Natural Hazards Research 4 (2024) 295–303

Contents lists available at ScienceDirect

Natural Hazards Research

journal homepage: www.keaipublishing.com/en/journals/natural-hazards-research

A comparative study on data mining models for weather forecasting: A case study on Chittagong, Bangladesh

Mohammad Sadman Tahsin [a,*], Shahriar Abdullah [b], Musaddiq Al Karim [c], Minhaz Uddin Ahmed [a], Faiza Tafannum [d], Mst Yeasmin Ara [c]

[a] Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[b] Department of Geography, University of Alabama, Tuscaloosa, AL, 35487, United States
[c] Department of Computer Science and Engineering, American International University Bangladesh, Dhaka, Bangladesh
[d] Department of Computer Science and Engineering, University of Texas at Arlington, Texas, TX, 76019, United States

ARTICLE INFO

Keywords:
Weather prediction
Temperature
Data mining
Weather forecasting

ABSTRACT

The primary focus of this study is to analyze and predict the patterns of this essential feature of the natural world. This study analyses and predicts the daily weather patterns of a specific urban area. This article utilizes weather data over 20 years to analyze the climate patterns of Chittagong city. A total of 12 distinct Data Mining models were employed to predict daily weather patterns. The algorithms can be categorized into three distinct types, namely rules-based, tree-based, and function-based. To evaluate the effectiveness of the models, various performance metrics were computed, including precision, recall, accuracy, F-measure, and the area under the receiver operating characteristic curve (ROC area). Based on the results obtained, it can be concluded that among the 12 algorithms evaluated, J48 exhibits the highest level of performance and accuracy. The J48 classifier demonstrated an accuracy of 82.30%, precision of 82.40%, recall of 82.20%, f-measure of 84.20%, and a ROC area of 97.8%. Furthermore, a comprehensive analysis of the confusion matrix for all twelve algorithms was conducted to facilitate further evaluation.

1. Introduction

The role of weather has been of utmost significance in the survival and sustainability of life since the beginning. The determination of water and food availability is of utmost importance for the sustenance of various forms of life. From time to time, the weather has exhibited varying manifestations, displaying both benevolence and severity. Hence, the examination, comprehension, and anticipation of meteorological conditions have been demonstrated to hold considerable importance in ensuring the long-term survival and well-being of the human species on our planet. Early civilizations historically relied on archaic techniques to predict rainfall, although these approaches were often characterized by limited efficiency and accuracy (Scofield, 2010). As a result, human beings were confronted with numerous existential challenges, including food insecurity, as well as enduring the adverse effects of prolonged droughts and floods. The advancement of technology and the availability of real-time satellite data have significantly facilitated the analysis and comprehension of weather patterns, becoming them more accessible and accurate (Denardini et al., 2016). Despite this, the exponential advancement of technology has not only been confined to the mere prediction of rainfall. Still, it has also extended to comprehending weather patterns in a manner that allows for the accurate anticipation of weather phenomena and climatic fluctuations.

Meteorologists have developed numerical weather forecasting models that employ mathematical equations to simulate the behavior of the atmosphere, specifically in terms of variations in temperature, pressure, and moisture over time (Al-Yahyai et al., 2010; Bauer et al., 2015; Gneiting and Raftery, 2005). The equations are programmed into a computer system, which is provided with data regarding the prevailing atmospheric conditions. Various atmospheric models exist to simulate the Earth's atmosphere, with each model offering a distinct interpretation of its characteristics (Skamarock and Klemp, 2008). The forecaster acquaints themselves with the idiosyncrasies of each model and prioritizes those that have superior predicting capabilities for certain weather phenomena.

Climate change is currently seen as a significant issue of concern. The phenomenon has led to a significant rise in both the quantity and intensity of natural disasters, resulting in the unfortunate demise of countless individuals. Based on the findings of the Intergovernmental

* Corresponding author.
E-mail address: sadman.tahsin3@gmail.com (M.S. Tahsin).

https://doi.org/10.1016/j.nhres.2023.12.014
Received 4 October 2023; Received in revised form 18 December 2023; Accepted 18 December 2023
Available online 22 December 2023

Panel on Climate Change (IPCC) Second Assessment Report, which was published in 1995, it is indicated that the global mean surface air temperature has seen an increase ranging from 0.3 to 0.6 °C over the preceding century (IPCC, 2007). Based on their estimates, it is projected that the mean global surface temperature is expected to increase by a range of 1.0–3.5 °C by the year 2100. It is projected that the global mean sea level could experience a range of increase between 15 cm and 95 cm by the year 2100, with a rise of 50 cm being the most probable outcome (Meehl et al., 2005). One of the effects brought about by this warming would be a shift in world wind and rainfall patterns.

In this paper, 12 data mining algorithms such as OneR, Decesion Tree, J48, Random Forest etc were used to predict weather. Many state of the art evaluation metrics were used to measure the performance of the algorithms. A dataset was collected from the NASA Power for this experiment.

2. Literature review

To effectively address the issue of climate change, it is imperative to have a comprehensive understanding of weather patterns and develop precise forecasting techniques. The utilization of advanced techniques and technologies such as machine learning, deep learning, and artificial intelligence has facilitated a more comprehensive exploration of research endeavors, resulting in enhanced precision in weather pattern prediction. Machine learning models are widely employed in the field of weather prediction, in comparison to other methodologies. For instance, authors of this paper (Markovics and Mayer, 2022) evaluated 24 machine learning models for one-day power forecasting using 15-min datasets from 16 Hungarian PV plants. Another study (Chandrayan et al., 2022) used libraries like NumPy, Pandas, Keras, Git, TensorFlow, Matplotlib, and platforms like Anaconda for weather forecasting. The accuracy rates of RNN, MLP, and RBF were respectively 94.3%, 91.5%, and 92.2%. respectively. This paper (Gad and Hosahalli, 2022) studied the most frequent machine learning techniques to construct a durable long-term weather forecasting model. In addition, all potential combinations of model parameters for simulations were studied, and the performance of each technique was evaluated using a 10-fold cross-validation process. The experimental findings of the classifiers demonstrated that the decision tree CART, XGBoost, and AdaBoost models had higher classification accuracy than the other strategies. A study (Yang et al., 2022) suggested an artificial intelligence-based approach, where the authors of this study utilized actual data from a wind farm in northern China to evaluate the example calculation. According to the findings, the accuracy of the suggested method's predictions is comparable to that of the conventional purely physical model. After a few months, the accuracy of the prediction approaches that of the purely physical model, and the efficiency of the calculation is much improved. Another research paper (Tahsin et al., 2021) examined the contributions of many key machine learning methods to the forecasting of weather patterns. Having access to 20 years of data from a single station has provided a consolidated perspective to evaluate and conclude whether weather forecasting system is superior. They used AdaBoosting, XGboosting, Stacking KNN, Stacking Neural Network, etc. in order to evaluate their performance.

