ISSN 2347 - 3983
Volume 13. No.11, November 2025
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter0113112025.pdf
https://doi.org/10.30534/ijeter/2025/0113112025

Weather Forecasting for Short and Medium Weather Patterns through Machine Learning

McDonald Owuor Adero[1], Charles Juma Katila[2], Erick Odhiambo Omuya[3]
[1]School of Computing and Mathematics – The Cooperative University of Kenya
[2]School of Engineering and Technology, Machakos University, Kenya

Received Date: September 28, 2025 Accepted Date: October 30, 2025 Published Date : November 07, 2025

ABSTRACT

Traditional data prediction techniques like AutoRegressive Integrated Moving Average (ARIMA) and Numerical forecasting have experienced decreased accuracy and reliability over time. This has been largely attributed to their inefficiency in handling the changing climatic patterns. These methods have challenges in accommodating non-linear and complex interactions between different climatic factors. They are highly sensitive to data quality and resolution; hence, small errors in the initial conditions result in low precision forecasts. They also struggle with forecasting small-scale weather phenomena like localized storms based on their limited temporal and spatial resolution. There is therefore the need to develop more robust and accurate techniques for weather forecasting to enhance safety and preparedness. This study employs machine learning (ML) models namely Random Forest (RF), Support Vector Machines (SVM) and Gradient Boosting (GB) for weather forecasting. It focuses on short- and medium-term weather patterns specific to humidity, windspeed and daily temperatures targeting Nairobi County. Experimental results show that Gradient Boosting (74.28%) perform better than the other models in predicting temperature, followed closely by Random Forest (72.72%). For humidity, SVM (68.60%) and GB (68.37%) performed well, while ARIMA performed poorly across both variables. These findings highlight the superior performance of ensemble and kernel-based machine learning algorithms for weather forecasting compared to the traditional methods.

Key words: Machine Learning, Support Vector Machines, Gradient Boosting, ARIMA model, Random Forest, Forecast, Accuracy, Traditional Weather Forecasting.

1. INTRODUCTION

Accurate weather prediction is important for areas with growing population and urbanization that increase vulnerability to extreme weather conditions. For instance, Nairobi County experienced unusual extreme rainfall between April and May, an issue that is attributed to inaccurate and unreliable weather forecast by the current methods [1]. With such extreme conditions, urban areas risk disruption in their socioeconomic activities citing the need for accurate forecasts. Precise weather forecasts help urban managers to plan, enhance safety, mitigate delays in traffic, and make informed adaptive strategies encountering climate change [1]. Such strategies include resource allocation, infrastructural development, and policy formulation tailored at improving the county's resilience to climate related risks.

Traditional weather methods, Numerical weather prediction (NWP) and synoptic forecasting, have setbacks accommodating non-linear and complex interactions between different climatic factors. NWP depends on technical mathematical models that simulate atmospheric behavior through fluid dynamics and thermodynamics principles which is resource intensive [2]. Equally, the NWP models' accuracy diminishes when forecasting short-range, climate-enhanced weather events more so in areas with complex geographical set up like Nairobi [4]. The technique is highly sensitive to data quality and resolution; hence, small errors in the initial conditions result in low precision forecasts, a scenario known as butterfly effect. The method also struggles in forecasting small-scale weather phenomena like localized storms based on their limited temporal and spatial resolution. Synoptic weather prediction is subject to recognition of nature patterns and dependence on historical analogs making the technique less precise more so with the changing climatic and weather conditions. The rising variability in patterns of weather because of climate change makes this technique less precise. The third method, climatology, lacks precision as it provides general trends making it less useful; it does not accurately forecast extreme weather conditions, making it unreliable [2]. Finally, the analogue method presents a challenge as getting exact analogs in unique and unprecedented weather conditions is difficult [3]. There is therefore, the need to deploy better techniques that solves the unreliability and inaccuracy challenges.

Machine learning can be used to fill these gaps. The flexibility of machine learning models in adapting non-linear relationships and high dimensional data makes them suitable in weather prediction [5]. Models such as Support Vector Machine, Random Forest, and Gradient Boosting have shown great promise in weather prediction applications. Random Forest has high precision considering the multiple decision trees. The method constructs multiple decision trees on numerous data subsets and averages their predictions, thus mitigating overfitting and improving reliability of the forecast [6]. On the other hand, gradient boosting sequentially builds models with every new model amending the errors from the previous ones thus improving the overall accuracy and robustness. Support Vector Machines (SVM), SVM constructs hyperplanes that maximize the margin between various data classes; this makes it relevant and accurate at establishing patterns and trends within large, complex datasets. In this study, the three ML models (SVM, GB, and RF) are selected for analysis because of the weather data that is non-linear and complex, a problem that they solve. In addition, the techniques are known for their robustness and capability to manage high-dimensional data without overfitting. The models as well offer flexibility in hyperparameter tuning, enabling tailored optimization of models based on the given weather dataset characteristics [7]. While machine learning models are computationally complex, advancement in cloud computing and the existence of open-source libraries like scikit-learn, statsmodels, matplotlib, seaborn, and joblib have made them more feasible to implement even with limited resources [8].

The aim of this research is to use machine learning models to enhance weather forecasting for short and medium weather patterns. In this research, we assess and compare the accuracy and reliability of forecasting short and medium range weather patterns in Nairobi County using Machine Learning and ARIMA models. We conduct review of existing prediction methods and modes, develop and implement machine learning models namely RF, GB, and SVM for weather forecasting. We also evaluate the effectiveness of the machine learning models in terms of performance in comparison with other traditional models. The study therefore addresses the drawbacks of traditional weather prediction techniques in Nairobi County. This will enhance preparedness for disaster and improve risk management. It will also help reduce the effect of extreme weather conditions on social activities, healthcare service provision, and educational services, transport within and outside the county, and general economic activities. By accurately predicting weather, the county team will make proper and timely planning that will help avoid disasters like the recently experienced floods. Finally, through this study, advancement of weather prediction methods will be enhanced through the provision of models that can be replicated in other regions with similar problems.

2. LITERATURE REVIEW

In current years, the application of machine learning (ML) in weather forecasting has gathered significant attention. This attention has specifically been driven by the need for more accurate and reliable prediction amidst changing climate change. Several research works originally acknowledged traditional prediction techniques as better in giving accurate forecasts. This has transformed over time considering the irregular weather patterns being experienced. As such, current studies display the challenges of previous results by justifying the traditional techniques' incapacity to accurately forecast the patterns of short- and medium-term weather. In response, current studies advocate for the use of more robust ML algorithms to accurately predict patterns and trends in weather. This literature review evaluates how ML methods can enhance precision in forecasting weather. It explores the capacity by Random Forest, Gradient Boosting, and Support Vector Machines (SVM) to enhance accuracy prediction in comparison to Autoregressive Integrated Moving Average models.

2.2 Traditional Weather Forecasting Methods

There are many traditional methods used in predicting weather. These include Synoptic forecasting, Numerical Weather Prediction (NWP), climatology, persistence, and analog predictions [9]. The methods are discussed in the following subsections.

2.2.1. Numerical Weather Prediction

Numerical Weather Prediction technique employs complex mathematical models to simulate atmospheric patterns in accordance with the observed initial conditions. The method forecasts future weather through solving fluids and thermodynamics equations. The NWP models require high-quality and high-resolution data; they are also computationally intensive [3] making it difficult to employ them in forecasting. NWP is highly dependent on the initial atmospheric state, showing that observational data errors can multiply and expand during the prediction phase [2]. Categorically, this is mostly observed in areas with sparse observational networks and complex topography, like Nairobi County, resulting in low prediction precision using NWP [9]. In support of the challenge, [38], being specific on Africa, emphasized that NWP had low precision, making the forecast unreliable. This, as they claimed, is based on its northerly biases in the upper troposphere, meridional wind inconsistencies and cool prejudice in the northern latitudes that need not be employed on predicting weather in dense regions like Nairobi County.

2.2.2. Synoptic Forecast

Synoptic weather prediction involves weather patterns and systems analysis via upper-air and surface observations. Meteorologists interpret patterns in this technique to predict future events with respect to their skills and experiences. Findings highlighted that synoptic prediction, when integrated with progressive observational data, can substantially improve short-term weather forecasts through providing comprehensive insights into weather systems [5]. Nonetheless, the method remains limited to individual explanations and the integral complexity of meteorological patterns, requiring the incorporation of machine learning models to improve prediction reliability and accuracy. This method gives important insights; nevertheless, its precision is limited to the subjective nature of identifying patterns and dependence on historical analogs [7]. Besides, growing weather inconsistency in weather patterns as a result of climate change obscures further the synoptic technique as historical comparison growingly becomes less dependable [11]. In justification to this limitation, depending on long-term historical weather to predict short- and medium-term conditions results in low accuracy because of the ever-changing weather patterns, a factor that applies in Nairobi weather conditions [9].

2.2.3. Climatology

Climatology depends on historical weather data to forecast future weather events based on long-term averages and patterns. This method has been shown to be effective in providing baseline forecasts, as demonstrated by studies such as those by [5], which utilized historical climate data to generate reliable long-term weather predictions. However, climatology can fall short in capturing short-term variability and recent climatic shifts, which can be addressed by integrating more dynamic techniques like machine learning to enhance forecast accuracy and adapt to recent changes in weather patterns. This is mainly useful in long term predictions like forecasting seasonal patterns and trends [8]. Nevertheless, its precision reduces for short term predictions in areas that experience substantial climatic changes [15]. Assuming that future weather conditions will be like the historical average usually fails in the case of rising climate variability[37]. With respect to Ethiopia and East Africa region, climatology technique as obsolete in a period where rapid changes in climatic and weather conditions are experienced due to urbanization. Specifically, they state that with the construction of skyscrapers, the prevailing weather patterns adjust and cannot thus be forecasted on historical averages.

