JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EARLY WARNING SYSTEM FOR WEATHER PREDICTION: FOG AND CLOUD COMPATIBLE MODEL

Rekha
Mtech Scholar
Golden College of Engineering and Technology
Gurdaspur

Mohit Trehan
Assistant Professor
Golden College of Engineering and Technology
Gurdaspur

Abstract

Disasters often led to economic, and human loses. Early predictions corresponding to disaster can allow administration to take preventive and precautionary measures Weathers are common and uncertain disasters that can occur due to disturbance in normal slope stability. Weathers often accompany earthquakes, rain, or eruptions. This research proposed an early warning system for Weather. Entire framework associated with proposed system consists of sensor, fog and cloud layer. Data acquisitions employed within sensor layer collects the data about the soil and land through sensors. Furthermore, pre-processing will be performed at sensor layer. Pre-processing mechanism remove any noise from the dataset. Fog layer contains feature reduction mechanism that is used to reduce the size of data to conserve energy of sensors during transmission of data. Furthermore, predictor variables selected within energy conservation mechanism will be used for exploratory data analysis (EDA). Main characteristics of data will be extracted using EDA. Furthermore, principal component analysis applied at fog layer analyses the dependencies between the attributes. Dependencies are calculated using correlation. Negatively skewed attributes will be rejected thus dimensionality of dataset is reduced further. All the gathered prime attributes are stored within cloud layer. K means clustering is applied to group the similar entities within same cluster. This step will reduce the overall execution time of prediction. Formed clusters are fed into ARIMA(Auto regressive integrated moving averages) for predictions. Relevant authorities can fetch the result by logging into the cloud. The effectiveness of proposed approach is proved at different levels using metrics such as classification accuracy and F-score.

Keywords: Fog computing, Weather prediction, energy efficiency, K means clustering, PCA, ARIMA

1. Introduction

Disasters can be of any volume leading to devastating effect on human life, environment, and economic conditions of the country. Disasters can be categorized either as natural or generated through activities performed by humans. To this end, dedicated effort by researchers yielding mechanisms and models for early detection and prediction corresponding to Weathers. (Thein et al. 2020) conducted a survey of Weathers in Myanmar. Real time monitoring, and early warning systems was developed using machine learning based approach. The prediction was based upon the parameters like moisture levels within soil and slope. (Juyal and Sharma 2021) discussed a Weather susceptibility using machine learning approach. The predictor variables used for detection includes moisture levels only. Classification accuracy through this approach was less. (Hartomo, Yulianto, and Maruf 2017) proposed exponential smoothing method using google API for the early prediction of Weathers. Applications of fog computing was rarely used to store the information regarding Weathers and generating appropriate warnings for the relevant authorities(Sun et al. 2015). This work proposed a fog-based model for early detection and prediction of Weathers ensuring least loss on terms of financial as well as human resources(Ayalew, Yamagishi, and Ugawa 2004).

The proposed work is portioned into multiple layers. In the first layer noise handling mechanisms are applied to handle the missing values and outliers. The normalized data will be fed into the second layer(Rau, Jhan, and Rau 2013). The second layer contains mechanism for reducing the size of extracted features. EDA will be applied at this layer for exploratory analysis. The cloud layer will be used to store the result produced through fog layer(Komac 2006).

Rest of the paper is organized as under. Section 1 presented the analysis of mechanisms used for prediction of Weathers along with definition of proposed mechanism. The section 2 gives indepth analysis of existing mechanisms used for prediction of Weathers at early stage. The datasets used are also explored through this section. Section 3 gives the methodology of the proposed work along with explanation of each phase. Section 4 gives the performance analysis and result section. Last section gives the conclusion and future scope.

