynamics by utilizing cutting-edge analytical techniques and putting forth a unique ISMR forecast model.

V. H et al [41] investigated Dengue Fever (DF) data from 1997 to 2016 that were collected from the National Institute of Health and Environment (NIHE). A dataset with 12 meteorological parameters that were gathered at the same period is included in the study. Data cleansing, feature collecting, performance assessment, and epidemic detection are all steps in the study process. Some of the models used for detection and prediction include Convolutional Neural Networks (CNNs), Transformer models, Long Short-Term Memory networks (LSTM), Seasonal Autoregressive Integrated Moving Average (SARIMA), Poisson regression, Support Vector Regression (SVR), and an enhanced version of SVR called SVR-L. We present and demonstrate the prediction performance improvement of the LSTM with Attention mechanism (LSTM-ATT). The study shows how accurate LSTM-ATT is at predicting epidemics of dengue fever. Wider applications of the LSTM-ATT model in public health research are indicated by the recommendation to utilize it for the prediction of additional climate-sensitive diseases like influenza, malaria, and diarrhea.

Arjariya et al [42] reported using 113 years of annual climatic data from Varanasi, India. Many time series forecasting methods are applied, including moving average, exponential smoothing, autoregressive integrated moving average (ARIMA), k-nearest neighbor (kNN), spline interpolation, and cubic spline interpolation. The study looks into the application of hybrid models, which combine statistical models with interpolation techniques, and standalone models, such as kNN and ARIMA. Comparative research is conducted to evaluate the working of individual and hybrid methods. The analysis demonstrates that the ARIMA model has extremely good error measurements, suggesting that Varanasi's annual climate data may be predicted with high accuracy. The findings emphasize how important it is to double-check these models, especially when working with time series data. The study highlights the need to select appropriate models for accurate forecasts and enhances our understanding of various modeling strategies for climate data forecasting.

2.2 Literature Review Table

In this chapter, we have seen the work done in the field of forecasting and the use of time series techniques. Still, there is minimal work done on the use of time series techniques. Here is the data table for the literature review.

