An Intelligent Ensemble Learning Framework for Visibility and CAVOK Prediction to Support Logistics Operations

Omar Alghushairy[1], Raed Alsini[2], Ahmed Alamri[3], Saud Yonbawi[4], Ayman Yafoz[5], and Xiaogang Ma[6]
Department of Information Systems and Technology-College of Computer Science and Engineering,
University of Jeddah, Jeddah 23890, Saudi Arabia[1, 3]
Department of Information Systems-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah 21589, Saudi Arabia[2, 5]
Department of Software Engineering-College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia[4]
Department of Computer Science, University of Idaho, Moscow, ID 83844, USA[6]

***Abstract*—The supply chain operation is fully dependent on weather conditions due to the damage that produce may face during bad weather or accidents that may occur because of cloud cover, especially in Saudi Arabia, which experiences frequent cloud cover throughout the year. In this study, a new weather prediction model based on machine learning and optimization algorithms is proposed to forecast visibility distance and Cloud and Visibility OK (CAVOK), a term indicating good weather, while also estimating cloud cover, which affects flight operations. The proposed system follows two main steps: feature selection and prediction, to predict CAVOK and visibility distance. In the feature selection phase, Fick's Law Algorithm (FLA) is used to choose the optimal parameters that can serve as indicators in the prediction phase, enhancing the proposed system's performance. In the prediction phase, a new two-stage ensemble algorithm is used to predict sky CAVOK and visibility distance. In the first stage, a new voting ensemble algorithm consisting of Bagging Regressor and Extra Trees Regressor algorithms is proposed to predict the visibility distance. The predicted visibility distance value is then used alone to predict sky CAVOK using the CatBoost Classifier in the second stage. The experimental results show that the proposed system demonstrates reasonable predictive performance, obtaining 529.7506 m, 1174.6118, and 0.3233 for Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2, respectively, in predicting visibility distance. Additionally, the system achieved competitive results while maintaining robust predictive capability in predicting sky CAVOK.**

***Keywords*—Weather forecasting; sky CAVOK; extra tree; supply chain; visibility distance; Saudi Arabia**

I. Introduction

Weather forecasting is a technology that aims to predict the state of the atmosphere at any given time. Air temperature, wind direction, wind strength, cloud cover, and Visibility OK (OK/CAVOK) and visibility distance are examples of atmospheric conditions that weather forecasting predicts [1][2]. The supply chain and logistics are affected by daily weather changes, making weather prediction a crucial factor in supply chain operations [3]. Extreme weather conditions can disrupt supply chains because of flight delays due to cloud cover and lead to financial losses and operational difficulties across various industries [4] [5]. For example, weather changes in 2024 caused widespread supply chain disruptions worldwide, resulting in companies losing over 100 billion dollars [6]. Additionally, weather-related supply chain damage increased by 30% in the first half of 2024 compared to 2023 [7]. It's predicted that the loss in the supply chain may reach 26 trillion dollars due to changes in the weather by 2050 [8]. Although 26.6% of organizations have their supply chains affected by bad weather in 2024, 36.7% of supply chain organizations do not analyze or predict weather conditions to protect themselves [9].

Machine learning and optimization algorithms are used to build a reliable system capable of predicting the weather and identifying the days when supply chain operations can be carried out [10]. There are many recent works focused on developing and designing weather prediction systems using machine and deep learning algorithms [11][12]. Many of these works aim to predict wind speed, temperature, and the probability of rain [13]. However, only a few recent studies concentrate on developing weather prediction systems that forecast visibility distance and CAVOK metrics, which help supply chain organizations determine the amount of cloud cover and assess whether they can perform their supply chain operations on a given day [14]. Unfortunately, these systems face many challenges [15]. Weather prediction systems that rely on traditional machine learning algorithms often produce low-accuracy results, while systems that use deep learning techniques achieve high accuracy but require a long time to generate predictions [15]. To address these challenges, this study proposes a new machine learning-based weather prediction system that combines traditional machine learning algorithms, which do not consume excessive time, with an optimization algorithm that selects the optimal features to improve accuracy. The system predicts visibility distance and CAVOK conditions to enhance forecasting reliability. The main objectives of this study are:

- Propose a new real-time weather prediction system that forecasts visibility distance and CAVOK conditions, enabling the identification of suitable days for supply chain operations.

- Propose a new two-stage ensemble algorithm that utilizes Bagging Regressor and Extra Tree Regressor to predict visibility distance in the first stage and then uses the predicted visibility distance value to determine CAVOK conditions using the Extra Tree Classifier.
- Apply Fick's Law Algorithm (FLA) to identify optimal features for predicting visibility distance and CAVOK, enhancing the accuracy and efficiency of the proposed system.

The rest of the study is organized as follows: Section II presents related works. Section III introduces the proposed system. Section IV discusses the experimental results of the proposed system. Finally, the study is concluded in Section V.

II. Related Work

There are many related works that have used machine learning and deep learning to predict the weather. However, only a few recent studies focus on forecasting weather by predicting visibility distance and CAVOK conditions. For example, Kim et al. [16] developed a weather prediction system using the Extreme Gradient Boosting (XGB) algorithm to predict visibility distance in Seoul, South Korea. To evaluate the performance of their system, they collected a dataset consisting of 152,488 cases from the Korea Meteorological Administration. The experimental results show that their system achieved acceptable performance, with a Mean Absolute Error (MAE) of 2.04 km and a Root Mean Square Error (RMSE) of 2.94.

Zhang et al. [17] proposed a weather prediction system that predicts visibility distance using Principal Component Analysis (PCA) and machine learning algorithms. The PCA algorithm is used to select optimal features that can serve as indicators for identifying visibility distance. Six machine learning algorithms—Decision Tree, Naïve Bayes, K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Linear SVM, and Neural Networks—were used to predict visibility distance. To evaluate the performance of their system, they used a dataset collected from ground observations and radiosonde data in Chengdu, Sichuan, containing hourly weather data recorded from 2016 to 2020. The experimental results show that the neural network outperformed the other machine learning algorithms, achieving 88.5% accuracy in predicting visibility distance for the next two hours and 79.5% accuracy for the next four hours.

Ortega et al. [18] proposed a weather prediction system capable of classifying visibility distance into one of three categories: Low, Moderate, or Good, in Florida, USA. To achieve this, the authors used five machine learning algorithms: SVM, KNN, Naïve Bayes, Classification and Regression Tree (CART), and Artificial Neural Networks (ANN). The authors evaluated their system using a dataset collected from weather stations in Florida. The experimental results show that their system achieved high accuracy, with the ANN performing the best, reaching an accuracy of 89.71%.

Zhen et al. [19] proposed a weather prediction system based on traditional machine learning to predict visibility distance during flight and supply chain operations. They also introduced a new stacking fusion model that combines eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) to enhance system performance. The experimental results show that the proposed fusion model outperformed traditional machine learning algorithms, including XGBoost, LightGBM, Random Forest, and SVM, achieving an RMSE of 3.82.