2.2.4. Analog forecasting

Analog forecasting entails the comparison of current weather patterns with the historical patterns that gave similar weather conditions. Here, forecasters check on historical analogs to forecast the possible outcome of existing conditions. This technique has been effectively used in studies [19], which demonstrated that identifying historical analogs can provide valuable insights into future weather events [6]. However, analog forecasting can be limited by its reliance on historical data, which may not account for recent climatic changes or emerging trends, making it beneficial to combine this approach with machine learning techniques to capture more complex and current weather dynamics. Whereas analog technique can be important, it is limited by the existence of same historical patterns and the ability by the predictor to recognize and interpret accurately these analogs [33]. They further state that some weather events are unique, such as extreme floods, which render the technique ineffective. Justifying the above limitation the swiftly changing weather patterns in urban areas within East Africa renders analog forecasting less effective as current weather patterns are to a better percentage, not like the historical patterns [32].

2.2.5. Autoregressive Integrated Moving Average (ARIMA)

ARIMA modeling has extensively been employed in weather forecasting specifically on linear time series data. ARIMA model forecasted visibility based on existing historical data [12]. Another application was done by Tektas (2010) in Instabul for general weather predictions [13]. ARIMA modeling is designed for univariate data and focusses on a single variable such as rainfall, wind speed or humidity [10]. The model, as explained by [45], incorporates autoregressive, integrated, and moving average components to capture trend and seasonality patterns making it important for predicting short term weather. The model dependence on past data allows it to forecast coming weather conditions with moderate to high accuracy in case of consistent conditions [17]. Nonetheless, ARIMA lacks the capacity to handle non-linear weather conditions with sudden changes like storms, hurricanes and atmospheric pressure [12]. Thus, being a linear model, ARIMA struggles to precisely forecast complex and non-linear interactions. In a different study, ARIMA modelling, being univariate, does not account for interdependence among numerous variables, which is key in accurate prediction [28]. The study indicated that while ARIMA can account for time series trends, it cannot establish relationships between weather variables such as temperature vs humidity or rainfall vs hurricanes hence another limitation. ARIMA assumes stationarity, as such, it only works best with data having constant statistical properties over time, a case that is rare considering the volatile weather patterns [13].

2.3 Machine Learning

Machine learning models are classified into three including supervised learning, unsupervised learning, and reinforcement learning [18]. The three categories are applied depending on the available data and nature of the task. In supervised learning, models are trained based on labeled datasets. Models learn through associating input data with corresponding output labels [19]. Supervised learning is effective in regression and classification tasks where the target is to forecast results of new, unseen data. Some of the commonly used supervised models include support vector machines, decision trees, Random Forest, Neural Networks and Gradient Boosting. Supervised learning is applied in different domains considering its capacity to generalize from the training data [23]. Nonetheless, it requires large, labeled datasets which is difficult to get in fields like weather forecasting hence a limitation. On the other hand, unsupervised learning algorithms identifies patterns and relationships in unlabeled dataset without predefined output labels. Some of the used unsupervised methods include dimensionality reduction and clustering to explore data structures [11]. While unsupervised machine learning is important in exploratory data analysis, is less effective in making actionable forecasts thus a limitation [28]. The final machine learning technique is reinforcement learning, where an agent learns through environmental interaction and feedback received in the form of penalties or rewards. Therefore, reinforcement learning is mainly important in tasks requiring sequential decision-making since the agent performance improves continuously based on previous experiences [24]. Reinforcement learning is successful in domains such as robotics and games [13]. Nonetheless, the study established that the application of reinforcement learning in predicting weather is limited based on the sequential nature of prediction in which future weather patterns relieve past patterns. For this study, supervised machine learning is selected because it gives room for developing predictive models based on labeled datasets, in which historical weather data is associated with known outcomes. The technique is specifically effective in establishing complex patterns and associations within the data, improving forecasting accuracy for short and medium-term weather conditions. Unlike unsupervised learning seeking to find hidden structures in unlabeled data, and reinforcement learning depending on feedback from actions taken in an environment, supervised learning gives a direct mapping from inputs to outputs making it relevant for this study [11]. The direct design, alongside its ability to fine-tune algorithms and validate them using performance metrics, makes supervised machine learning models robust and reliable for making forecasts in a real world context, thus highly suitable for weather forecasting.

2.3.1 Classification Techniques in Machine Learning

Classification techniques are critical in machine learning for tasks such as classification and prediction of weather patterns. Three supervised machine learning techniques including Random Forest, Gradient Boosting, and Support Vector Machine are discussed in this section.

i. Random Forest

Random Forest (RF) is an ensemble machine learning method that builds numerous decision trees during training; it outputs the classification mode or mean regression (prediction) of the single trees. RF is specifically efficient in mitigating overfitting and capturing non-linear associations in the data [21] The method has widely been applied in weather forecasting in developed nations to forecast weather conditions as precipitation, temperature, and wind speed [8]. In the United States for example, a blend of the Storm Prediction Center's (SPC's) convective outlooks and RF outlooks significantly outperformed the SPC outlooks alone, suggesting that use of RFs can improve operational severe weather forecasting throughout the Day 1–3 period, which is short term [44]. In the same study, random forest machine learning model recorded minimum error values of 0.750 MSE and R2 score of 0.97 relative to regression models and SVM ML model showing that this technique is more accurate in short-, medium- and long-term weather patterns [44]. Therefore, it is worth noting that the application of RF in forecasting short- and medium-term weather patterns within Nairobi County will result in a more accurate result. While the technique offers robust solution, it has computational limitations including high cost and memory usage due to the large number of decision trees, issues that can be solved through hyperparameter tuning to optimize the number of trees. It also requires significant training time when using complex data and may experience parallelization challenges that further affect the efficiency of real-time weather forecasts [22].

ii. Gradient Boosting

Gradient Boosting (GB) is a powerful machine learning method that sequentially builds models. Here, every new model corrects errors of the previous ones thus enhancing the overall accuracy and robustness of the previous model, specifically in handling complex data relationships and outliers. Friedman introduced the Gradient Boosting Machines (GBMs) concept in 2001 [23]. This has since been used to forecast meteorological variables with high precision. In a study by Anwar et al. (2021) involving 7-year period historical weather, the gradient boosting model produced an accurate daily rainfall forecast with training RMSE of 2.7 mm and the testing MAE of 8.8 mm. Nonetheless, they experienced longer duration of implementation which is a challenge. In a different study on wind speed forecast, the results displayed that after using 300 features in different layers and heights, GB resulted in more accurate results than weather and research forecasting (WRF) models like decision tree regression (DTR) and multi-layer perceptron regression (MLPR) [27]. Particularly, GB reduced root mean squared error (RMSE) of the predicted wind speed from 2.7-3.5 ms^{-1} in the original WRF. It also improved the index of agreement (IA) by 0.10-0.18 and Nash-Sutcliffe efficiency (NSE) by 0.06 – 06. Stating the technique limitation, they experienced high computational cost and memory usage due to sequential training of trees. In a different study on SYM-H index, GB resulted in more accurate and consistent prediction with physical understanding [16]. Specifically, GB yielded a more statistically significant improvement in root mean squared error relative to Burton equation and black-box neural network schemes. Like the last two studies, they also experienced problems with computation time and cost rendering the technique resource intensive. Despite these limitations, this technique is among the important machine learning algorithms that should be employed in forecasting short- and medium-term weather patterns within Nairobi County.

iii. Support Vector Machines

Support Vector Machines (SVM) are supervised learning models; SVM analyzes data for regression and classification analysis. The technique is effective for high-dimensional spaces; it can also handle non-linear data through kernel functions [24]. In weather prediction, SVMs have been used to forecast phenomena like temperature, precipitation, and wind speed. Introduced by Vapnik (1995), SVM algorithm has since been adapted in various applications, including meteorology. SVMs have the capacity to establish subtle patterns in large and complex sets of data that traditional methods might miss, thus enhancing forecasts involving short and medium-range weather conditions [25]. SVM model improved the spatial representation of hourly precipitation, increasing correlation coefficient from 0.39 to 0.49 in January and 0.24 to 0.30 in July in the cross-validation experiment in Japan [21]. The technique nonetheless underestimated the hourly precipitation variability same as extreme precipitation events, a factor attributed to the non-linearity of precipitation data, suboptimal feature selection, sensitivity to noise, and complex atmospheric interactions that are hard to model accurately. In a different study, forecasting humidity, wind speed, temperature, and rainfall in Indonesia, a precision value of 82% rain was obtained with an evaluation ROC score of 0.74 further showing the ML algorithm's enhancement of forecasting accuracy [19]. The method, nonetheless, recorded high cost of computation, a factor that was solved using kernel approximation to reduce complexity, as reported by the authors, giving an insight that machine learning models are generally expensive to employ in weather forecasting. Nonetheless, alongside Gradient Boosting and Random Forest, Support Vector Machine algorithms need to be employed in quest to establish the most accurate forecasting technique for Nairobi County Weather.

