A Detecting Technique for the Climatic Factors that Aided the Spread of COVID-19 using Deep and Machine Learning Algorithms

Waad Al-Sharari[†], Mahmood A. Mahmood[†,††,*], A. A. Abd El-Aziz[†,††], and Nesrine A. Azim[††]

[†] Dept. of Information Systems, College of Computer and Information Sciences, Jouf University, KSA.
[††] Dept. of Information Systems and Technology, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt
[*]Corresponding Author: Mahmood A. Mahmood: mahmoodissr@cu.edu.eg

Abstract
Novel Coronavirus (COVID-19) is viewed as one of the main general wellbeing theaters on the worldwide level all over the planet. Because of the abrupt idea of the flare-up and the irresistible force of the infection, it causes individuals tension, melancholy, and other pressure responses. The avoidance and control of the novel Covid pneumonia have moved into an imperative stage. It is fundamental to early foresee and figure of infection episode during this troublesome opportunity to control of its grimness and mortality. The entire world is investing unimaginable amounts of energy to fight against the spread of this lethal infection. In this paper, we utilized machine learning and deep learning techniques for analyzing what is going on utilizing countries shared information and for detecting the climate factors that effect on spreading Covid-19, such as humidity, sunny hours, temperature and wind speed for understanding its regular dramatic way of behaving alongside the forecast of future reachability of the COVID-2019 around the world. We utilized data collected and produced by Kaggle and the Johns Hopkins Center for Systems Science. The dataset has 25 attributes and 9566 objects. Our Experiment consists of two phases. In phase one, we preprocessed dataset for DL model and features were decreased to four features humidity, sunny hours, temperature and wind speed by utilized the Pearson Correlation Coefficient technique (correlation attributes feature selection). In phase two, we utilized the traditional famous six machine learning techniques for numerical datasets, and Dense Net deep learning model to predict and detect the climatic factor that aide to disease outbreak. We validated the model by using confusion matrix (CM) and measured the performance by four different metrics: accuracy, f-measure, recall, and precision.

Keywords:
Coronavirus; COVID-19; Machine learning; Deep learning, Climate factors changes.

1. Introduction

Covid's are a huge group of infections that can make extreme ailment the individual. The first realized serious scourge is severe acute respiratory syndrome (SARS) happened in 2003, while the second flare-up of extreme disease started in 2012 in Saudi Arabia with the Middle East Respiratory Syndrome (MERS). The ongoing flare-up of ailment due to Covid is accounted for in late December 2019. This new infection is exceptionally infectious and has rapidly spread around the world. On January 30, 2020, the World Health Organization (WHO) pronounced this episode a Public Health Emergency of International Concern (PHEIC) as it had spread to 18 nations. On Feb 11, 2020, WHO named this COVID-19. On March 11, as the quantity of COVID-19 cases has expanded multiple times separated from China within excess of 118,000 cases in 114 nations and north of 4,000 passings, WHO proclaimed this a pandemic [1] [2].

WHO has adopted a three-pronged strategy to upgrade the limit of the analysis of this illness and prescribed activities to manage COVID-19. The infection is quickly spreading among nations from one side of the planet to the other, dangerous, quick, and enraged. Covid-19 has become worldwide pandemic in under four months since it was first revealed, coming to a 3.3 million affirmed cases and 238,000 passing as of May Second, 2020. Because of its exceptionally infectious nature and absence of fitting treatment and immunizations, early detection of COVID-19 turns out to be progressively vital to forestall further spreading and to level everything out for legitimate portion of restricted clinical assets [2]. Therefore, a couple of analysts have been started to focus on detection of Covid-19 by utilizing computerized reasoning techniques to estimate, foresee and adjust learning for the new Covid, and others demonstrated the advantage of man-made consciousness in pandemics and plagues forecasts [3].

The role of occasional and geographic environment varieties in tweaking the transmission of the infection has gotten expanding consideration. Concentrates on utilizing a relapse structure play tracked down a part for temperature and relative and explicit moistness in the transmission of SARS-CoV-2, recommending that cold and dry circumstances increment the transmission of the infection [4].

Environment influences the transmission of a few straightforwardly communicated microbes. Explicit humidity has been demonstrated to be significant for flu transmission in both research center settings and populace level studies [4].