Zhang et al. [20] proposed a new meteorological visibility prediction system capable of predicting visibility distance using machine learning algorithms. They introduced a new multimodal fusion model that combines XGBoost and LightGBM to enhance prediction accuracy. To evaluate the performance of the proposed system, they collected a dataset containing atmospheric data from meteorological observation stations in Beijing, Tianjin, and the Hebei region, covering the period from 2002 to 2018. The experimental results show that the new multimodal fusion model outperformed single machine learning algorithms such as XGBoost, LightGBM, SVM, and ANN, achieving an RMSE of 6.71.

As seen above, traditional machine learning algorithms do not achieve very high accuracy when predicting visibility distance. Therefore, recent works have shifted towards using deep learning algorithms to improve prediction performance. For example, Zang et al. [21] proposed a new deep learning-based weather prediction system that predicts visibility distance using a modified version of the Recurrent Neural Network (RNN) called SwiftRNN. This modified version incorporates a frame-hopping transmission gate, reverse scheduled sampling, and a feature fusion module. The purpose of the frame-hopping transmission gate is to reduce the training time required by the RNN algorithm, while reverse scheduled sampling is designed to improve the accuracy of traditional RNN models. Additionally, the feature fusion module is used to detect both local and global optimal features. To evaluate the SwiftRNN model, the authors applied it to a dataset containing atmospheric visibility conditions in China, collected from the Meteorological Information Center from the beginning of 2018 until the end of 2020. The experimental results show that SwiftRNN achieved an average Structural Similarity Index (SSIM) of 0.444 and an average Learned Perceptual Image Patch Similarity (LPIPS) of 0.289.

Ortega et al. [22] proposed a deep learning-based weather prediction system capable of forecasting visibility distance. To predict the visibility distance for each hour, they collected a dataset containing climatological data from two weather stations in Florida, USA. The authors used five deep learning algorithms: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Fully Convolutional Neural Network (FCNN), Multi-Input Convolutional Neural Network (MICNN), and Long Short-Term Memory (LSTM) network to predict visibility distance. The experimental results show that MLP, CNN, FCNN, MICNN, and LSTM achieved RMSE values of 1.170, 1.136, 1.175, 1.201, and 1.247, respectively, in predicting visibility distance for the next nine hours.

Chen et al. [23] proposed a new weather prediction system that uses deep learning algorithms to predict visibility distance from both spatial and temporal perspectives. Their system follows two main steps: feature selection and forecasting. In the feature selection step, the authors used the Maximum Relevance Minimum Redundancy (mRMR) algorithm to select the optimal features that serve as indicators in the forecasting step. In the forecasting step, they proposed a new deep learning model called GCN-GRU (Graph Convolutional Network and Gated Recurrent Unit), which is composed of two main algorithms: GCN and GRU, to forecast visibility distance. To evaluate the system, the authors used a dataset containing weather data collected from Jiangsu Province between 2017 and 2018. The experimental results show that their system outperformed LightGBM, Random Forest, and XGBoost, achieving RMSE and MAE values of 17.52 and 16.53, respectively.

Qasem [24] presents a deep learning-based framework for forecasting long-term temperature trends in Saudi Arabia to support climate change analysis. The study compares three recurrent neural network models—LSTM, GRU, and BiLSTM—using historical temperature records from 1861 to 2013. After preprocessing the data through interpolation, normalization, and time-series splitting, the models are evaluated using MAE, MRE, RMSE, and R^2 metrics. Experimental results demonstrate that the GRU model with 50 hidden units achieves the highest forecasting accuracy, outperforming LSTM, BiLSTM, and traditional methods such as ARIMA and Support Vector Regression. The proposed approach provides an effective tool for analyzing long-term temperature trends and can support climate adaptation and environmental planning in Saudi Arabia.

Yuying et al. [25] proposed a deep learning-based weather prediction system capable of predicting visibility distance. To achieve this, they used PhyDNet-ATT (Physical Dynamics Network with Attention). To evaluate their system, they collected a dataset containing weather data from Jiangsu, Anhui, Henan, and Shandong Provinces, sourced from the Precision Weather Analysis and Forecast System. They compared their system with the European Centre for Medium-Range Weather Forecasts, and their system outperformed it by reducing RMSE and MAE by 201% and 310%, respectively.

Suleiman. [26] proposes a hybrid intelligent framework that integrates the Adaptive Neuro-Fuzzy Inference System (ANFIS) with Geographic Information Systems (GIS) to forecast long-term weather elements in Jordan. Using over 30 years of historical meteorological data (1985–2015) collected from 26 weather stations, the study predicts annual rainfall as well as minimum and maximum temperatures for the period 2016–2025. The proposed approach combines fuzzy logic and neural networks to model the nonlinear behavior of weather data, while GIS is used to visualize and analyze the spatial distribution of the predictions. Experimental results indicate that ANFIS provides reliable forecasting performance, predicting a decline in annual rainfall and noticeable changes in temperature patterns across Jordan. The study demonstrates that integrating ANFIS with GIS is an effective tool for long-term climate forecasting and supports environmental planning and climate change assessment in semi-arid regions.

Table I summarizes recent works, highlighting the machine learning algorithms used and the results achieved.

Table I presents all recent works focused on developing weather prediction systems that predict visibility distance. However, no related work has been found that specifically predicts CAVOK metrics. Additionally, no recent studies have focused on predicting either visibility distance or CAVOK metrics in Saudi Arabia.

TABLE I. RELATED WORK COMPARISON

Related work	Year	Machine learning algorithms	Forecast Variables	Results
Kim et al. [16]	2022	Extreme Gradient Boosting (XGB)	Visibility distance	Their system achieves 2.04 km and 2.94 as MAE and RMSE, respectively.
Zhang et al. [17]	2022	Decision Tree, Naïve Bayes, K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Linear SVM, and Neural Networks	Visibility distance	Their system achieves 88.5% accuracy in predicting visibility distance for the next two hours and 79.5% accuracy for the next four hours when the neural network algorithm is used.
Ortega et al. [18]	2019	SVM, KNN, Naïve Bayes, CART, and ANN.	Visibility distance	ANN outperformed other machine learning algorithms, achieving an accuracy of 89.71%.
Zhen et al. [19]	2023	a new stacking fusion model that combines XGBoost and LightGBM	visibility distance	The experimental results show that the proposed fusion model outperformed traditional machine learning algorithms, including XGBoost, LightGBM, Random Forest, and SVM, achieving an RMSE of 3.82.
Zhang et al. [20]	2019	new multimodal fusion model that combines XGBoost and LightGBM	visibility distance	The experimental results show that the new multimodal fusion model achieves 6.71 RMSE.
Zang et al. [21]	2023	RNN	visibility distance	Their system outperformed traditional deep learning models, achieving an average SSIM of 0.444 and an average LPIPS of 0.289.
Ortega et al. [22]	2023	MLP, CNN, FCNN, MICNN, and LSTM	visibility distance	The experimental results show that MLP, CNN, FCNN, MICNN, and LSTM achieved RMSE values of 1.170, 1.136, 1.175, 1.201, and 1.247, respectively, in predicting visibility distance for the next nine hours.
Chen et al. [23]	2024	GCN-GRU, LightGBM, Random Forest, and XGBoost	visibility distance	Their system achieves RMSE and MAE values of 17.52 and 16.53, respectively.