2.3.2 Data Representation in Machine Learning

Representing multidimensional data that change over time and across geographical location involves the use of time series, spatial and tabulation. Time series data is key in weather prediction; it captures temporal progression of atmospheric variables including wind speed, precipitation, and temperature among others. Time series representation allows machine learning models to account for longitudinal changes, thus important for forecasting future weather patterns [15]. Over time, traditional models like ARIMA have widely employed time series to forecast weather considering its ability to model time dependencies and linear relationships [44]. Nonetheless, the models struggle to capture non-linear and complex relationships thus limiting the forecasting accuracy [42]. The use of machine learning techniques like Long Short-Term Memory (LSTM) networks that are particularly designed to solve both short and long-term fluctuations in sequential data solves this problem. Through effectively learning the weather patterns and handling longer data sequences, LSTMs have enhanced accuracy in weather prediction [14]. The second representation, spatial data, captures geographic variability in the weather system more so for phenomenon like cloud formation or storms. Spatial information is extracted from radar data, satellite imagery, and meteorological stations, providing data on atmospheric conditions within given locations [30].

Handling spatial data has majorly been done using Convolutional Neural Networks (CNNs) that process images [44]. CNN is preferred as they identify local patterns of weather; it also forecasts the changes or movements in the patterns over time. CNN successfully identifies complex weather patterns such as hurricanes and storms, thus enhancing accuracy in region-specific predictions [21]. The technique, however, experience a problem in integrating spatial information with temporal dependencies, presenting a gap for research [24]. The final representation is tabular data that represents weather components in structured format. Here, every row corresponds to a specific observation and each column a variable. Tabular data is mostly employed for simpler machine learning models such as support vector machine and decision trees [14]. Despite being easy to manage, extensive preprocessing is required alongside feature engineering to capture relationships between variables [43].

2.3.3 Evaluation Metrics for Classification Models

Model evaluation in machine learning is critical in establishing their performance, reliability and accuracy. Evaluating machine learning models entails different techniques including bootstrap, cross validation, and performance metrics like precision, accuracy, recall, F1-score, and root mean square error (RMSE)[9]. Cross validation is employed in evaluating generalizability in machine learning models. According to the study by Roberts et al. (2021), cross validation involves splitting data into k folds, training the model on k-1 subsets and testing it using the remaining sets. The process is repeated k times, then the mean performance metric reported. In weather forecasting, as reported by Merz et al. (2020), cross validation is used to mitigate overfitting, a crucial aspect considering the complex nature of weather data. However, according to Lamptey et al. (2024), K-fold requires keenness when using time series data to ensure there is no data leakage between training and test subsets. As reported by Roberts et al. (2021), the keenness can be attained through models like time series cross validation, in which data is divided along temporal axis. The second evaluation metric, bootstrap, is a resampling method that repeatedly sample data with replacement, hence applicable in evaluating the variability of model forecasts. The technique was introduced by Efron and Tibshirani (1994) to solve small, noisy datasets, making it important in cases of limited or incomplete datasets [26]. While bootstrap reduces bias, it is computationally intensive when applied to large dataset; the method may not capture fully variability in cases of complex weather datasets [24]. Another measure of performance is accuracy is the percentage of correctly forecasted instances of the possible outcomes [27]. Accuracy has been employed in meteorology to assess the performance of machine learning models and to make comparisons between different models. The technique alone, however, can be misleading when data is imbalanced like during storms and tornadoes. In such cases, models can attain higher percentage accuracy without effectively capturing critical but rare conditions [28]. Therefore, additional metrics like precision and recall are used to comprehensively forecast weather patterns. In an explanation, while precision measures the percentage of correct positive predictions, recall measures the actual number of correctly predicted positive cases [29]. Precision and recall are important when forecasting extreme weather events such as hurricanes and thunderstorms unlike bootstraps and cross validation [33]. Another evaluation technique, F1-score, refers to harmonic precision and recall mean, thus balancing the tradeoffs between false negatives and positives. F1score is applicable to rare but impactful weather events like flash floods where over-forecasting an event or missing it have substantial impact [28]. For regression specific tasks, metrics like mean absolute error (MAE), and root mean squared error (MAPE) are used [29]. The methods are sensitive to large errors capturing extreme deviations like in cases of excess rainfall. Hence, RMSE is crucial in forecasting outlier weather conditions that causes significant disruption such as extreme weather patterns [27]. With different shortcomings, evaluating machine learning models requires integrating various techniques to provide a holistic view [24].

2.3.4 Feature Selection

In machine learning, feature selection entails establishing and choosing the most relevant features from a dataset. The process is important in high-dimensional datasets like weather where several weather condition variables are recorded. In such cases, feature selection eliminates redundant or irrelevant variables thus reducing model complexity, enhance accuracy and improve interpretability [30]. Feature selection methods include filter, wrapper, and embedded methods. Discussing the filter techniques, researchers stated that the method assesses feature relevance by their intrinsic properties; this method is mainly used as a preprocessing step [31]. Common examples of filtering techniques include correlation coefficients, chi-square tests and mutual connection. Another study reported that while filtering methods are effective computationally, they do not consider model-feature interactions [30]. The next feature selection method is the wrapper; this evaluates subset features by assessing model performance using those subsets [14]. Wrapper techniques such as Recursive Feature Elimination (RFE) and forward/ backward selection iteratively remove features that are least significant and evaluate model performance to establish most significant ones [36]. In predicting weather, RFE is better than filter methods as it considers feature interactions [14]. The third method, embedded, integrates feature selection into the process of model training, enabling concurrent feature selection and model training [38]. The commonly used embedded featuring selection techniques include decision trees and Lasso (Least Absolute Shrinkage and Selection Operator). Another study reports that Lasso employs L1 regularization to punish features that are less important; this has long been used in handling high-dimensional data like in weather forecasts [30]. On the other hand, decision trees such as Random Forest naturally conduct feature selection via assessing feature importance during the process of tree building [37]. The study report further showed that feature selection done using Random Forest can guide in choosing relevant variables thus mitigating overfitting while enhancing forecasting accuracy. Combination of feature selection techniques and Principal Component Analysis (PCA) in extracting features from weather dataset successfully extracted main features thus improving model accuracy and efficiency [24]. Further, they established that integrating feature selection techniques with Machine Learning models such as RF and SVM results in enhanced weather forecasting outcome by mitigating the effect of irrelevant features; this also reduces the cost of computation. While beneficial, feature selection experiences the challenge of balancing tradeoff between retaining important features and reducing dimensionality [34]. Here, over-selection may result in losing important data whereas under-selection may lead to missing important predictors. Therefore, they suggested a hybrid technique integrating numerous feature selection techniques is recommended to achieve an effective and comprehensive selection process.

2.3.5 Machine Learning in Weather Forecasting

Integration of machine learning in weather forecasting has resulted in significant improvements towards attaining accuracy [5]. In countries like the US, a proven forecast effectiveness for short term precipitation, temperature and wind speed conditions through Random Forest, which aggregates numerous decision trees to make robust forecasts [44]. The method, however, requires a high-level interpretability, making it hard for meteorologists to extract important information from the output. As such, it is recommended tha development of hybrid models involving Random Forest and dimensionality reduction methods to improve interpretability while reducing the cost of computation [27]. The second widely used method is Gradient Boosting Machine (GBM), specifically XGBoost applies an iterative approach by refining forecasts based on previous errors; this makes it important in forecasting extreme weather patterns [15]. In a practical scenario using employed XGBoost in forecasting heavy rainfall which showed a significant improvement in accuracy after iteratively learning the process [34]. The method, however, requires extensive hyperparameter tuning and is computationally extensive [12]. Another ML technique used in forecasting is support vector machines that classify weather events and forecast storm occurrences. A study in Asia employed SVM to classify weather patterns based on different conditions thus improving forecasting accuracy for storm events [34]. They however reported that while SVM managed to handle complex, nonlinear data, it struggled with very large sets and required careful tuning of parameters. Nevertheless, advanced kernel and approximation methods overcome these limitations thus making SVMs more efficient and more scalable while handling large climate data [36]. Another method that has been employed in weather forecasting is Neural Networks, which includes deep learning models capable of capturing complex patterns in weather [25]. A different study employed Neural Networks and Long Short-Term Memory (LSTM) networks to forecast for typhoons in Japan; these resulted in enhanced prediction accuracy [39]. The models, however, required extensive cost of computation and large datasets hence limitations. Whereas individual ML models have high accuracy and precision in weather forecasting, a different study recommended the use of ensemble method combining forecasts from several models to attain the best prediction [38]. The combination of techniques including SVMs, GBMs, and RF can enhance accuracy by averaging the forecasts from different models thus reducing individual bias; this also solves individual cases of underfitting, and overfitting.

2.3.6 Machine Learning Models vs. Traditional Methods

Table 1 compares Machine Learning Models and Traditional Foresting techniques based on strengths and weaknesses.