Manuscript received June 5, 2022
Manuscript revised June 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.6.20

The predecessor SARS-CoV-1 quickly lost reasonability at higher temperatures and higher humidity [5]. The in vitro solidness of SARS-CoV-2 tests has shown that the infection is exceptionally steady at 4◦C yet is touchy to hotness and SARS-CoV-2 loses infectivity at typical center internal heat level (37◦C) [6]. Nonetheless, little decreases at temperatures near 37◦C may considerably build its viral soundness [7]. Many researches have proposed a relationship between weather factors and the COVID-19 pandemic likewise to other viral contaminations like flu. Nonetheless, a few different research have revealed inconsistent outcomes showing that meteorological circumstances may not as a matter of fact be related with the COVID-19 extension. A portion of these research have considered just meteorological elements and others have included other significant factors, for example, populace density, which has been demonstrated to be vital in viral transmissions [8].

As the flare-up of the COVID-19 has turned into an overall pandemic, the ongoing investigations of epidemiological information are expected to set up the public.

In this paper, we used machine learning techniques and deep learning techniques for analyzing what is going on utilizing countries shared information and for detecting the climate factors that effect on spreading Covid-19, such as humidity, sunny hours, temperature and wind speed for understanding its regular dramatic way of behaving alongside the forecast of future reachability of the COVID-2019 around the world. We utilized data collected and produced by Kaggle and the Johns Hopkins Center for Systems Science. The dataset has 25 attributes and 9566 objects. Our Experiment consists of two phases. In phase one, we preprocessed dataset for DL model and features were decreased to four features humidity, sunny hours, temperature, and wind speed by utilized the Pearson Correlation Coefficient technique. In phase two, we utilized the traditional famous eight machine learning techniques with 10-fold cross validation for numerical datasets, and Dense Net deep learning model to predict and detect the climatic factor that aide to disease outbreak.

The utilized machine learning techniques are support vector machine (SVM) for regression, logistic regression (LR), K-Nearest neighbor (KNN), Decision tree (J48), random forest (RF), and Naïve Bayes (NB). The used deep learning techniques are Dense Convolution Network (DenseNet). The main contributions of this work are as follows:

- We proposed a detecting technique for the climatic factors that aided the spread of COVID-19 using DL and ML algorithms.
- The effectiveness of the proposed technique is verified utilizing a recent publicly available dataset.
- We compared the performance of the proposed technique, utilizing four different measures, with similar techniques. The experimental results describe the practicality and superiority of the proposed technique.

The reminder of this paper is formulated as follows. Section 2 reviews related works. Section 3 explains the detailed phases and the utilized techniques in the proposed technique. Section 4 presents the proposed approach. Section 5 shows the results and discussion. Finally, Section 6 presents the conclusion of our work and highlights our future research directions.

2. Related Work

In [10], authors concentrated on the relationship of normal encompassing temperature with ensuing COVID-19 mortality in the Organization for Economic Cooperation and Development (OECD) nations and the individual United States (US), while representing other significant meteorological and non-meteorological co-variates. The openness of interest was normal temperature and other atmospheric conditions, estimated at 25 days earlier and 25 days after the first detailed COVID-19 passing was gathered in the OECD nations and US states. The result of interest was combined COVID-19 mortality, surveyed for every area at 25, 30, 35, and 40 days after the primary revealed passing. Examinations were performed with negative binomial relapse and adapted to other weather patterns, particulate matter, sociodemographic factors, smoking, weight, ICU beds, and social removing. The outcomes were hearty for COVID-19 mortality at 25, 35 and 40 days after the principal demise, as well as other responsiveness investigations. The outcomes give steady proof across different models of an opposite relationship between higher normal temperatures and resulting COVID-19 death rates after representing other meteorological factors and indicators of SARS-CoV-2 disease or demise. This proposes possibly diminished viral transmission in hotter locales and throughout the mid-year season.

In [11], authors concluded that weather conditions, like air temperature and relative mugginess, were not connected with the COVID-19 frequency during the principal wave in the icy and subarctic winter and spring. The surmising depends on a somewhat modest number of cases and a confined time span.