Table: 1 Meta Data Table

Ref	Dataset	Preprocessing	Methods Involved	Novelty	Results	Future Work
[1]	-Wind projection(CMIP6) -Avg wind speed -Near-surface wind data -SSP2-4.5 and SSP5-8.5 -All satellite image data (sentinel) -Monthly wind power	-MME -Data validation -GCMs -60% similar points merger	- NorESM2-MM -CESM2-watch - GFDL-ESM4 -GFDL-CM4 -EC-Earth3	-	-General increases of 15% and 30% predicted in SSP2-4.5 and SSP5- 8.5 -	-Regional level studies -New policies till 2100
[2]	-HCMC air data -AQMN -Meta info (location and timestamp) -Acquired data -Statistical summaries -Technical specifications -Exploratory data	-Time series data -Data cleaning -Measurements of elements -Outlier elimination -Data collection of raw air pollution	-Correlation analysis -N-Beats architecture -Neural network-based architecture -Random Search	-Forecasting model developed for 5 climatic variables	-Average RMSE values lower than the average standard deviation -Stable accuracy -Health and asthmatic recommendations	-Forecasting NO2, CO, SO2, and O3 simultaneously with single a single model -Dispersion model -Air pollution forecasting model -Machine learning techniques
[3]	-Weather dataset -OpenWeather data -Data from (January 1, 2000, to April 21, 2020) -Hourly recorded values of (Temperature, Pressure, Humidity, Wind speed & Direction) -Percentage of clouds -Volume of snow, & Volume of rain	-Data separation by city -Nullness percentage of features -Sea & ground level measurement removal -ISO features removal - Weather dataset (43-dimensional feature vector structure) -Correlation Matrix & feature selection	- Rainfall prediction approach - LSTM-Network -Stacked-LSTM -Network -Bidirectional-LSTM Networks -AutoMl tools -XGBoost model -TPOT tool	-Bidirectional LSTM model	-Bidirectional-LSTM Network as an alternate for StackedLSTM Network -Stacked-LSTM Network presented the worst performance	-Gap b/w predicted values & observed rainfall volumes -Calculation of lag features -Forecasting with fuzzy time-series -Hybrid methods of LSTM-Networks -Grey wolf optimizer
[4]	-Wheat yield data -Meteorological data -SPI drought index -SPEI drought index	-Pearson's correlation -Computation of SPEI & SPI drought index -Wheat cultivation trends	-SPEI time-series meteorological representation -Pearson's correlation -Computation of SPEI & SPI drought index -Correlation analysis	-SPI drought trend data	-Reduction dependence on fossil fuels from 80% to 50% to achieve negative carbon emission	-Use of indices (Quantitative assessment)
[5]	-Time Series Data (1961 to 2019) -Meteorological data (190 countries and 37 other territories	-Temperature change around the world -Normalization data -Standard Deviation	- Extra tree (ET) - Light gradient boosting machine - Random forest - k nearest neighbors - Gradient boosting - Bayesian ridge		-Extra Trees RMSE = 0.3998 C◦ in 47.62 seconds, -Highest error was 3.5 ◦C -Bayesian Ridge algorithm weakest RMSE = 0.5293 C◦ The highest error of the algorithm is 5 ◦C	
[6]	-1200 wind power series data	-Beta model -PSO model -LSTM model	-Beta-LSTM model and LSSVM model -Beta IM -Beta PSO	-Combination of Beta-PSO-LSTM model	- Beta-PSO-LSTM model - Best result 4/6 indexes	-Prediction interval in power systems applications. -Integration in optimal scheduling problem of hydro-thermal-wind system