Table 1: ML-Traditional models' Summary table

Criteria	Traditional Weather Forecasting Models	Machine Learning Models
Data Requirements	ARIMA: Requires less data, works with smaller datasets [5].	Large Data Needs [25]
Scalability	ARIMA: Handles smaller datasets effectively [13].	Efficient Processing of Large Datasets [31]
Enhanced Accuracy in Nonlinear Systems	ARIMA: May not perform well in highly volatile weather scenarios [26].	Enhanced Performance in Complex Weather Phenomena [25]
Handling Complex and Nonlinear Patterns	ARIMA: Assumes linear relationships and stationary data [34].	Capability to Model Complex Dynamics [15]
Computational Demand	ARIMA: Computationally less demanding [30].	High Resource Requirements [15]
Predictive Framework Based on Historical Data	ARIMA: Effective for well understood linear time series data [30].	Robustness in Well-Defined Scenarios [30]
Interpretability	ARIMA: Provides clear, interpretable results with model parameters [12].	Black Box Nature [35]
Risk of Overfitting	ARIMA: Less prone to overfitting with simpler models [4].	Potential for Overfitting [8]

2.4. Conceptual framework

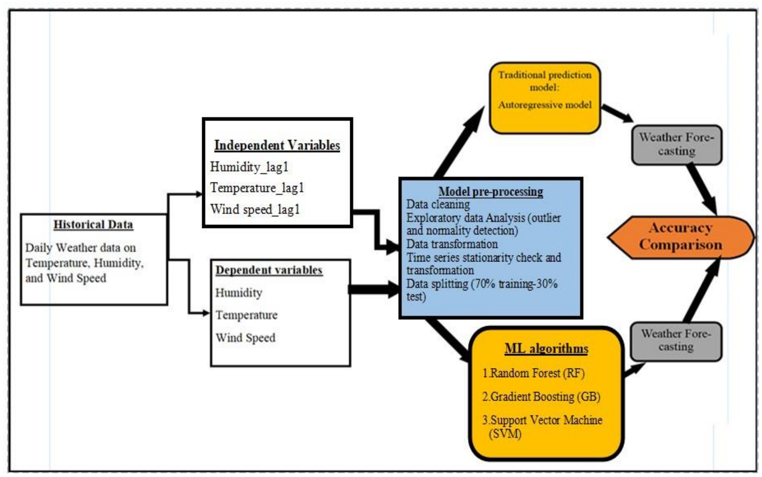

Figure 1: Conceptual framework

The above conceptual framework displays the study variables and the process taken to attain study objectives.

3. METHODOLOGY

In this section, the methodological techniques used to attain study objectives, that is to enhance short- and medium-term weather forecasts, are discussed. It provides details regarding research paradigm, design, and population, sample size, sampling techniques used, models to be created and ethical consideration to be considered. The section seeks to provide a systematic framework for collecting, analyzing, and interpreting data to accurately develop weather forecasting models for Nairobi County.

3.1 Research Paradigm

The study adopted a positivist research paradigm. The paradigm is based on the belief that reality can be objectively observed and measured using empirical evidence. Positivism highlights how quantitative techniques are used in data collection and analysis

[40]; this made it convenient for the study goal which is to enhance weather forecasting accuracy within Nairobi County. By integrating machine learning algorithms and statistical analysis, the study targeted to create models that can give accurate and dependable weather forecasts. Positivism supports the hypothesis-testing methodology, in which machine learning algorithms are trained and assessed against traditional prediction methods to establish their effectiveness. It also ensures a systematic and scientific technique, initiating the generation of replicable and generalizable results [41]. Hence, positivism aligned with the objective of this study of integrating advanced forecasting methods to solve real-world problems in weather prediction, eventually adding to enhanced risk management and disaster preparedness in Nairobi County.

3.2 Research Design

The study adopted experimental research design. The technique involves manipulating one or more independent variables to observe their effect on a dependent variable, typically under controlled conditions, to establish cause-and-effect relationships [42]. In this study, the same data was modelled using SVM, Random Forest, Gradient Boost and ARIMA modelling to establish which one provides the most accurate forecast in different weather categories.

The first step involved collection of historical weather data on wind speed, temperature and humidity based on days in Nairobi County. Here, the dependent variables included historical data on wind speed, temperature, and humidity while the independent variable included windspeed_lag1, temperature_lag1, and humidity_lag1 respectively; the independent variables served as proxies for temporal effects. The next step was model development where Autoregressive Integrated Moving Average (ARIMA), Gradient Boosting (GB), Random Forest (RF), and Support Vector Machines (SVM) were created and trained using data. Thereafter, the developed models were evaluated using metrics such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Finally, comparative analysis was done to establish the most suited approach in handling weather data forecasting within Nairobi County.

3.3 Dataset

The dataset of this study consists of all historical weather records for Nairobi County; this includes information on wind speed, temperature, and humidity. Periodically, the population covers every recorded weather data specific to Nairobi County. The data was sourced from a secondary database containing publicly available weather dataset, Weather Query Builder (https://www.visualcrossing.com/weather/weather-dataservices/). The site is selected considering that it has a real time update on daily weather values. Also, Visual Crossing provides historical weather data dating back to January 1, 1970, ensuring a robust dataset for long-term analysis. Thirdly, the service offers global geographic availability, making it suitable for studies focusing on diverse locations, including Nairobi County. Moreover, Visual Crossing integrates data from over 100,000 weather stations, employing advanced interpolation methods to maintain data accuracy and reliability, even in areas with sparse station coverage. Additionally, the platform provides detailed weather metrics, including wind speed, humidity, and temperature, which are crucial for in-depth climatic analyses. Finally, Visual Crossing is recognized for its established credibility, as it is widely used in both academic research and industry applications, making it a trusted source for meteorological data analysis. Reliability of the data source was attained through cross-referencing sampled weather statistics with those from Kenya Meteorological department to ascertain precision. The data is acquired for the period between 1994 and 2024 (30 years) translating to 11312 observations measured in days. The data was downloaded in CSV format and is available for re-experiment.

The study employed stratified sampling technique to extract three weather events of interest (humidity, temperature, and wind speed). These elements were selected considering that they take place daily and hence daily data available with minimum missing values hence data quality. This is unlike factors like rainfall and other precipitations that occur seasonally. From the three conditions, a sample spanning 30 years (1994-2024) was extracted. This is considering that the period represents short to medium term weather data needed for this study. For the stated period, data was collected on daily frequency resulting in a sample size of 11312. The size was chosen considering that it is large enough hence capable of building robust models.

3.5 Modelling Tools

Several libraries were utilized in forecasting short and medium-term weather conditions in Nairobi County using Gradient Boosting (GB), Random Forest (RF), Support Vector Machine (SVM), and ARIMA models in Python. Gradient Boosting, Random Forest, SVM, and ARIMA were selected due to their strong performance in regression tasks, ability to model nonlinear relationships, and suitability for structured and time-series data. NumPy and Pandas were used to facilitate numerical operations and data manipulation, whereas Scikit-learn provided the required tools for implementing RF, GB, and SVM, as well as evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared. On the other hand, Statsmodel was utilized for ARIMA modelling while Matplotlib and Seaborn helped visualize weather trends and model performance. Other relevant libraries SciPy for statistical computations. Finally, Joblib library was used to initiate model saving and efficiency during training. The above libraries ensured robust data analysis, modeling, and assessment of forecasting accuracy.

3.6 The Weather Forecasting Model

The developed model design, shown in figure 2, contains a robust workflow for short- and medium-term weather prediction in Nairobi County. The diagram starts with data collection and preprocessing, entailing cleaning of data and feature engineering to improve model efficiency. It shows a data split of A 70%-30% train-test to ensure robust evaluation.

It then shows model selection phase containing four efficient algorithms: Random Forest, Support Vector Machine, Gradient Boosting, and ARIMA, all optimized through cross validation and hyperparameter tuning. It then shows evaluation metrics such as MSE, MAPE, MAE, and RMSE that assesses performance. Visualization of results and accuracy gives insights into model precision and trends, whereas techniques for handling computational complexity are important for scalability and efficiency throughout the modeling process. The contribution to knowledge is reflected in the comparative integration of machine learning models (Random Forest, Gradient Boosting, SVM) with a traditional time series model (ARIMA) within a unified framework for localized weather forecasting. In addition, the model design introduces a systematic training and evaluation pipeline that emphasizes reproducibility and objective comparison through standardized train-test splits and the application of appropriate error metrics. Performance enhancement was achieved by tuning model hyperparameters and ensuring data integrity through preprocessing steps, which collectively improved prediction accuracy without reliance on manual or heuristic interventions. By situating the model specifically in the context of Nairobi County, the study addresses a contextual research gap in the application of hybrid predictive algorithms to weather forecasting in Sub-Saharan urban settings.

