Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(2), 1583-1590 | ISSN:2251-6727

An Covid-19 Disease Prediction and Analysis Using Sprint Algorithm in Machine Learning Technique

Mrs.V.Ramya
Research scholar Department of Computer Science VISTAS Pallavaram, Chennai, India.
Dr.K.S.Thirunavukkarasu
Assistant Professor Department of Computer Science VISTAS Pallavaram, Chennai, India. thirukst@gmail.com

(Received: 07 January 2024 Revised: 12 February 2024 Accepted: 06 March 2024)

KEYWORDS
Covid-19, Machine learning, Classification, Accuracy, Data processing

Abstract
Technology innovation, social media platforms, and online communication play a vital role in balancing human life during the COVID-19 epidemic. Scientists have struggled to predict this disease accurately because of its uniqueness and rapid spread. This is partly caused by variations in human behavior and environmental elements that affect the spread of diseases. However, it spread untruths and misinformation about the disease and the vaccine.The accuracy in determining the covid-19 patients, real news, and fake news becomes very challenging due to widespread and enormous data generation. In our research, we proposed an improved SPRINT algorithm for classifying the correct and incorrect predictions in terms of accuracy.The Histogram and the attribute table are two data structures that the enhanced SPRINT method specifies. The three components of the property sheet are the indexing of attribute value, class identification, and data records. The proposed algorithm is executed in MATLAB code to generate the training model. The accuracy of the proposed system is evaluated with the AdaBoost algorithm with the Random Forest model [16] and with machine learning methods SVM and KNN [17]. The improved SPRINT algorithm shows better results on all the performance metrics than the others. The accuracy obtained by the proposed improved SPRINT algorithm is 99.5% which is far better than the others.

Introduction:
The advent of COVID-19 rapidly expanded into a global pandemic and significantly impacted everyone in the world. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a public health emergency with a potential pandemic [1]. Due to the pandemic's rapid spread, immediate action was needed to minimize the consequences. Regardless of the severity of the illness, all COVID-19-positive individuals have been obliged to be hospitalized from the start of the pandemic. As a result of the substantial rise of cases worldwide, hospitals are now at 100% occupancy, placing undue pressure on the available medical resources. However, it is essential to have methods that can quickly identify people at a high risk of developing severe and non-severe symptoms of illness [2-3].
The age distribution and the total active COVID-19 cases are just two examples of the protected data that many conventional tools for disease analysis use to differentiate one region from another. While this might be adequate to predict disease spread across wide areas, it is inadequate at a more detailed level [4-5]. The data constraints on which each projection of the COVID-19 spread is based must be considered. The total cases from big regions dominate forecasts at the national level (states or counties). Due to the scarcity of case studies and the diversity of each site, projections for less populated areas become more challenging. Additionally, these less populated locations frequently have the worst COVID-19 case planning [6].
Since 2021, various coronaviruses have appeared and spread over numerous nations. About 6.2 million deaths and 485 million confirmed cases were reported in 2021 [7-8]. Social media platforms have become more critical for keeping people informed and preserving their interpersonal connections, productivity and safety since the coronavirus spread globally. News organizations and medical institutions have used the opportunity to disseminate information about the infection. Medical journals, World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), and healthcare organizations have additionally helped publish and update the data about the virus, its treatments, prevention, and spreading. But since this issue involves aspects of human existence related to

health and medicine, spreading false information about the virus on social media platforms and Internet has become an additional obstacle to overcome. The nation's economy was harmed, people's faith in their governments was diminished, particular items were promoted to earn enormous profits, and incorrect information on avoiding and treating the virus was disseminated [9].

According to the literature, "Fake News" is news that is designed to manipulate the reader and lead them to assume that it is real and accurate [10], as well as"scam," "misinformation," "disinformation," and "rumor" [11]. "Disinformation," which is deliberately spread by those with ulterior motives. Disinformation propagated about the COVID-19 epidemic blamed ethnic groups, undocumented immigrants, and even governments for the virus's spread. Some political parties desire to create disorder to further their agenda [11–12]. "Misinformation," which is false yet spreads innocently. Misinformation regarding the COVID-19 pandemic includes misconceptions about the virus, a desire for untrue remedies, and fictitious conclusions drawn from the virus's transmission mode. This manifests as speculative writing by well-intentioned individuals who present their views as fact.