2. Literature Survey

This section puts a light on different techniques used for the detection and prediction of Weathers at early stage. (Dai et al. 2021) proposed ensemble-based approach for the prediction of Weathers. The ensembles-based approach uses KNN, random forest, SVM and decision tree for the prediction process. The overall process detects the maximum true positive values predicted through classifiers. The highest prediction becomes result. The classification accuracy through this approach was in the range of 90s. real time dataset was employed for the detection and prediction process. (Azmoon et al. 2021) proposed image-based slope stability analysis using deep learning mechanism. The layered based approach works on real time dataset. The prediction of Weathers depends greatly upon clarity of the extracted image. The result was presented in the form of prediction accuracy. (Amit and Aoki 2017)proposed disaster detection using aerial images. Spatial mechanism employed to tackle the noise from the images. The boundary value analysis detects image boundary accurately and rest of the image segment was eliminated. Result of the proposed approach was expressed in the form of classification accuracy. (Jana and Singh 2022) discuss the impact of climate and environment on natural disasters in various countries. Official datasets available on the government websites were explored for this purpose. (Sarwar and Muhibbullah 2022)proposed mechanism to explore the issue of Weathers within the Chittagong. The real time dataset corresponding to Hill region of Bangladesh was presented in this analysis.(Marjanović et al. 2011) discussed the Weather susceptibility detection and prediction using support vector machine. Only two hyperplanes were used in this case. The prediction was oriented towards Weather detected or not detected. Classification accuracy through this approach was poor due to high degree of misclassification. (Lee 2005) discussed the applications of logistic regression in the detection and prediction of Weather. The prediction model used real time dataset and high degree of misclassification causes this model to perform adversely incase of large dataset collection. (Lee 2007)proposed fuzzy based model for the early detection of Weathers using benchmark dataset derived from Kaggle. The result of the system was expressed in the form of classification accuracy.

The suggested literature indicates that dataset used in most of the existing models was real time. Fog computing was rarely implemented in the existing models. To overcome the issue, proposed system implements fog-based model for the early detection and prediction of Weathers. Next section discussed the methodology corresponding to the proposed work.

3. Methodology of proposed work

The methodology of proposed work starts from dataset acquisition. The dataset was collected corresponding to state of Jammu and Kashmir. The structure of the dataset is presented in table1.

Table 1: Dataset description

Field	Description
Event_Date	Date at which Weather occurred
Category	Indicates types of disaster
Weather_trigger	Cause of Weather
Size	Indicates size of destruction

Setting	Indicates location of the event
Latitude	Indicates latitude of location
Longitude	Indicates longitude of location
Dew/Frost point at 2mtrs	Indicates amount of water vapors' presents within the air.
Earth skin temperature	Indicating temperature of the earth
Temperature 2mtrs range	Water vapors temperature
Specific humidity	Humidity present within the air.
Relative humidity	Relative humidity of environment
Precipitation	Amount of Precipitation due to temperature
Surface pressure	Pressure on the surface where event occurred
Wind speed	Wind speed during the event
Surface soil wetness	Wetness could be critical for Weathers
Root zone soil wetness	Zone at which disaster occurred
Profile soil moisture	Indicates the soil moisture that is compared against the threshold

The data acquisition layer will receive this dataset and perform initial analysis. The details of the used layers is given as under

- Data Acquisition layer

This layer is critical in the operation of the fog based Weather prediction model. This layer receives the dataset and removes the noise if any from the dataset. The noise in terms of missing and unnamed values will be tackled through replacement with '0'(Rosi et al. 2018). The outliers indicating extreme value will be tackled by the use of box plot method. The values lying inside the box plot will be retained and rest of the values will be outliers. These outliers will be handled using the median values. The pre-processed dataset will be fed into the fog layer(Ercanoglu and Gokceoglu 2002).

- Fog Layer

The primary purpose of this layer is to conserve energy of the sensors(Ermini, Catani, and Casagli 2005). This is possible only if dimensionality reduction mechanism is in place. For dimensionally reduction principal component analysis is used. Exploratory data analysis is used for determining the highest correlated values. These highest correlated values will be used as a predictor variable. The fog layer thus has two tasks, first task is associated with dimensionality reduction and then identifying predictor variables with EDA(Catani et al. 2013).

