TOWARDS BUILDING AN ENHANCED WEATHER FORECASTING USING OPTIMIZED FUZZY LOGIC: AN OVERVIEW

Supported by tetfund TERTIARY EDUCATION TRUST FUND

Sunday Eneji Samuel, Taiwo Kolajo,
Department of Computer Science, Federal University, Lokoja
samuel.sunday-pg@fulokoja.edu.ng

Received: September 21, 2022 **Accepted:** November 12, 2022

Abstract:

The numerical weather prediction relies strongly on the precipitation forecast because of the use to civil protection agencies, enterprises, daily activities of people around the world, and the reduction of economic and social damages. However, there is the need to evolve better and more accurate parameterization of physical processes in order to improve on the outcomes of forecasts generated. This seminar paper discusses the weather forecasting issues, methods and gaps of existing solutions. In particular, evolutionary computing and fuzzy logic techniques are being investigated for the purpose of developing an effective model that could guarantee better accuracy and reliability of outcomes when applied for weather uncertainty problems in real-world situation. It was found that, the fuzzy logic approach has low accuracy, which needs to be improved with rule-lists optimization. The genetic algorithms hold promise in overcoming these tasks, by providing important information on weather and state of the atmosphere in certain places and periods through the application of fuzzy logic and genetic algorithms for improved outcomes of weather forecasts.

Keywords: Weather, forecasting, fuzzy logic, optimization, rules-lists, accuracy, redundancy, genetic algorithms

Introduction

The concept of climate change explains the fact that, the frequency and/or intensity of extreme weather and climate events are increasing across parts of the world. Extreme weather and climate events are among the primary causes of infrastructure damage causing large-scale cascading power outages, or shifts in the end-use electricity demands leading to supply inadequacy risks in many nations of the world (Mukherjee, Nateghi, & Hastak, 2018). To this end, several variables have to be accounted for including: air pressure, rainfall, tide level, wave characteristics, water and air temperature, and cloud cover (Panidhapu, Li, Aliashrafi, & Peleato, 2019).

Weather depicts the condition of air on earth at specific place and period of time. It is a continuous, data-intensive, dynamic and disorganized technique. Forecasting is the procedure of evaluation in indefinite circumstances from historical data. Weather forecasting is systematically and technically demanding issues across the globe in the past century. Weather forecasting is one field of traction for many scholars and researchers, which seeks to ascertain the present state of atmosphere gets varied. Though, the tasks of predicting forecasts are daunting due to their unpredictable and chaotic nature. These have been applied to diverse scenarios including severe weather alerts and advisories for transportation, agricultural production and development and forest fire minimizations (Pooja & Balan, 2019).

Also, weather nowcasting is a short-time weather prediction, which involves weather analysis and forecast for the next 0-6 hrs. Presently, the role of nowcasting in crisis management and risk prevention is increasing, as more and more severe weather events are expected. Enormous piles of meteorological data obtained from radar, satellite and weather stations' observations, are held by meteorological institutes and available for purpose of analysis. Radars and weather stations are constantly collecting real-time data, while data about cloud patterns, winds, temperature are continuously gathered by weather-focused satellites. Consequently, there is large amount of meteorological related data available to be analyzed using machine learning (ML) based algorithms for enhancing the accuracy of short-term weather-prediction techniques (Czibula, Mihai, & Mihule, 2021).

Processing of the weather-station data in the cloud requires efficient algorithms and approaches which are able to extract relevant features and make precise classification of the inputs in a computationally efficient way such as deep learning approaches (Sokolov, Vlaev, & Chalashkanov, 2020). Aside this, many weather forecasting models have been combined to improve the accuracy of outcomes (Chen, Wang, & Zhou, 2021). This paper utilizes optimized fuzzy logic technique for forecasting weather events in more accurate manner.

In case of the weather forecasting problems, mathematical or traditional modeling is less-effective because of the following points: (i) the process might be too complex for mathematical reasoning; (ii) it might contain some uncertainties during the process, or; (iii) the process might simply be stochastic in nature. Artificial Intelligence (AI) has influenced multiplicity of fields of endeavour including robotics, programming languages, expert systems, visual systems, intelligent database, and games (Aldahwan *et al.*, 2020).

Artificial Intelligence or computational intelligence, by means of nature-inspired technique, attempts to develop intelligent systems that imitate aspects of human behaviour, such as learning, perception, reasoning, evolution, and adaptation. This way, AI provides algorithms for solving real-world problems analogous to the problems solved by natural systems. Though, while human intelligence distinguishes fundamental adaptation in new circumstances, AI follows a methodological algorithm of adaptability to the environment.

Fuzzy logic is an AI approach which utilizes an approximate reasoning instead of exact reasoning by incorporating a certain degree of uncertainty through a reasoning procedure. Again, genetic algorithm is an evolutionary computing that relies on the capability of certain algorithms to learn data and behaviours for effective resolution of uncertainty imprecision (Helena, Moraes, Diego, Claudia, & Hono, 2020). These two approaches hold promise in improving weather forecasting accuracy and reliability on the basis historical data exploiting the learning theory.

The numerical weather prediction relies strongly on the precipitation forecast because of the use to civil protection agencies, enterprises, daily activities of people around the world, and the reduction of economic and social damages. However, there is the need to evolve better and more accurate parametrization of physical processes in order to improve on the outcomes of forecasts generated. There is still the problem of existing weather forecasting techniques to produce location precise, time efficient and intensity of weather related events (Mazzarella, Ferretti, Picciotti, & Marzano, 2021).

Previously, the main procedure for forecasting weather, that is, the state of atmosphere over a particular place, involves the use of statistical and empirical methods by means of the principle of physics, but fall short of higher accuracy and relatively short-time required (Pooja & Balan, 2019). Subsequently, the renew calls for machine learning and ensemble methods, which utilizes complex computerized mathematical models for desirable outcomes.

The birth of AI, big data analytics and machine learning techniques offer the opportunities for planners and policy makers to understand the implications of diverse weather conditions as well as allocating resources in the case of extreme weather-related systems disruptions. Nonetheless, researchers and scholars are making efforts to increase the accuracy and reliability of the modelling systems. In fact, the next stage of the applications of these innovative solutions is to continue to minimize errors of forecasts offered by solutions including fuzzy logic, deep neural networks, decision trees and many other machine learning inspired methods (Chen et al., 2021; Coulibaly, Kamsu-foguem, & Tangara, 2020).

However, the accuracy of daily weather classification relies on both the applied classifiers and the training data (Chen, Wang, & Mi, 2018).