[7]	-CDPC - PMD -Monthly data (precipitation, maximum temperature (°C), humidity (knots) & wind speed (Karachi, Lahore, Peshawar, Islamabad & Quetta) 1990–2017)	- Decomposition	- AR-Model - Augmented Dickey-Fuller (ADF) Test - Autoregressive Moving average ARMA / Autoregressive Integrated Moving average ARIMA	-RMSE -MAE -MAPE	-Chi-Square test results -5% significance (n-1) -AR models have min least performance measures	
[8]	-Wind Database (2011-2021) -One year of daily temperature data (2017-2018)	-Support Vector Regression (SVR) -Nonlinear mapping -Primal-Dual Algorithm	-Random Forest (RF) -LSTM -RNN		-ARIMA with LSTM -SVR for short-term temperature forecast.	
[9]	-Monthly rainfall, air pressure, temperature, moisture, wind speed & direction (four stations from HelioClim-1) - HelioClim-1 data (2010-2020)	-ARIMA - Shapiro-Wilks test	-Statistical analysis	- MERRA-2	-Similarity 65.59 %, with temperature 24.290C -98 % confidence level	
[10]	-GRAPES-3km model prediction data -January 2019 to December 2020	-Outlier & missing values removal -Data interpolation -Data standardization	-LightGBM -Linear Regression -Fully Convolutional Network (LSTM-FCN)	-LSTM-FCN model on GRAPES-3	-RSTM (21.4% at 1.95 ◦C) -LSTM-FCN (21.5% at 1.92 ◦C) -LightGBM (25.2% at 1.48 ◦C)	
	-sample size of (731 × 2) twice a day -Observational data -Model forecast & station location data					
[11]	-Solar radiation (Aug 2017 to Aug 2018) -Air temperature -Wind direction -Humidity uncertainty -Barometric pressure	-Dataset Validation -Machine Learning Techniques	-K-Nearest Neighbors -CNN -Support Vector Machine -Logistic Regression -Random Forest Classifier -Support Vector Classifier	-(RMSE) (rRMSE) (MBE) (MABE)	-RMSE 6.96 to 37.5 MJ/m2 -MAPE 0.18	
[12]	-Electricity Load Data(208 & 2019) -Weather Data (ECMWF)	-Exploratory data process -	-GRNN -SVR	-Generalized Regression Neural Network (GRNN)	-GRNN model (CC value of 0.956) -RMSE (value of 28.82) -SVR (CC value of 0.965) -RMSE value of 44.40.	-Improvement in accuracy -Deep learning models -Correlation between parameters -New technique for feature selection
[13]	-CMIP5 and CMIP6 (1850 to 2005 & 1850 to 2014)	- Super ensembling	- Projections verification - Mediterranean hotspot evaluation - Weighting method -Quantification	- HighResMIP -Temperature & precipitation projection	-Decline ranges from -49% to -16% in CMPI6 - Decline ranges from -47% to -22% in CMPI5	-Weighting method identification
[14]	- Indian annual rainfall less than 1,000 mm from 2,500 mm parts (1901 to 2014)	- MLR - Wavelet-based approach -WT -MRA based DWT	-Traditional Multiple Linear Regression (TMLR) models -Wavelet-based Multiple Linear Regression (WMLR) models -Pyramid algorithm	-Significant Predictor Variables -MRA-based MLR	-WMLR model outperforms traditional TMLR model on original time series	
[15]	-Tree ring data -10 sites ranging 9.15° to 61.30°N latitude	-Identifying primary climate drivers (Climwin) -Combining climate drivers (GLS model)	-Generalized least square model			
[16]	-Monthly rainfall data(1977 to 2020) -26 weather stations	-Data cleaning -Inconsistencies removal	-A fuzzy logic model		-Accuracy (100 RMSE)	-Large data volumes