In [12], authors utilized the non-direct least squares strategy to appraise the effect of temperature on COVID-19 cases for every million out of 43 nations, isolated into three groups as follows: the primary group is made out of thirteen nations that declared the principal COVID-19 cases in January 2020, while the second and third groups contain thirteen and seventeen nations, individually, that saw the pandemic without precedent for February and March of that very year. The study showed that a backwards connection in the three groups of nations under study, in all the four-time period since the primary case was accounted for.

In [13], authors found that one element that might add to more quick popular development in the upper aviation routes is the outstanding expansion in SARS-CoV-2 solidness that happens with decreases in temperature, as estimated in vitro. Since SARS-CoV-2 habitually starts contamination in the upper aviation routes prior to spreading through the body, expanded upper aviation route viral development from the get-go in the infection course might bring about more fast movement of sickness and possibly add to additional extreme results. Also, higher SARS-CoV-2 viral titer in the upper aviation routes probably upholds more effective transmission. Alternately, the conceivable meaning of air temperature to upper aviation route viral development recommends that drawn out conveyance of warmed air could address a protection measure and prophylactic treatment for Covid illness 2019.

In [14], authors utilized COVID-19 cases reported until February 29th, 2020, and showed that higher normal temperature was unequivocally connected with lower COVID-19 frequency for temperatures of 1°C and higher. Nonetheless, temperature made sense of a moderately humble measure of the all-out variety in COVID-19 occurrence. These starter discoveries support rigid control endeavors in Europe and somewhere else.

In [15], authors concluded that what the investigation discovered, to specific degree, temperature could massive change COVID-19 transmission, and there may be a best temperature for the viral transmission, which may mostly make sense of why it originally broke out in Wuhan. It was proposed that nations and districts with a lower temperature take on the strictest control measures to forestall future inversion.

In [16], authors showed that COVID-19 favors cool and dry circumstances and was to a great extent missing under incredibly cold and extremely hot and wet circumstances. This illuminate deciding for the timing and extent of the probable public intercessions to alleviate the antagonistic outcomes of the Covid on general wellbeing.

In [17], authors showed that climate factors showed a little impact on Covid transmission, and no connection can be removed between the effect of climate and affirmed cases in all areas. In certain regions, temperature showed a positive connection according to affirmed cases and dampness exhibited a negative relationship. In different regions, no relationship was found.

In [18], authors showed that warm and dry weather conditions was great for the endurance of the infection with a temperature scope of 13-24˚C, a stickiness scope of 50-80%, a precipitation of 30 mm/month or less. Cold air for over seven days affects SARS-CoV-2.

In [19], Considering the ongoing information on the spread of COVID-19, the authors, theorized that the lower number of cases in tropical nations may be because of warm sticky circumstances, under which the spread of COVID-19 may be slower than had been noticed for other viruses.

In [20], authors showed that outright humidity and temperature were related with nearby outstanding development of COVID-19 across areas in China and other impacted nations. Outright humidity and temperature yielded a positive relationship and a slight negative relationship individually. Changes in weather conditions alone won't be guaranteed to prompt decreases if counts without the execution of broad general wellbeing intercessions.

In [21], temperature showed a negative relationship, demonstrating that higher temperatures seemed to have lower Coronavirus transmission. Outright humidity showed a negative relationship, demonstrating that areas with higher outright humidity experienced lower transmission. Changes in climate alone won't be guaranteed to prompt decreases if count without the execution of broad general wellbeing mediations. Table 1 in the appendix presents a comparison between the state-of-the-art recent existing systems:

From the review of the related work, we found that no one of them considered the climate factors, humidity, sunny hours, temperature, and wind speed, all together. Therefore, in our research, we utilized machine learning and deep learning techniques for analyzing what is going on utilizing countries shared information and for detecting the climate factors that effect on spreading Covid-19, such as humidity, sunny hours, temperature, and wind speed for understanding its regular dramatic way of behaving alongside the forecast of future reachability of the COVID-2019 around the world.

3. Material and Methods

This section includes browsing the most relevant topics which are machine learning techniques, deep learning model and benchmarks datasets. The experiment work was performed in two phases to investigate and detect the climatic factors that aided to spread the COVID-19.