This study compares the performance of several models using different assessment metrics, enabling users or AI developers to specify criteria or weight metrics according to their significance, select the model that best suits their needs, or rank the models using multiple metrics. The multi-criteria decision-making has the potential to be used to manage the identification of fake news [13–15]. In this process, users rank deep learning algorithms following their interests, such as the greatest appropriate prediction with the greatest F1 score and the quickest training time. In this research, we proposed an improved SPRINT algorithm to predict covid-19 patients accurately.

Problem Statement:

- Lack of accuracy in determining the correct predictions and incorrect predictions
- Inefficiency in handling huge amounts of data from various sources
- Existing approaches consume more time consumption for prediction
- Lack of enhanced approaches like cluster mechanism in determining the accuracy
- Need for an enhanced system for testing the machine learning-based models for predicting the covid-19 severity

Research Contribution:

- Designed and implemented the proposed improved SPRINT algorithm for enhancing classification accuracy.
- Enhanced approach for identifying and classifying the Covid-19 severities
- Implementing MATLAB code for generating and recreating the trained model.
- Proposed algorithm execution with the covid -19 dataset provided by ECDC
- Determining the efficiency of the proposed algorithms by comparing them with existing algorithms in terms of accuracy

Organization of the paper: This research is organized as follows; in section 1 introduction is discussed. Section 2 contains various works on improving the classification accuracy and other machine learning approaches designed especially for covid-19. In section 3, the implementation of the proposed system is discussed in detail. The observation of the proposed work and comparison work with the existing approaches are discussed in the result and section 4. Finally, the conclusion part is mentioned in section5.

Related Work:

Iwendi et al. [16] applied the AdaBoost algorithm and RF model to predict the severity of COVID-19 disease. Based on geographic and demographic data, health, and travel from the COVID-19 patients, the accuracy of the enhanced RF model was assessed. On the used dataset, the enhanced RF model yields 94% of accuracy and 86% of F1-Score.

Brinati et al. [17] implement machine learning methods such as Decision Tree, KNN, SVM, Naïve Bayes, Logistic Regression, Random Forest, and Extremely Randomized Trees to detect the covid-19 patients. The simulation used routine blood tests taken from 279 patients. The outcomes showed that the Random Forest algorithm is feasible and useful by reaching AUC with 84%, accuracy at 82%, sensitivity at 92%, precision at 83%, and specificity at 65%.

Yassine Meraihi et al. [18] conducted a survey on covid-19 diagnosis, identification, and prevention using Machine Learning (ML). In this work, the author analyzed about 160 ML-based mechanisms developed during the impact of covid-19. The main intention of these algorithms is to analyze huge amounts of data to address the complexities of the detection and diagnosis of covid-19. The author stated that about 79% uses deep learning, 65% uses Convolutional Neural Network

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(2), 1583-1590 | ISSN:2251-6727