		-Outlier data removed				
[17]	-Greenhouse real data -Indoor temperature -Indoor humidity -Illumination	-Matrix factorizations -RMSE	-LSTM -GRU	-GCP -LSTM	-Performed well with abnormal data -Robustness	-Accuracy of prediction -Climate prediction -Sensor anomaly detection
[18]	-SAT(HadCRUT4) -Thetae_sfc	-OLR Trend -PR Trend	-WBGT -Reanalysis			
[19]	-Daily precipitation, minimum and maximum temperature data (CPC) -14 climate indices		-PRCTOT (Precipitation & Temperature)	-GCMs and Eta	-Model uncertainty -Initial and boundary conditions -Scenario uncertainty -Internal variability of the climate system.	
[20]	-SCWA 2001 to 2016 (low heterogeneity) -Slope range 1.75 to 2.24 (Average value of −0.04)	-Mechanism analysis -Correlation & Regression model -Spatial overlay analysis	-Trend analysis	-Arid & semi-arid transition zone -Temperate continental climate	-SCWA fluctuated -No significant change -SCWI's downward trend	
[21]	-Rainfall (monthly rainfall dataset) -Temperature (Jan 1796–Aug 2013) - Time series 1352 monthly values	-Data division (Meteorological & International)	-Trend Analysis - Correlation Analysis -Tuning CDLSTM Model -LSTM Model	-High memory bandwidth - CDLSTM model -FB-Prophet model -PWL algorithm	-Trend Analysis increased around 1.07◦ C, 1796–2013). -Correlation (0.98) - Temperature Forecasting (CDLSTM and FB-Prophet) - Rainfall Forecasting (CDLSTM)	- CDLSTM model development -Computational efficiency improvement
[22]	-Visual time series data, 2D images., 1D numeric vectors -Approx 235,000 23,000 yearly time series samples	-Time series pre-processing -CNN -Training DL network	-Encoder -Regressor -Generalizing of CNN	-CNN	-ForCNN-SD faster than N-BEATS &DeepAR by approx 24 and 23 h	-High-frequency data. -Recurrence and colored plotting -DL models for time series analysis
[23]	-ECL1 -Weather -Exchange -ETT -Wind -AirDelay -9 Baselines -Time-Determined Lengths	-Input representation block -Encoder & decoder architecture -Normalizing flow	-Multivariate TF setting	-SIRN	-A normalizing flow framework -Wide empirical studies on six real-world datasets -Long-term time-series forecasting	
[24]	-5 Major cities of the UK (Hourly climatic measurements) -Hourly recorded weather measurements of climate variables with the percentage of clouds, the volume of rain and snow. -Dataset records of city name, latitude and longitude coordinates time zone code.	-Separate weather data -Nullness percentage -Sea and ground-level measurements removal. -Iso features -Removing missing values of rain -Separate weather condition codes. -Correction of weather condition codes. -Compute encoding for weather ID codes. -Structuring feature vector rows - Training, validation, and testing sets	-Correlation Matrix & feature selection -Rainfall prediction approach -XGBoost model -Automated machine learning -LSTM and Stacked-LSTM	-	-Bi-LSTM Network used as a rainfall forecast method -S-LSTM Network with 10 hidden layers presented the worst performance	-Fine-tuning the parameters and hyperparameters -Comprehensive analysis of features and the inclusion of other weather factors
[25]	-Image & graphical data (European beech forest from 234 sites)	-Predictive growth model - Climate variables CHELSAcruts	-Tree-ring network		-Growth rate decline from 1955 to 2016	
[26]	35 papers from multiple articles databases -Google Scholar -IEEE -Science Direct Etc.	-Paper Selection -Multi-label classification -Multi-objective optimization -	-Artificial Fish Swarm Algorithm (AFSA) - Non-heuristic and heuristic algorithms.	- Scalability - Integration & accurate forecasts -	-24.2%, accuracy with 16.5%, & scalability with 12.1%, MSE, RMSE, precision, & reliability with 5.5%	- Data before 2014 & after 2020 - Data mining for MapReduce solution
[27]	-Ground meteorological station data, 1981 to 2020 -Air quality monitoring data, 2016 to 2020 -Manual air quality monitoring station data, 2015 to 2020 -Synoptic process data, 2015 to 2020 - Meteorological Station Data - Data on temperature	-	- Time series analysis - Machine learning algorithms - Performance analysis -	-	-Approx 80%–90% of the heating period exceeded PDK values, particularly for NO2, NO, & HCOH.	-Conduction of numerical analyses -ML and deep ML approaches
[28]	-Single treatment group -Multiple-group compared data	- Single-group ITSA - Multiple-group ITSA -Multiple treatment periods	- ITSA -Multiple-group analysis -Single-group ITSA -Multiple-group ITSA	-	- With ITSA, complex models may be estimated with ease by adding more factors.	
[29]	-MEDLINE keywords -Web of Science citation -English language articles (interrupted time series methods)	-Increase in the number of applications of interrupted time series analysis.	-Segmented regression -ARIMA models -Linear regression	-Interrupted time series analysis for ARIMA	-Healthcare intervention -Methodological and reporting standards development	-Data variability and appropriateness -Relevant ceiling or floor effects & outliers -Direction and magnitude
[30]	-Database 500 scientific articles data (published since 2018) -Numerical and weather prediction (250 papers) -Machine learning (250 papers) -Climate and machine learning (250 papers)	-K means clustering	-tidy text -R package	-Meteorology and synoptic climatology -	-Methods (RF, XGB, ANN, and DL)	-
[31]	-Rainfall measurements (2020) -Crop data from Hungary	-Image processing classification & selection techniques	- Random forest regression -Spatial prediction with TSA	-RFR robust model (developed)	-Regression RMSE values ranging from 121.9 and 284.5 kg/ha.	- Pixel level identification