Sufficient works with deep leaning models on weather prediction is also done. A group of researchers (Pérez et al., 2004) developed a deep learning technique using trained convolutional neural networks. It gave a confidence rating to medium-range predictions created from a new meteorological situation, revealing if predictability is above or below average for the season. This strategy was less accurate than ensemble weather prediction models but computationally efficient and beat non-numerical alternatives. Using a conditional deep convolutional generative adversarial network, this paper (Bihlo, 2021) forecasted Europe's geopotential height, 2-m temperature, and total precipitation for the following 24 h. The geopotential height and the temperature at 2 m may be adequate for data-only weather forecasting. The author developed a Monte-Carlo dropout-based deep learning ensemble weather prediction system. This computationally inexpensive method enhances the forecasting model by evaluating the present weather forecast's uncertainty. In another (Huang et al., 2020) article, a two-phase weather management system was presented. The system offers real-time weather monitoring in buses and stations and makes weather forecasts using predictive models by combining information processing, bus mobility, sensors, and deep learning. The proposed system provides consistent weather monitoring and an accurate one-day weather prediction using trained models. Author of this paper (Sharma and Sharma, 2022) used deep learning algorithms to mimic an updated general circulation model's dynamic, improving weather forecasts and long-term climate time series. This study predicted weather using deep learning techniques. Using historical data from multiple climate stations and Deep Learning models, this study offered a better weather predicting method. Though machine learning and deep learning techniques have been used by several researchers, data mining methods are not yet used adequately in weather forecasting.

Data mining is a technique in which a large group of data is collected to gain knowledge, understand patterns in the collection of data so that a general understanding of that particular data model can be achieved (Calders and Custers, 2013). There are several methods of data mining, for example: Classification, Clustering Regression, Genetic Algorithms, Rule Induction, Nearest Neighbor method, etc. This paper focuses wholly on classification techniques to derive accurate results regarding weather forecasts. Classification algorithms are supervised learning methods which determine outputs based on previous observations and classifies outputs as distinct labels (Umadevi and Marseline, 2017). Data mining approaches, in contrast to conventional statistical methods, look for interesting information without requiring a priori preconceptions; the kind of patterns that might be found depend on the data mining tasks used. Data mining tasks may be broadly divided into two categories: descriptive tasks that explain the basic qualities of the current data and predictive tasks that aim to make predictions based on inference from the data. In comparison to statistical procedures, these strategies are frequently more potent, adaptable, and effective for exploratory analysis (Casas et al., 2009). Some research has been conducted on data mining and weather prediction, and it has proven to be a highly useful tool for the prediction of various weather phenomena, such as rainfall (Casas et al., 2009; Asklany et al., 2011), thunderstorm, evaporation, wind speed (Pabreja and Datta, 2012; KAYA et al., 2013). The other weather parameters like temperature are not yet extensively studied by using data mining.

By understanding the gap in weather prediction by data mining, the paper uses 12 models to forecast temperature which can be divided into 3 categories of classifiers such as rules based, tree based and function-based classifiers. The parameters of the used models were also discussed in this paper. The dataset that has been used in this paper has been collected from NASA power which consists of weather. The data timeline of this work's is 20 years. The location that was considered for this research was Chawkbazar, Chittagong. For evaluation, some state-of-the-art performance evaluation metrics were used as accuracy, precision, recall, ROC score, confusion matrix etc. to validate the performance. The main goal of this paper is to find out which is the best data mining model for a specific location. The paper also describes which parameters were used in the algorithms. The finding of this study will help to select attributes and approaches to forecasting temperature.

3. Methodology

3.1. Data Acquisition

The meteorological data of center of Chawkbazar, Chittagong city (latitude: 22.3572, longitude: 91.8302) was collected from NASA Power (https://power.larc.nasa.gov/data-access-viewer) (NASA POWER Prediction Of Worldwide) for 20 years (from 01/01/2000 to 31/12/2020). A total of 10 attributes were selected for our study excluding the data time frame. The resolution of the meteorological data is $\frac{1}{2}^{\circ} \times \frac{5}{8}^{\circ}$ latitude/longitude grid. The timescale is one day into the future. The sample of data and the details of the attributes are shown in Fig. 1 and Table 1:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	YEAR	MO	DY	WS10M_MAX	WS10M_MIN	WS10M	QV2M	RH2M	PRECTOTC	PS	T2M_MAX	T2M_MIN	T2M
2	2000	1	1	4.04	2.55	3.23	8.97	73.88	0	100.94	24.21	12.47	17.57
3	2000	1	2	3.88	2.12	3.09	8.42	69.69	0	100.9	24.17	11.67	17.44
4	2000	1	3	3.23	1.69	2.77	8.42	70.62	0	100.79	23.97	12.01	17.32
5	2000	1	4	3.59	2.3	2.91	8.54	72.31	0	100.83	23.87	11.92	17.27
6	2000	1	5	2.98	1.99	2.45	8.42	71.69	0	100.89	23.69	12.12	17.27
7	2000	1	6	3.15	1.34	2.56	8.12	69.81	0	100.8	22.62	12.16	16.91
8	2000	1	7	2.91	1.75	2.48	8.3	74	0	100.92	22.49	10.73	16.49

Fig. 1. Sample dataset.

Table 1
Attributes used for the prediction.

SL No	Attributes	Details
1	WS10M MAX	This displays the highest possible wind speed measured 10 m above ground.
2	WS10M MIN	This displays the lowest possible wind speed measured 10 m above ground.
3	WS10M	Average hourly wind velocity at a height of 10 m above ground.
4	QV2M	The amount of water vapor in the air as a fraction of the total mass of the air at 2 m (g water/kg total air).
5	RH2M	Partial pressure of water vapor, expressed as a percentage of the partial pressure at saturation.
6	PRECTOTCORR	The weighted average, after accounting for bias, of all rain that falls on Earth's surface (includes water content in snow).
7	PS	The standard atmospheric pressure above Earth's surface.
8	T2M MAX	Here, the highest temperature is shown at a height of 2 m above the ground.
9	T2M MIN	This is the lowest hourly air temperature recorded at a height of 2 m above ground for the relevant time frame. We scored this quality on a 7-point scale.
10	T2M	The typical dry-bulb air temperature 2 m above the ground.