3.1 Support vector machine (SVM)

SVM gives more models that can go past linear decision limits. SVM can be utilized for both classification and regression. SVM takes the inputs and changes it to another higher dimensional space, where it turns out to be a lot simpler to predict the change to data utilizing a linear regressor. This thought of changing the inputs to another component space where a linear regressor can be effectively applied is an exceptionally broad and strong one. There are loads of various potential changes that could be applied to data and the various kernels accessible for the kernelized SVM relate to various changes, for example, radial basis function kernel (RBF) and polynomial Kernel. Table1 presents the advantages and disadvantages of SVM [22].

Table 1: The advantages and disadvantages of SVM

Advantages	Disadvantages
1- work on a range of datasets.	1-Efficiency decreases as the size of the dataset increases.
2-Various kernel functions can be utilized.	2- Need normalization for inputs data.
3-Perform well with high and low dimensional data.	3- Difficult to explain its predictions.

3.2 K-Nearest Neighbors Classification (KNN)

KNN can be utilized for classification and regression. KNN classifiers are an illustration of memory-based or instance-based supervising learning. It implies that instance-based learning strategies retain the labels found in the training dataset and afterward use those learned guides to classify new items.

The K in KNN represents the number of the nearest neighbors that the classifier will acquire and use to make a classification. Whenever K is small, such as k=1, the classifier makes a respectable showing of learning the classes for individual places in the dataset, however with a decision limit, the dataset becomes divided and variable. This is on the grounds that when K = 1, individual data focuses are more powerless to outliers, mislabeled data and noise, and different wellsprings of unpredictability. The regions allocated to classes get smoother and less divided as K increments, and they become stronger to commotion in individual focuses. Subsequently, the worth of K affects the classifier's precision. The KNN technique, specifically, involves three stages that can be determined [22].

3.3 Logistic Regression

Logistic regression is the least difficult and most generally involved statistical method for prescient modelling. It is a method for modelling the relationship between two inputs and outputs variables. It essentially gives us a logistic regression function that can be utilized to make expectations about information, where we have our attributes as independent variables, on which our objective variable is reliant upon. The case of one input variable is called simple logistic regression; for more than one input variables, the process is called multiple logistic regression [23] [24].

3.4 Random Forest

A generally utilized and viable strategy in AI includes making learning models known as ensembles. An ensemble takes numerous singular learning models and joins them to deliver a total model that is more impressive than any of its singular learning models alone.

Random forests are an illustration of ensembles applied to decision trees. Random forests are broadly utilized and accomplish generally excellent outcomes on a wide assortment of issues. They can be used as classifiers or as regressors. one drawback of utilizing a single decision tree was that decision trees will be overfitted on the training dataset. A random forest makes heaps of individual decision trees on a training dataset, frequently on the request for many trees. The idea of the random forest is that every one of the singular decision trees in a random forest ought to get along admirably at predicting the objective label in the training dataset however ought to likewise be developed to be different somehow or another from different trees in the forest.

This random variation during tree creating occurs in two ways. First, the data utilized to build each tree is chosen randomly. Second, the attributes are selected in each split test are also randomly chosen [23] [26].

3.5 DenseNet

DenseNet is one of the new revelations in neural networks for visual item recognition. DenseNet is very like ResNet for certain basic distinctions. ResNet utilizes an added substance technique that consolidates the past layer (personality) with the future layer, while DenseNet connects the result of the past layer with the future layer [27].

3.6 Dataset

Several environmental factors, according to the WHO [29] can impact the spread of communicable diseases that might create epidemics. Water supply, sanitation facilities, food, and climate are the most fundamental of these. The basic hypothesis is that the number of cases and transmission of prior contagious viruses follow seasonal cycles that are influenced by climate, and Covid-19 should follow suit. Furthermore, seasonal changes in temperature and humidity have an impact on the amount of virus outbreaks.

Beginning with data collected and produced by Kaggle and the Johns Hopkins Center for Systems Science [30], the dataset is compiled from official case reports from several countries. The COVID-19 data was collected between January 22, 2019, and March 21, 2020, from the start of the epidemic. The dataset has 25 attributes and 9566 objects divided as shown in table 2. On a country-by-country basis, we collect various sorts of data on the dissemination of COVID-19. For example, country-level aggregates are used in the rest of world (52 counties from 6 regions). From Africa 9 countries, Americas 8 countries, Middle East 6 countries, Europe 17 Countries, Southeast 3 Countries, and Western 9 Countries.