III. Methodology

In this study, a new weather prediction system is proposed to forecast the numerical value of the visibility distance that the supply chain can observe. The predicted visibility distance is then used to classify sky CAVOK conditions as either cloudy or not cloudy. The proposed system consists of three main steps: preprocessing, feature selection, and a two-stage prediction model to predict both visibility distance and sky CAVOK. Fig. 1 illustrates the architecture of the proposed system.

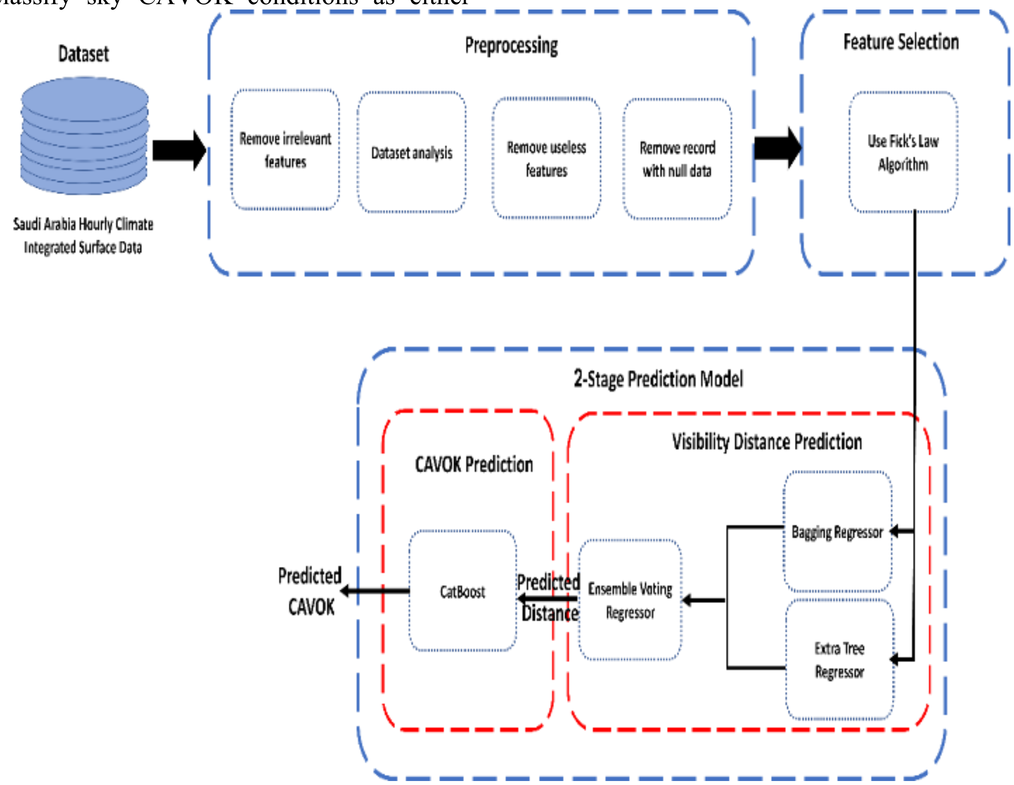

Fig. 1. Proposed system architecture.

As shown in Fig. 1, a Saudi Arabian dataset called Saudi Arabia Hourly Climate Integrated Surface Data is used, which contains historical weather data collected from 2019 to the present. This dataset records hourly weather conditions in Saudi Arabia. Fig. 1 also illustrates that the dataset undergoes analysis to examine the distribution of each feature, identifying irrelevant features for visibility distance and CAVOK prediction while removing rows with null data during the preprocessing step. After preprocessing, Fick's Law algorithm is applied in the feature selection phase to identify the optimal features that serve as indicators for predicting visibility distance and CAVOK. Finally, a new two-stage weather prediction model is introduced for identifying visibility distance and CAVOK conditions. In the first stage, a new ensemble algorithm, which combines Bagging Regressor and Extra Trees Regressor, is developed to predict the visibility distance. In the second stage, the predicted visibility distance value is then fed into a CatBoost Classifier to classify the sky condition as either CAVOK (clear skies) or non-CAVOK (cloudy). Algorithm 1 outlines the steps taken to predict the visibility distance and CAVOK condition in the proposed system.

Algorithm 1: Visibility and CAVOK prediction system algorithm

Input: Saudi Arabia Hourly Climate Integrated Surface Data (Dataset)

Output: visibility distance and CAVOK

Remove irrelevant features for visibility distance and CAVOK
Remove rows with null data
For i=1 to number of features
    Analyze and compute the distribution of values for each fea
    Remove features with bad distribution
End
Final features=Apply Fick's Law algorithm to find the optimal feature for visibility distance.

Visibility Distance 1 = Apply Bagging Regressor to predict visib distance.

Visibility Distance 2 = Apply Extra Trees to predict visibility dis

Final Visibility Distance = $\frac{\text{Visibility Distance 1} + \text{Visibility Distance2}}{2}$

CAVOK = Apply CatBoost classifier to Final Visibility Distance to predict CAVOK

Return Final Visibility Distance and CAVOK

A. Dataset

As shown in Fig. 1, the Saudi Arabia Hourly Climate Integrated Surface Data dataset is used to evaluate the system's performance. This dataset contains 1,760,587 records of historical weather data collected from Saudi Arabia by the National Oceanic and Atmospheric Administration (NOAA) [27][28]. The dataset records 15 different features, describing the weather conditions in Saudi Arabia on an hourly basis. These features include latitude, longitude, station_id, station_name, wind_direction_angle, wind_type, wind_ speed_rate, observation_date, elevation, visibility_distance, air_temperature_dew_point, air_temperature, atmospheric_se_ level_pressure, sky_ceiling_height, and sky-CAVOK. Latitude and longitude represent the decimal values of the geographical coordinates of the station recording the weather data. Station_id and station_name correspond to the unique identifier and name of the weather station. Wind_type categorizes the wind condition into one of three values: calm, normal, or variable. Wind_direction_angle measures the wind's direction relative to true north, while wind_speed_rate records the speed of the wind. Observation_date logs the date and time when the weather data was recorded. Elevation represents the station's height above sea level, measured in meters. Visibility_distance indicates the distance that can be seen under the recorded weather conditions. Air_temperature and air_temperature_ dew_point measure the recorded air temperature and the temperature at which air becomes saturated, respectively. Atmospheric_sea_level_pressure denotes the pressure value at the time of recording. Sky_ceiling_height measures the height from the ground to the cloud base, in meters. Finally, sky-CAVOK indicates whether the weather is favorable or if cloud conditions could impact supply chain operations, with values of yes or no.