3.2. Classification

The proposed approach classifies the Chittagong weather data using multiple Data mining models. A number of feature selection techniques are used in attribute reduction. For these experiments, WEKA workbench (Hall et al., 2009) was used. For all classifiers, default parameters are used. The experiments employ the meteorological data collection retrieved from the NASA Power database. The acquired data from the database was in numeric format. The performance of classifiers is assessed using multiple evaluation metrics.

After collecting the dataset, the dataset was converted to categorical form. Table 2 describes how the conversion was done. In the dataset, "T2M" is the target value. This attribute has 6 classes such 'very low', 'low', 'moderately low', 'moderate', 'high', 'very high' (Table 2). Rest attributes have 3 unique values except "T2M_MIN", "T2M_MAX". These two attributes also have six unique values like T2M.

Table 2
Dataset description.

Attribute name	Description
T2M	13.5 ≥ very low <20, 20 ≥ low <24, 24 > moderately low <27, 27 ≥ moderate <28 28 ≥ very high <29, 29 ≥ High <34
T2M_MIN	7 ≥ very low <16, 16 ≥ low <21 21 ≥ moderately low <25, 25 ≥ moderate <26 26 ≥ high <26.5, 26.5 ≥ very high <29
T2M_MAX	19 ≥ very low <26, 26 ≥ low <29 29 ≥ moderately low <30, 30 ≥ moderate <31 30 ≥ high <33, 33 ≥ very high <39
PS	98.5 ≥ low <100.5, 100.5 ≥ moderate <101, 101 ≥ high <102
PRECTOTCORR	0 = low, 0.1 ≥ moderate <20, 20 ≥ high <500
RH2M	33 ≥ low <75, 75 ≥ moderate <85, 85 ≥ high <95
QV2M	5 ≥ low <12, 12 ≥ moderate <18, 18 ≥ high <24
WS10M	0.5 ≥ low <3, 3 ≥ moderate <5, 5 ≥ high <12.5
WS10M_MIN	0.1 ≥ low <1.5, 1.5 ≥ moderate <2.5, 2.5 ≥ high <8.5
WS10M_MAX	1 ≥ low <3.5, 3.5 ≥ moderate <5, 5.5 ≥ high <7

3.3. Classification methods

Data mining classification involves the mapping of each tuple in a data collection to specific output variables. The output variables are discrete and are commonly denoted as class labels or categories. The mapping functions are used to forecast the class label for each tuple based on the provided observation. The classifiers' performance is assessed based on their ability to accurately and effectively predict the class label for new test data. In the WEKA tool, categorization strategies can be broadly grouped into several approaches depending on their functionality, including rules, trees, and functions. Tables (3–14) contains the details of the parameters and their values.

3.3.1. Classification based on rules

OneR: One rule is a basic and accurate classifier that generates rules based on the value of a single prediction. Each predictor generates a single rule from the data and selects the rule with the lowest error cost (Mahajan and Ganpati, 2014). For each predictor, a frequency table is compiled for the purpose of formulating a rule.

Table 3
Details of OneR algorithm.

Parameter name	value
Batch Size	100
Debug	False
minBucketSize	6
numDecimalPlaces	2

Decision Table: Decision tables are the simplest hypothesis spaces and simplest to read. Experimental findings reveal that IDTM, an algorithm inducing decision tables, may beat cutting-edge algorithms like C4.5 on simulated and real-world domains with discrete information. Remarkably, some datasets with continuous features perform well, suggesting that many machine learning datasets either do not require them or have few values for them in a study (Kohavi, 1995).

Table 4
Details of Decision Table algorithm.

Parameter name	value
Batch Size	100
crossVal	1
displayRules, doNotCheckCapabilities, uselBk	False
evaluationMeasure	Accuracy (Discrete Class) RMSE (Numeric Class)
numDecimalPlaces	2
Search	BestFirst – D1 –N5

JRIP: William W. Cohen's Repeated Incremental Pruning to Generate Error Reduction (RIPPER) administers JRIP, a learning approach that improves IREP (Waseem et al., 2013). JRIP develops an initial set of rules due to the fact that error reduction increases class size. Rule independently. Divide occurrences of rule formation into two-thirds growth and one-third pruning. Order by size Using precedents, establish rules for the lowest class.

Table 5
Details of JRIP algorithm.

Parameter name	value
batchSize	100
checkErrorRate, usePruning	True
Debug, doNotCheckCapabilities	False
Folds	3
minNo	2.0
numDecimalPlaces	2
Optimizations	2
Seed	1

Ridor: The RIDOR classifier develops basic rules with the lowest possible error rate. For each exception that generates exceptions in a tree structure, the optimal exceptions are generated. The exception subset of rules anticipates class labels distinct from those of the fundamental rules. Error trimming is lowered progressively to create exceptions (Thepade et al., 2013).

Table 6
Details of Ridor algorithm.

Parameter name	value
batchSize	100
debug, doNotCheckCapabilities, majorityClass, wholeDataErr	False
Folds	3
minNo	2.0
numDecimalPlaces	2
Seed	1
Shuffle	10

3.3.2. Classification based on trees

J48: J48 is an updated version of Ross Quinlan's WEKA decision tree classifier C4.5. The majority of predictive learning algorithms determine the target value by assessing the predictor attribute values of the input dataset (Cufoglu et al., 2009). Also, a method translates the vector of attribute values to the existing data's underlying target class. The data set is subdivided at each node using recursive divide-and-conquer to produce a decision tree. The data set consists of a vector of attributes, and each tuple contains a class label indicating whether it belongs to a category. Partitioning a data set, also known as attribute selection measure, is a heuristic technique for picking a characteristic to divide the data set (Kandhasamy and Balamurali, 2015).

Table 7
Details of J48 algorithm.