(CNN), 17% applies Specialized CNN, and 16% applies Regression algorithms, Support Vector Machine (SVM), and Random Forest. All these approaches use clinical data and examine with evaluation metrics like AUC, F1-score, accuracy, specificity, sensitivity, and precision.
Mariam Laatif et al. [19] analyzed various approaches in predicting the severity risk of covid-19. The main purpose of this work is to evaluate the machine learning approaches with severity prediction. The parameters taken for consideration are platelets, D-dimers, and C-reactive protein. The proposed topological data-based Uniform Manifold Approximation and Projection (UMAP) results in better aspects of the ROC curve, sensitivity, specificity, and accuracy.
Charles Nicholson et al. [20] proposed a machine learning and clustering-based mechanism for predicting the covid -19 analysis at the county level. In this work, the author employs supervised and unsupervised approaches to identify the essential county-level demographic, medical capacity, weather, mobility, and health-associated county-level characteristics. Next, the author combined counties into significant clusters using this feature subspace to assist more in-depth disease analysis efforts.
Yahya Tashtoush et al. [21] proposed a Deep Learning Framework for identifying Fake news about COVID-19 on Social Media Platforms. In this work, the author has taken both real and fake news about covid 19 posted on social platforms. Then applied, deep neural algorithms such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), LSTM networks, hybrid of CNN, and Bi-directional LSTM. From the observation, the author states among these algorithms, CNN has achieved 94.2 % accuracy than the others.
Jiang et al. [22] developed a classifier using Bi-LSTM deep learning models for identifying the real news. A dataset for fact-checking was used to assess the proposed approach. Additionally, they employed a variety of evaluation criteria, including execution time, F1 measure, accuracy, and recall, to show the effectiveness of the proposed model.
Umer et al. [23] identified fake news using hybrid deep neural network architecture. The hybrid model is a combination of LSTM and CNN. Before sending the feature vectors to the classifier, they utilized two methods to minimize their dimensions: Principal Component Analysis (PCA) and Chi-Square. The Fake News Challenges (FNC) dataset contains news stories classified into four categories: Discuss, Agree, Disagree, and Unrelated. More contextual characteristics are produced for the fake news identification task by feeding non-linear data into the PCA and chi-square. The findings demonstrate that PCA performed better than Chi-square in terms of accuracy.
To forecast the severity of COVID-19 patients, Wang et al. [24] employed a data collection of 296 patients from the First People's Hospital of Jiangxia District in Wuhan, China. The machine learning method X GBoost was used to generate the model. The clinical model was based on the age, coronary heart disease, and hypertension history records. The model was built using D-dimer, oxygen saturation (SpO2), hs CRP, lymphocyte count, neutrophil, and neutrophil.
Banoei et al. [25] used the statistical method SIMPLS to forecast hospital mortality. Identifying the high-risk and low-risk patients with COVID-19 cluster formation is done using Latent class analysis (LCA). The SIMPLS system could correctly predict patient hospital mortality using training and validation data (AUC > 0.85).

3. PROPOSED METHODOLOGY

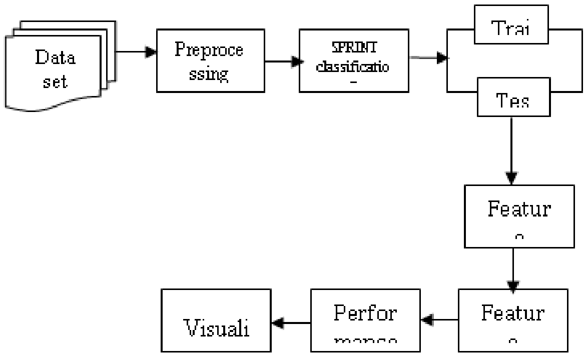

Figure -1 Proposed architecture

Data preprocessing

Data preprocessing is converting raw data into useful information that a machine learning model can use.It is the initial and most important step in developing a machine-learning model. Weneed access to clean, well-formatted data while developing a machine learning project.Additionally, data must always be cleaned and formatted before being used in any activity. Hence data preprocessing is considered a vital task.

SPRINT classification

The histogram and attribute table are the two data structures proposed in the enhanced SPRINT method.The three components of the property sheet are data indexing, attribute value, and class identification. In the memory section, the hard disc, instead of recorded data, contains the attribute list only. The attribute table was divided with node expansion, and each relevant

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(2), 1583-1590 | ISSN:2251-6727

child node was linked. To describe the attribute node distribution types, Histogram is linked with the nodes. In the class distribution, as per the numerical properties' nodes are linked with two histograms such as Cabove and Cbelow. The former can identify the sample type, which deals with distribution.

Additionally, it specifies the distribution pattern for untreated samples and two samples' values when considered together with updates. In the discrete distribution class, the Histogram's attribute is expressed by one node only. The minimum description length idea is applied to SPRINT pruning for achieving efficient performance.

Feature extraction is the mechanism used to minimize the amount of resources required for executing data preprocessing with minimum loss of relevant information. Additionally, feature extraction can help analysis by reducing the duplicated data. The machine's efforts to generate variable features and the minimization of the data optimize the learning and generalization stages of the machine learning process.