Table 2: The Data set Statistics

WHO Region	Number of objects
Africa	1404
Americans	1350
Eastern	936
Europe	3172
Southeast	416
Western	2288

Furthermore, weather data is gathered from an archived weather database [30][31]. For each site where infection data exists, including nation, longitude, latitude, date, confirmed, fatalities, recovered, and current cases, we gathered minimum and maximum daily temperatures, humidity, precipitation, snowfall, moon illumination, sunshine hours, UV index, cloud cover, wind speed and direction, and pressure data for the weather data. We also used Demography's population density data.

4. **The Proposed Approach**

The proposed approach as shown in figure 1 has three stages are preprocessing, learning, verification. In Preprocessing stage, the dataset is cleansing, normalized and decrease the outlier's data. In Learning stage, the approach uses traditional machine learning techniques such as support vector machine (SVM), K- nearest neighbor (KNN), Naive Bayes (NB), Decision tree (J48), Logistic regression (LR), and Random Forest (RF), and use Dense net convolution neural networks (Dense Net CNN). In Verification stage, the precision, recall, f-measures, and accuracy are used to metric the learning classifiers.

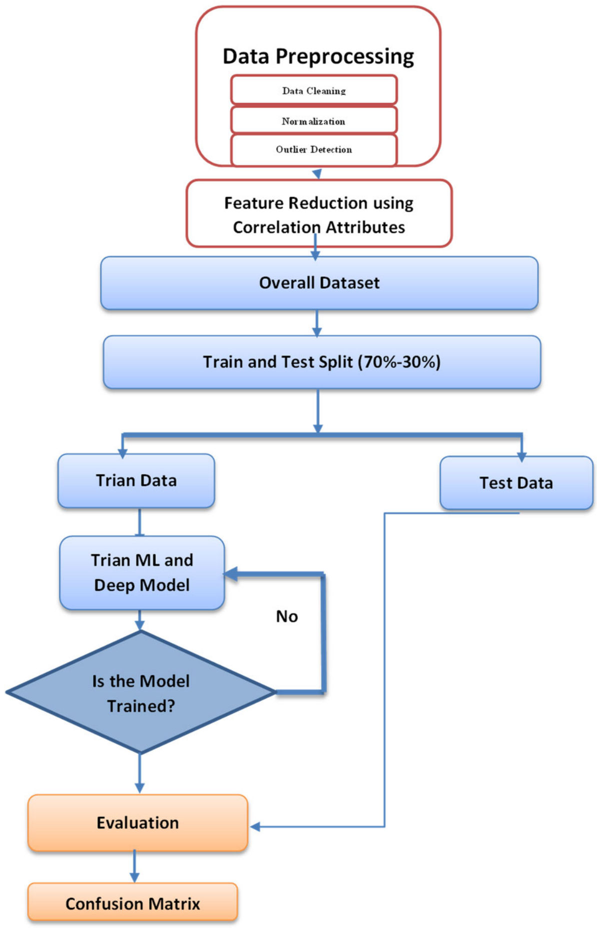

Figure 1: Proposed Approach

4.1 Preprocessing

Data Cleaning It is also known as scrubbing. This task involves filling of missing values, smoothing, or removing noisy data and outliers along with resolving inconsistencies, Data Transformation: This involves normalization and aggregation of data according to the needs of the data set, Data Reduction: During this step data is reduced. The number of records or the number of attributes or dimensions can be reduced. Reduction is performed by keeping in mind that reduced data should produce the same results as original data, Data Discretization: It is considered as a part of data reduction. The numerical attributes are replaced with nominal ones.

4.2 Feature selection

Correlation attributes feature selection is used to determine the attributes that effects on the decision by Pearson correlation equation as shown in figure 2.