Parameter name	value
batchSize	100
binarySplits, debug, doNotCheckCapabilities, doNotMakeSplitPointActualValue, reducedErrorPruning, saveInstanceData, Unpruned, useLaplace	False
collapseTree, subtreeRaising, useMDLcorrection	True
confidentialFactor	0.25
minNumObj, numDecimalPlaces	2
numFolds	3
Seed	1

LMT: The Logistic Model Tree classifier makes use of logistic regression and decision tree learning. LMT can accommodate two-class, n-class, numerical, nominal, and missing target variables. LMT is a mix of the basic regression model and the complicated tree structure, providing there are enough data to support the structure. LMT provides explicit class probability estimates as opposed to classifying (Chen et al., 2018).

Table 8
Details of LMT algorithm.

Parameter name	value
batchSize	100
Debug, doNotCheckCapabilities, doNotMakeSplitPointActualValue, errorOnProbabilities, useAIC, splitOnResiduals	False
fastRegression	True
minNumInstances	15
numBoostingInterations	−1
numDecimalPlaces	2
weightTrimBeta	0.0

Random Forest: Breiman and Pal created Random Forest Classifier using several classification trees (Breiman, 2001; Pal, 2005). Each tree classifies with a weight. The output is the sum of all trees' classifications. Each tree votes for the best class to classify the vector input when the classifier is constructed from random input samples. A tree is formed by combining or randomly selecting node characteristics. Bagging random samples from the training data set. Classifying the input tuple uses the class with the highest votes from all tree predictors. Metric-based feature selection and pruning create the decision tree (Pandey et al., 2020).

Table 9
Details of Random Forest algorithm.

Parameter name	value
batchsize	100
braeakTimeRandomly, calcOutOfBag, computeAttributeImportance, Debug, doNotCheckCapabilities, printClassifiers, outputOutOfBagComplexityStatistics, storeOutOfBagPredictions	False
maxDepth, numFeatures	0
numDeccimalPlaces	2
numExecutionSlots, seed	1
numIterations	100

CART: Classification and Regression Trees (CART) is a machine learning decision tree prediction model (Loh, 2011). By dividing input data and choosing the optimum classification model for each split, a classifier is built. Using decision trees, forks may be represented. Study examples to create decision trees. Using queries to form a decision tree, input samples for learning are broken down into little components. CART computes the optimum split, which divides the tree into two identical halves, based on all given attributes and values (Crawford, 1989).

Table 10
Details of CART algorithm.

Parameter name	value
batchSize	100
debug, sidesPer1.0useOneSe, doNotCheckCapabilites,	False
heuristic, usePrune	True
minNumObj	2.0
numDecimalPlaces	2
numFoldsPuning	5
seeds	1

3.3.3. Classification based on function

MLR: Multinomial logistic regression is an extension of binomial logistic regression that is used to forecast outcomes for multiclass variables (Bohning, 1992). Given one or more independent factors, it predicts nominal dependent variables. The maximum likelihood estimate is used to determine the probability of belonging to a category.

Table 11
Details of MLR algorithm.

Parameter name	value
batchSize	100
debug, doNotCheckCapabilities, resume, useEstimatedPriors, useResampling	False
numDecimalPlaces, numFoldsPuning, minNumObj	2
nominalToBinaryFilter, normalizeAttributes	True
seed	0

MLP: MLP is a deep feedforward neural network variant. Perceptron, according to Rosenblatt's MLP, is a linear classifier (Pinkus, 1999). A perceptron produces distinct outputs depending on the linear combination of its real-valued inputs and its weight. MLP is composed of multiple perceptrons with an arbitrary number of hidden layers, an input layer, and an output layer (Pandey and Barai, 1995). The input–output layer training teaches layer dependencies. Training reduces error by adjusting the weights and focus bias. Backpropagation is used to alter the weights and biases of the network.

Table 12
Details of MLP algorithm.

Parameter name	value
GUI, debug, decay, doNotCheckCapabilities, resume	False
autoBuild, reset, nominalToBinaryFilter, normalizeAttributes, normalizeNumericClass	True
batchSize	100
hiddenLayers	a
learningRate	0.3
momentum	0.2
numDecimalPlaces	2
seed, validationSetSize	0
trainingTime	500
validationThreshold	20
weightThreshold	100

LogitBoost: The LogitBoost method is developed for binary classification situations in which one of two probable outcomes must be predicted (Cai et al., 2006). It works by constructing an ensemble of weak classifiers, each of which predicts based on a subset of the given characteristics. Using the gradient boosting approach, the weak classifiers are merged to create a powerful classifier. Each iteration of the LogitBoost method adds a new decision stump to the ensemble that minimizes the negative log-likelihood of the training data, a measure of how well the current model matches the data. The weights of the training examples are modified to highlight instances that were incorrectly identified by the present ensemble of weak classifiers (Kamarudin et al., 2017).

Table 13
Details of LogitBoost algorithm.

Parameter name	value
ZMax	3.0
batchSize	100
classifier	DecisionStump
debug, doNotChecckCapabilities, resume, useEstimatedPriors, useResampling	False
likelihoodThreshold	−1.7076
numDecimalPlaces	2
numIterations	10
numThreads, poolSize, seed, shrinkage	1

SMO: Platt presented Sequential Minimum Optimization (SMO) in 1998 as the shortest quadratic programming-based optimization method for linear SVM with sparse data (Huang et al., 2015). Extreme decomposition and the resolution of the smallest optimization issue in each phase are used to handle SMO. Using analytic approaches, it optimizes two Lagrange multipliers in succession. Eliminating the iterative usage of quadratic programming optimizer is the method's strength (Dantas Dias and Rocha Neto, 2016).

Table 14
Details of SMO algorithm.

Parameter name	value
batchSize	100
buildCalibrationModels, checksTurndeOff, doNotCheckCapabilities	False
calibrator	Logistic
epsilon	1.0E-12
filterType	Normalize Training Data
kernel	PolyKernal
numDecimalPlaces	2
numFold	−1
randomSeed	1
toleranceParameters	0.001

3.4. Performance evaluation parameter

It is difficult to assess the performance of data mining algorithms with little data. When there are few data points and the dataset is in categorical form, cross-validation (Witten et al., 2005) is the most appropriate technique for study (Ayres-de-Campos et al., 2000). After splitting meteorological data into training and test sets, tenfold cross-validation is used to assess performance. The training set and test set are used to construct and validate a categorization model.