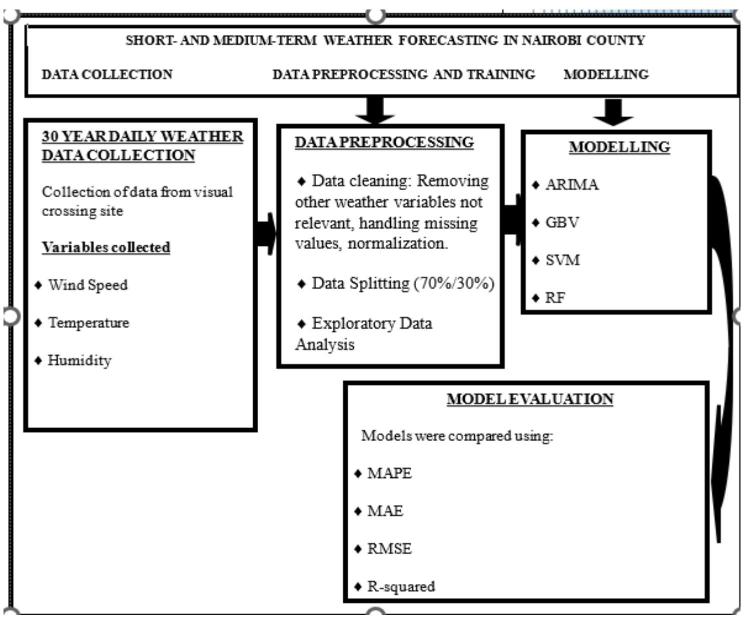

Figure 2: Model process

3.7 Experimental Setup

Model development involved a hybrid strategy, integrating both machine learning and classical time series methods [39], to evaluate and compare prediction performance across different techniques. The first step involved performing Augmented Dickey-Fuller (ADF) stationarity test, a critical assumption for time series modeling to understand whether temperature, humidity, and windspeed series exhibited unit roots. The test resulted in p-values lower than 5% level of significance ($p < .001$) for all three, justifying the use of ARIMA models without differencing. A further automated grid search was performed using SARIMAX to establish the optimal ARIMA(p,d,q) orders for each dependent variable based on the lowest AIC. This resulted in the ARIMA order 1,0,1 for all the three weather conditions, the order that was selected while performing ARIMA modelling. Next, the data was split into 70% training and 30% testing set, a split that was chronological. Time-based split was key in preserving temporal ordering of observations, thus mitigating data leakage and simulating real-world prediction conditions where models forecast future values based on historical trends. The next step involved introducing lag features to capture temporal dependencies that are integral in weather to enhance the predictive capacity of machine learning models. For every weather variable (temperature, humidity, and windspeed), a lag of one-time step (temp_lag1, humidity_lag1, and windspeed_lag1) which served as predictor variables for forecasting their respective current values. Lag 1 was selected considering the short-term temporal dependency assumption usually observed in daily weather patterns. Further, an autocorrelation test was done on individual variables to assess the presence of temporal dependence. This was visualized as displayed in figure 3 which displayed strong autocorrelation for temperature and humidity and a weak one for windspeed.

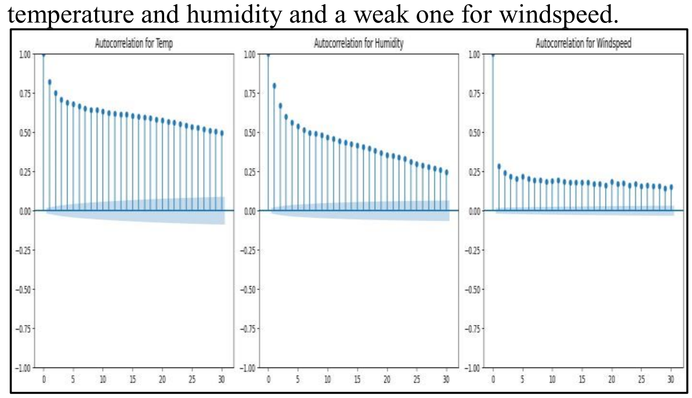

Figure 3: Autocorrelation results

The significance of autocorrelation was further tested for the three and this is displayed in table 2 below.

Table 2: Autocorrelation results

```
TEMP Lag 1 ACF:
   Lag       ACF  CI Lower  CI Upper  Significant
1    1  0.823299   0.80487  0.841728         True

HUMIDITY Lag 1 ACF:
   Lag       ACF  CI Lower  CI Upper  Significant
1    1  0.796851  0.778423   0.81528         True

WINDSPEED Lag 1 ACF:
   Lag       ACF  CI Lower  CI Upper  Significant
1    1  0.283742  0.265313   0.30217         True
```

Autocorrelation analysis revealed statistically significant lag-1 dependencies for temperature, humidity, and windspeed ($p < 0.05$), justifying the inclusion of one-period lag features as predictors in the modeling process. Using the 70% training set, Support Vector Regression (SVM), Random Forest (RF), and Gradient Boosting (GB) were trained independently for temperature, wind speed and humidity.

3.8 Model Evaluation

Model evaluation assesses the performance of the trained machine learning models. The effectiveness of each model was measured using four performance metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared. These metrics provided insights into the accuracy and reliability of the models' forecasts. The performance of machine learning models was compared to traditional forecasting methods to determine which approach offers the most accurate predictions for temperature, humidity, and wind speed. This evaluation ensured that the chosen models deliver reliable weather forecasts, contributing to better decision making and risk management.

4. RESULTS

The chapter presents the findings of the prediction models employed to forecast short and medium-term weather patterns in Nairobi County, covering between 1994 and 2024. The study applied four models including Gradient Boosting (GB), Support Vector Machine (SVM), Random Forest (RF), and ARIMA models to project humidity, temperature, and windspeed. The models were trained and tested through lagged input variables extracted from historical weather data, with a clear distinction between training and testing periods; this ensured unbiased evaluation. Evaluating model performance was done using four primary metrics including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the R-squared (R^2) statistic. Of the four, the best-performing machine learning model was established for temperature, humidity, and windspeed separately, then compared to the primary traditional method, ARIMA values. each weather variable, followed by a comparative analysis with ARIMA forecasts. Using best performing models, prediction visualizations were created for every weather condition compared to actual values. Besides, the section displays forecasted weather patterns for the first 10 days of 2025 and the average monthly forecasts for the entire 2025. Besides, the section chapter discusses the relevance of model selection criteria, including consistency, accuracy, and interpretability. Issues experienced during modeling, like the existence of negative R^2 values, are addressed to give a transparent account regarding the modeling process.

4.1 Exploratory Data Analysis

Summary statistics

The daily values for the three elements were summarized using mean, standard deviation and five-number summary as displayed in table 3 below.

Table 3: Descriptive Statistics results

Weather Element	Count	Mean	Std	Min	25%	50%	75%	Max
Temperature (F)	11312	66.83	2.98	55.4	64.8	66.9	68.9	77.3
Humidity	11312	70.33	9.73	34.5	63.9	70.9	77.4	97.3
Windspeed	11312	15.7	6.1	0	12	15	18.3	135.8

In the last 30 years, Nairobi County experienced an average temperature of 66.83 (SD = 2.98)^{0}C. The lowest temperature was 55.4 recorded in 7/2/2004 while the highest temperature was 77.3 as recorded on 3/2/1998, suggesting an inconsistency in temperature change patterns. While the value of standard deviation is low (2.98) suggesting a small variation around the mean, this is large enough to disrupt ecosystems by accelerating biodiversity loss, changing species distributions, and intensifying climate-related stresses like drought, wildfires, and habitat degradation, calling for accurate forecasts. The county experienced an average humidity level of 70.33 (9.73) with the value of standard deviation expressing a wider variation around the mean, suggesting over-time disparity in daily humidity values. This was further displayed by the minimum and maximum values that were 34.5 (2/22/2000) and 97.3 (11/4/2001) respectively, resulting in a range of 62.8 in a span of 1 year. This shows a huge difference in a short time, suggesting the need for accurate models that can handle sharp spikes in weather patterns over time. Finally, the county had an average windspeed of 15.7 (SD = 6.1) ms^{-1}, which is way higher than the recommended figures. The county recorded a single day with 0m/s (1/15/1998) and 4795 days within and below the desired 10.8 – 13.8 m/s range suggesting that in most cases (6516 days), it experienced higher speeds that may pause danger. For instance, some days like 3/14/2011 experienced 135.8m/s windspeed, a similar case with 4/29/2010 (117.4), 2/5/2014 (115.9), 9/25/2022 (114.1), 2/26/2016 (110.9), 7/6/2018 (108.3), 8/29/2015 (100.9), and 10/22/2015 (100.8), values that are extreme and dangerous for the county environment and infrastructure. It can be insighted from the date figures that no linear trend was followed, calling for forecasting models that can handle non-linear patterns effectively.

4.2 Seasonality trend Analysis

Daily trends

Establishing trends for temperature, humidity, and windspeed was done as shown in figure 4. Here, there is no clear linear pattern (negative or positive with respect to time). This suggests irregularity in weather patterns that require advanced models that can capture nonlinear weather conditions. The highest daily disparities were recorded for windspeed that had highest and lowest spikes, suggesting greatest difference. Besides, humidity also recorded similar trends with lowest disparity emanating from daily temperatures. The results suggest that Nairobi County weather patterns specific to Temperature, Humidity, and Windspeed are irregular with no specific trend, making it difficult to project using traditional patterns that do not capture non-linear complex data.