B. *Preprocessing*

In the preprocessing step, the features in the dataset are analyzed to determine the distribution of values for each feature. This process consists of four main steps: removing irrelevant features, dataset analysis, removing useless features, and eliminating records with null data. In the removing irrelevant features step, features that do not significantly impact the prediction of visibility distance or CAVOK are removed. These include station_name, station_id, and observation_date, as they mainly provide metadata about the recording station rather than contributing to weather prediction. After this step, the number of features is reduced to 12. In the dataset analysis step, numerical analysis is performed to examine the value distribution of each feature. This helps identify features that may not contribute effectively to predicting visibility distance due to poor data distribution. Seven statistical metrics are calculated: count, mean, 25% median, 50% median, 75% median, minimum, and maximum. The results of this statistical analysis are presented in Table II. Table II reveals that atmospheric_sea_level_pressure and sky_ceiling_height have the same values for the 25% median, 50% median, 75% median, and maximum. This indicates that these features contain numerous records with identical values, which could be default values for missing data. Consequently, these features are unlikely to be useful indicators for predicting visibility distance or CAVOK. The distribution of feature values is further visualized in Fig. 2, which confirms that sky_ceiling_height and atmospheric_sea_level_pressure are concentrated around the same value for most records. This supports the decision to remove these features in the removing useless features step. After eliminating these features, the number of features in the dataset is reduced to 10.

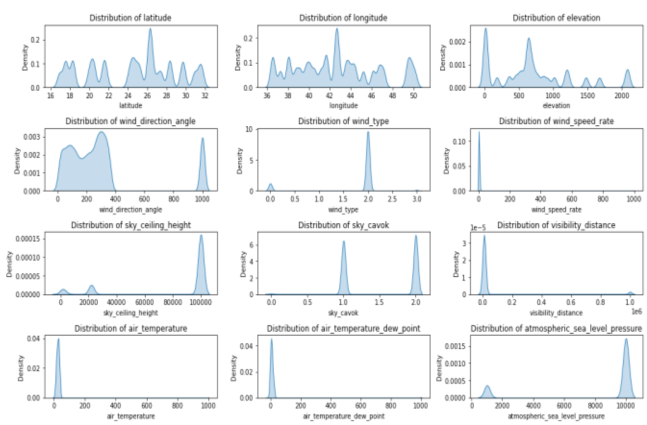

Fig. 2. Feature distributions before applying the preprocessing step.

TABLE II. NUMERICAL FEATURES ANALYSIS BEFORE APPLYING THE PREPROCESSING STEP

Attributes	Count	Mean	Std	Min	25%	50%	75%	Max
Latitude	1760587	24.48343	4.43772	16.901	20.504	25.285	27.438	31.693
Longitude	1760587	42.56574	3.86388	36.476	39.7051	42.621	45.2	50.152
Elevation	1760587	701.7991	578.068	1	179.22	628.49	923	2090.31
wind_direction_angle	1760587	289.0228	276.587	10	110	240	320	999
wind_type	1760587	1.795372	0.63406	0	2	2	2	3
wind_speed_rate	1760587	4.104066	22.0863	0	2.1	3.1	5.1	999.9
sky_ceiling_height	1760587	82487.49	34861.59	0	99999	99999	99999	99999
sky_cavok	1760587	1.515148	0.510321	0	1	2	2	2
visibility_distance	1760587	48884.27	193561.84	0	9900	9900	10000	999999
air_temperature	1760587	26.53717	24.843410	-6	19.1	26	32.6	999.9
air_temperature_dew_point	1760587	9.474121	48.888199	-28	1	6	12	999.9
atmospheric_sea_level_pressure	1760587	8447.8209	3397.8368	990.2	9999.9	9999.9	9999.9	9999.9

In the step of removing records with null data, records containing missing values are identified and removed from the dataset. The number of null records for each feature is counted before elimination. It is observed that wind type has 7 missing records, sky CAVOK has 9,381 missing records, wind direction angle has 204,426 missing records, wind speed rate has 855 missing records, air temperature has 999 missing records, air temperature dew point has 4,156 missing records, and visibility distance has 70,014 missing records. Since missing values can negatively impact the prediction model, all records containing null values are removed. After this step, the total number of records in the dataset is reduced to 1,481,842. After removing records with null data, the dataset is analyzed again to calculate seven statistical metrics: count, mean, 25% median, 50% median, 75% median, minimum, and maximum. This analysis helps to understand the distribution of feature values after eliminating useless features and records with missing data. Table III presents the calculated statistical metrics for each feature, while Fig. 3 illustrates the distribution of values for each feature after preprocessing. Table IV presents the number and names of features before and after preprocessing, along with the total number of records in the dataset before and after the preprocessing steps. This comparison highlights the impact of data cleaning and feature selection on refining the dataset for improved model performance (Algorithm 2).

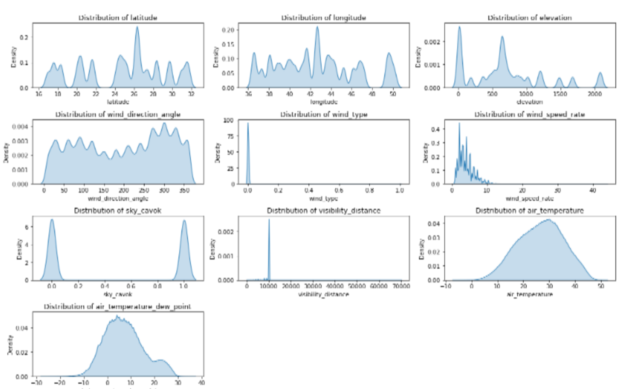

Fig. 3. Feature distributions after applying the preprocessing step.

Algorithm 2: Preprocessing algorithm

Input: Saudi Arabia Hourly Climate Integrated Surface Data (Dataset)

Output: Cleaned Dataset (C-Dataset)

```
For i=1: 15
    If feature(i) is an irrelevant feature
        Remove feature(i) from dataset
    End if
End for
Count=0
For i=1: num of features
    Min=feature[i][0]
    Max=feature[i][0]
    For j=1: num of records
        Count = Count +1
        Sum[i][j]=sum[i][j] + feature[i][j]
        If feature [i][j] > Max
            Max = feature[i][j]
        End if
        If feature [i][j] <Min
            Min = feature[i][j]
        End if
    End for
    Calculating means = Sum[i][j] / Count        (2)
    Calculate the 25% median
    Calculate the 50% median
    Calculate the 75% median
If means == 25% median == 50% median== 7% median== Max
    Remove feature(i) if it has bad distribution
End if
End for
For i=1: num of features
    For j=1: num of records
        If feature[i][j] == null
            Remove this record
        End if
    End for
End for
Return C-Dataset
```

TABLE III. NUMERICAL FEATURES ANALYSIS AFTER APPLYING THE PREPROCESSING STEP

Attributes	Count	Mean	Std	Min	25%	50%	75%	Max
Latitude	1481842	24.63173	4.39709	16.901	20.504	25.285	28.335	31.693
Longitude	1481842	42.54232	3.865	36.476	39.705	42.586	45.2	50.1520
Elevation	1481842	685.4775	560.794	6.09	179.22	628.49	922.32	2090.31
wind_direction_angle	1481842	195.9219	107.827	10	100	210	290	360
wind_type	1481842	0.004463	0.066659	0	0	0	0	1
wind_speed_rate	1481842	4.091146	2.222666	0.5	2.1	3.6	5.1	43.8
sky_cavok	1481842	0.495779	0.499982	0	0	0	1	1
visibility_distance	1481842	9485.737	1436.765	0	9900	9900	9999	70001
air_temperature	1481842	26.57518	8.830703	-6	20	27	33	51
air_temperature_dew_point	1481842	7.179006	8.34327	-27	1	6	12	36

C. Feature Selection

After the dataset is cleaned, the optimal features that can be used as indicators to predict visibility distance and CAVOK are selected in the feature selection step. In this step, the FLA algorithm is used to choose the most relevant features. FLA applies the same concept as Fick's Law, which is based on the diffusion mechanism that explains how materials move from regions of high concentration to low concentration. Therefore, FLA is used as a feature extraction algorithm for predicting visibility distance due to its ability to select relevant features that influence the movement of fog or particles, ultimately reducing visibility.