- Cloud layer

Cloud layer stores the generated predictions. To generate the predictions, first we have applied KNN clustering and after that ARIMA model is applied for forecasting. The forecasted result will be accessed with the help of accounts within the cloud.(Althuwaynee, Pradhan, and Lee 2012) The early prediction can help the governments to initiates the preventive steps to save from financial and human loses.

The algorithm corresponding to KNN clustering is given as under

KNN_Clustering

- Receives the dataset with the predictor variables.
- Set the value of K=P where K is the distance metric and P is the static values corresponding to the distance
- Repeat the following steps until all the values within dataset is checked for inclusion within cluster
  - If (distance <K)
    Include within cluster
    End of if
  - Move to next value within dataset
- End of loop
- Return Clusters

The clustering mechanism will give the groups corresponding to parameters possessing similar nature. Clustering will cause faster result propagation. The KNN clustering preferred over K means clustering primarily due to random values of k in centroid prediction. The result of K means and KNN clustering is presented in this section

Parameters	KNN	Kmeans
Optimal Clustering	4 indicating four separate locations with similar characteristics	2 indicating two different locations with similar characteristics
Convergence Rate	10 out of 10 simulation	8 out of 10 simulation
Execution speed	Fast with presented dataset	Slow as compared to KNN as size of dataset increased

Table 2: KNN vs Kmeans

The validation process revealed that kmeans clustering results are close but KNN results are more accurate. Hence K means clustering approach is used in the proposed work.

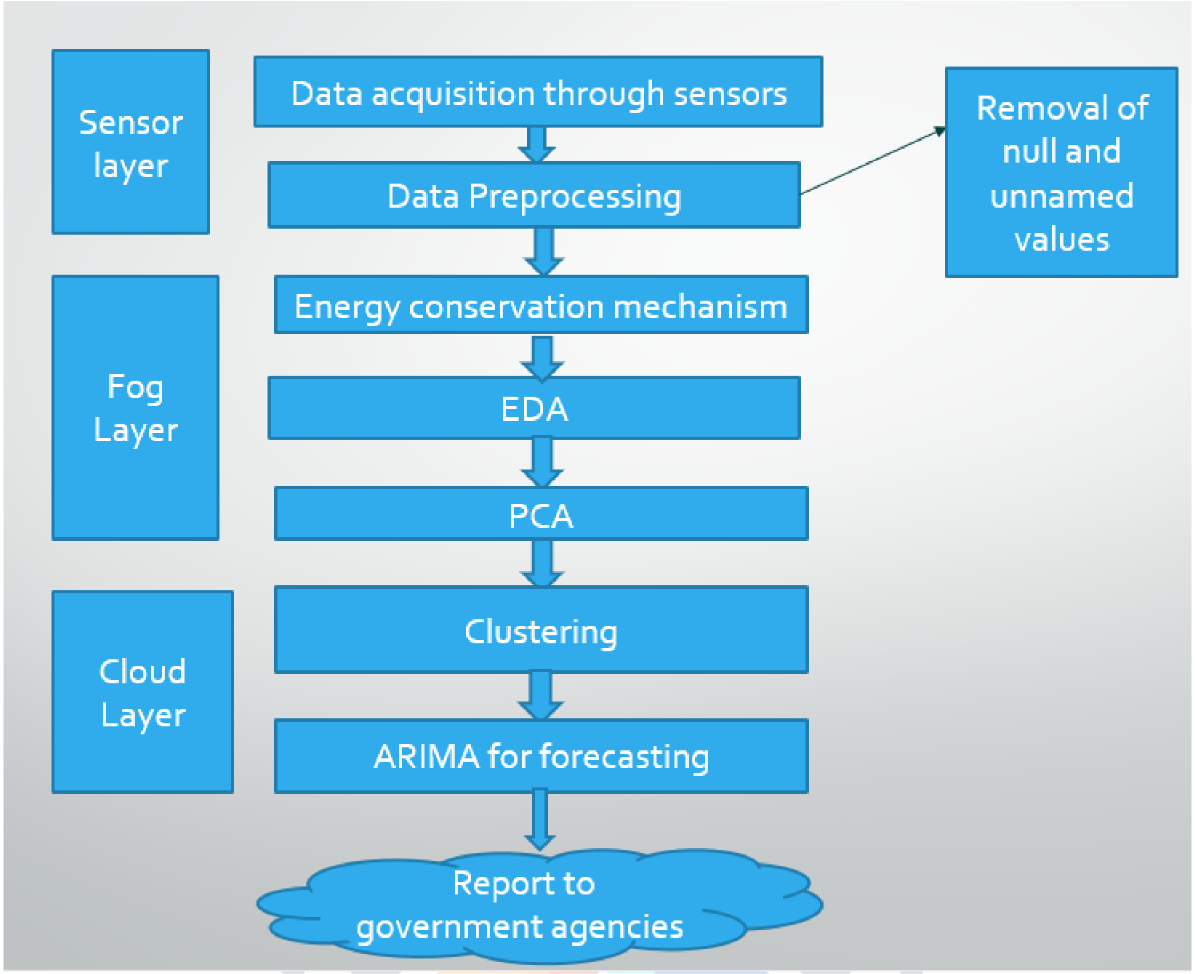
The obtained clusters will be fed within ARIMA model to generate the predictions corresponding to Weather.