	-Six months of cloud-free satellite images of the Sentinel-2 -Sunflower yielding data (Sentinel-2 satellite imagery)					
[32]	- Power spectral values -Time series data from spectral analysis techniques	- Spectral LOMB-Scargleperiodogram (SLOMBS) permutation test for confidence levels	- Multivariate time series analysis - Terrain analysis - Spectral analysis	-Module-based developed algorithms - "LR04" stack - "LR04" unturned	-Code is useful for authenticating the results derived from time-frequency techniques, such as the wavelet, to refine and/or calibrate	
[33]	-Forecasted time series dataset on neural architecture.	-AR model -MA model -ARMA model -ACF and PCF plots -ARIMA model	- Independently Recurrent Neural Networks (IndRNN) - LSTM - Recurrent Neural Network (RNN) -ANN	-Time series pipeline optimization (TSPO)	- AutoML automatically finds realistic ML models and optimizes AutoML context	-Empirical study
[34]	-Published articles in PubMed	-Duplicate articles were removed -Articles were screened based on weather factors -33 selected articles	-PRISMA flow		-Predetermined lag lengths often supported by literature review -Immune population was rarely addressed in the study model	-More diseases can selected for further research. -use of methods other than GLMs and GAMs.
[35]	-ERA5 Dataset from ECMWF	-Data interpolated onto a spherical grid -13 year time period restricted -Construction of cross-validation experiments -Sampling	-CNN -ConvLSTM -Variationalautoencoder -SAVP model -Generative adversarial network	-Combination of VAE and GAN	-Iterative CNN beats the forecasts up to 4hr lead time -Global variability by SSIM is better captured with SAVP -Local spatial variability is scarcely better than the ConvLSTM model	-Contemporary NWP models still considerable -Deep neural network and classic NWP models.
[36]	-Economic and financial data for forecasting. (Brazilian Stock Market) -major indexes (S&P500 (US), FTSE100 (UK), DAX (Germany), Hang Seng (Hong Kong), &Nikkei (Japan) -IID data		-Nonlinear models(shallow NN, deep NNs, regression tree, random forest) -Bagging, complete subset regression	-Use of both linear & nonlinear methods	-Empirical evidence (nonlinear ML models combined with large datasets are extremely useful for economic forecasting) -RFs did not perform well -NNs outperform the benchmark	-Use of Support Vector Regressions, autoencoders, nonlinear factor models, and many more
[37]	-Output of ten main crops in Pakistan (2000-2019)	-Panel unit root analysis	-Panel pooled mean group (PMG) technique. -Sensitivity analysis	-Dynamic ordinary least square (DOLS) & completely modified ordinary least square (FMOLS)	-Rising temperatures have a considerable detrimental impact on agricultural yield -Crop productivity is not significantly affected by higher-quality seeds	-Examining the effects of temperature variations in the Himalayan and Karakorum mountain ranges as well
[38]	-55 Articles from multiple databases(Web of Science, Google Scholar, Scopus Index Journals, Etc.	-Excluded 40 articles out of 95 based on title and data, removal of duplications, filtration after reading	-IRENA's model		-Analyzed climate and sector mitigation measures used in the study -Decline in cereal productivity	-Seasonal variations -New varieties of crops -Other input factors -Socio-economic adaptation
[39]	-Research databases(Scopus, Elsevier, Wiley, & IEEE) -Research studies between Jan 2015 – Feb 2022	-Clustering -Dimensionality reduction -Classification -Regression	-Deep Neural Networks(LSTM, BiLSTM, ConvLSTM) -Statistical Models -TRLs Models -DRLs Models -Ensemble methods(Boosting, Bootstrapping) -AI models for unsupervised, supervised, semi-supervised, and reinforcement learning	-Edge computing -Fog computing -Hybrid computing	-ML, IoT, and new connectivity capabilities have a serious role in shaping the future of BAMSs	-Advancements in building automation systems and equipment standards -Market infiltration, regulatory, security, and privacy preservation, interoperability, and scalability, and struggle barriers.
[40]	-Greenhouse gas emission data from the IPCC	-Grouping based on common factors	-Content analysis and systematic review techniques	-Thorough analysis of approaches used	-Significant uncertainty in climate projections, which	-Refining the resolution of climate models, enhancing input
	-Research from Scopus and Web of Science		-Harmonising methods		affects estimates of renewable energy	data, and evaluating the overall electricity generation system.
[41]	-Previous year data from (AIRI, CIRI, EIRI, and NOAA,)	-NCEP-NCAR analysis -Time-delayed function -Cross-correlation function	-IN-DEGREE TIME SERIES -CN construction	-Dynamical and physical climate networks based on the global surface air temperature field -Forecast skill of .0.5, comparable to the best statistical and dynamical forecast models -Increase of the in-degree index and a change in the structure of the network	-ISMR forecast model is better than operational forecast models -NetworkENSO and ENSO monsoon -SWAS index	
[42]	DF data from NIHE (1997-2016) -12 meteorological factors were collected (1997-2016)	-Data cleaning -Feature collection -Performance evaluation -Outbreak detection	-CNNs -LSTM -Transformer model -SARIMA -Poisson regression, SVR, and SVR-L	-LSTM ATT	-Improved prediction performance	-Use of LSTM-ATT in forecasting other climate-sensitive diseases such as influenza, diarrhea, and malaria
[43]	113 years of annual climatic data from Varanasi, India	-KNN -Moving average, exponential smoothing -ARIMA -Cubic spline interpolation	-Spline interpolation -KNN -ARIMA -Hybrid models(interpolated with statistical model)	-Comparison of individual and hybrid models	-Very good error measurements for the ARIMA model	-further validation of the models for time series data