The algorithm starts by treating the problem space as a finite-dimensional space and searches for the optimal solution by applying the diffusion mechanism to all particles. The initial population is divided into two equal groups. Each particle updates its position based on the principle of diffusion, which progresses through three key steps: exploration, transition from exploration to exploitation, and exploitation. In the exploration phase, the algorithm searches across a wide area to avoid premature convergence to local solutions and aims to identify the most promising regions in the search space. In the transition phase, particles gradually shift from broad search to a more focused search within the best identified regions. In the exploitation phase, the algorithm optimizes solutions by refining the search within the region that is most likely to contain the optimal solution. The steps performed by the FLA algorithm are shown in Algorithm 3.

Algorithm 3: FLA feature extraction algorithm

```
Input: Population Size (N), Number of features (D), and Maximum
number of iterations (max)
  Output: features (F)
    Generate the initial population (I) with size N randomly
    Evaluate the fitness function for each feature using MAE that
    predicted by the model
    F=Identify the best global feature
    While (i<max)
        Divide the population into two equal groups
        For each practice in the first group
            Apply the diffusion mechanism of Fick's law to
            update feature selection
            Evaluate the fitness function for each feature
            F= add or remove the features
        End For
        Adjust diffusion rate dynamically to balance global and
        local search
        For each practice in the second group
        Perform local refinement
            Evaluate the fitness function for each feature
            F= add or remove the features
        End For
        F= add or remove the features
        i=i + 1
    End While
 Return F
```

After applying the FLA algorithm to the cleaned dataset, six features were identified as optimal indicators for predicting visibility distance: longitude, elevation, wind direction angle, wind speed rate, air temperature, and air temperature dew point. Additionally, FLA was applied after predicting visibility distance to identify the optimal feature for classifying sky CAVOK in weather data. The algorithm was used to select the most relevant feature from the ten existing features in the cleaned dataset, along with the predicted visibility distance. The FLA algorithm ultimately selected a single feature—the predicted visibility distance—as the key indicator for classifying sky CAVOK. Table IV presents the number and names of features before and after applying the FLA algorithm to identify the optimal features for predicting visibility distance and sky CAVOK.

TABLE IV. ATTRIBUTES USED IN EACH STAGE

Step	Number of rows	Number of features	Attributes
Dataset before preprocessing	1760587	15	station_name, station_id, observation_date, Latitude, longitude, Elevation, wind_direction_angle, wind_type, wind_speed_rate, sky_ceiling_height, sky_cavok, visibility_distance, air_temperature, air_temperature_dew_point, and atmospheric_sea_level_pressure
preprocessing	1481842	10	latitude, longitude, Elevation, wind_direction_angle, wind_type, wind_speed_rate sky_cavok, visibility_distance, air_temperature, and air_temperature_dew_point.
After applying the FLA algorithm	1481842	6	Longitude, elevation, wind_direction_angle, wind_speed_rate, air_temperature, air_temperature_dew_point
After applying the FLA algorithm	1481842	1	Predicted visibility distance

D. 2-Stage Prediction Model

As shown in Fig. 1, the prediction of visibility distance and sky CAVOK is organized in the newly proposed two-stage prediction model. In the first stage, a new ensemble algorithm, which is a combination of bagging and Extra Trees Regressor, is proposed to predict the visibility distance from weather data. The ensemble algorithm follows two steps to make the prediction. In the first step, bagging and extra tree regressor are applied separately to predict visibility distance. Then, in the second step, the mean value of the two predicted visibility distances (from bagging and extra tree regressor) is calculated to produce the final predicted visibility distance. In the second stage of the prediction model, the final predicted visibility distance is used to classify sky CAVOK into one of two values: "Yes" (indicating favorable weather conditions) or "No" (indicating cloud conditions). The CatBoost classifier is used for this classification. To determine the optimal hyperparameters for the bagging regressor, extra tree regressor, and CatBoost classifier, a random search strategy is used. This approach selects random points from the set of available values and evaluates the system by calculating MAE until the minimum MAE is achieved.

IV. Result and Discussion

Since predicting two different weather variables—visibility distance and sky CAVOK condition—requires different approaches, visibility distance depends on predicting a specific value, while sky CAVOK classifies each weather data point as either "Yes" (indicating favorable weather conditions) or "No" (indicating cloud conditions). Therefore, two different evaluation metrics are needed for each prediction. To evaluate the performance of the proposed system in visibility prediction, RMSE, R^2, and MAE are estimated using the following equations.

$$RMSE = \sqrt{\frac{\sum_{w=1}^{n}(DV_w - O_w)^2}{n}} \quad (3)$$

$$MAE = \frac{\sum_{w=1}^{Nn}|\, O_w - DV_w|}{n} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{w=1}^{n}\left(DV_w - O_j\right)^2}{\sum_{w=1}^{n}(DV_w - \overline{DV}_w)^2} \quad (5)$$

where, $DV_w, O_w, n, and\ \overline{DV}_w$ represent the actual visibility distance, observed visibility distance, number of weather data points predicted, and the average value of all actual visibility distances.

To evaluate the performance of the proposed system in identifying the sky CAVOK condition, precision, recall, F1-score, and accuracy. Precision (P_i), recall (R_i), F1-score ($F1_i$) are estimated for each category i as follows:

$$P_i = \frac{Tp_i}{Tp_i + Fp_i} \quad (6)$$

$$R_i = \frac{Tp_i}{Tp_i + FN_i} \quad (7)$$

$$F1_i = \frac{2\ x\ P_i\ x\ R_i}{P_i + R_i} \quad (8)$$

where, the Tp_i, $Fp_i, and\ FN_i$ represent the true positive, false positive, and false negative for each category i.

The overall accuracy (A) of the proposed system is estimated as follows:

$$A = \frac{Tp+TN}{Tp+TN+FP+FN} \quad (9)$$

where, $Tp, TN, FP, and\ FN$ represent the true positive, true negative, false positive, and false negative for the yes or no CAVOK condition.