ARIMA_Prediction(Clusters)

- Stores clusters
- Repeat the following steps corresponding to test datasets for predictions
  - Perform regression analysis.
  - Perform integration by obtaining difference with raw observations to make the time series to become stationery.
  - Calculate moving averages by evaluating the error by subtracting observations from the actual values.
  - Generate prediction
- End of loop

The flow of the proposed model is given in figure 1

Figure 1: Flow of the proposed model



4. Performance Analysis and results

The result using improved Weather prediction system using differential approach is given within this section. All the four classes are predicted using the proposed mechanism. The result in terms of classification accuracy is elaborated first. Classification accuracy is obtained using the equation 1

$$CLasification_{Acc} = \frac{TrueP + TrueN}{TrueP + TrueN + FalseP + FalseN}$$

Equation 1

TrueP indicates true positive values and TrueN indicates true negative values. FalseP indicates false positive values and FalseN indicates false Negative values.

Dataset Size	Classification Accuracy(%) using Weather prediction without ARIMA	Classification Accuracy(%) using Weather prediction with ARIMA
1000	85	95
2000	83	94.2
3000	82	94
4000	79	93.5
5000	78	93

Table 3: Classification accuracy result with varying dataset size

The train dataset values are normalized between 0 and 1 to reduce the complexity of operation. The visualization corresponding to the classification accuracy differ from the existing work without ARIMA by 5-6% that is significant and proves worth of study.

The visualization result corresponding to traffic prediction is given within figure 2