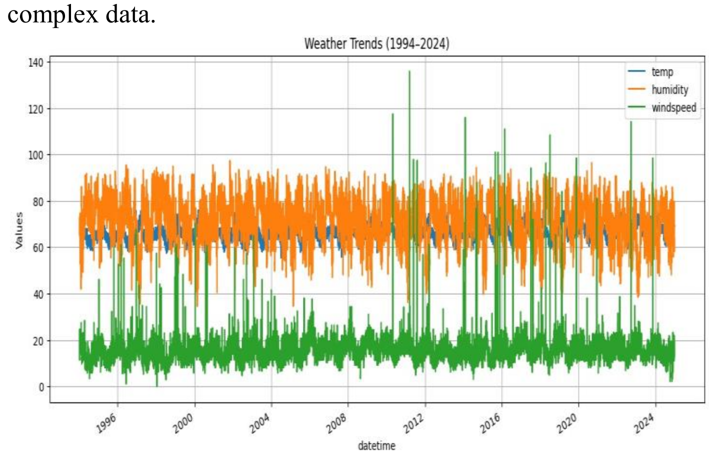

Figure 4: Daily weather (Temp, Humidity, Windspeed) trends over time (1994 – 2024)

Monthly Seasonal Analysis

Post normalization seasonality analysis was done establish monthly weather patterns
(temperature, humidity, and wind speed) over time across different periods in Nairobi County using boxplots as shown in figure 5. The results displayed that in the past 30 years, temperatures were lowest between July and August and highest in February, March, and October with outliers suggesting that the temperatures were not consistent throughout the year, calling for advanced analytics to establish the non-linear seasonal patterns. For humidity, the highest scores were recorded in April, May, November, and December with the lowest scores being in February and September. The presence of seasonal outliers

exhibited no significant trends making the data complex. Finally, windspeed recorded similar trends to that of temperature with highest values recorded in the first and last quarters of the year and lowest between May and August. Generally, the three weather patterns recorded outliers throughout the twelve months suggesting inconsistencies in their recorded patterns hence complexities, calling for robust models that captures non-linear relationships over time.

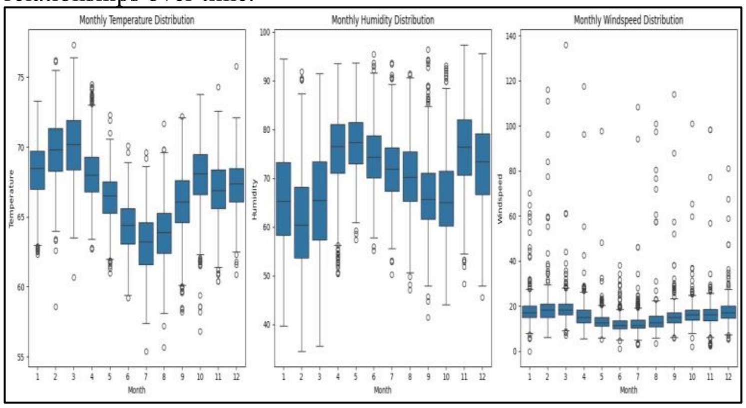

Figure 5: Seasonal variation of Temperature, Humidity, and Wind Speed

Normalization

The next step involved visualizing data normality, a property required to ensure that all input variables contributed equally to the machine learning models. This involved boxplots for temperature, humidity and windspeed as expressed in figure 6. Based on the boxplots, windspeed was strongly skewed to the right showing that most values were low with the presence of outliers, suggesting non-normality as fulfilled by the Shapiro Wilk test result ($p < .001$). Also, Humidity was skewed to the left showing that most values were high. A further normality test (Shapiro Wilk), resulted in a p-value of <.001, confirming nonnormality. Finally, temperature recorded outliers on both tails with a Shapiro Wilk p-value of <.001 suggesting non-normality. This non-normality, combined with differences in units and scales, necessitated the normalization of data prior to modeling. Such differences can bias models that are sensitive to feature scaling, like Support Vector Machines (SVM) and Gradient Boosting (GB).

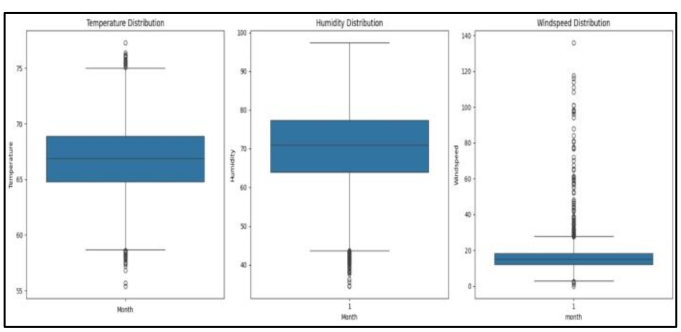

Figure 6: Temperature, Humidity, and windspeed distribution boxplots

Addressing skewness and non-normality involved the use of Min-Max normalization where all features were rescaled to a standard range of [0, 1]. The method preserved the shape of distribution while aligning the scales, giving room for the model to more efficiently converge during training and enhancing general predictive performance.

Thereafter, histograms were constructed to confirm data distribution as shown in figure 6. The results showed attainment of normality in temperature and humidity variables. While there existed positively skewed values in windspeed, bell shaped was attained, allowing for the assumption of normality.

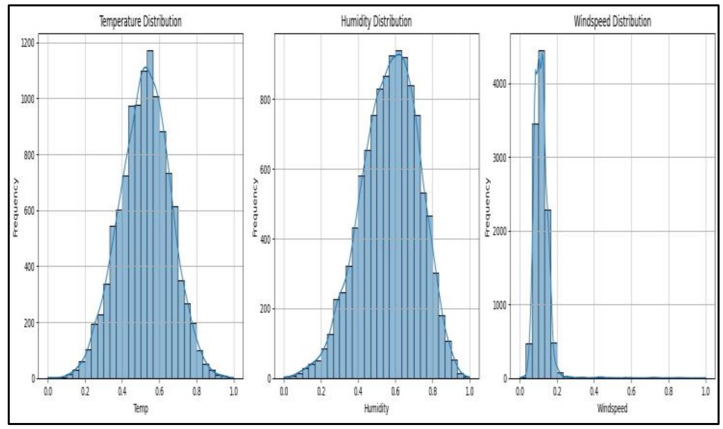

Figure 7: Temperature, Humidity, and windspeed distribution boxplots after normalization

4.3 Model Evaluation

The ML models, alongside ARIMA, were then evaluated using four metrics (MAE, MAPE, RMSE, and R-squared) as displayed in table 4 below.

Table 4: Model Evaluation results

Model		**SVM**	**RF**	**GB**	**ARIMA**
Variable	Metric				
Humidity	MAE	0.07	0.08	0.07	0.13
	MAPE	14.95%	16.29%	14.97%	31.65%
	RMSE	0.09	0.10	0.09	0.16
	R^2	68.60%	63.22%	68.37%	-3.48%
Temp	MAE	0.06	0.06	0.05	0.12
	MAPE	10.75%	11.27%	10.45%	21.76%
	RMSE	0.07	0.07	0.07	0.15
	R^2	72.72%	69.43%	74.28%	-17.86%
Windspeed	MAE	0.05	0.03	0.03	0.02
	MAPE	46.32%	23.05%	21.07%	24.03%
	RMSE	0.06	0.05	0.05	0.05
	R^2	-53.47%	-8.33%	0.31%	-0.79%

The evaluation results displayed better performance among Machine Learning models relative to ARIMA. The performance stood out specifically for humidity and Temperature. For humidity, lower MAE values were recorded in SVM (.07), GB (.08), and RF (.08) compared to ARIMA which recorded .13.

Besides, the three ML models recorded Mean Absolute Percentage Errors (MAPE) below the required maximum of 20% while ARIMA recorded 31.65% which is way higher. From the metric, the best performing models were SVM (14.95%) and GB (14.97%) followed by RF (16.29%). The same trend was confirmed in Root Mean Squared Error where GB and SVM recorded .09, RF recorded .10 and ARIMA recorded [16]. The final comparison was on the coefficient of determination where the highest model variation explained was extracted from SVM (68.60%) followed by GB (68.37%) and RF (63.22%). ARIMA recorded a negative R-squared which shows it is poor in predicting short- and medium-term humidity. Therefore, while seeking to establish accurate humidity forecasts, it is recommended to apply SVM having outstood in the four metrics. The trained model was applied in predicting the testing data and this is visualized in figure 8 below.

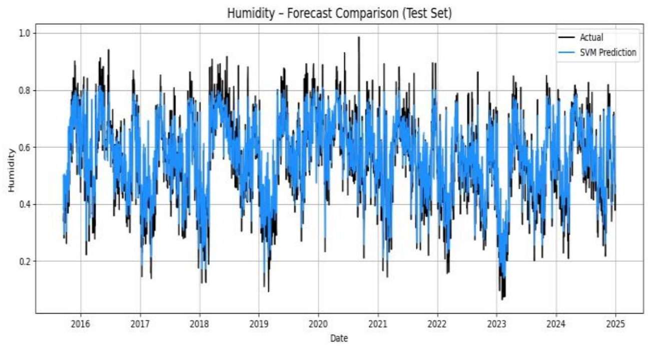

Figure 8: SVM humidity forecast trend using testing set

For temperature, ML models outperformed ARIMA in terms of MAE, MAPE, RMSE, and R-squared. The best performing model was GB with a lower MAE of .05, MAPE of 10.45%, RMSE of .07 and percentage of the model variation explained of 74.28%. This was closely followed by SVM with a MAE of .06, MAPE of 10.75%, RMSE of .07 and the percentage variance of the model explained of 72.72%. Closely following model was RF with a MAE of .06, MAPE of 11.27%, RMSE of .07, and the percentage variance explained of 69.43%. ARIMA on the other hand, recorded a higher MAE of .12, MAPE of 21.76% which is above 20%, RMSE of .15 and a negative coefficient of determination, suggesting it is not appropriate forecasting temperature. Therefore, while seeking to establish accurate humidity forecasts, it is recommended to apply GB having outstood in the four metrics. The trained model was applied in predicting the testing data and this is visualized in figure 9 below.