True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are used to calculate the accuracy of classifiers (FN) (Chicco and Jurman, 2020). Precision and recall act as the validation's statistical metrics. Precision indicates the classifier's positive predictive value. Precision is the ratio of precisely predicted positive observations to the total number of accurately predicted positive observations, while Recall is the fraction of positive test results.

$$Precision = (TP \,/\, (TP + FP)) \times 100 \tag{1}$$

$$Recall\ (Sensitivity) = (TP(TP + FN) \times 100) \tag{2}$$

$$Accuracy = ((TP + TN) \,/\, (TP + TN + FP + FN)) \times 100 \tag{3}$$

The receiver operating curve (ROC) is a graphical depiction of classifier performance as thresholds increase (Fawcett, 2006). The ROC curve is shown between each specificity and sensitivity value. The relevance of ROC is related to the number of thresholds tested. F-measure is an extra statistical indicator for performance evaluation.

$$F - measure = ((2 \times TP) \,/\, (2 \times TP + FP + FN)) \tag{4}$$

The numbers in a confusion matrix indicate the difference between expected and actual values. The “TN' output indicates the number of properly identified negative samples. Similarly, “TP,” which stands for True Positive, represents the number of positive events that were properly identified. In contrast, “FP” is the number of incorrectly classified positive classes, while “FN” is the number of incorrectly classified negative classes. In the context of classification, accuracy is a crucial performance parameter (Kulkarni et al., 2020).

4. Result

4.1. Model performance

To evaluate the performance of the prediction models, the dataset is divided into two sets: the training set and the testing set. The k-fold cross-validation method is the most well-known and trustworthy evaluation methodology. k-fold cross-validation is a method that separates the dataset into k distinct subgroups and performs the hold-out procedure k times (Yadav and Shukla, 2016). One of the k-subsets acts as the testing set during each phase, while the other k-1 subsets serve as training sets (Huang et al., 2007). The average performance and error are determined for all k trials. In this investigation, the number of trials for k is set at 10 since this maximizes training efficiency and minimizes generalization error (Yadav and Shukla, 2016). The weather dataset is separated into 10 about equal-sized chunks. One of the subsets acts as the testing set each time, while the residuals serve as the training set. For each subgroup, the method is repeated 10 times.

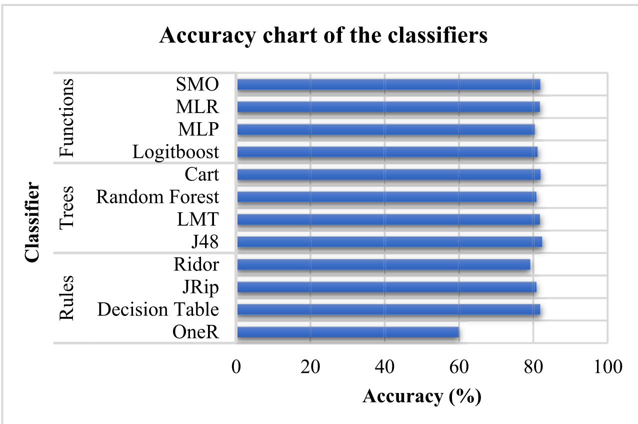

Fig. 2. Accuracy chart of the Classifiers.

The rule-based classifiers demonstrate a wide range of overall accuracy, ranging from 59.92% to 81.79% (refer to Fig. 2). The OneR, Decision Table, JRIP, and RIDOR classifiers achieve accuracies of 59.92%, 81.79%, 80.84%, and 79.09%, respectively. Among all the rule-based classifiers that were evaluated, Decision Tables provided the highest classification accuracy, achieving a value of 81.79%. Nevertheless, the OneR model had the lowest level of effectiveness when compared to all other models.

The research results demonstrate that the tree-based classifiers shown a notable level of performance, ranging from 79% to 83.5%. The J48 algorithm has exhibited the highest level of potential in terms of forecast accuracy, with a rate of 82.30%. The Logitboost classifier had the highest performance with an accuracy of 81.10%, followed by MLP with 80.37%, MLR with 81.67%, and SMO with 81.86%.

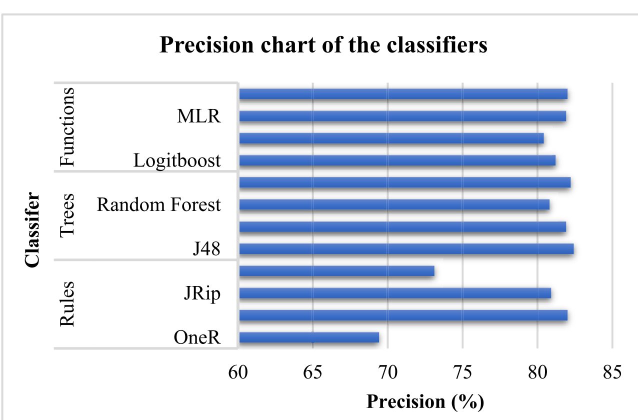

Fig. 3. Precision chart of the classifiers.

The categorization skills are characterized by a high level of recall and accuracy, as shown by the examination of the acquired data for numerous parameters. The accuracy and recall of tree-based classifiers are higher, as seen in Table 3. It indicates that all tree-based classifiers functioned well with comparable results. The overall performance of the classifiers was dominated by tree-based classifiers, with J48 having the highest performance, followed by LMT, Random Forest, and CART. Rule-based classifiers have shown good performance with the weather data set, similar to RIDOR, JRIP, Decision Table, and OneR. The function-based classifier assisted with the classification of data inside the MLP, SMO, MLR, and Logitboost hierarchies. The precision and recall of rule-based, tree-based, and function-based classifiers are shown in Figs. 2 and 3.

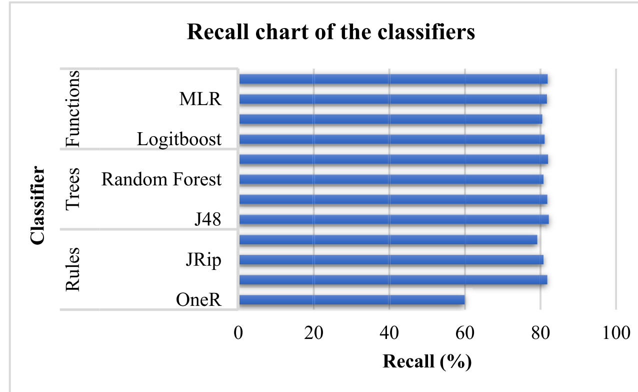

Fig. 4. Recall chart of the classifiers.