This study is to to analyze existing weather forecasting approaches and identify weaknesses, and, thereafter, to propose a framework that will enhance the accuracy of weather forecasting.

Background and Related Works

Information about future natural events or happenings is key to efficient and effective planning. In the quest to propose building an enhanced weather forecasting in determining the future occurrence of these natural events, this session presents some key facts on weather forecast, fuzzy logic system, genetic algorithm optimization and research efforts underpinning weather forecasting.

Weather Forecasting

Weather is the atmospheric condition of a particular place within a given time, driven by air pressure, temperature, and relative humidity. Numerical Weather Prediction (NWP) based methods have been utilized for forecasting applications for many years and perform best for the time horizons from 6hours to 2weeks. The problem of reliability of the models of weather predictions is a complex question, because it depends on many parameters and the technical infrastructure which supports them. More so, the reliability of the observations is essential for the numerical reasoning and the quality of the simulation, since there are various sources of problems such as uncertainties, error measurements, and load forecasts (Coulibaly et al., 2020; Helena et al., 2020).

Fuzzy Logic Systems

Fuzzy logic aims to model the approximate mode of reasoning by trying to mimic the human ability to make rational decisions in an environment of uncertainty and imprecision. Therefore, fuzzy logic is an intelligent technique that provides a mechanism to manipulate inaccurate information such as low/high, good/bad, very hot/cold, and allows inferring an approximate answer to a question from the knowledge that is inaccurate, incomplete, or not fully reliable. Due to its intrinsic characteristics, fuzzy logic is able to incorporate both objective knowledge (from numerical data) and subjective knowledge (from linguistic information). It is applied in control and decision support systems where the description of the problem (rules of production) cannot be made precisely (Helena et al., 2020).

A British Engineer, Ebrahim Mamdani, is the foremost person to make use of fuzzy sets in a concrete control system unpremeditatedly. In early 1970s, Mamdani was working on an automated control system for stream engine by means of human operator's expertise on the basis of Bayesian decision principle (that is, a technique of defining possibilities in uncertain situation) that focused on happenings after the fact to adjust prediction about upcoming outcomes. Consequent upon the poor performance of the work that artificial intelligence method of enhancing the learning controller known as rule-based expert system through the combination of human skills with sequence of logical rules for utilizing the knowledge (Kayacan, 2011).

The main components of fuzzy logic systems include:

Fuzzification: To construct a fuzzy rule-based model, the input and output fuzzy variables are specified. For every variable of fuzzy, its units are defined; the universe of discourse and a group of membership functions are specified in order to explain the distinct fuzzy theories linked with the fuzzy variable. The challenge of the model is the selection of technical indicators required to form the inputs to the decision support system (Gamil, El-fouly& Darwish, 2007; Delnavaz, 2014).

Membership Functions (MF): It is a curve defining ways in which every point in the input region is plotted to a value of membership (or membership degree) usually between the range 0 and 1. MF can be referred to as universe of discourse (Kayacan, 2011).

Logical Operations: The fuzzy logical reasoning concept is an enlarged collection of a standard Boolean logic (Kotta, 2015). The common fuzzy logic operations between two-valued and multivalued of maximum and minimum functions for input memberships function include: AND (interjection), OR and NOT. Other classical operators are: OR = *max,* AND = *min,* and NOT = additive complement (Kotta, 2015).

If Then Rules: The fuzzy sets and operators represent the subjects (antecedents/consequents) and verbs (conditions) of fuzzy logic, which are utilized in formulating the conditional statements that makeup the fuzzy logic.

The weaknesses of fuzzy logic systems*:* (Kotta, 2015)

a) Fuzzy system cannot be easily modelled.
b) It is easier to design and prototype, but requiring several simulations and fine tuning before releases.
c) The rule-based concept employed by fuzzy logic make it less favourable over mathematical precision or crisp system and linear models.

Genetic Algorithm Optimizations

Genetic Algorithm is an evolutionary computing paradigm, inspired by the Darwinian principle of species evolution and genetics. It is a probabilistic algorithm which provides a parallel and adaptive search mechanism based on the survival principle of the fittest in the reproduction. The mechanism is obtained from a population of individuals (solutions), represented by chromosomes (binary words, vectors, matrices, etc.), each associated with an aptitude (evaluation of the solution in the problem), that are submitted to a process of evolution (selection, reproduction, crossing-over, and mutation) for several cycles (Helena et al., 2020).

The features of genetic algorithms:

a) Population of points search are carried concurrently and not about single points.
b) Derivative information and other secondary knowledge are not required by its operators.
c) Concept is based to probabilistic transition rules rather than deterministic rules.
d) Encoding of parameter set instead of the set of parameters itself with few exceptions in cases where individuals with real-value are deployed.

Objective function is used to offer a degree of individuals' execution within the problem area. The associated objective function having the lowest numerical value is the best fit individual for minimization problems, which relatively measure the individuals at intermediate stage only. Usually, the fitness function converts the value of objective function into a quantity of best fitness given by Equation 1: (Kayacan, 2011)

$$Q(x)Q = c(b(x) \quad (1)$$

where:
b = objective function
c = convert objective function value to a non-negative quantity
Q = the resultant relative fitness

Related Works

A number of works have been carried out with regards to forecasting weather, climatic and cloud conditions of places around the world, which are discussed in this subsection.

The need to turn to more environmentally friendly sources of energy has led energy systems to focus on renewable sources of energy. Wind power has been a widely used source of green energy. However, the wind's stochastic and unpredictable behavior has created several challenges to the operation and stability of energy systems. Forecasting models have been developed and excessively used in recent decades in order to deal with these challenges. Bazionis and Georgilakis (2021) investigated renewable source energy unpredictability by utilizing deterministic and probabilistic models for wind power forecasting based on climatic conditions. This offered point predictions for daily operations of power systems with considerable accuracy. Again, advanced probabilistic and deterministic forecasting models could be used in order to facilitate the energy systems operation and energy markets management. However, there is the need to advance on accuracy of the models obtained.

Czibula, Mihai and Mihule (2021) proposed Ensemble of deep learning techniques known as NowDeepN for weather nowcasting under Radar products' values and parameterizations. The values predicted by NowDeepN may be used by meteorologists in estimating the future development of potential severe phenomena and would replace the time-consuming process of extrapolating the radar echoes. NowDeepN is intended to be a proof of concept for the effectiveness of learning from radar data relevant patterns that would be useful for predicting future values for radar products based on their historical values. For assessing the performance of NowDeepN, a set of experiments on real radar data provided by the Romanian National Meteorological Administration is conducted. The outcomes show the error rate of less than 4%. It was found that no relationships exit between normal and adverse metrological products' values, though, the models rely majorly on historical dataset cleaning.