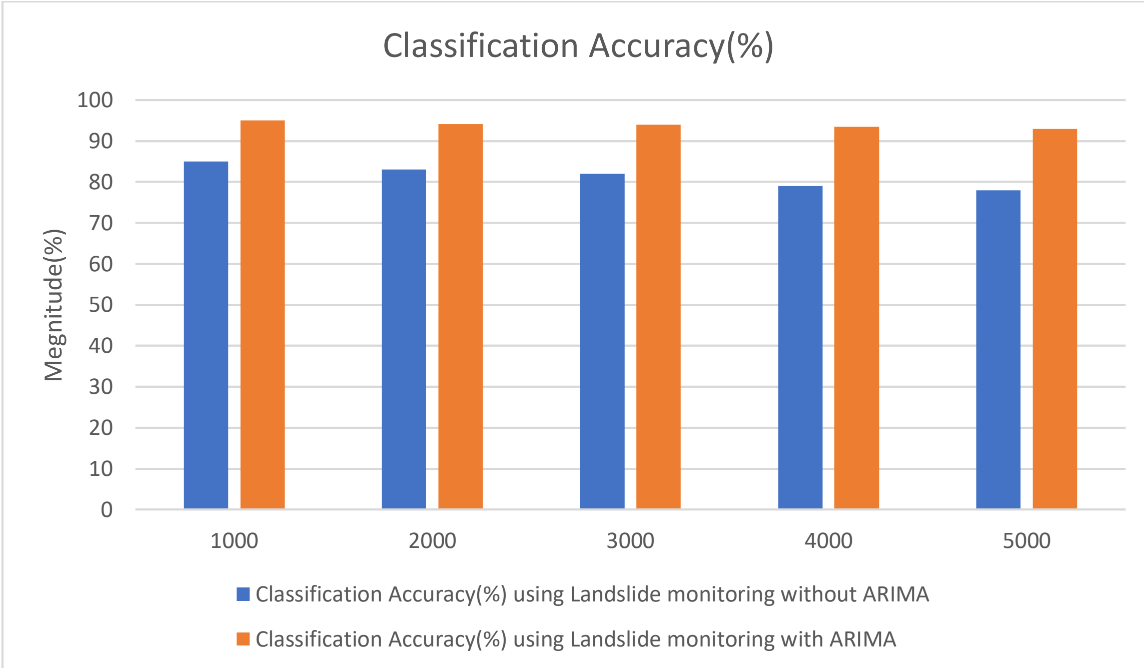


Figure 2: Visualization result corresponding to classification accuracy

The result in terms of sensitivity is considered next. This metric indicates the percentage of correctly classified instances positively into any class. The sensitivity result is given through following equation

$$Sensitivity = \frac{TrueP}{TrueP + FalseN}$$

Equation 2

The result of sensitivity is given in table 4

Dataset Size	Sensitivity(%) using Weather prediction without ARIMA	Sensitivity(%) using Weather prediction with ARIMA
1000	72	75
2000	70	73.6
3000	65	73.2
4000	64	72
5000	63	71