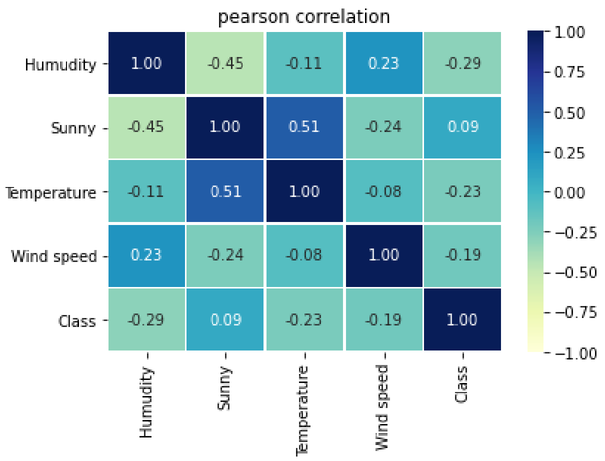

Figure 2: Feature Selection Correlation Attributes

5. Experimental Results and Discussion

The proposed approach was implemented using python 3.7 on the free cloud service "Kaggle GPU." Precisely, we utilized the popular TensorFlow 2.4 machine learning library. 347. We also utilized the open-source Python library for the preprocessing steps and the machine learning and deep learning Python open-source library "TFLearn" for classification.

Table 3: The average resultss of SVM,KNN,DT,LR,RF,NB,Dense Net

	Precision	**Recall**	**F-Measure**	**Accuracy**
SVM	57.7%	100%	73.3%	57.7%
KNN	69.1%	84.7%	76.1%	69.3%
DT	78.7%	79.3%	79%	75.7%
LR	67.9%	79%	73%	66.3%
RF	**78.3%**	**83.8%**	**81%**	**77.2%**
NB	69.1%	75.3%	72.1%	66.3%
Dense Net CNN	58%	93%	71.4%	57.1%

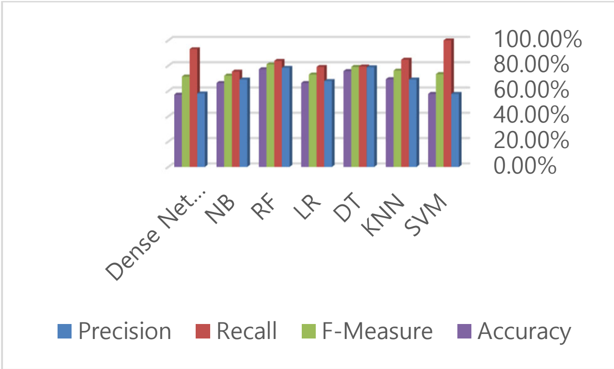

Figure 3: Average results of Precision, Recall, F-measure, Accuracy

Figure 3 and table 3 presents the experimental results for the comparison of the state of art models for active cases. The highest performance was registered by the Random Forest algorithm when evaluated using mean absolute error and root mean square error. The result of F-measure means that the humidity, sunny hours, wind speed, and temperature aide to spread COVID-19 with 81%.

6. Conclusion

This research provides an analysis-based scientific approach to detect the climatic factors that aide to spread COVID-19 in the rest of world by using the machine learning classification process, and feature selections techniques for example, Pearson Correlation. Using the Pearson correlation coefficient method will enable the identification of four main climatic factors that effects on spread COVID-19 in the world. the conclusion of this approach is the decrease of humidity and temperature and wind speed aid to spread COVID-19 but the increase of sunny hour increases the spread of COVID-19.

Acknowledgement

The authors would like to thank the Deanship of Graduate Studies at Jouf University for funding and supporting this research through the initiative of DGS, Graduate Students Research Support (GSR) at Jouf University, Saudi Arabia".

References

[1] N. S. Punn, S. K. Sonbhadra and S. Agarwal, “COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms,” MedRxiv, preprint, 2020.
[2] Y. Oh, S. Park and J. C. Ye, “Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets,” IEEE TRANSACTIONS ON MEDICAL IMAGING, vol. 39, no. 8, 2020.

[3] N. M. Ghazaly, M. A. Abdel-Fattah and A. A. Abd El-Aziz, “Protein secondary structure prediction: A survey of the state of the art,” International Journal of Advanced Science and Technology, vol. 29, no. 5, pp. 1831 - 1849, 2020.

[4] R. E. Baker, W. Yang, G. A. Vecchi, J. E. Metcalf and B. T. Grenfell, “Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic,” Science, vol. 369, no. 6501, pp. 315-319, 2020.