Gupta *et al.* (2021) investigated the potential ability for weather to affect SARS-CoV-2 transmission has been an area of controversial discussion during the COVID-19 pandemic. It applied the concepts of machine learning (such as Gradient Descent SVM, TD-IDF) for weather impact on COVID-19 outbreak based on users' twitter feeds collected from 166,005 English tweets posted between January 23 and June 22, 2020 and filtered the relevant tweets, classify them by the type of effect they claimed, and identify topics of discussion. The outcomes showed correctly classified users' a claim based on their tweets was 95% AUC-PR and AUC-ROC. This approach that is effective in measuring population perceptions and identifying misconceptions, which can inform public health communications. Future works must consider classifier ineffectiveness on other languages.

The prediction of the occurrence of infectious diseases is ofcrucial importance for public health, as clearly seen in the ongoing COVID-19 pandemic. Hochman *et al.* (2021) used a stepwise linear regression model, clinical and laboratory tests for cyclonic weather regimes effect on seasonal influenzas. The study analyzed the relationship between the occurrence of a winter low- pressure weather regime in Cyprus Lows as well as the seasonal Influenza in the Eastern Mediterranean. This reveals that the weekly occurrence of Cyprus Lows is significantly correlated with clinical seasonal Influenza in Israel in recent years ($R = 0.91$; $p < .05$) It was found that Climate changes aggravates health risks of people using regression and root mean square difference. But, there are large inaccuracies from datasets.

The impact of severe weather events on the operational performance of transportation systems, which existing lackscapability to forecast to what extent the performance of different transportation systems may vary under various conditions. To this end, Chen, Wang and Zhou (2021) advanced an integrated modeling framework that allows us to predict the performance of weather-induced delays of different transportation systems, including HSR and aviation. It used machine learning modeling (Ensemble, Linear, SVM, and Decision Tree) of weather events for Weather-associated delays in transportation sector. The prediction of severe and disruptive weather events was reported indicative that, the modeling framework provides important implications for the improvement of transportation system resilience to various severe weather-related disruptions. Though, there is the need for accuracy improvement for the majority of the models especially the Ensemble method.

The precipitation forecast over the Mediterranean basin is still a challenge because of the complex orographic region 10 which amplifies the need for local observation to correctly initialize the forecast. Mazzarella *et al.* (2021) experimented Fraction Skill Score (FSS) for Weather radar reflectivity towards flood events. This applied the context the data assimilation techniques as play a key role in improving the initial conditions and consequently the timing and position of precipitation pattern. The radar reflectivity measured by the Italian ground radar network is assimilated in the WRF model to simulate an event occurred on May 3, 2018 in central Italy. The results showed AUC of 91% for the predictive model. However, the reliability of forecasts can be improved for the area under the curve is 0.91 compared to 20 the 0.88 of control experiment without data assimilation.

Natural disasters ravage the world's cities, valleys, and shores on a monthly basis. Havingprecise and efficient mechanisms for assessing infrastructure damage is essential to channel resourcesand minimize the loss of life. Therefore, Chen (2021) worked on the CNN-based building damage image classification using interpretability of satellite imagery. It relied on dataset that includes labeled pre- and post- disaster satellite imagery to train multiple convolutional neural networks to assess building damage on a per-buildingbasis.The outcomes revealed accurate classification of building damages through images. But, there is the need to consider Pre-and-post-disaster images modeling in subsequent works on growing humanitarian crisis, heightened by climate change.

In many of the application sectors , it will be very difficult to collect the various weather parameter due to the use of wires and analog devices. Chinchawade and Lamba (2021) utilized Internet of Everything such as sensors for measuring weather parameters for smart weather reporting system. The proposed model provided weather information in an effective manner through the measuring various parameters of weather of a particulate area and sharing the information on World Wide Web with messaging facility on Mobile phones. Often, the Internet enabled approach for weather data collection is problematic due to poor Internet connectivity.

The value of machine learning as an accelerator for the parameterisation schemes of operational weather forecasting systems need to be assessd, specifically the parameterisation of non-orographic gravity wave drag. Chantry *et al.* (2021) conducted weather forecasting with gravity wave drag emulation using machine learning based neural networks. However, the emulators of this scheme can be trained that produce stable and accurate results up to seasonal forecasting timescales. Though, more complex networks produce more accurate emulators. This approach increased speed and accuracy of models. Nevertheless, the Neural network algorithms are less-effective on CPU.

Numerical weather forecasting on high-resolutionphysical models consume hours of computations on supercomputers. Application of deep learning and machine learning methods in forecasting revealed new solutions in this area. Tekin *et al.* (2021) carried out Spatio-temporal weather forecasting based on Convolutional LSTMs. The forecasts of high-resolution numeric weather data used both input weather data and observations (pressure levels and temperature) by providing a novel deep learning architecture, which offered superior MSE and performance. There is the need to explore other datasets rather than Spatio-temporal datasets utilized.

Removing turbine clutter from weather radar observations has become an essential problem in the community since Wind Turbine Clutter (WTC) signals cannot be filtered using traditional clutter filtering. Dutta, Chandrasekar and Ruzanski (2021) attempted to minimize turbine clutter based on weather radar data using Generalized Likelihood Ratio Test for identifying signal subspace and gates impacted by WTC. The results offered better prediction due to overlap in datasets utilized to detect the range gates affected, and to mitigate turbine clustering. However, subsequent efforts must improve on local information about precipitation and filtered radar IQ.

The process of calculating storm time electric fields from time-independent impedance tensors computed from an array of magnetotelluric (MT) sites and storm time magnetic fields recorded at geomagnetic observatories or assumed from line current models in the US. A paucity of direct measurements of storm time electric fields has restricted validation of these different techniques for nowcasting electric. Simpson and Bahr (2021) utilized Magnetotelluric data of geomatic storms for the purpose of Nowcasting of extreme space weather events during, and after the September 2017 magnetic storm. The preferred technique involves frequency domain multiplication of magnetic field spectra from a regional site with a local impedance tensor that has been corrected for horizontal magnetic field gradients present between the local site and the regional site using perturbation tensors derived from geomagnetic depth sounding (GDS) It makes use of bivariate approach for polarization of storm time electric fields. The weakness is in the short-lived magnetic field of storm for spatial and temporal events.