The performance of the proposed system is evaluated by dividing the dataset into 80% for training the model and 20% for testing the model. Additionally, the system's performance is assessed both with and without using the FLA algorithm for feature extraction. Table V presents the results of the proposed system in the first stage of the prediction model, where visibility distance is predicted with and without the feature extraction step. Table V shows that various machine learning algorithms, including Gradient Boosting, AdaBoost, Bagging, Extra Trees, Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, KNN, Decision Tree, XGBoost, and LightGBM, are used to compare their performance against the proposed system. The results indicate that the proposed system outperforms individual machine learning algorithms, whether FLA is applied or not, achieving RMSE and MAE values of 1178.8224 and 530.6069 without FLA, and 1174.6118 and 529.7506 with FLA. This demonstrates that the proposed system performs better when the FLA algorithm is used to select the optimal features. Furthermore, the results show that most machine learning algorithms perform better when FLA is applied, except for the LightGBM algorithm, which performs better without FLA. The experimental findings also indicate that Bagging and Extra Trees outperform other machine learning algorithms, achieving RMSE values of 1218.5085 and 1192.1272 without FLA, and 1216.1397 and 1191.6472 with FLA. This justifies the selection of these two algorithms for developing the proposed ensemble model. Additionally, Table V highlights that KNN and Decision Tree achieve relatively good results, with RMSE values of 1257.6711 and 1508.1836 without FLA, and 1256.6323 and 1509.4436 with FLA. However, AdaBoost is identified as the worst-performing algorithm, achieving RMSE values of 1419.2078 without FLA and 1590.2311 with FLA. Moreover, the Decision Tree algorithm exhibits a negative R^2 value, indicating its poor capability in predicting visibility distance.

TABLE V. Proposed System Performance When Predicting Visibility Distance with and Without Using Feature Extraction Algorithm

Feature extraction step	Machine learning algorithm	MAE	RMSE	R^2
Without using FLA algorithm	Gradient Boosting	676.6967	1296.1086	0.1789
	AdaBoost	832.4155	1419.2078	0.0155
	Bagging	565.1363	1218.5085	0.2743
	Extra Trees	505.8370	1192.1272	0.3054
	Linear Regression	760.5695	1378.1543	0.0717
	Ridge Regression	760.5694	1378.1543	0.0717
	Lasso Regression	760.4274	1378.1627	0.0716
	ElasticNet	757.6100	1378.3685	0.0714
	KNN	575.2487	1257.6711	0.2269
	Decision Tree	558.2505	1508.1836	-0.1118
	XGBoost	617.6881	1228.1535	0.2627
	LightGBM	634.8607	1245.9460	0.2412
	Proposed system	**530.6069**	**1178.8224**	**0.3208**
Using FLA algorithm	Gradient Boosting	678.2282	1297.3529	0.1773
	AdaBoost	861.2343	1590.2311	-0.2361
	Bagging	563.8108	1216.1397	0.2771
	Extra Trees	505.4231	1191.6472	0.3059
	Linear Regression	760.1859	1378.4346	0.0713
	Ridge Regression	760.1859	1378.4346	0.0713
	Lasso Regression	760.0959	1378.4329	0.0713
	ElasticNet	757.3761	1378.6370	0.0710
	KNN	574.8747	1256.6323	0.2282
	Decision Tree	557.8073	1509.4436	-0.1137
	XGBoost	617.5476	1227.9163	0.2630
	LightGBM	636.0922	1247.5079	0.2393
	Proposed system	**529.7506**	**1174.6118**	**0.3233**

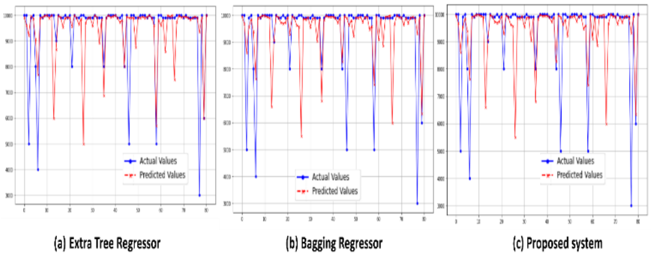

Fig. 4. Visibility distance prediction without FLA.

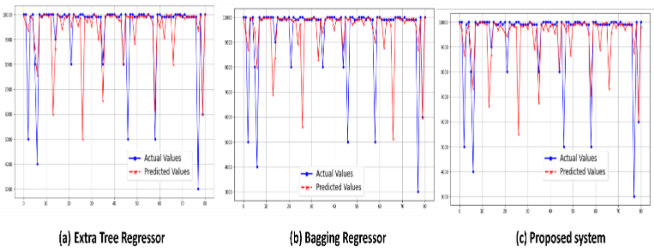

Fig. 5. Visibility distance prediction using FLA.

Fig. 4 and 5 illustrate the predicted values generated by Extra Trees, Bagging, and the proposed system, compared to the actual values for each weather data point, both without and with the FLA algorithm, respectively. These figures show that using the FLA algorithm enhances the predictive ability of the proposed system, Extra Trees, and Bagging in forecasting visibility distance. Furthermore, the results indicate that the proposed system outperforms the other two algorithms.

The performance of the proposed system in the second stage is evaluated when using all features to predict sky CAVOK or when using only the predicted visibility distance feature to predict sky CAVOK. Table VI presents the performance of the proposed system in the second stage of the prediction model when using all features or only the predicted visibility distance for predicting sky CAVOK. The table also includes a comparison with 11 other machine learning algorithms: Gradient Boosting, AdaBoost, Logistic Regression, Bagging, KNN, Gaussian Naïve Bayes, Decision Trees, Multi-Layer Perceptron (MLP), Extra Trees, Quadratic Discriminant Analysis (QDA), and Linear Discriminant Analysis (LDA), all of which are compared against the proposed system, which utilizes CatBoost. Table VI shows that the proposed system outperforms other machine learning algorithms, achieving 100% accuracy when using only the predicted visibility distance and 77.38% accuracy when using all features to predict sky CAVOK. These results indicate that the proposed system performs better when using only the predicted visibility distance. The table also demonstrates that all machine learning algorithms perform better in predicting sky CAVOK when using only the predicted visibility distance rather than all features. Additionally, Table VI highlights that Bagging achieves a strong accuracy performance, reaching 71.09% accuracy when using all features and 100% accuracy when using only the predicted visibility distance. Decision Trees, Gradient Boosting, Extra Trees, KNN, and AdaBoost also show acceptable accuracy levels, achieving 67.72%, 65.50%, 60.35%, 60.90%, and 62.22% accuracy, respectively, when using all features, and 100% accuracy for all of them when using only the predicted visibility distance. On the other hand, Table VI reveals that Logistic Regression, MLP, QDA, and LDA are the weakest machine learning algorithms in this context. They achieve relatively poor accuracy results of 60.41%, 63.65%, 60.41%, and 60.38%, respectively, when using all features, and 62.02%, 67.28%, 52.03%, and 64.50% when using only the predicted visibility distance.