Table 4: Result of sensitivity

The visualization result corresponding to sensitivity is given within figure 3

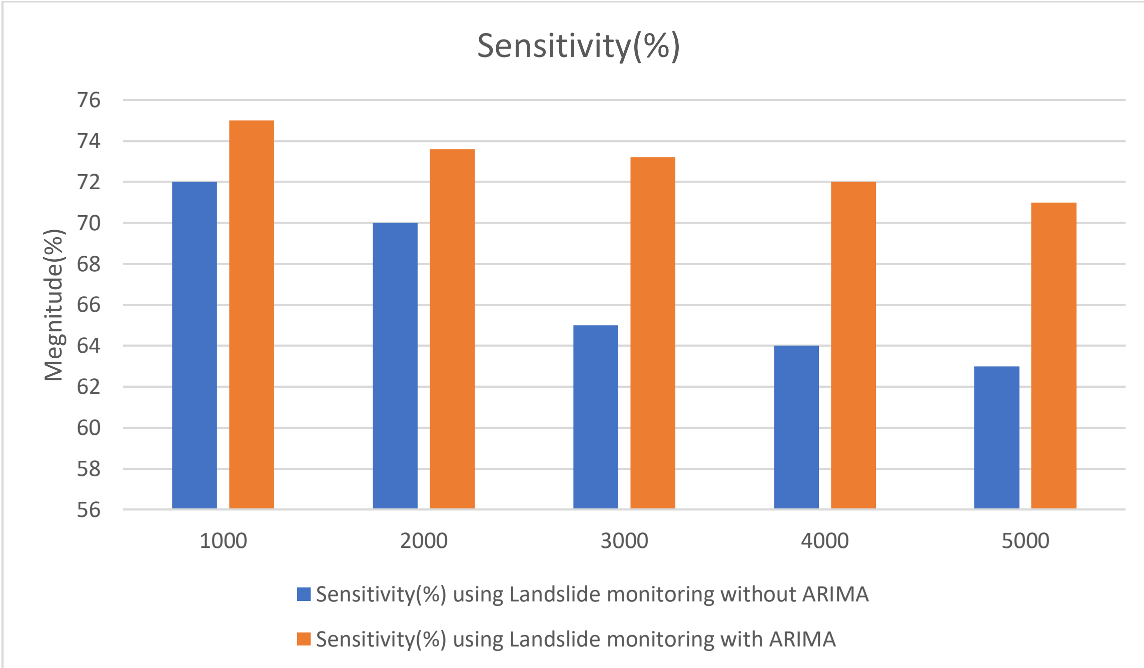


Figure 3: Sensitivity by varying dataset size

The last result is in the form of specificity which is the result in terms of correctly negatively classified instances from the dataset. The specificity is given through equation 3

$$Specificity = \frac{TrueN}{TrueN + FalseP}$$

Equation 3

The result corresponding to specificity is given by table 5

Dataset Size	Specificity%) using Weather prediction without ARIMA	Specificity(%) using Weather prediction with ARIMA
1000	28	25
2000	30	27
3000	35	27
4000	36	28
5000	37	29