The rapid growth of demand for electrical energy and the depletion of fossil fuels opened the door for renewable energy; with solar energy being one of the most popular sources, as it is considered pollution free, freely available and requires minimal maintenance. Nkambule *et al.* (2020) proposed machine learning techniques for solar photovoltaic system forecast under several weather factors. It identified certain contributions to the field of PV systems and ML based systems were made by introducing nine (9) ML based MPPT techniques, by presenting three (3) experiments under different weather conditions. Decision Tree (DT), Multivariate Linear Regression (MLR), Gaussian Process Regression (GPR), Weighted K-Nearest Neighbors (WK-NN), Linear Discriminant Analysis (LDA), Bagged Tree (BT), Naïve Bayes classifier (NBC), Support Vector Machine (SVM) and Recurrent Neural Network (RNN) performances are validated and the weighted-KNN outperforms other ML approaches investigated. But, energy efficiency and high error rates are shortcomings identified.

Salvador *et al.* (2020) classified main synoptic meteorological patterns of atmosphere based on Particle formulation analysis of air quality. It effectively determined weather scenarios that consider a large amount of interconnected variables can be very useful for characterizing atmospheric situations that have a strong influence on atmospheric pollution processes in other regions Future works to consider applicability to particle formulation and air quality prediction.

The Climate change trains regularly some phenomena threatening directly the environment and the humanity. Meteorology plays a more important role in the control of these phenomena. Hence, it is important to search resources allowing to contribute to the improvement of the numerical model for the predictions of weather and climate. The weaknesses of the models in the simulation of exchanges between the surface and the atmosphere is important. Consequently, Coulibaly *et al.* (2020)

experiment machine learning with rule base approach such as K-NN, ARIMA for weather data knowledge mining. This improved the quality of concomitant factors prediction. Nonetheless, high errors occurred during simulation of weather reports.

Weather forecasting plays a significant role in different aspects of life such as in the operation of hydro-power plants, renewable energy, flood management, and agriculture. Recently, machine learning techniques have been used for weather forecasting for large periods of time, as it is more accurate than models based on physical principles. Gad & Hosahalli (2020) performed National Climatic Data Center (NCDC) weather data classification and predictive models using machine learning based models including CART, AdaBoot, Decision Tree, and XGBoost. The outcomes revealed that KNN, Random forest and XGBoost have higher accuracy. Overfitting and smaller datasets impacted on performance. Therefore, the selected methods do not show significant differences in their performance and further tests can be done by using different algorithms to better understand the results.

Integration of photovoltaics into power grids is difficult as solar energy is highly dependent on climate and geography; often fluctuating erratically. This causes penetrations and voltage surges, system instability, inefficient utilities planning and financial loss. Forecast models can help; however, time stamp, forecast horizon, input correlation analysis, data pre and post-processing, weather classification, network optimization, uncertainty quantification and performance evaluations need consideration. Ahmed, Sreeram, Mishra, & Arif (2020) studied Photovoltaic (PV) solar power forecasting based on climatic conditions. It utilized particle swarm optimization and genetic algorithms. Again, optimization of inputs and network parameters was performed with genetic algorithm and particle swarm optimization. Thereafter, the performance evaluation metrics MAE, RMSE and MAPE were adopted. The results showed that, CNN Deep learning model best for determining PV power. The hybrid algorithms can be experimented in future studies.

General circulation models (GCM) have been used by researchers to assess the effect of climate change in different fields of study. In the case of building energy performance, GCMs can be used to evaluate future building energy performance through simulations. However, a key issue with the use of GCM data in building energy simulation is the inadequate resolution and bias of the data. Hosseini, Bigtashi, & Lee (2020) carried a study on weather files for building energy designs optimization using machine learning based on hybrid model of regression and classification. In the model, workflow utilized the observed weather data to determine similar weather patterns from historical data and use it to generate future weather data, contrary to previous studies, which use artificially generated data. The outcomes showed that, it generated highly accurate weather files subsequently. The location and climate change events and applications not considered especially, extreme weather characteristics are preserved for extreme or reliability analysis and design optimization.

Comparisons between anomaly and full-field methods have been carried out in weather analysis and forecasting over the last decade. Evidence from these studies has demonstrated the superiority of anomaly to full-field in the following four aspects: depiction of weather systems, anomaly forecasts, diagnostic parameters and model prediction. Qian, Du, & Ai (2020) carried out weather forecasting using numerical weather prediction approach. It uses full-field weather system to perform anomaly weather forecasts accurately. Combining the anomaly forecast with an ensemble is emphasized, where a societal impact index is considered. The outcomes may be inaccurate and misleading without full-field data especially in predicting typhoon tracks with clearer physical explanation.

The state of the weather became a point of attraction for researchers in recent days. It influences activities in many fields as agriculture; the country determines the types of crops depend on state o f the atmosphere. It is important to know the weather in the coming days in order to take precautions. Forecasting the weather in future especially rainfall won the attention of many researchers, to prevent flooding and other risks arising from rainfall. Abdul-Kader, Abd-el-Salam, & Mohamed (2020) conducted rainfall forecasts based on a hybrid machine learning model of PSO and Feed Forward Neural Network. It improves outcomes of forecasts for rainfall. In future study, there is the need to increase accuracy.

Storing the data in the cloud is an important part and we provide our consideration for implementing a particular open source platform that could reduce costs fulfil the requirements for weather data storage. The designed system is expected to be cloud-provider agnostic, especially suitable for purposes such as local weather forecast; disaster prediction: hurricane, severe storms, floods, tornadoes, etc.; severe droughts with fire dangers; prediction of landslides. Sokolov *et al.* (2020) used an LSTM based neural network model for automated weather data processing. The proposed approach predicts local weather events such as hurricane, flood, severe storm, etc. This study did not explore more parameters of soils forecasts to ascertain its effectiveness.

One of data mining techniques is classification, which is used to predict relationships between data on a dataset. The prediction performed by classifying data into several different classes considering certain factor. Classification is a performance of Supervised Learning application where the training data already has a label when entered as input data. Classification is an approach of empirical techniques that can be utilized for short-term weather prediction. Prasetya & Ridwan (2019) relied on the classification tree, KNN, and Naïve Bayes to perform weather prediction. The input data is synoptic data of Kemayoran Meteorological Station, Jakarta (96745) for 10 years (2006 - 2015) consists of 3528 datasets and 8 attributes. The outcomes showed that, the Naïve Bayes had 77.1% accuracy as best. More data consisting of weather observational data over stations are needed to be used to ascertain its level of accuracy.