3. Discussion

The topic of big data climate change forecasting using time series approaches has seen the emergence of several innovative concepts and methodologies that have substantially expanded our understanding of this crucial field. Notable innovations include the shift to regional-level research, which makes it possible to examine the effects of climate change in greater depth, and the discussion of policies that will be in place until 2100, which emphasizes the importance of long-term planning. Furthermore, the ability to forecast many pollutants simultaneously with a single model—like NO2, CO, SO2, and O3—presents a complete picture of the quality of the air and facilitates broad pollution control measures. Combining machine learning methods with dispersion models for air pollution forecasting has revolutionized our capacity for future prediction. The forecast robustness and accuracy have been significantly enhanced by technologies such as fuzzy time series, hybrid models, and the Grey Wolf Optimizer. The use of prediction intervals and indicators for quantitative assessment recognizes the importance of standardization and uncertainty in predicting. In addition, real-world implications for resource optimization result from the integration of climate data into the hydro-thermal-wind systems optimal scheduling problem and from the focus on improving accuracy and deep learning models. Since more precise and location-specific projections are necessary for effective policy-making and climate change mitigation programs, all of these advancements contribute to the field of climate change forecasting. By combining these methods with continuous model optimization, it is possible to make progress in the field of climate forecasting and address one of the most pressing issues of our day.

4. Conclusion

The review study summarized the results of a wide range of investigations that covered different geographic regions, climates, and methodologies. Different significant discoveries on climate change and forecasting, assessment of its impact on the environment, and forecasting models and methods are found after a thorough examination of various research papers. Different sets of data, for example, crop production statistics; weather indicators, satellite imaging, and atmospheric observations exhibited the complicated nature of climate research. Several preprocessing techniques were applied to develop and understand relationships within the data such as trend and correlation analysis, and application of machine learning techniques or algorithms.

The investigation of climate change because of its consequences on various areas or regions was a recurrent concept in different studies. Scholars traced changes in rainfall patterns, temperature, and crop production at various places as diverse as European beech woods and the Himalayan region exhibiting globally changing climate phenomena. In addition, advanced models were established to increase the accuracy and effectiveness of climate predictions using deep learning methods such as Convolutional Neural Networks (CNNs), Long-Short-Term Memory (LSTM), and innovative architectures for example, CNN and SIRN demonstrating positive results in time series prediction and pattern recognition.

The research also spotlights the significance of multidisciplinary association, citing a study that incorporates climate data with power demand, agriculture production, and disease epidemiology. These interdisciplinary approaches offer a broad understanding of the complicated relationship that falls between societal schemes and climate procedures. The study also highlights the necessity of future research to overcome the existing challenges and accelerate further in the field of climate research. Succeeding studies cloud concentrate on improving and investigating complex deep learning architectures, modeling procedures, integrating high-frequency data, and assessing the efficiency of mitigation and adaptation policies.

To summarize all the above, integrating outcomes from different studies provides intuitive information on the state of climate research today and highlights the obstacles and advancements that need to be implemented. Progressing the thoughtful understanding of climatic variations and forming evidence-based plans for sustainable growth and resilience to ecological effects can be attained using advanced methods, diverse datasets, and multidisciplinary collaboration.

References

[1]. Sánchez-Morales, J., Pardo-Igúzquiza, E., & Rodríguez-Tovar, F. J. 105342, s.l. :Terrain methods on spectral analysis for paleoclimate interpretations: A novel visualization technique using Python. Computers & Geosciences, 175, (2023).

[2]. Fathi, M., Haghi Kashani, M., Jameii, S. M., & Mahdipour Big data analytics in weather forecasting: A systematic review., E. 1247-1275, s.l. : Archives of Computational Methods in Engineering, 29(2), (2022).

[3]. Haq, M. A. CDLSTM: A novel model for climate change forecasting. 2363-238, s.l. : Comput. Mater. Contin, 71(2), (2022).

[4]. M., Zang, C. S., Buras, A., Hacket-Pain, A., Esper, J., Serrano-Notivoli, R.,& Tzvetan, Z. 5(1), s.l. :Climate-change-driven growth decline of European beech forests. del Castillo Edurne, Communications Biology, (2022).