[5] K. H. Chan, J. S. M. Peiris, S. Y. Lam, L. L. M. Poon and K. Y. Yuen, “The effects of temperature and relative humidity on the viability of the SARS coronavirus,” Advances in Virology, 2011.

[6] A. W. H. Chin, J. T. S. Chu, M. R. A. Perera, K. P. Y. Hui, H. L. Yen et al., “Stability of SARS-CoV-2 in different environmental conditions,” The Lancet Microbe, vol. 1, no. 1, 2020.

[7] D. Kang and C. Ellgen, “The role of temperature in COVID-19 disease severity and transmission rates,” Preprints, vol. 93, pp. 5358-5366, 2021.

[8] Á. B. Redón and Á. S. Aroca, “The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques,” Physical Geography, vol. 44, no. 5, pp. 591–604, 2020.

[9] P. B¨uhlmann, “Bagging, Boosting and Ensemble Methods,” Springer Handbooks of Computational Statistics book series, pp.985–1022, 2011.

[10] C. A. Christophi, M. S. Prieto, F. Y. Lan, M. D. Velandia, V. Efthymiou et al., “Ambient temperature and subsequent COVID-19 mortality in the OECD countries and individual United States,” Sci Rep, vol. 11, no. 8710, 2021.

[11] H. Behzad, W. Wenge, R. R. I. Niilo, D. Francesca , D. Alan et al., “Weather Conditions and COVID-19 Incidence in a Cold Climate: A Time-Series Study in Finland,” Frontiers in Public Health, vol. 8, 2021.

[12] A. Z. E. Kassem, “Does Temperature Affect COVID-19 Transmission?,” Frontiers in Public Health, vol. 8, 2020.

[13] D. Kang, C. Ellgen and E. Kulstad, “Possible effects of air temperature on COVID-19 disease severity and transmission rates,” Medical Virology, vol. 39, no. 9, pp. 5358-5366, 2021.

[14] M. B. Tyrrell1, A. Meyer, C. Faverjon and A. Cameron, “Preliminary evidence that higher temperatures are associated with lower incidence of COVID-19, for cases reported globally up to 29th February 2020,” medRxiv:20036731v1 [Preprint], 2020.

[15] M. Wang, A. Jiang, L. Gong, L. Lina, G. Wenbin et al., “Temperature significant change COVID-19 Transmission in 429 cities,” medRxiv: 20025791v1 [Preprint], 2020.

[16] M. B. Arau´jo and B. Naimi, “Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate,” medRxiv: 20034728v1 [Preprint], 2020.

[17] N. Al-Rousan and H. Al-Najjar, “Nowcasting and forecasting the spreading of novel coronavirus 2019-nCoV and its association with weather variables in 30 Chinese Provinces: A case study,” SRRN [Preprint], 2020.

[18] J. Bu, D. Peng, H. Xiao, Y. Qian, H. Yan, et al., “Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020,” medRxiv: 20022715v1 [Preprint]. 2020.

[19] Q. Bukhari and Y. Jameel, “Will coronavirus pandemic diminish by summer?,” SRRN. [Preprint], 2020.

[20] W. Luo, M. S. Majumder, D. Liu, C. Poirier, K. D. Mandl et al., “The role of absolute humidity on transmission rates of the COVID-19 outbreak,” medRxiv: 20022467v1 [Preprint], 2020.

[21] C. Poirier, W. Luo, M. Majumder, D. Liu, K. Mandl et al., “The role of environmental factors on transmission rates of the COVID-19 outbreak: An initial assessment in two spatial scales,” SRRN. [Preprint], 2020.

[22] M. Parsian, “Data algorithms: recipes for scaling up with hadoop and spark,” in O’Reilly Media, Inc., Gravenstein Highway, North, Sebastopol, USA, 2015.

[23] S. Sridhar and M. Vijayalakshmi, “Machine learning,” in Oxford University Press, 1st ed., vol. 1. New Delhi. India, 2021.

[24] E. Frank E and J. R. Harrell, “Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis,” In Springer Series in Statistics, 2nd ed., pp. 1-582, 2015.

[25] R. Kohavi, “Machine Learning: ECML-95,” Springer Berlin Heidelberg, Lecture Notes in Computer Science, pp. 174-189, vol 912, 1995.