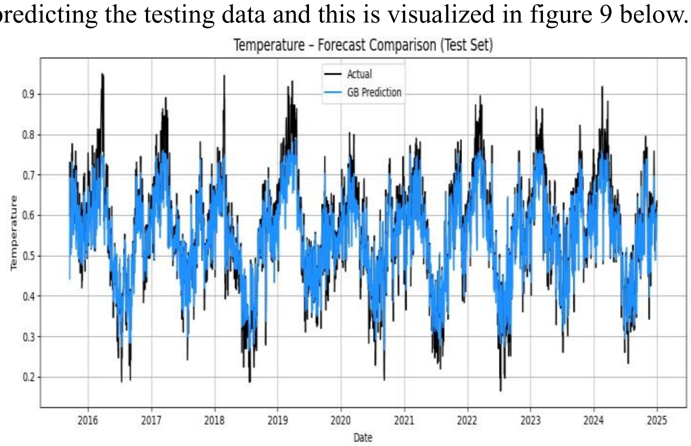

Figure 9: GB temperature forecast trend using testing set

While all the ML models were high achievers here, GB is recommended for predicting daily short- and medium-term temperature conditions in Nairobi County.

Whereas the first two weather conditions realized better forecasting accuracy, no model was effective in predicting windspeed, a factor attributable to the high disparity (spikes) in the recorded daily conditions even after data standardization. The four models performed better in terms of MAE. Nonetheless, SVM was the worst performer in terms of MAPE (46.32%) same as RF (23.05%), GB (21.07%) and ARIMA (24.03%). This was the same case with coefficient of determination where the percentage variance of the model explained was highest for GB (.31%) and negative for SVM (-53.47%), RF (- 83%), and ARIMA (-.79%), suggesting the need for other ML models that will accurately capture the data patterns of wind data.

5. DISCUSSION OF RESULTS

The study sought to explore the predictive power of machine learning models (RF, GB, and SVM) in short- and medium-term weather forecasting in comparison to a classical time series approach (ARIMA). This was done with specific focus Nairobi County's daily temperature, humidity, and windspeed data, covering a period of 30 years (1/1/1994 – 31/12/2024). The analysis was guided by a rising body of research that acknowledges the capacity of machine learning in handling nonlinear, complex data with wider flexibility relative to traditional statistical techniques. The study results displayed that, generally, the three ML models outperformed ARIMA modelling in predicting short- and medium-term humidity and temperature within this period. For the two weather variables, the ML models recorded Mean Absolute percentage errors of less than the recommended 20% maximum while ARIMA recorded values above 20%. Besides, The ML models explained stronger coefficient of determination than ARIMA. Specifically, SVM outstood in predicting humidity followed closely by GB, suggesting their capacity to handle complex and irregular short- and medium-term weather data of Nairobi County. For temperature, GB stood higher in performance followed closely by SVM and RF, further stressing on the strength to handle complex data. While the ML models outstand, it is worth highlighting that they performed differently for dissimilar conditions. Based on this, SVM and GB models were respectively selected to be the best predictors of humidity and temperature. The four models, nonetheless, struggled in predicting short- and medium-term wind speed patterns with all recording MAPE values above 20% hence not accurate. This was further confirmed by the lower value of R-squared where SVM, RF, and ARIMA recorded negative values and GB a variation close to 0.00%. The struggling of models predicting windspeed can be attributed to its low autocorrelation, high variability, and sensitivity to localized, transient environmental factors that are hard to model. For instance, while most dates recorded wind speed of between 5 – 20 m/s, some recorded higher above 100 resulting in a wide disparity even after normalization. Also insighted is that of the three weather conditions selected, temperature was the most accurately predicted with lowest Mean Absolute Percentage Error and strongest percentage weather variation explained by all the four models, a factor attributable to the strongest autocorrelation and low data variability recorded.

Discussing the study findings to the existing works of literature, then the results largely support the effectiveness of supervised machine learning models, specifically SVM, RF, and GB in predicting short- and medium-term weather conditions such as temperature and humidity. The results align with assertions in the existing literature that supervised learning models exhibit substantially better performance in complex forecasting tasks when there is adequate historical labeled data [11;19]. Referring to temperature, GB model performed the best, recording the lowest RMSE and highest R^2, a finding that is consistent with that of Anwar et al. (2021) who established high accuracy of GB in predicting rainfall over a 7-year period, alongside the finding by [11] who reported significant enhancements in wind speed prediction using GB over WRFbased models. Nonetheless, as reported by [18], GB only works well in situations where data is not highly seasonal or noisy, a factor that cites model instability and performance degradation under such conditions as observed in wind speed prediction performance. For SVM, strong performance was demonstrated in forecasting humidity in the study, strongly aligning with [20] and Hayaty et al. (2023) who established that SVMs are effective and efficient in detecting subtle patterns in non-linear and complex datasets for short- and medium-term forecasts. Nonetheless, the model recorded a higher MAPE for humidity relative to temperature prediction, suggesting the existence of nonlinear dynamics or feature limitations, consistent with the findings by Yin et al. (2022), who highlighted that SVM struggle in handling nonlinearity and extreme events in precipitation forecasts. The third ML model, RF, while exhibiting robustness in predicting temperature, underperformed relative to GB and SVM across all variables. This result confirms observations that emphasized the strength of RF in modeling nonlinearity and handling overfitting but also highlighted its computational demands [44]. Specifically, this study recorded slightly lower values of R^2 for RF relative to those found by Hill et al. (2020), who attained an R^2 of 0.97 in severe weather forecasting, a factor attributable to data or regional differences in weather inconsistency between Nairobi and the United States. While GB and SVM outperformed RF in making humidity and temperature forecasts, some literature including [46] and [47], established that RF models perform more reliably than boosting methods in small to moderate datasets, a factor attributed to reduced variance and simpler interpretability, categorically in resource-limited environments. In contrast, ARIMA models underachieved across temperature, humidity, and wind speed, a finding consistent with literature, which suggests that statistical models struggle to capture complex and nonlinear patterns, and interactions present in meteorological systems [17]. The finding is further justified by the fact that that ARIMA modelling, being univariate, does not account for interdependence among numerous variables, which is key in accurate prediction thus the low performance in forecasting non-linear wind speed, temperature, and humidity [28]. Nevertheless, certain studies express that ARIMA models, when integrated with exogenous regressors (ARIMAX) or within hybrid frameworks, can outperform or match ML models in stable climates [12], something that is not the case with Nairobi County temperature, wind speed and humidity patterns over time.

From the comparison, it is clear that whereas machine learning models give clear advantages in short- and medium-term weather prediction, selectin of models should be specific to variable of interest. Besides, model enhancements such as hybridization and feature tuning may further enhance predictive performance in future research. In general, this study results authenticate the rising acknowledgement that machine learning techniques provide superior precision and adaptability in short- and medium-term weather predictions in urban settings, specifically in tropical cities like Nairobi. The research insights, other than reinforcing theoretical basis for adopting ML in meteorological prediction, provide practical justification for its usage in sub-Saharan African settings with limited infrastructure for complex physical modeling.

6. CONCLUSION AND FUTURE WORK

The research sought to examine the applicability and performance of three machine learning (ML) models, SVM, RF, and GB in predicting short- and medium-term daily weather conditions in Nairobi County compared with classical ARIMA. Here, daily recordings for humidity, temperature, and windspeed as the target variables. The results displayed that ML models outperformed ARIMA in most prediction activities, specifically in forecasting humidity where SVM emerged the best followed by GB and temperature, where GB achieved the highest accuracy followed by SVM; these were evaluated across multiple metrics (RMSE, MAE, MAPE, and R^2). The study, however, showed the selected ML models as poor in predicting windspeed. The low performance of ARIMA across the three weather patterns can be attributed to its inability to capture non-linear complex data unlike ML models.

The findings stress the ML models' flexibility in learning complex temporal patterns and interactions between variables, which is a clear advantage over ARIMA's linear and univariate setting. The study also highlighted the value of incorporating lagged variables to capture temporal dependencies in the data, enabling the models to learn from past behavior. Besides, RF performed lower than SVM and GB in making predictions. By attaining superior prediction performance using ML methodologies, the study successfully addressed its main objectives. It displayed that machine learning models can be practicable and possibly superior substitutes to classical time series techniques in the setting of localized weather forecasting. The study findings, other than contributing to the increasing body of literature on ML-based prediction, also provide real-world insights for disaster preparedness, urban planning, and climate informed decision-making in developing nations like Kenya. Future studies should build on these findings by exploring multi-step forecasting strategies, additional predictors, and advanced deep learning or ensemble architecture.