The integration of both water quality measures and 425 current/historical weather for prediction of fecal coliforms and Escherichia coli levels was attained through BBNs. For 4-bin classification of fecal coliform levels, BBNs increased prediction accuracy by 25% to 54% compared to other previously used techniques including logistic regression, Naïve Bayes, and Random Forests. Binary. In particular, Panidhapu *et al*. (2019) used Bayesian Belief Networks for weather conditions-based water quality prediction. Water surrogates are determinants for water quality prediction using this proposed approach. But, higher accuracy models are required for safety of drinking water.

Weather forecast in an area is unpredictable because it is based on human factors which cannot predict accurately. Statistical classification method are useful for the process of determining the probability of a class membership to a large extent. Findawati *et al.* (2019) considered the use of three models namely: Naïve Bayes, C.45 and KNN, for weather forecasts. It was observed that KNN produced highest accuracy (71.59%) of forecasts. Though, the input criteria and constraints are inconsistent, which impact on the performance of the weather forecasting models.

In the field of computer vision, multi-class outdoor weather classification is a difficult task to perform due to diversity and lack of distinct weather characteristic or features. Oluwafemi & Zenghui (2019) proposed a novel framework for identifying different weather scenes from still images using heterogeneous ensemble methods. It selected a Selection Based on Accuracy Intuition and Diversity (SAID) of ensemble algorithms for a multi-class weather classification. In the method, the extraction of histogram of features from different weather scenes was performed. Then, blending and boosting of different weather features were achieved with the stacked ensemble algorithms in order to increase the recognition rate of various weather conditions as against other classification and ensemble method. SAID outperformed other algorithms in classifying weather images. New studies to explore computer vision for weather classification tasks need to be carried out.

Kwon, Kwasinski, & Kwasinski (2019) selected a Naïve Bayes classifier for Solar irradiance forecast based on weather variables. NB's independence assumption-based on simple Bayes' theorem makes the process speed faster and less constrained than other classification algorithms. Based on publicly available weather forecasting information about temperature, relative humidity, dew point, and sky coverage, they are used as a training dataset in NB classification with hourly resolution. The outcomes show improved accuracy for real-time weather. But, the training data is relatively small.

Weather forecasting is the prediction of atmosphere state for particular location by using principles of physics provided by many statistical and empirical techniques. Weather forecasts are frequently made by collecting quantitative data about current state of atmosphere through scientific understanding of atmospheric processes to illustrate how atmosphere changes in future. The existing techniques failed to predict the weather with higher accuracy and lesser time. Pooja & Balan (2019) attempted to perform weather forecasting with machine learning and ensemble methods. This approach increased the accuracy and speed of the weather forecasting. New works to utilize classification and clustering approaches to better the performance of the model.

Severe weather-induced power outages affect millions of people and cost billions of dollars of economic losses each year. The National Association of Regulatory Utility Commissioners have recently highlighted the importance of building electricity sector's resilience, and thereby enhancing service-security and long-term economic benefit. Mukherjee *et al.* (2018) considered a two-stage Hybrid risk estimation model for weather-based major power outage forecasts. The model was trained using publicly available information on historical major power outages, socio-economic data, state-level climatological observations, electricity consumption patterns and land-use data. It identified risks to be associated with different factors influencing power outage risk as a function of various factors such as the type of natural hazard, expanse of overhead T&D systems, the extent of state-level rural versus urban areas, and potentially. However, this approach can be applied to resilience of power systems forecasting in the future times.

Accurate solar photovoltaic (PV) power forecasting is an essential tool for mitigating the negative effects caused by the uncertainty of PV output power in systems with high penetration levels of solar PV generation. Weather classification based modeling is an effective way to increase the accuracy of day-ahead short-term (DAST) solar PV power forecasting because PV output power is strongly dependent on the specific weather conditions in a given time period. Wang *et al.* (2018) identified the accuracy of daily weather classification as relying on both the applied classifiers and the training data. Therefore, the work performed a weather-based Solar PV power forecasting with KNN, and SVM classifiers using operation data from a grid-connected PV plant in Hohhot, Inner Mongolia, China. It was observed that, the SVM produced the best accuracy of forecasts. It was found that, the SVM performs well with small sample scale, while KNN is more sensitive to the length of the training dataset and can achieve higher accuracy than SVM with sufficient samples. But, it did not extend to other models such as K-Means, Random Forest, etc., which can be investigated in subsequent works.

Rainfall prediction is one of the challenging tasks in weather forecasting. Accurate and timely rainfall prediction can be very helpful to take effective security measures in advance regarding: ongoing construction projects, transportation activities, agricultural tasks, flight operations and flood situation, etc. Aftab *et al.* (2018) identified data mining approaches to effectively extract the hidden patterns, buried with historical weather data, for rainfall forecasts. It adopted weather data extrapolation for determining rainfall patterns. Though, the study analyzed previous works related to data mining techniques alone. In the future, the models can be optimized and integrated with data mining techniques for better accuracy and performance.

Artificial intelligence (AI) continues to push the innovative technological changes across diverse areas of human endeavour globally. This is often referred to as intelligent behaviour depicted by machinery rather than natural human intelligence. The research is on-going in the use of AI and its systems to effectively undertake tasks hitherto performed by humans. Presently, AI has established itself as critical discipline, which includes a range of a well-recognized and advanced areas: Expert Programs, Fuzzy Logic, Genetic Analytics, Language Analysis, and Logic Programming (Aldahwan and Alsaeed, 2020).

However, fuzzy logic like any other AI techniques have problems of poor accuracies of its outcomes, which led to the concept of optimization with other forms of AI and machine learning algorithms such as Genetic analytics (Aftab *et al.*, 2018), (Abdul-Kader, Abd-el Salam and Mohamed, 2020), which are adopted in this seminar paper.

The outcomes of weather forecasts approach using the AI methods rely heavily on the historical dataset collected from the environment, which are analyzed through

classification and prediction techniques. The process of mining and extracting important attributes and their relationships in datasets remain a top challenge, even in the advent of empirical tools such as decision tree, artificial intelligence, regression, fuzzy logic, artificial neural networks and evolutionary algorithms (Prasetya and Ridwan, 2019).
There are no considerations for fuzzy logic analytics and systems in performing weather and related parameters forecasting. This is an area to be explored in the seminar work, which is motivated by the study (Alfa, Misra, *et al.*, 2020) and (Alfa, Yusuf, *et al.*, 2020). These studies utilized genetic algorithms optimization of fuzzy logic rules-lists antecedents and consequents for stock prices forecasting tasks. In the case of (Alfa, Yusuf, *et al.*, 2020), the forecasts accuracy improved from 32.424% to 67.576% after removing redundancy in fuzzy inference system's antecedents utilized for stock price prediction. Again, the consequent optimization for the fuzzy inference system based on genetic algorithms increase the accuracy to 70.37% from 29.63%. This seminar paper will adopt both methodologies for city-wide weather forecasts as explained in next section.