[26] C. Sheppard, “Tree-based machine learning algorithms: decision trees, random forests, and boosting,” CreateSpace Independent Publishing Platform, 2nd ed., USA. 2019.

[27] R. Shanmugamani, “Deep learning for computer vision: expert techniques to train advanced neural networks using tensorflow and keras,” O’Reilly Media, Inc., 2018.

[28] .

[29] (WHO), W.H.O. World report on vision. https://www.who.int/publications/i/item/9789241516570, 2019. Accessed on 02 456 October 2021.

[30] Kaggle, “covid19 global weather data. Kaggle 2020,” https://www.kaggle.com/ winterpierre91/covid19- global-weather- data.

[31] Imdevskp, “covid-19 jhu data web scrap and cleaning,” github 2020, https: //github.com/imdevskp/covid _ 19 _ jhu _ data _ web _ scrap _ and _ cleaning.

Appendix

Table 1: presents a comparison between the state-of-the-art recent existing systems:

Reference	Variables	Dataset	Statistical analysis	Results
[14]	Temperature	Open-source list of confirmed COVID-19 cases.	Generalized linear Regression framework, ratio Tests and pseudo R-squared values.	There might be occasional changeability in transmission of SARS-CoV-2, yet this investigation doesn't suggest that temperature alone is an essential driver of COVID-19 transmission. The beginning of hotter climate in the northern side of the equator may unobtrusively diminish pace of spread.
[15]	Temperature	Official websites of health authorities of overseas countries.	Descriptive statistics, Log transformation, restricted cubic spline function and generalized linear mixture model.	Temperature altogether affects the transmission of COVID-19. There may be a nonlinear portion reaction relationship between the two, it is a to show that the best temperature adding to its transmission and that low temperature is helpful to the viral transmission. For nations and locales with a lower temperature, severe counteraction, and control measures ought to be kept on forestalling future inversion of the plague.
[16]	Temperature and humidity.	WHO and Johns Hopkins University.	Descriptive statistics.	COVID-19 favors cool and dry circumstances and was to a great extent missing under incredibly cold and extremely hot and wet circumstances. This illuminate making arrangements for the timing and extent of the probable public intercessions to alleviate the antagonistic outcomes of the Covid on general wellbeing.
[17]	Temperature and relative humidity.	WHO, Johns Hopkins University, ECDPC and CDC.	Pearson correlation coefficient.	Climate factors showed a little impact on Covid transmission, and no connection can be removed between the effect of climate and affirmed cases in all areas. In certain regions, temperature showed a positive connection according to affirmed cases and dampness exhibited a negative relationship. In different regions, no relationship was found.
[18]	Temperature, mean humidity.	WHO website and other public sources.	Descriptive statistics.	Warm and dry weather conditions was great for the endurance of the infection with a temperature scope of 13-24˚C, a stickiness scope of 50-80%, a precipitation of 30 mm/month or less. Cold air for over seven days affects SARS-CoV-2.
[19]	Temperature, absolute and relative humidity and wind speed.	WHO and Johns Hopkins University.	Descriptive statistics.	Considering the ongoing information on the spread of COVID-19, the authors, theorized that the lower number of cases in tropical nations may be because of warm sticky circumstances, under which the spread of COVID-19 may be slower than had been noticed for other viruses.
[20]	Temperature, absolute and relative humidity.	WHO, Johns Hopkins, CDC, ECDPC, DXY-COVID-19-Data, USCDCP and NHC.	Proxy for the reproductive number R, Clausius Clapeyron equation, Loess regression, exponential fit and linear model.	outright humidity and temperature were related with nearby outstanding development of COVID-19 across areas in China and other impacted nations. Outright humidity and temperature yielded a positive relationship and a slight negative relationship individually. Changes in weather conditions alone won't be guaranteed to prompt decreases if counts without the execution of broad general wellbeing intercessions.
[21]	Near-surface air temperature and absolute humidity		Proxy for the reproductive number R, linear model with the local Rproxy and Loess regression	temperature showed a negative relationship, demonstrating that higher temperatures seemed to have lower Coronavirus transmission. Outright humidity showed a negative relationship, demonstrating that areas with higher outright humidity experienced lower transmission. Changes in climate alone won't be guaranteed to prompt decreases in the event that count without the execution of broad general wellbeing mediations.