TABLE VI. PROPOSED SYSTEM PERFORMANCE WHEN PREDICTING SKY CAVOK

Features used	Machine learning algorithm	Precision	Recall	F1-score	Accuracy
All selected features of the clean dataset	Gradient Boosting	68.51%	58.45%	63.08%	65.50%
		63.23%	72.68%	67.63%	
	AdaBoost	63.57%	58.75%	61.06%	62.22%
		61.05%	65.76%	63.32%	
	Logistic Regression	61.04 %	59.36%	60.19%	60.41%
		59.80%	61.47%	60.62%	
	Bagging	71.03%	72.05%	71.54%	71.09%
		71.15 %	70.11%	70.63%	
	KNN	62.14 %	57.44%	59.70%	60.90%
		59.81%	64.41%	62.03%	
	Gaussian Naive Bayes	0.6224	0.4744	0.5384	58.99%
		56.95%	70.73%	63.10%	
	Decision Trees	66.80%	71.56%	69.10%	67.72%
		68.81%	0.6383	0.6623	
	MLP	67.16%	54.61%	60.24%	63.65%
		61.21%	72.84%	66.52%	
	Extra Trees	60.97 %	59.35%	60.15%	60.35%
		0.5975%	61.36%	60.54%	
	QDA	64.45%	47.91%	54.96%	60.41%
		57.99%	7312%	64.68%	
	LDA	61.02%	59.28%	60.14%	60.38%
		59.76%	61.49%	60.61%	
	Proposed system	**83.65%**	**68.55%**	**75.35%**	**77.38%**
		72.97%	**86.37%**	**79.11%**	
Predicted Visibility distance	Gradient Boosting	100%	100%	100%	100%
		100%	100%	100%	
	AdaBoost	100%	100%	100%	100%
		100%	100%	100%	
	Logistic Regression	100%	.01%	.01%	49.58%
		49.58%	100%	66.29%	
	Bagging	100%	100%	100%	100%
		100%	100%	100%	
	KNN	100%	100%	100%	100%
		100%	100%	100%	
	Gaussian Naive Bayes	100%	100%	100%	100%
		100%	100%	100%	
	Decision Trees	100%	100%	100%	100%
		100%	100%	100%	
	MLP	100%	25.68 %	40.87 %	62.53 %
		56.95%	100%	72.57 %	
	Extra Trees	100%	100%	100%	100%
		100%	100%	100%	
	QDA	51.24%	100%	67.76%	52.03%

		100%	03.24%	06.27%	
	LDA	67.95%	56%	61.40%	64.50%
		62.04%	73.14%	67.13%	
	Proposed system	**100%**	**100%**	**100%**	**100%**
		100%	**100%**	**100%**	

The proposed system achieved an accuracy of 100% when the visibility distance feature was used. However, when using all features without visibility distance, the model accuracy decreased to approximately 77.38%. Furthermore, using visibility distance as the sole predictor still resulted in 100% accuracy. This observation suggests that the feature has an exceptionally strong association with the target variable (SKY CAVOK). Since CAVOK is partially defined by visibility conditions in aviation meteorology, the model may have benefited from information that is intrinsically related to the target definition. Consequently, the reported perfect performance should be interpreted with caution, as it may indicate the presence of potential target leakage (data leakage) rather than genuine predictive capability.

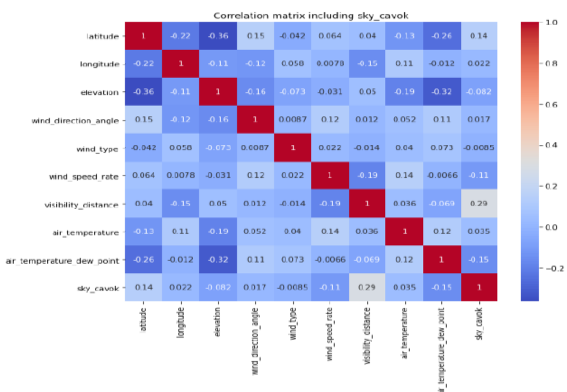

Fig. 6. Correlation matrix of features.

Fig. 6 illustrates the correlation between the features present in the cleaned dataset. It also shows that sky CAVOK is correlated with visibility distance, with a correlation of 29%. This correlation supports the conclusion that using only visibility distance can achieve accuracy comparable to using all other features combined.

V. Discussion

Table V and Table VI show that the proposed system outperformed machine learning algorithms such as Gradient Boosting, AdaBoost, Logistic Regression, Bagging, KNN, Gaussian Naïve Bayes, Decision Trees, MLP, Extra Trees, Quadratic QDA, Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, XGBoost, LightGBM, and LDA. The proposed system achieved an MAE of 529.7506 and an RMSE of 1174.6118 in predicting visibility distance, along with 77.38% accuracy in predicting the sky CAVOK condition.

In contrast, the machine learning algorithms achieved an MAE in the range of 505.4231 to 861.2343 and an RMSE in the range of 1191.6472 to 1590.2311 for predicting visibility distance, as well as an accuracy range of 58.99% to 71.09% in predicting sky CAVOK, which is lower than what the proposed system achieved.

TABLE VII. Comparison Between Related Works and the Proposed System

Visibility distance prediction metrics				Sky CAVOK prediction metrics
Related work	**Prediction**	**MAE**	**RMSE**	**Accuracy**
Kim et al. [16]	Visibility distance	2.04 km	2.94 km	-
Zhen et al. [19]	Visibility distance	-	3.82km	-
Zhang et al. [20]	Visibility distance	-	6.71km	-
Ortega et al. [22]	Visibility distance	-	1.175km	-
Chen et al. [23]	Visibility distance	17.52km	16.53km	-
Proposed system	**Visibility distance and sky CAVOK**	**.05297506 km**	**1.1746118 km**	**100%**

Also, the proposed system is compared with recent works, and this comparison is shown in Table VII. Table VII demonstrates that the proposed system is capable of predicting the sky CAVOK condition with 100% accuracy, whereas other related works do not predict the sky CAVOK. The table also shows that the proposed system outperforms recent works in predicting visibility distance, achieving an RMSE of 1.1746118 km, while other recent works report RMSE values in the range of 1.175 km to 6.71 km. The proposed system also utilizes a dataset that contains weather data specific to Saudi Arabia, whereas other related works do not use a dataset that includes weather data from Saudi Arabia.

VI. Conclusion

The success of supply chain operations in Saudi Arabia fully depends on weather conditions, particularly visibility distance and sky CAVOK. In this study, a new weather prediction system is proposed to predict visibility distance and sky CAVOK. The system follows three main steps: preprocessing, feature extraction, and prediction. In the preprocessing step, the dataset is cleaned by removing irrelevant features and rows with missing data. In the feature extraction step, the FLA algorithm is used to select the optimal features for the prediction process. Finally, in the prediction step, a new two-stage prediction algorithm is proposed. The first stage combines Bagging Regressor and Extra Trees Regressor to form a new ensemble algorithm for predicting visibility distance. The predicted visibility distance is then fed into the CatBoost algorithm to predict the CAVOK condition. The experimental results show that the proposed system achieves very good performance, outperforming other machine learning algorithms and recent studies. In future work, the system may be evaluated on additional datasets, and its capability can be extended to predict other weather variables. In addition, the system can be enhanced by using advanced optimization techniques and refining its architecture to achieve improved predictive performance.