Research Methodology

Description of the Proposed Model

The weather forecasting system proposed in this paper is mainly composed of the input, process and output components, whose complete structure or layout is presented in Figure 3.1

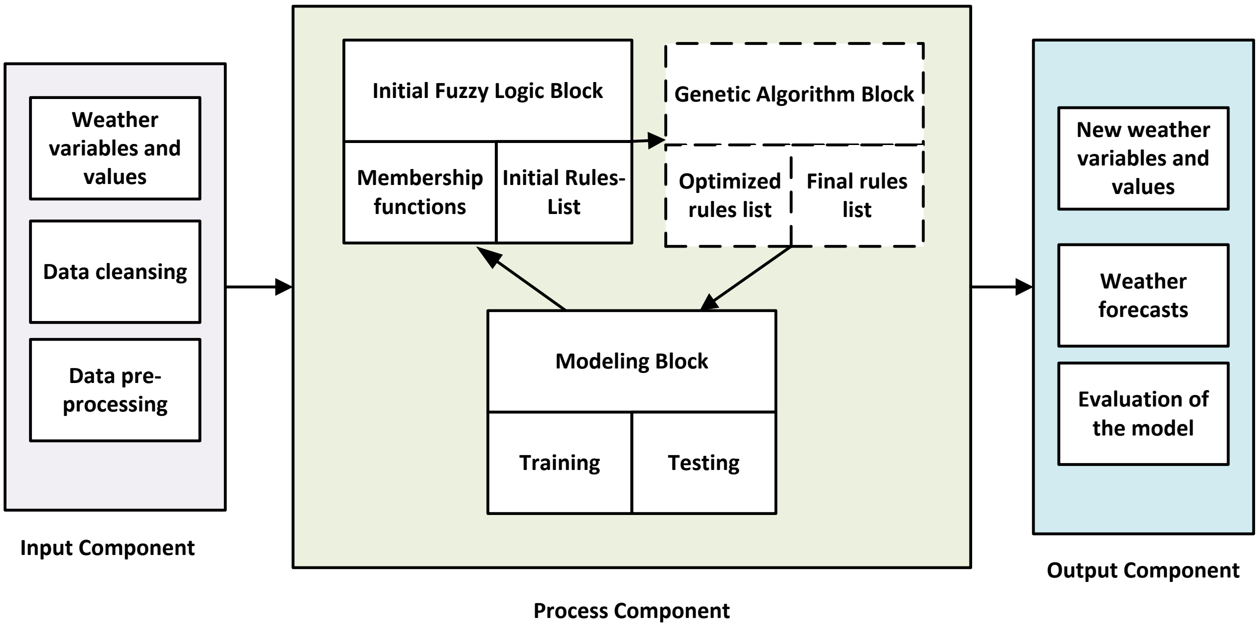

Figure 3.1: A High-level Overview of the Proposed Weather Forecasting Framework.

From Figure 3.1, the various subcomponents of all the main components of the proposed weather forecasting framework are discussed in the subsequent subsections.

Input Component

The input component is concerned with accepting and formatting of the raw weather information generated from the environmental factors including the relative temperature, relative humidity, and rainfall. Thereafter, the raw data is preprocessed into the target, test and train datasets for further processing using the following steps.

Data cleansing: This is the first subcomponent of the input component that is responsible for identifying, removing, correcting incomplete or corrupt elements in the dataset, table, or records. These are irrelevant for the subsequent processing of the data by replacing, deleting and modifying dirty portions.

Data pre-processing: This is the second subcomponent of the input component that involves data exploration and understanding through manual review of the dataset with the view of avoiding mistakes during modeling and analysis tasks. It is used to create Data Frame composed of names of columns, data types, and description of target labels. Before the data elements are fed into the machine learning algorithms, the missing values and categorical attributes are modified to be adapted for further processing. In particular, numerical and non-missing values are expected to make up the preprocessed dataset modeling bound. Again, scaling and normalization are often applied to specify the relationships in the quantities or elements in the datasets for appropriate understanding of the chosen machine learning algorithm.

Process Component

The process component undertakes the tasks of genetic optimization, training, and learning of the model using the preprocessed datasets collected from the previous component. The detail of the optimization based on genetic algorithm is presented in this subsection.
The human expert rules lists are converted to the set of membership functions whose indices are rearranged to form chromosomes required for the genetic algorithm optimization procedure.

The Fuzzy Inference System Modeling: The rules list from the genetic algorithm optimization are restructured (that is, the chromosomes) to serve as the antecedents and consequents (or membership functions) in the fuzzy logic inference system rule base model. Thereafter, the best combination of consequents after generating the rule base is used to produce the optimized FIS capable of accurately forecasting weather data forecasting for Nigerian cities.

Output Component
The outcome component offers the complete evaluation results of the weather forecasts using the required datasets. The main subcomponents of the output component include: new weather variables and values, weather forecasts, and evaluation of the model. These are discussed as follows:
The new weather variables and values: The weather variables associated values to be predicted, which is extracted from the pre-processed dataset as testing dataset, are adopted at this stage in the model output component. In fact, the testing dataset has all the weather variables and labels as those of the training dataset except the target variable.
The weather forecasts: This level of weather output component is the most important because; the weather forecasting model is put to actual use by applying all the weather parameters and related values to ascertain its behaviour. The input dataset is accepted by the weather forecasting to generate the corresponding output after learning the hidden relationships and patterns.
The evaluation of the model: The weather forecasting model is evaluated on the basis of model output values realized from the input data supplied at the previous stage. To this end, the model performance is adjudged using selected metrics such as Mean Square Error, Mean Absolute Percentage Error, amongst others.