[5]. Long, B., Tan, F., & Newman, M. Forecasting the Monkeypox Outbreak Using ARIMA,

Prophet, Neural Prophet, and LSTM Models in the United States. 127-137, s.l. : Forecasting, 5(1), (2023).

[6]. Semenoglou, A. A., Spiliotis, E., & Assimakopoulos, Image-based time series forecasting: A deep convolutional neural network approach. V. 39-53., s.l. : Neural Networks, 157, (2023).

[7]. Linden, A. Conducting interrupted time-series analysis for single-and multiple-group comparisons. 480-500., s.l. : The Stata Journal, 15(2), (2015).

[8]. Bochenek, B., & Ustrnul, Z. Machine learning in weather prediction and climate analyses—applications and perspectives. 180., s.l. : Atmosphere, 13(2), (2022).

[9]. A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. Barrera-Animas, A. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Barrera-Animas, 100240, s.l. : Machine Learning with Application, 7, (2022).

[10]. Alsharef, A., Aggarwal, K., Sonia, Kumar, M., & Mishra, A. Review of ML and AutoML solutions to forecast time-series data. 5297-5311. Archives of Computational Methods in Engineering., 29(7), (2022).

[11]. Amankulova, K., Farmonov, N., & Mucsi, L. 100098., Time-series analysis of Sentinel-2 satellite images for sunflower yield estimation. Smart Agricultural Technology, 3, (2023).

[12]. Li, Y., Lu, X., Xiong, H., Tang, J., Su, J., Jin, B., & Dou, D. Towards Long-Term Time-Series Forecasting: Feature, Pattern, and Distribution. arXiv preprint arXiv: 2301.02068., (2023).

[13]. Martinez, A., & Iglesias, G. 150580., Climate change impacts wind energy resources in North America based on the CMIP6 projections. s.l. : Science of The Total Environment, 806, (2022).

[14]. Vietnam. Rakholia, R., Le, Q., Ho, B. Q., Vu, K., & Carbajo, R. S. Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, 107848., s.l. : Environment International, 173, (2023).

[15]. Kimothi, S., Thapliyal, A., Gehlot, A., Aledaily, A. N., Bilandi, N., Singh, R., ... & Akram, S. V. Spatio-temporal fluctuations analysis of land surface temperature (LST) using Remote Sensing data (LANDSAT TM5/8) and multifractal technique to characterize urban heat islands. Sustainable energy technologies and assessment, 55, (2023).

[16]. Malakouti, S. M. 100312., s.l. : Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a Global Temperature Change Prediction Model. Case Studies in Chemical and Environmental Engineering, 7, (2023).

[17]. Yuan, X., Chen, C., Jiang, M., & Yuan, Y. 105550., s.l. : Prediction interval of wind power using parameter optimized Beta distribution based LSTM model. Applied Soft Computing, 82, (2019).

[18]. Vambol, S., Soomro, R., Ghauri, S. P., Marri, A. A., Dung, H. T., Manzoor, N., & Lutsenko, Y. Viable forecasting of monthly weather data using time series methods. 1-16., Ecological Questions, 34(1), (2023).

[19]. China. Li, B., & Loskot, P. Integrating ARIMA with machine learning for temperature prediction in major cities in In Third International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI 2022), 12509, (2023).

[20]. Aweda, F. O., Olufemi, S. J., & Agbolade, J. O. 80-91., Meteorological parameters study and temperature forecasting in selected stations in Sub-Sahara Africa using MERRA-2