Acknowledgment

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-23-RSP-4). The authors, therefore, thank the University of Jeddah for its technical and financial support.

REFERENCES

[1] M. Bonavita, "On some limitations of current machine learning weather prediction models," *Geophysical Research Letters*, vol. 51, no. 12, Jun. 2024. doi:10.1029/2023gl107377

[2] X. Ren et al., "Deep learning-based Weather prediction: A survey," Big Data Research, vol. 23, p. 100178, Feb. 2021. doi:10.1016/j.bdr.2020.100178.

[3] I. J. Orji and F. I. Ojadi, "Redefining supply chain sustainability: Introducing the context of extreme weather events," *Business Process Management Journal*, vol. 30, no. 6, pp. 1943–1975, Jun. 2024. doi:10.1108/bpmj-09-2023-0740.

[4] H. Liu, R. Xie, H. Qin, and Y. Li, "Research on dangerous flight weather prediction based on machine learning," *Journal of Physics: Conference Series*, vol. 2870, no. 1, p. 012020, Oct. 2024. doi:10.1088/1742-6596/2870/1/012020.

[5] Q. Li, R. Jing, and Z. S. Dong, "Flight delay prediction with priority information of weather and non-weather features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7149–7165, Jul. 2023. doi:10.1109/tits.2023.3270743.

[6] B. De Muynck, "Weather's wrath: Supply chains reel from 2024's extreme events," FreightWaves, https://www.freightwaves.com/news/weathers-wrath-supply-chains-reel-from-2024s-extreme-events (accessed Feb. 15, 2025).

[7] American Journal of Transportation, "Supply chain disruptions up 30% in first half of 2024," Resilinc, https://resilinc.ai/in-the-news/supply-chain-disruptions-up-30-in-first-half-of-2024/ (accessed Jan. 17, 2025).

[8] "Infographic: How climate change is impacting your supply chain," EarthScanTM, https://www.earth-scan.com/blog/climate-impact-supply-chain (accessed Dec. 12, 2024).

[9] C. M. Rebecca Mathews, "Climate challenges: Enhancing supply chain resilience to weather events," The Business Continuity Institute (BCI), https://www.thebci.org/news/climate-challenges-enhancing-supply-chain-resilience-to-weather-events.html (accessed Jan. 30, 2025).

[10] L. Chen *et al.*, "Machine learning methods in weather and climate applications: A survey," *Applied Sciences*, vol. 13, no. 21, p. 12019, Nov. 2023. doi:10.3390/app132112019.

[11] A. McGovern *et al.*, "A review of machine learning for Convective Weather," *Artificial Intelligence for the Earth Systems*, vol. 2, no. 3, Jul. 2023. doi:10.1175/aies-d-22-0077.1

[12] O. Alghushairy, R. Alsini, A. Yafoz, S. Yonbawi, and X. Ma, "Two-stages weather forecasting system for optimizing supply chain operations in Saudi Arabia using BIGRU and Fox metaheuristics algorithms," *Arabian Journal for Science and Engineering*, vol. 51, no. 5, pp. 5721–5738, Jul. 2025. doi:10.1007/s13369-025-10464-8.

[13] D. Fister, J. Pérez-Aracil, C. Peláez-Rodríguez, J. Del Ser, and S. Salcedo-Sanz, "Accurate long-term air temperature prediction with machine learning models and data reduction techniques," *Applied Soft Computing*, vol. 136, p. 110118, Mar. 2023. doi:10.1016/j.asoc.2023.110118.

[14] K. Ait Ouadil *et al.*, "Atmospheric visibility estimation: A review of Deep Learning Approach," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 36261–36286, Sep. 2023. doi:10.1007/s11042-023-16855-z

[15] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine learning methods for weather forecasting: A survey," *Atmosphere*, vol. 16, no. 1, p. 82, Jan. 2025. doi:10.3390/atmos16010082

[16] B.-Y. Kim, M. Belorid, and J. W. Cha, "Short-term visibility prediction using tree-based machine learning algorithms and Numerical Weather Prediction Data," *Weather and Forecasting*, vol. 37, no. 12, pp. 2263–2274, Dec. 2022. doi:10.1175/waf-d-22-0053.1

[17] Y. Zhang *et al.*, "Visibility prediction based on machine learning algorithms," *Atmosphere*, vol. 13, no. 7, p. 1125, Jul. 2022. doi:10.3390/atmos13071125

[18] L. Ortega, L. D. Otero, and C. Otero, "Application of machine learning algorithms for visibility classification," *2019 IEEE International Systems Conference (SysCon)*, pp. 1–5, Apr. 2019. doi:10.1109/syscon.2019.8836910

[19] M. Zhen *et al.*, "Application of a fusion model based on machine learning in visibility prediction," *Remote Sensing*, vol. 15, no. 5, p. 1450, Mar. 2023. doi:10.3390/rs15051450

[20] C. Zhang *et al.*, "Weather visibility prediction based on multimodal fusion," *IEEE Access*, vol. 7, pp. 74776–74786, 2019. doi:10.1109/access.2019.2920865

[21] Z. Zang *et al.*, "A modified RNN-based deep learning method for prediction of atmospheric visibility," *Remote Sensing*, vol. 15, no. 3, p. 553, Jan. 2023. doi:10.3390/rs15030553

[22] L. C. Ortega, L. D. Otero, M. Solomon, C. E. Otero, and A. Fabregas, "Deep learning models for visibility forecasting using climatological data," *International Journal of Forecasting*, vol. 39, no. 2, pp. 992–1004, Apr. 2023. doi:10.1016/j.ijforecast.2022.03.009

[23] H. Chen *et al.*, "Visibility forecast in Jiangsu province based on the GCN-GRU model," *Scientific Reports*, vol. 14, no. 1, Jun. 2024. doi:10.1038/s41598-024-61572-8

[24] S. N. Qasem and S. M. Alzanin, "An effective forecasting approach of temperature enabling climate change analysis in Saudi Arabia," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, 2024. doi:10.14569/ijacsa.2024.0150372

[25] Z. Yuying, Z. Yu, and Z. Bei, "Visibility forecast based on PhyDNet-ATT deep learning algorithm," *Journal of Applied Meteorological Science*, vol. 35, no. 6, 2024. doi:10.11898/1001-7313.20240603

[26] O. Suleiman, "Long-term weather elements prediction in Jordan using adaptive neuro-Fuzzy Inference System (ANFIS) with GIS techniques," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, 2018. doi:10.14569/ijacsa.2018.090213

[27] KAPSARC Data Portal, Saudi Arabia Hourly Climate Integrated Surface Data, Natl. Ocean. Atmos. Adm. (2024).

[28] P.H. Silverstone, National Oceanic and Atmospheric Administration (Noaa), in: Navy 21st Century, 2001-2022, 2024. https://doi.org/10.4324/9780367808259-12.