Problem Definitions
Definition 1: The raw datasets are collected from the environment to obtain meaningful numerical information as depicted in Equation 1.

$$y = |x^n|_1^j \tag{1}$$

Where,
n is the number of terms for the environmental parameters collected such as temperature, rainfall and relative humidity,
j is the number samples or events for environment from the first instance,
x is the input term realized from the environment independently, and
y is the raw dataset collected from the environment.
Definition 2: The preprocessing is undertaken to remove redundancy and noise from the raw data as indicated Equation 2.

$$Z = \bigcup t(y) \tag{2}$$

Where,
Z is the preprocessed dataset after denoising and cleaning processes,
t is the transformation function of raw dataset into meaningful information, and
y is the raw dataset collected from the environment.

Definition 3*:* Fuzzy logic control is used for the modeling of the weather forecasting system by generating the membership functions and other parameters as represented in Equation 3.

$$R = \sum_{i}^{m} Z\emptyset\left(\frac{G.w}{n}\right) \tag{3}$$

Where,
G is the initial independent rule,
R is the initial rules list,
w is the membership function,
$\emptyset$ is the level of fuzziness of the rules,
i is the first fuzzy rule instance, and
m is the last fuzzy rule instance.

Definition 4: The optimization of fuzzy logic is performed to improve on the outcomes of the weather forecasting system through genetic algorithm operations are shown in Equation 4.

$$OpFR = \left|\left(\partial R * \int b(i) * r_k\right)^{p*h}\right|_n \tag{4}$$

Where,
b is the objective function of the genetic algorithm procedure,
i is the chromosomes of the membership functions arranged sequentially from consequent (h) and antecedents (p) after optimization.
$OpFR$ is the optimized fuzzy rule.
k is the rule list identifier after optimization,
∂R is the first optimized rules lists with redundancy rules lists, and
r is the reduced rules list with redundancy in rules lists.
Definition 5: The fuzzy logic-based weather system is trained using the training dataset in order to learn the properties and characteristics laying within the dataset as given by Equation 5.

$$OpFS = \left|\prod OpFR\,(Z * h * p)_v^n\right| \tag{5}$$

Where,
v is the first rule-list in the fuzzy engine,
$OpFS$ is the optimized fuzzy inference system.

Definition 6: The proposed weather forecasting system is evaluated with standard parameters to ascertain the effectiveness and accuracy as shown in Equation 6.

$$C = OpFS(Z_e, Z_f) \tag{6}$$

Where,
C is the output of the weather forecasting using the $OpFS$,
Z_e is the target dataset from model prediction for the consequent, and
Z_f is the target dataset for the consequent (actual values).

Conclusion and Future Works
The concept of optimizations has been widely proposed for fuzzy logic controls and inference systems to overcome the weaknesses identified. The inconsistencies and large inaccuracies in the outcomes realized from forecasting tasks make them undesirable. The process of determining the weather scenarios around the cities and townships in Nigeria has inspired diverse approaches from numerical weather prediction, regression analysis, computational and deep learning techniques. However, there is the need to obtain quicker and accurate weather conditions over period of time because of the influence exacted across various fields of human endeavours. This

proposed approach will assist stakeholders in improving the planning and policy formulation directed at enhancing the livelihood of people. The next stage of this research is to completely implement the proposed weather forecasting model with real-life datasets of selected Nigerian cities and townships. Also, there is the need to effectively measure the performance and weaknesses of this proposed weather forecasting model.

References

Abdul-Kader, H., Abd-el Salam, M., & Mohamed, M. (2020). Hybrid Machine Learning Model for Rainfall Forecasting. *Journal of Intelligent Systems and Internet of Things*, *1*(1), 5–12. https://doi.org/10.5281/zenodo.3376685

Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., & Nawaz, Z. (2018). *Rainfall Prediction using Data Mining Techniques : A Systematic Literature Review*. *9*(5), 143–150.

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, *124*(June 2019), 109 792. https://doi.org/10.1016/j.rser.2020.1092

Ajayi, G. O., & Wang, Z. (2019). Multi-Class Weather Classification from Still Image Using Said Ensemble Method. *Proceedings - 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa, SAUPEC/RobMech/PRASA 2019*, 135–140. https://doi.org/10.1109/RoboMech.2019.8704783

Aldahwan, N. S., & Alsaeed, N. I. (2020). Use of Artificial Intelligent in Learning Management System (LMS): A Systematic Literature Review. *International Journal of Computer Applications*, *175*(13), 16–26. https://doi.org/10.5120/ijca2020920611

Alfa, A. A., Misra, S., Bumojo, A., Ahmed, K. B., Oluranti, J., & Ahuja, R. (2020). Comparative Analysis of Optimisations of Antecedents and Consequents of Fuzzy Inference System Rules Lists Using Genetic Algorithm Operations. In R. R. Chillarige (Ed.), *Lecture Notes in Networks and Systems, Advances in Computational Intelligence and Informatics* (Vol. 119, pp. 373–379). Springer Nature Singapore Pte Ltd.

Alfa, A. A., Yusuf, I. O., Misra, S., & Ahuja, R. (2020). Enhancing Stock Prices Forecasting System Outputs Through Genetic Algorithms Refinement of Rules-Lists. In *Lecture Notes in Networks and Systems* (Vol. 121). Springer Singapore. https://doi.org/10.1007/978-981-15-3369-3_49

Bazionis, I. K., & Georgilakis, P. S. (2021). Review of Deterministic and Probabilistic Wind Power Forecasting: Models, Methods, and Future Research. *Electricity*, *2*, 13–47. https://doi.org/https://doi.org/10.3390/electricity2010002

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, 1–23.

Chen, T. Y. (2021). Interpretability in Convolutional Neural Networks for Building Damage Classification in Satellite Imagery. *Technical Note*, 1–11. https://doi.org/10.20944/preprints202101.0053.v1

Chen, Z., Wang, Y., & Zhou, L. (2021). Predicting weather-induced delays of high-speed rail and aviation in China. *Transport Policy*, *101*(November 2020), 1–13. https://doi.org/10.1016/j.tranpol.2020.11.008

Chinchawade, A. J., & Lamba, O. S. (2021). Secure Communication In Internet Of Everything (IOE) Based Smart Weather Reporting Systems. *Journal of Information and Computational Science*, *14*(1), 46–51.