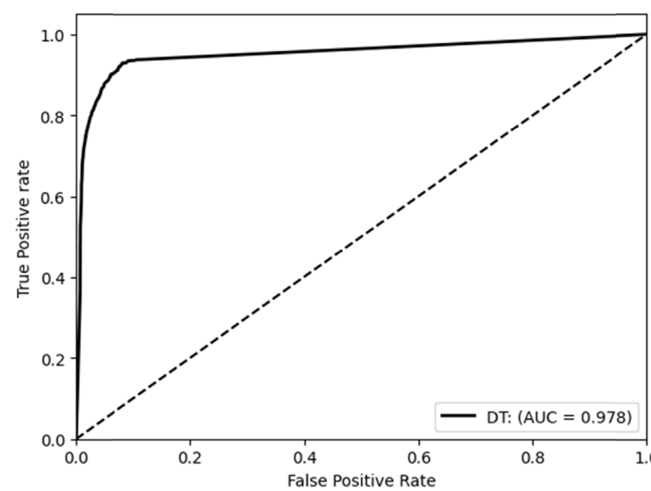

Fig. 5. ROC curve of J48 model.

The F-measure and ROC area values for the classifiers are shown in Table 15. ROC values for rule-based classifiers lie between 54% and 82%. The decision Table illustrates that rule-based classifiers have a ROC value of 96.9%. The ROC area value for tree-based classifiers falls between 82% and 85%, with J48, LMT, Random Forest, and CART having values of 84.2%, 81.7%, 80.8%, and 82%, respectively. J48's perfect ROC and F-measure scores of 97.8 and 84.2, respectively, demonstrate its supremacy over other classifiers. Similarly, the ROC area values for Cart, Random Forest, and LMT are 96%, 96.4, and 96.3%, respectively, suggesting exceptional performance in recognizing meteorological data according to their respective functions. J48 has definitely shown the greatest F-measure and ROC value, with 84.2% and 97.7%, respectively Fig. 4 and Table 16 (see Fig. 5).

4.2. *Confusion matrix*

Table 17 displays the resulting matrices. For a deep understanding, the OneR classifier is described here. Out of a total count of 1041 “Very low” class data, 1017 instances in the first column were correctly predicted to be in this category. In the second column, 951 out of a total of

Table 15
Performance of classifiers in terms of accuracy, precision and recall.

Classifier	Accuracy (%)	Precision (%)	Recall (%)
Rules			
OneR	59.92	69.40	59.9
Decision Table	81.79	82	81.8
JRip	80.84	80.9	80.8
Ridor	79.09	79.1	79.1
Trees			
J48	82.30	82.4	82.2
LMT	81.75	81.9	81.8
Random Forest	80.77	80.8	80.8
Cart	81.92	82.2	82
Functions			
Logitboost	81.10	81.2	81.1
MLP	80.37	80.4	80.4
MLR	81.67	81.9	81.7
SMO	81.86	82	81.9

Table 16
Performance of classifiers in terms of f-measure and ROC area.

Classifier	F-Measure	ROC area
Rules		
OneR	54.2	74.9
Decision Table	81.8	96.9
JRip	80.8	95.1
Ridor	79.1	87.2
Trees		
J48	84.2	97.8
LMT	81.7	96.3
Random Forest	80.8	96.4
Cart	82	96
Functions		
Logitboost	82.1	97.3
MLP	80.4	96.6
MLR	81.7	97.3
SMO	81.9	95.1

1224 cases (255 + 951+18) were correctly predicted to be in the “low” class. Similar to this, 1340 cases out of 1914 “Moderate” class data and 727 instances out of 1284 total counts in the “Moderately Low” class were correctly predicted. Finally, out of a total count of 683 cases in the “Very High” category, none were correctly predicted, and all of them were incorrectly assigned to the categories of “Moderately Low,” “Moderate,” and “High.” The same method was used to calculate each of the 12 classifiers' remaining confusion matrices.

5. Conclusion

The primary aim of this study was to ascertain and evaluate the efficacy of various data mining methods in predicting meteorological data. A distinct dataset was obtained from NASA's data access website, encompassing additional variables to forecast the mean temperature at a specific location within Chittagong City, Bangladesh. The dataset utilized in this study was subjected to an analysis of the performance of 12 distinct data mining techniques, each belonging to different approaches such as Trees, Functions, and Rules. The validation approach utilized the 10-fold cross-validation technique. Among the twelve classifiers examined, except OneR, all exhibited strong performance, with an approximate accuracy of 80% in their predictions of the average temperature. The classifiers demonstrated satisfactory results for many performance measure parameters, including ROC area, F-measure, and confusion matrix. The J48 classifier, implemented using the Tree method, demonstrated an accuracy of 82.3%, surpassing the other algorithms that were considered. Additionally, the results of the experiment exhibited a substantial ROC area score and F-measure score of 97.8 and 84.2, respectively. The OneR classifier, when applied using the Rule-based technique, failed to identify any temperature criterion within the ‘Very High’ class. The attained accuracy of the classifier was only 59.92%, in contrast to the other 11 classifiers which exhibited strong performance and indicated high accuracy ranging from 79.1% to 82.3%, in terms of

Table 17
Confusion matrix.

OneR		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**1017**	24	0	0	0	0
	Low	255	**951**	18	0	0	0
	Moderately low	0	379	**727**	178	0	0
	Moderate	0	14	406	**1340**	154	0
	High	0	2	189	771	**561**	0
	Very high	0	0	57	210	416	**0**
Decision Table		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**964**	77	0	0	0	0
	Low	73	**1094**	54	3	0	0
	Moderately low	7	76	**1024**	166	11	0
	Moderate	0	0	158	**1466**	290	0
	High	0	0	2	199	**1218**	104
	Very high	0	0	0	4	172	**507**
JRip		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**971**	70	0	0	0	0
	Low	84	**1079**	51	10	0	0
	Moderately low	0	75	**1029**	175	4	1
	Moderate	0	0	170	**1472**	271	1
	High	0	0	6	279	**1115**	123
	Very high	0	0	0	39	110	**534**
Ridor		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**972**	67	0	1	0	1
	Low	107	**1035**	73	6	2	1
	Moderately low	0	110	**993**	148	18	15
	Moderate	0	2	195	**1415**	280	22
	High	1	3	19	209	**1138**	153
	Very high	3	2	3	11	151	**513**