data. Nigerian Journal of Technological Development, 19(1), (2022).
[21]. Zhang, H., Wang, Y., Chen, D., Feng, D., You, X., & Wu, W. Temperature Forecasting Correction Based on Operational GRAPES-3km Model Using Machine Learning Methods. 362., Atmosphere, 13(2), (2022).
[22]. Dore, M. H. Climate change and changes in global precipitation patterns: what do we know? 1167-1181., Environment international, 31(8), (2005).
[23]. Alghamdi, H. A. A time series forecasting of global horizontal irradiance on geographical data of Najran Saudi Arabia. 928., Energies, 15(3), (2022).
[24]. Jan, F., Shah, I., & Ali, S. Short-term electricity price forecasting using functional time series analysis. 3423., Energies, 15(9), (2022).
[25]. Jan, F., Shah, I., & Ali, S. 3423., s.l. : Energies, 15(9), (2022).
[26]. Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., ... & Manjunatha, B. An MRA-Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. 1. 137-150, International Journal of Environment and Climate Change, 13(5), (2023).
[27]. Anderson-Teixeira, K. J., Herrmann, V., Rollinson, C. R., Gonzalez, B., Gonzalez-Akre, E. B., Pederson, N., & Zuidema, P. A. Joint effects of climate, tree size, and year on annual tree growth derived from tree-ring records of ten globally distributed forests. 1. 245-266., Global Change Biology, 28(1), (2022).
[28]. Zahran, B., Ayyoub, B., Abu-Ain, W., Hadi, W., & Al-Hawary, A fuzzy-based model for rainfall prediction. S. 97-106., International Journal of Data and Network Science, 7(1), (2023).
[29]. Liu, Y., Li, D., Wan, S., Wang, F., Dou, W., Xu, X., ... & Qi, L. A long short-term memory-based model for greenhouse climate prediction. 135-151., s.l. : International Journal of Intelligent Systems, 37(1), (2022).
[30]. Song, F., Zhang, G. J., Ramanathan, V., & Leung, L. R. Trends in surface equivalent potential temperature: A more comprehensive metric for global warming and weather extremes. e211783211, Proceedings of the National Academy of Sciences, 119(6), (2022).
[31]. Reboita, M. S., Kuki, C. A. C., Marrafon, V. H., de Souza, C. A., Ferreira, G. W. S., Teodoro, T., & Lima, J. W. M. 459-485., South America climate change revealed through climate indices projected by GCMs and Eta-RCM ensembles. Climate Dynamics, 58(1-2), (2022).
[32]. Imai, C., & Hashizume, M. A systematic review of methodology: time series regression analysis for environmental factors and infectious diseases. 1-9., Tropical medicine and health, 43(1), (2015).
[33]. Song, P., & Russell, C. T. Time series data analyses in space physics. 387-463., Space Science Reviews, 87(3-4), (1999).
[34]. Masini, R. P., Medeiros, M. C., & Mendes, E. F. Machine learning advances for time series forecasting. 76-111., Journal of economic surveys, 37(1), (2023).
[35]. Mariani, M., & Baggio, R. Big data and analytics in hospitality and tourism: a systematic literature review. 231-278., International Journal of Contemporary Hospitality Management, 34(1), (2022).
[36]. Abbas, S. Climate change and major crop production: evidence from Pakistan. 5406-5414, Environmental science and pollution research, 29(4), (2022).
[37]. Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., & Younis, I. A review

of the global climate change impacts, adaptation, and sustainable mitigation measures. 42539-42559., Environmental Science and Pollution Research, 29(28), (2022).

[38]. Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., & Amira, A. AI-big data analytics for building automation and management systems: a survey, actual challenges, and future perspectives. 4929-5021., Artificial Intelligence Review, 56(6), (2023).

[39]. Russo, M. A., Carvalho, D., Martins, N., & Monteiro, A. Forecasting the inevitable: A review on the impacts of climate change on renewable energy resources. 102283., Sustainable Energy Technologies and Assessments, 52, (2022).

[40]. Fan, J., Meng, J., Ludescher, J., Li, Z., Surovyatkina, E., Chen, X., & Schellnhuber, H. J. Network-based approach and climate change benefits for forecasting the amount of Indian monsoon rainfall. 1009-1020., Journal of Climate, 35(3), (2022).

[41]. Nguyen, V. H., Tuyet-Hanh, T. T., Mulhall, J., Minh, H. V., Duong, T. Q., Chien, N. V., ... & Son, M. T. Deep learning models for forecasting dengue fever based on climate data in Vietnam. e0010509., PLoS Neglected Tropical Diseases, 16(6), (2022).

[42]. Khatri, P., Arjariya, T., & Shivhare Mitra, N. Climate change forecasting using data mining algorithms. AQUA-Water Infrastructure, Ecosystems and Society,
Geree. zeersferer. 12344, asdf, p. 12, (2023).