Coulibaly, L., Kamsu-Foguem, B., & Tangara, F. (2020). Rule-based machine learning for knowledge discovering in weather data. *Future Generation Computer Systems*, *108*, 861–878. https://doi.org/10.1016/j.future.2020.03.012

Czibula, G., Mihai, A., & Mihule, E. (2021). NowDeepN: An Ensemble of Deep Learning Models for Weather Nowcasting Based on Radar Products' Values Prediction. *Applied Sciences*, *11*(125), 1–27. https://doi.org/https://doi.org/10.3390/app11010125

Dutta, A., Chandrasekar, V., & Ruzanski, E. (2021). A signal sub-space based approach for mitigating wind turbine clutter in fast scanning weather radar. *2021 United States National Committee of URSI National Radio Science Meeting, USNC-URSI NRSM 2021 - Proceedings*, *1*, 202–203. https://doi.org/10.23919/USNC-URSINRSM51531.2021.9336489

Findawati, Y., Astutik, I. R. I., Fitroni, A. S., Indrawati, I., & Yuniasih, N. (2019). Comparative analysis of Naïve Bayes , K Nearest Neighbor and C . 45 method in weather forecast Comparative analysis of Naïve Bayes , K Nearest Neighbor and C . 45 method in weather forecast. *4th Annual Applied Science and Engineering Conference*, *1402*, 1–7. https://doi.org/10.1088/1742-6596/1402/6/066046

Gad, I., & Hosahalli, D. (2020). A comparative study of prediction and classification models on NCDC weather data. *International Journal of Computers and Applications*, *0*(0), 1–12. https://doi.org/10.1080/1206212X.2020.1766769

Gupta, M., Bansal, A., Jain, B., Rochelle, J., Oak, A., & Jalali, M. S. (2021). Whether the weather will help us weather the COVID-19 pandemic: Using machine learning to measure twitter users' perceptions. *International Journal of Medical Informatics*, *145*(November 2020), 104340. https://doi.org/10.1016/j.ijmedinf.2020.104340

Helena, M., Moraes, B. De, Diego, D., Claudia, A., & Hono, B. (2020). *Use of artificial intelligence in precision nutrition and fitness*. https://doi.org/10.1016/B978-0-12-817133-2.00020-3

Hochman, A., Alpert, P., Negev, M., Abdeen, Z., Abdeen, A. M., Pinto, J. G., & Levine, H. (2021). The relationship between cyclonic weather regimes and seasonal influenza over the Eastern Mediterranean. *Science of the Total Environment*, *750*, 141686. https://doi.org/10.1016/j.scitotenv.

2020.141686

Hosseini, M., Bigtashi, A., & Lee, B. (2021). Generating future weather files under climate change scenarios to support building energy simulation – A machine learning approach. *Energy and Buildings*, *230*, 110543. https://doi.org/10.1016/j.enbuild.2020.110543

Kayacan, E. (2011). Single-Step ahead Prediction Based on the Principle of Concatenation using Grey Predictors. *Expert Systems with Applications 38(8)*:9499 9505. DOI:10.1109/GSIS. 2009.5408 302

Kwon, Y., Kwasinski, A., & Kwasinski, A. (2019). Solar Irradiance Forecast Using Naïve Bayes Classifier Forecasting Variables. *Energies*, *12*(1529), 1–13.

Mazzarella, V., Ferretti, R., Picciotti, E., & Marzano, F. S. (2021). Investigating 3D and 4D Variational Rapid-Update-Cycling Assimilation of Weather Radar Reflectivity for a Flash Flood Event in Central Italy. *Natural Hazards and Earth System Sciences*, 1–26. https://doi.org/https://doi.org/10.5194/nhess-2020-406

Mukherjee, S., Nateghi, R., & Hastak, M. (2018). A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. *Reliability Engineering and System Safety*, *175*(March), 283–305. https://doi.org/10.1016/j.ress.2018.03.015

Nkambule, M. S., Hasan, A. N., Ali, A., Hong, J., & Geem, Z. W. (2021). Comprehensive Evaluation of Machine Learning MPPT Algorithms for a PV System Under Different Weather Conditions. *Journal of Electrical Engineering and Technology*, *16*(1), 411–427. https://doi.org/10.1007/s42835-020-00598-0

Oluwafemi, A. G. & Zenghui, W. (2019) Multi-Class Weather Classification from Still Image Using SAID Ensemble Method, in *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*. IEEE, 135–140. doi: 10.1109/RoboMech.2019.8704783.

Panidhapu, A., Li, Z., Aliashrafi, A., & Peleato, N. M. (2020). Integration of weather conditions for predicting microbial water quality using Bayesian Belief Networks. *Water Research*, *170*, 115349. https://doi.org/10.1016/j.watres.2019.115349

Pooja, S. B., & Siva Balan, R. V. (2019). An investigation study on clustering and classification techniques for weather forecasting. *Journal of Computational and Theoretical Nanoscience*, *16*(2), 417–421. https://doi.org/10.1166/jctn.2019.7742

Prasetya, R., & Ridwan, A. (2019). Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbor Algorithm With Model Testing of Supervised Learning Probabilistic Brier Score, Confusion Matrix and ROC. *Journal of Applied Communication and Information Technologies*, *4*(2), 25–33.

Qian, W., Du, J., & Ai, Y. (2020). A review: anomaly based versus full-field based weather analysis and forecasting. *The American Meteorological Society*, 1–52. https://doi.org/10.1175/BAMS-D-19-0297.1.

Salvador, P., Barreiro, M., Gómez-Moreno, F. J., Alonso-Blanco, E., & Artíñano, B. (2021). Synoptic classification of meteorological patterns and their impact on air pollution episodes and new particle formation processes in a south European air basin. *Atmospheric Environment*, *245*, 118016. https://doi.org/10.1016/j.atmosenv.2020.118016

Simpson, F., & Bahr, K. (2021). Nowcasting and Validating Earth's Electric Field Response to Extreme Space Weather Events Using Magnetotelluric Data: Application to the September 2017 Geomagnetic Storm and Comparison to Observed and Modeled Fields in Scotland. *Space Weather*, *19*(1), 1–17. https://doi.org/10.1029/2019SW002432

Sokolov, S., Vlaev, S., & Chalashkanov, M. (2020). Technique for storing and automated processing of weather station data in cloud platforms Technique for storing and automated processing of weather station data in cloud platforms. *Series, IOP Conference Science, Materials*, *1032*, 1–7. https://doi.org/10.1088/1757-899X/1032/1/012021

Tekin, S. F., Karaahmetoglu, O., Ilhan, F., Balaban, I., & Kozat, S. S. (2021). *Spatio-temporal Weather Forecasting and Attention Mechanism on Convolutional LSTMs*. 1–13.

Wang, F., Zhen, Z., Wang, B., & Mi, Z. (2018). Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting. *Applied Sciences*, *8*(28), 1–23. https://doi.org/10.3390/app8010028