(*continued on next page*)

Table 17 (*continued*)

OneR		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
J48		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**966**	75	0	0	0	0
	Low	77	**1095**	52	0	0	0
	Moderately low	0	64	**1063**	146	11	0
	Moderate	0	0	163	**1446**	304	1
	High	0	0	5	190	**1207**	121
	Very high	1	0	0	2	153	**527**
LMT		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**971**	70	0	0	0	0
	Low	81	**1090**	52	0	0	1
	Moderately low	1	69	**1083**	120	11	0
	Moderate	0	0	206	**1409**	295	4
	High	0	0	13	199	**1204**	107
	Very high	0	0	0	3	167	**513**
Random Forest		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**982**	59	0	0	0	0
	Low	89	**1075**	57	0	0	0
	Moderately low	0	77	**1032**	175	16	2
	Moderate	0	0	169	**1476**	258	11
	High	0	0	10	258	**1108**	147
	Very high	0	0	0	11	153	**519**
Cart		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**962**	79	0	0	0	0
	Low	73	**1100**	48	0	3	0
	Moderately low	0	74	**1033**	164	13	0
	Moderate	0	0	163	**1454**	297	0
	High	0	1	4	193	**1217**	106
	Very high	0	0	0	0	161	**522**
Logitboost		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**982**	59	0	0	0	0
	Low	82	**1095**	47	0	0	0
	Moderately low	0	75	**1064**	136	9	0
	Moderate	0	0	178	**1444**	289	3
	High	0	0	11	204	**1198**	110
	Very high	0	0	0	5	164	**514**
MLP		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**972**	69	0	0	0	0
	Low	159	**1007**	53	0	5	0
	Moderately low	56	64	**942**	156	56	10
	Moderate	6	0	163	**1468**	246	31
	High	3	0	17	253	**1166**	84
	Very high	1	0	31	1	377	**273**
MLR		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**978**	63	0	0	0	0
	Low	97	**1072**	54	1	0	0
	Moderately low	0	79	**1005**	188	11	1
	Moderate	0	0	164	**1493**	252	5
	High	0	0	9	275	**1106**	133
	Very high	0	0	0	11	162	**510**
SMO		Actual					
		Very low	Low	Moderately low	Moderate	High	Very high
Predicted	Very low	**965**	76	0	0	0	0
	Low	77	**1099**	48	0	0	0
	Moderately low	1	85	**1015**	173	10	0
	Moderate	0	0	158	**1474**	282	0
	High	0	0	2	209	**1215**	97
	Very high	0	0	0	4	169	**510**

percentage. This study examined attributes that have a direct influence on temperature. The discovery made in this study will aid in the identification and selection of features and methodologies for predicting temperature. Nevertheless, this study does have a few shortcomings. This study exclusively utilized data from a limited geographical area within a large urban center. It is suggested that future research endeavors incorporate a larger number of data points sourced from diverse geographical locations to facilitate a more comprehensive empirical study.

Funding

This research was not supported by any organizations.

CRediT authorship contribution statement

Mohammad Sadman Tahsin: Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Shahriar Abdullah:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation. **Musaddiq Al Karim:** Visualization, Validation, Formal analysis, Data curation. **Minhaz Uddin Ahmed:** Writing – original draft. **Faiza Tafannum:** Writing – review & editing, Writing – original draft, Methodology. **Mst Yeasmin Ara:** Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Al-Yahyai, S., Charabi, Y., Gastli, A., 2010. Review of the use of numerical weather prediction (NWP) models for wind energy assessment. Renew. Sustain. Energy Rev. 14, 3192–3198.

Asklany, S.A., Elhelow, K., Youssef, I.K., Abd El-wahab, M., 2011. Rainfall events prediction using rule-based fuzzy inference system. Atmos. Res. 101, 228–236. https://doi.org/10.1016/j.atmosres.2011.02.015.

Ayres-de-Campos, D., Bernardes, J., Garrido, A., et al., 2000. SisPorto 2.0: a program for automated analysis of cardiotocograms. J. Matern. Fetal Med. 9, 311–318.

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525, 47–55. https://doi.org/10.1038/nature14956.

Bihlo, A., 2021. A generative adversarial network approach to (ensemble) weather prediction. Neural Network. 139, 1–16. https://doi.org/10.1016/j.neunet.2021.02.003.

Bohning, D., 1992. Multinomial logistic regression algorithm. Ann. Inst. Stat. Math. 44, 197–200. https://doi.org/10.1007/BF00048682.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Cai, Y.-D., Feng, K.-Y., Lu, W.-C., Chou, K.-C., 2006. Using LogitBoost classifier to predict protein structural classes. J. Theor. Biol. 238, 172–176. https://doi.org/10.1016/j.jtbi.2005.05.034.

Calders, T., Custers, B., 2013. What Is Data Mining and How Does it Work?, pp. 27–42.

Casas, D.M., JÁT, González, Rodríguez, J.E.A., Pet, J.V., 2009. Using Data-Mining for Short-Term Rainfall Forecasting, pp. 487–490.

Chandrayan, S.S., Singh, K., Bhoi, A.K., 2022. Atmospheric Weather Fluctuation Prediction Using Machine Learning, pp. 431–443.

Chen, W., Shahabi, H., Shirzadi, A., et al., 2018. A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. Geocarto Int. 33, 1398–1420. https://doi.org/10.1080/10106049.2018.1425738.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 21, 6. https://doi.org/10.1186/s12864-019-6413-7.

Crawford, S.L., 1989. Extensions to the CART algorithm. Int. J. Man Mach. Stud. 31, 197–217. https://doi.org/10.1016/0020-7373(89)90027-8.

Cufoglu, A., Lohi, M., Madani, K., 2009. A comparative study of selected classifiers with classification accuracy in user profiling. In: 2009 WRI World Congress on Computer Science and Information Engineering. IEEE, pp. 708–712.

Dantas Dias, M.L., Rocha Neto, A.R., 2016. Evolutionary support vector machines: a dual approach. In: 2016 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp. 2185–2192.