Table 5: Result of specificity

The visualization result is given in the figure 4

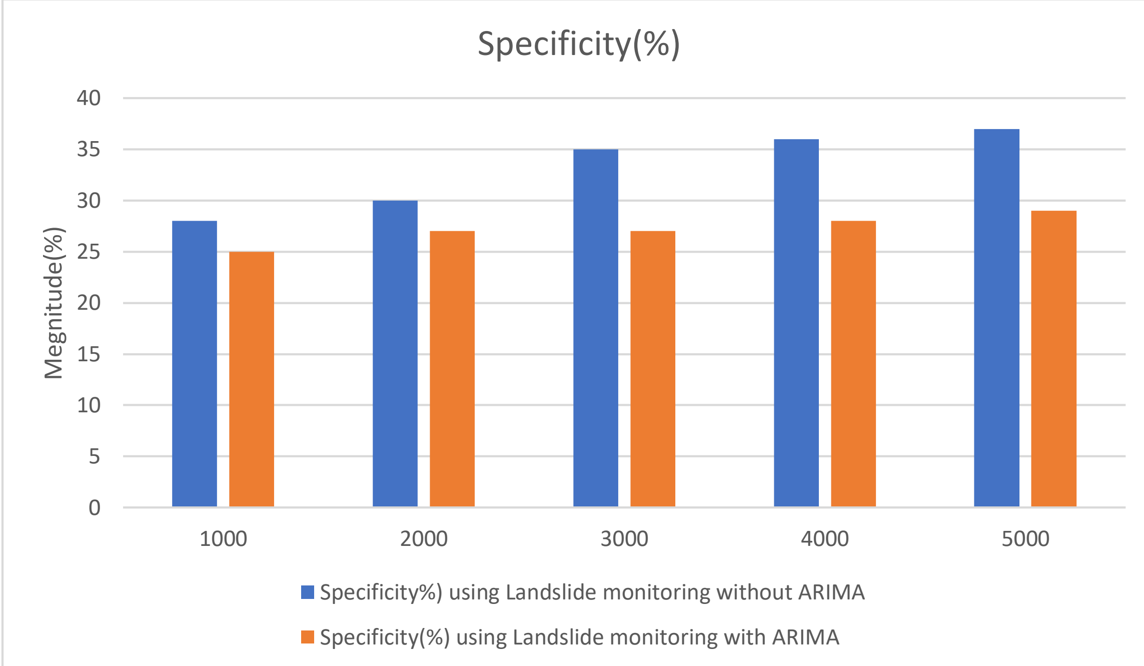


Figure 4: Specificity results visualization

5. Conclusion

This paper presented the fog-based model for the prediction of Weathers. The dataset for Weather prediction was derived from the benchmark website. The dataset pre-processing mechanism within acquisition layer will handle all the abnormality and finalized classification accuracy that is stored within the cloud layer. The data acquisition layer result is fed into the fog layer. The fog layer contains the mechanism of energy conservation that is achieved through reduction mechanism through principal component analysis. Exploratory data analysis mechanism reduce the size based upon correlation calculated through PCA. The obtained result of the fog layer will be entered within the cloud layer. The result from the cloud layer can be extracted by the administrators having account within cloud layer. The result of the classification accuracy in the range of 95% that better by almost 7% from existing model proves the worth of study.

References

1. **Althuwaynee, Omar F., Biswajeet Pradhan, and Saro Lee. 2012.** "Application of an Evidential Belief Function Model in Weather Susceptibility Mapping." *Comput Geosci* 44 (July): 120–135. doi:10.1016/j.cageo.2012.03.003.
2. **Amit, Siti Nor Khuzaimah Binti, and Yoshimitsu Aoki. 2017.** "Disaster Detection from Aerial Imagery with Convolutional Neural Network." *Proceedings - International Electronics Symposium on Knowledge Creation and Intelligent Computing, IES-KCIC 2017* 2017-January (December). Institute of Electrical and Electronics Engineers Inc.: 239–245. doi:10.1109/KCIC.2017.8228593.
3. **Ayalew, Lulseged, Hiromitsu Yamagishi, and Norimitsu Ugawa. 2004.** "Weather Susceptibility Mapping Using GIS-Based Weighted Linear Combination, the Case in Tsugawa Area of Agano River, Niigata Prefecture, Japan." *Weathers* 1 (1). Springer Verlag: 73–81. doi:10.1007/s10346-003-0006-9.
4. **Azmoon, Behnam, Aynaz Biniyaz, Zhen Liu, and Ye Sun. 2021.** "Image-Data-Driven Slope Stability Analysis for Preventing Weathers Using Deep Learning." *IEEE Access* 9. Institute of Electrical and Electronics Engineers Inc.: 150623–150636. doi:10.1109/ACCESS.2021.3123501.
5. **Catani, F., D. Lagomarsino, S. Segoni, and V. Tofani. 2013.** "Weather Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues." *Nat Hazards Earth Syst Sci* 13 (11): 2815–2831. doi:10.5194/nhess-13-2815-2013.
6. **Dai, Leiyu, Mingcang Zhu, Zhanyong He, Yong He, Zezhong Zheng, Guoqing Zhou, Chao Wang, et al. 2021.** "WEATHER RISK CLASSIFICATION BASED ON ENSEMBLE MACHINE LEARNING." *International Geoscience and Remote Sensing Symposium (IGARSS)* 2021-July. Institute of Electrical and Electronics Engineers Inc.: 3924–3927. doi:10.1109/IGARSS47720.2021.9553034.