Data set

The dataset is taken from European Centre for Disease Prevention and Control (ECDC). It has been a European Union (EU) public health organizationsince 2005. The ECDC aims to shield more than 500 million individuals from contagious diseases, which are mostly brought on by parasitesand germs.Data on over 50 infectious disease topics, includingCOVID-19, HIV/AIDS, TB, antibiotic resistance, measles, hepatitis, influenza, and vaccination, are gathered, analyzed, and shared by ECDC.Professionals from the ECDC evaluate the dangers to Europe and offer recommendations to help nations stop breakouts and deal with hazards to the public's health. ECDC will provide new COVID-19 cases and deaths around the EU/EEA country from 11 March 2021. They published these reports every Thursday worldwide. The data on COVID-19 casesis available weekly and daily and can be downloaded in CSV, XML, XLSX, and JSON formats.

Tools used

Data classification models are trained using the Classification Learner software. This system allows exploring with supervised machine learning utilizing multiple classification techniques.Here data can be analyzed and trained models, select features, defined validation procedures, and evaluate outcomes.Support vector machines, naive Bayes, ensemble, neural network classification, discriminant analysis, kernel approximation, nearest neighbors, decision trees, and logistic regression can be automatically trained to find the optimum classification model type.By presenting defined input data and known data responses, supervised machine learning (labels or classes) can be specified easily. The data can be employed to train the predicting model. Now the model can export to the workspace or develop MATLAB code to replicate the trained model, utilize it with new data, or learn more about programmatic classification.

Experimental results:

In this section, experimental work is described, and for this research covid-19, the dataset is taken from the ECDC repository. It contains various sets of datasets related to covid-19 and is the most commonly used repositoryamong researchers. We have taken the normal as well as ICU-admitted cases on a weekly and daily basis. It is implemented in MATLAB, and the obtained results are described in terms of Recall, Precision, and F-measure.

Confusion matrix

A confusion matrixexplicitly declares the classification problem's prediction results. Predictions are of two types such as correct and incorrect predictions.The correct or incorrect prediction result can be determined based on the broken down and count values. The finest thing it does is show the error type made by that specific classifier and how the classifier made the error.

Description of the Terms:

- Positive (P): Observation is positive
- Negative (N): Observation is not positive
- True Positive (TP): Observation is positive and is predicted to be positive.
- False Negative (FN): Observation is positive but is predicted negative.
- True Negative (TN): Observation is negative and is predicted to be negative.
- False Positive (FP): Observation is negative but is predicted positive.

Recall

The recall is determined by dividing the correctly categorized positive examples by the total positive samples. The minimum FN and strong recall values indicate that examples were correctly recognized. It is calculated as below;

$$Recall = \frac{TP}{TP + FN}$$

Precision

Precision is the result measured by dividing the total correctly classified positive samples by the total predicted positive samples. High precision values indicate positive results. It is calculated as below;

$$Precision = \frac{TP}{TP + FP}$$

F-measure

F-measure is obtained by calculating the Recall and Precision. F-measure is measured using Harmonic Mean rather than Arithmetic Mean. It is because Harmonic Mean works effectively with high values. Among these, F-measure is always less than the Recall or Precision. It is calculated as below;

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

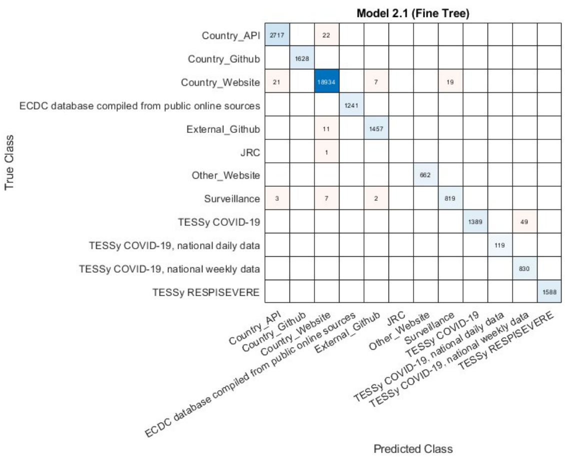

Figure 2 (a) Predicated class.