Denardini, C.M., Dasso, S., Gonzalez-Esparza, J.A., 2016. Review on space weather in Latin America. 3. Development of space weather forecasting centers. Adv. Space Res. 58, 1960–1967. https://doi.org/10.1016/j.asr.2016.03.011.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Gad, I., Hosahalli, D., 2022. A comparative study of prediction and classification models on NCDC weather data. Int. J. Comput. Appl. 44, 414–425. https://doi.org/10.1080/1206212X.2020.1766769.

Gneiting, T., Raftery, A.E., 2005. Weather forecasting with ensemble methods. Science 310 (80), 248–249. https://doi.org/10.1126/science.1115255.

Hall, M., Frank, E., Holmes, G., et al., 2009. The WEKA data mining software. ACM SIGKDD Explor. Newsl. 11, 10–18. https://doi.org/10.1145/1656274.1656278.

Huang, C.-L., Chen, M.-C., Wang, C.-J., 2007. Credit scoring with a data mining approach based on support vector machines. Expert Syst. Appl. 33, 847–856. https://doi.org/10.1016/j.eswa.2006.07.007.

Huang, X., Shi, L., Suykens, J.A.K., 2015. Sequential minimal optimization for SVM with pinball loss. Neurocomputing 149, 1596–1603. https://doi.org/10.1016/j.neucom.2014.08.033.

Huang, Z.-Q., Chen, Y.-C., Wen, C.-Y., 2020. Real-time weather monitoring and prediction using city buses and machine learning. Sensors 20, 5173. https://doi.org/10.3390/s20185173.

IPCC, 2007. Intergovernmental panel on climate change. World Meteorol. Organ 52.

Kamarudin, M.H., Maple, C., Watson, T., Safa, N.S., 2017. A LogitBoost-based algorithm for detecting known and unknown web attacks. IEEE Access 5, 26190–26200. https://doi.org/10.1109/ACCESS.2017.2766844.

Kandhasamy, J.P., Balamurali, S., 2015. Performance analysis of classifier models to predict diabetes mellitus. Procedia Comput. Sci. 47, 45–51. https://doi.org/10.1016/j.procs.2015.03.182.

KAYA, E., BARUTÇU, B., MENTEŞ, Ş.S., 2013. A method based on the Van der Hoven spectrum for performance evaluation in prediction of wind speed. Turk. J. Earth Sci. https://doi.org/10.3906/yer-1207-1.

Kohavi, R., 1995. The Power of Decision Tables, pp. 174–189.

Kulkarni, A., Chong, D., Batarseh, F.A., 2020. Foundations of data imbalance and solutions for a data democracy. In: Data Democracy. Elsevier, pp. 83–106.

Loh, W., 2011. Classification and regression trees. WIREs Data Min. Knowl. Discov. 1, 14–23. https://doi.org/10.1002/widm.8.

Mahajan, A., Ganpati, A., 2014. Performance evaluation of rule based classification algorithms. Int. J. Adv Res. Comput. Eng. Technol. 3, 3546–3550.

Markovics, D., Mayer, M.J., 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew. Sustain. Energy Rev. 161, 112364.

Meehl, G.A., Washington, W.M., Collins, W.D., et al., 2005. How much more global warming and sea level rise? Science 307 (80), 1769–1772.

NASA POWER prediction of Worldwide energy resources (POWER). https://power.larc.nasa.gov/data-access-viewer/. (Accessed 1 February 2023).

Pabreja, K., Datta, R.K., 2012. A data warehousing and data mining approach for analysis and forecast of cloudburst events using OLAP-based data hypercube. Int. J. Data Anal. Tech. Strat. 4, 57–82.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Rem. Sens. 26, 217–222. https://doi.org/10.1080/01431160412331269698.

Pandey, P.C., Barai, S.V., 1995. Multilayer perceptron in damage detection of bridge structures. Comput. Struct. 54, 597–608. https://doi.org/10.1016/0045-7949(94)00377-F.

Pandey, S.K., Mishra, R.B., Tripathi, A.K., 2020. BPDET: an effective software bug prediction model using deep representation and ensemble learning techniques. Expert Syst. Appl. 144, 113085. https://doi.org/10.1016/j.eswa.2019.113085.

Pérez, I.A., García, Má, Sánchez, M.L., de Torre, B., 2004. Autocorrelation analysis of meteorological data from a RASS sodar. J. Appl. Meteorol. 43, 1213–1223. https://doi.org/10.1175/1520-0450(2004)043<1213:AAOMDF>2.0.CO;2.

Pinkus, A., 1999. Approximation theory of the MLP model in neural networks. Acta Numer. 8, 143–195. https://doi.org/10.1017/S0962492900002919.

Scofield, B., 2010. A History and Test of Planetary Weather Forecasting. University of Massachusetts Amherst.

Sharma, U., Sharma, C., 2022. Deep learning based prediction of weather using Hybrid_stacked Bi-long short term memory. In: 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, pp. 422–427.

Skamarock, W.C., Klemp, J.B., 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. J. Comput. Phys. 227, 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037.

Tahsin, M.S., Karim, M Al, Ahmed, M.U., et al., 2021. Comparative analysis of weather prediction using ensemble learning models and neural network. In: 2021 19th OITS International Conference on Information Technology (OCIT). IEEE, pp. 325–330.

Thepade, S., Das, R., Ghosh, S., 2013. Performance comparison of feature vector extraction techniques in RGB color space using block truncation coding for content based image classification with discrete classifiers. In: 2013 Annual IEEE India Conference (INDICON). IEEE, pp. 1–6.

Umadevi, S., Marseline, K.S.J., 2017. A survey on data mining classification algorithms. In: 2017 International Conference on Signal Processing and Communication (ICSPC). IEEE, pp. 264–268.

Waseem, S., Salman, A., Muhammad, A.K., 2013. Feature subset selection using association rule mining and JRip classifier. Int. J. Phys. Sci. 8, 885–896. https://doi.org/10.5897/IJPS2013.3842.

Witten, I.H., Frank, E., Hall, M.A., et al., 2005. Practical machine learning tools and techniques. In: Data Mining.

Yadav, S., Shukla, S., 2016. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). IEEE, pp. 78–83.

Yang, H., Yan, J., Liu, Y., Song, Z., 2022. Statistical downscaling of numerical weather prediction based on convolutional neural networks. Glob Energy Interconnect 5, 217–225. https://doi.org/10.1016/j.gloei.2022.04.018.