7. **Ercanoglu, Murat, and Candan Gokceoglu. 2002.** "Assessment of Weather Susceptibility for a Weather-Prone Area (North of Yenice, NW Turkey) by Fuzzy Approach." *Environ Geol* 41 (6): 720–730. doi:10.1007/s00254-001-0454-2.
8. **Ermini, Leonardo, Filippo Catani, and Nicola Casagli. 2005.** "Artificial Neural Networks Applied to Weather Susceptibility Assessment." *Geomorphology* 66 (1-4 SPEC. ISS.): 327–343. doi:10.1016/j.geomorph.2004.09.025.
9. **Hartomo, Kristoko Dwi, Sri Yulianto, and Joko Maruf. 2017.** "Spatial Model Design of Weather Vulnerability Early Detection with Exponential Smoothing Method Using Google API." *Proceedings - 2017 International Conference on Soft Computing, Intelligent System and Information Technology: Building Intelligence Through IOT and Big Data, ICSIIT 2017* 2018-January (July). Institute of Electrical and Electronics Engineers Inc.: 102–106. doi:10.1109/ICSIIT.2017.37.
10. **Jana, N. C. (Narayan Chandra), and R. B. Singh. 2022.** "Climate, Environment and Disaster in Developing Countries," 536. Accessed May 18.
11. **Juyal, Amit, and Sachin Sharma. 2021.** "A Study of Weather Susceptibility Mapping Using Machine Learning Approach." *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, February. Institute of Electrical and Electronics Engineers Inc., 1523–1528. doi:10.1109/ICICV50876.2021.9388379.
12. **Komac, Marko. 2006.** "A Weather Susceptibility Model Using the Analytical Hierarchy Process Method and Multivariate Statistics in Perialpine Slovenia." *Geomorphology* 74 (1–4): 17–28. doi:10.1016/j.geomorph.2005.07.005.
13. **Lee, Saro. 2005.** "Application of Logistic Regression Model and Its Validation for Weather Susceptibility Mapping Using GIS and Remote Sensing Data." *Int J Remote Sens* 26 (7). Taylor and Francis Ltd.: 1477–1491. doi:10.1080/01431160412331331012.
14. **Lee, Saro. 2007.** "Application and Verification of Fuzzy Algebraic Operators to Weather Susceptibility Mapping." *Environ Geol* 52 (4): 615–623. doi:10.1007/s00254-006-0491-y.
15. **Marjanović, Miloš, Miloš Kovačević, Branislav Bajat, and Vít Voženílek. 2011.** "Weather Susceptibility Assessment Using SVM Machine Learning Algorithm." *Eng Geol* 123 (3): 225–234. doi:10.1016/j.enggeo.2011.09.006.
16. **Rau, Jiann Yeou, Jyun Ping Jhan, and Ruey Juin Rau. 2013.** "Semiautomatic Object-Oriented Weather Recognition Scheme from Multisensor Optical Imagery and Dem." *IEEE Trans Geosci Remote Sens* 52 (2): 1336–1349. doi:10.1109/tgrs.2013.2250293.
17. **Rosi, A., V. Tofani, L. Tanteri, C. Tacconi Stefanelli, A. Agostini, F. Catani, and N. Casagli. 2018.** "The New Weather Inventory of Tuscany (Italy) Updated with PS-InSAR: Geomorphological Features and Weather Distribution." *Weathers* 15 (1). Springer Verlag: 5–19. doi:10.1007/s10346-017-0861-4.
18. **Sarwar, Md. Iqbal, and Muhammad Muhibbullah. 2022.** "Vulnerability and Exposures to Weathers in the Chittagong Hill Region, Bangladesh: A Case Study of Rangamati Town for Building Resilience." Springer, Singapore, 391–399. doi:10.1007/978-981-16-6966-8_21.
19. **Sun, Q., L. Zhang, X. L. Ding, J. Hu, Z. W. Li, and J. J. Zhu. 2015.** "Slope Deformation Prior to Zhouqu, China Weather from InSAR Time Series Analysis." *Remote Sens Environ* 156 (January). Elsevier Inc.: 45–57. doi:10.1016/j.rse.2014.09.029.
20. **Thein, Thin Lai Lai, Myint Myint Sein, Ken T. Murata, and K. Tungpimolrut. 2020.** "Real-Time Monitoring and Early Warning System for Weather Preventing in Myanmar." *2020 IEEE 9th Global Conference on Consumer Electronics, GCCE 2020*, October. Institute of Electrical and Electronics Engineers Inc., 303–304. doi:10.1109/GCCE50665.2020.9291809.