From the study results, the following recommendations are made regarding the accurate forecasting of short- and medium-term weather conditions in Nairobi County, particularly temperature, windspeed, and humidity: Incorporation of Machine Learning into National Weather Systems , Application of Hybrid and Ensemble Machine Learning Models for enhanced Wind Speed forecasting, Institutionalize Comparative Evaluation of Forecasting Models, Replicate and Scale ML-Based Forecasting Systems Across Counties

REFERENCES

1. Jiménez, J. (2024, April 24). Flooding Inundates Kenya, Killing at Least 32 and Displacing Thousands. The New York Times. https://www.nytimes.com/2024/04/24/world/africa/kenya-flooding-nairobiphotos.html
2. Pagano, T. C., Casati, B., Landman, S., Loveday, N., Taggart, R., Ebert, E. E., Khanarmuei, M., Jensen, T. L., Mittermaier, M., Roberts, H., Willington, S., Roberts, N., Sowko, M., Strassberg, G., Kluepfel, C., Bullock, T. A., Turner, D. D., Pappenberger, F., Osborne, N., & Noble, C. (2024). Challenges of operational weather forecast verification and evaluation. *Bulletin of the American Meteorological Society*, *105*(4), E789–E802.
3. Yano, J., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., & Wyszogrodzki, A. A. (2018). Scientific Challenges of Convective-Scale Numerical Weather Prediction. *Bulletin of the American Meteorological Society*, *99*(4), 699–710.
4. Zhang, J., Qiu, X., Li, X., Huang, Z., Wu, M., & Dong, Y. (2021). Support vector machine weather prediction technology based on the improved quantum optimization algorithm. *Computational Intelligence and Neuroscience*, *2021*, 1– 13.
5. Lamptey, B., Abed, S. S., Gudoshava, M., Mutemi, J., Bopape, M., Adefisan, E. A., Igri, Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., Domeisen, D. I. V., Feser, F., Koszalka, I., Kreibich, H., Pantillon, F., Parolai, S., Pinto, J. G., Punge, H. J., Rivalta, E., Schröter, K., Strehlow, K., Weisse, R., & Wurpts, A. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, *58*(4).

1. 6.Men, H., Fu, S., Yang, J., Cheng, M., Shi, Y., & Liu, J. (2018). Comparison of SVM, RF and ELM on an electronic nose for the intelligent evaluation of paraffin samples. *Sensors*, *18*(1), 285.
2. Meenal, R., Michael, P. A., Pamela, D., & Rajasekaran, E. (2021). Weather prediction using random forest machine learning model. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(2), 1208881).

6. Balehegn, M., Balehey, S., Fu, C., & Liang, W. (2019). Indigenous weather and climate forecasting knowledge among Afar pastoralists of northeastern Ethiopia: Role in adaptation to weather and climate variability. *Pastoralism*, *9*(1).
7. Kniffka, A., Knippertz, P., Fink, A. H., Benedetti, A., Brooks, M. E., Hill, P. G., Maranan, M., Pante, G., & Vogel, B. (2020). An evaluation of operational and research weather forecasts for southern West Africa using observations from the DACCIWA field campaign in June–July 2016. *Quarterly Journal of the Royal Meteorological Society*, *146*(728), 1121–1148.
8. Anwar, M. T., Winarno, E., Hadikurniawati, W., & Novita, M. (2021). Rainfall prediction using Extreme Gradient Boosting. *Journal of Physics. Conference Series*, *1869*(1), 012078.
9. Walsh, J. E., Brettschneider, B., Kettle, N. P., & Thoman, R. L. (2021). An analog method for seasonal forecasting in northern high latitudes. *Atmospheric and Climate Science*, *11*(03), 469–485.
10. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538.
11. Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog Forecasting of Extreme-Causing Weather Patterns using Deep Learning. *Journal of Advances in Modeling Earth Systems*, *12*(2).
12. Salman, A. G., & Kanigoro, B. (2021). Visibility forecasting using Autoregressive Integrated Moving Average (ARIMA) models. *Procedia Computer Science*, *179*, 252–259.
13. Tektaş, M. (2010). Weather forecasting using ANFIS and ARIMA MODELS. *Environmental Research Engineering and Management*, *51*(1), 5–10.
14. Sarker, I. H. (2021). Machine learning: algorithms, Real-World applications and research directions. *SN Computer Science*, *2*(3).
15. Xu, W., Ning, L., & Luo, Y. (2020). Wind speed forecast based on post-processing of numerical weather predictions using a gradient boosting decision tree algorithm. *Atmosphere*, *11*(7), 738.
16. Iong, D., Chen, Y., Toth, G., Zou, S., Pulkkinen, T., Ren, J., Camporeale, E., & Gombosi, T. (2022). New findings from explainable SYM-H forecasting using gradient boosting machines. *Space Weather*, *20*(8).
17. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal Promoting Communications on Statistics and Stata*, *20*(1), 3–29.
18. Duffy, J., & Puzzello, D. (2021). The Friedman rule: experimental evidence. *International Economic Review*, *63*(2), 671–698
19. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215.
20. Mariotti, A., Bader, D. C., Bauer, S. E., Danabasoglu, G., Dunne, J., Gross, B., Leung, L. R., Pawson, S., Putman, W. R., Ramaswamy, V., Schmidt, G. A., & Tallapragada, V. (2024). Envisioning U.S. climate predictions and projections to meet new challenges. *Earth S Future*, *12*(6).
21. Yin, G., Yoshikane, T., Yamamoto, K., Kubota, T., & Yoshimura, K. (2022). A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. *Journal of Hydrology*, *612*, 128125.
22. Hayaty, N., Kurniawan, H., Rathomi, M. R., Chahyadi, F., & Bettiza, M. (2023). Rainfall Prediction with Support Vector Machines: A Case Study in Tanjungpinang City, Indonesia. *Bio Web of Conferences/BIO Web of Conferences*, *70*, 01003.
23. Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, *148*, 461–468.
24. Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine Learning Methods in Weather and Climate Applications: a survey. *Applied Sciences*, *13*(21), 12019.
25. Roberts, A. J., Fletcher, J. K., Groves, J., Marsham, J. H., Parker, D. J., Blyth, A. M., Adefisan, E. A., Ajayi, V. O., Barrette, R., De Coning, E., Dione, C., Diop, A., Foamouhoue, A. K., Gijben, M., Hill, P. G., Lawal, K. A., Mutemi, J., Padi, M., Popoola, T. I., . . . Woodhams, B. J. (2021). Nowcasting for Africa: advances, potential and value. *Weather*, *77*(7), 250–256.

26. Efron, B., & Tibshirani, R. (1994). An introduction to Bootstrap. In *Chapman and Hall/CRC eBooks*.
27. Safia, M., Abbas, R., & Aslani, M. (2023). Classification of weather conditions based on supervised learning for Swedish cities. *Atmosphere*, *14*(7), 1174.
28. Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, *12*(1).
29. Kumar, A. (2024, August 28). *Accuracy, Precision, Recall & F1-Score - Python Examples*. Analytics Yogi. https://vitalflux.com/accuracy-precision-recall-f1-scorepython-example/
30. Malathi, T., & Manimekalai, M. (2020). Feature selection techniques for weather forecasting models using machine learning techniques. *International journal of electrical engineering & technology*, *11*(4).
31. Sankar, S. R., & Panchapakesan, M. (2024). Hybrid Feature Selection Model for Accurate Wind Speed Forecasting from Numerical Weather Prediction Dataset. *Expert Systems with Applications*, *248*, 123054.
32. Nnamoko, N. A., Arshad, F., England, D., & Norman, J. (2014). Evaluation of filter and wrapper methods for feature selection in supervised machine learning. *ResearchGate*.
33. Effrosynidis, D., & Arampatzis, A. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics*, *61*, 101224.
34. Shaiba, H., Marzouk, R., Nour, M. K., Negm, N., Hilal, A. M., Mohamed, A., Motwakel, A., Yaseen, I., Zamani, A. S., & Rizwanullah, M. (2022). Weather forecasting prediction using ensemble machine learning for big data applications. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, *73*(2), 3367–3382.
35. Dong, J., Zeng, W., Wu, L., Huang, J., Gaiser, T., & Srivastava, A. K. (2023). Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. *Engineering Applications of Artificial Intelligence*, *117*, 105579.
36. Shawe-Taylor, J., & Sun, S. (2014). Kernel methods and support vector machines. In *Academic press library in signal processing* (pp. 857–
37. Baik, J., & Paek, J. (2000). A neural network model for predicting typhoon intensity. *Journal of the Meteorological Society of Japan Ser II*, *78*(6), 857–869.
38. Wu, H., & Levinson, D. (2021). The ensemble approach to forecasting: A review and synthesis. *Transportation Research Part C Emerging Technologies*, *132*, 103357.
39. Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., & Brinne, B. (2022). Data management for production quality deep learning models: Challenges and solutions. *Journal of Systems and Software/the Journal of Systems and Software*, *191*, 111359.
40. Park, Y. S., Konge, L., & Artino, A. R. (2019). The positivism paradigm of research. *Academic Medicine*, *95*(5), 690–694.
41. Lim, W. M. (2023). Philosophy of science and research paradigm for business research in the transformative age of automation, digitalization, hyperconnectivity, obligations, globalization and sustainability. *Journal of Trade Science*, *11*(2/3), 3–30.
42. Turner, D. P. (2020). Experimental study designs. *Headache the Journal of Head and Face Pain*, *60*(8), 1501–1507.
43. Drolet, M., Rose-Derouin, E., Leblanc, J., Ruest, M., & Williams-Jones, B. (2022). Ethical issues in research: perceptions of researchers, research ethics board members and research ethics experts. *Journal of Academic Ethics*, *21*(2), 269–292.
44. Hill, A. J., Herman, G. R., & Schumacher, R. S. (2020). Forecasting Severe Weather with Random Forests. *Monthly Weather Review*, *148*(5), 2135–2161.
45. Lai, Y., & Dzombak, D. A. (2020). Use of the Autoregressive Integrated Moving Average (ARIMA) model to forecast Near-Term regional temperature and precipitation. *Weather and Forecasting*, *35*(3), 959–976.
46. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*
47. Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging Clinical and Experimental Research*, *35*(11), 2363–2397.