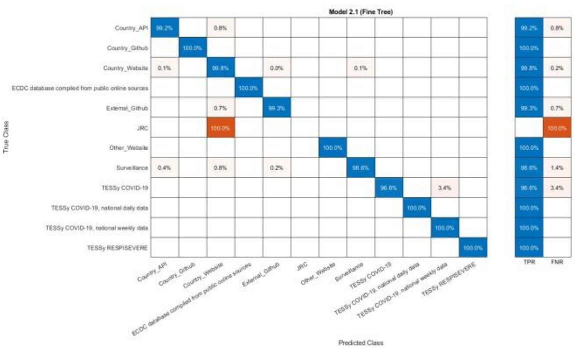

Figure 2 (b) confusion matrix predicated class.

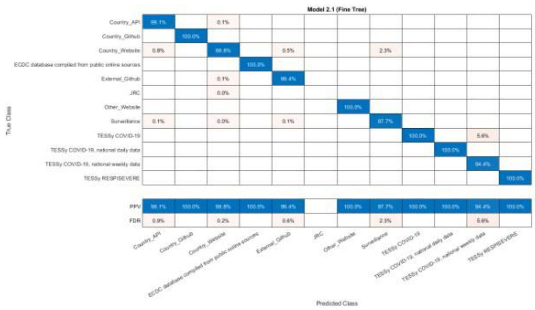

Figure 2(c)

Figure 2(a) represents a fine tree model with predicted class and true class. Figures 2(b) and 2(c) illustrate the confusion matrix predicted class with true class. It evaluates how well a machine learning classification algorithm performs with the output containing two or more classes. It is quite helpful for determining the accuracy and efficiency of the categorization model.As shown in figure 2(b), TPR – True Positive Rates denote the prediction as true and positive. The FNR – False Negative Rates represents the wrong results with false values. The fine tree structure exhibits accuracy with the respective training model. In figure 2(c), two additional parameters, such as positive predictive value (PPV) and False Discovery Rate (FDR), are considered. Here, PPV is the ratio of total predictions for the specific class divided by the total records, or correctly classified records, for the specified class.FDR is the calculation of accuracy when several hypotheses are being evaluated simultaneously, such as when several metrics or variations are being evaluated in a single experiment.

Scatter plot:

A scatter plot is a mathematical figure or plot that displays results for usually two variables for data collection. It uses Cartesian coordinates. The horizontal axis represents the value of one variable, and the vertical axis represents its position by the value of the other

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(2), 1583-1590 | ISSN:2251-6727

variable. The points are represented as size or color, or shape.

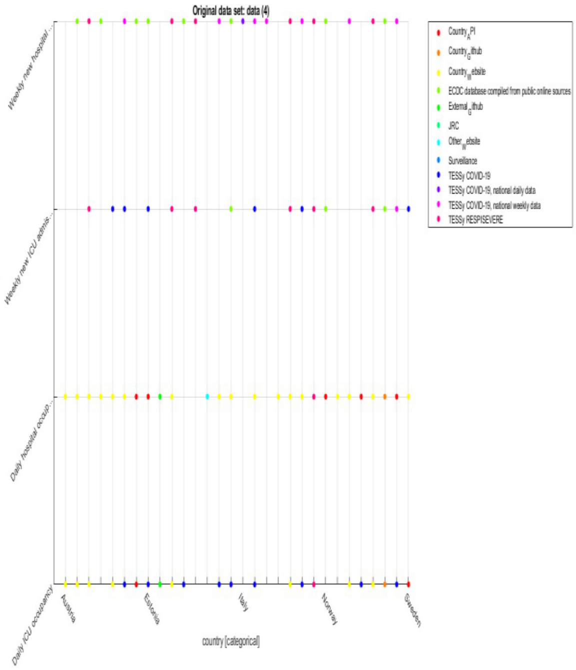

Figure 3(a) scatter plot 1

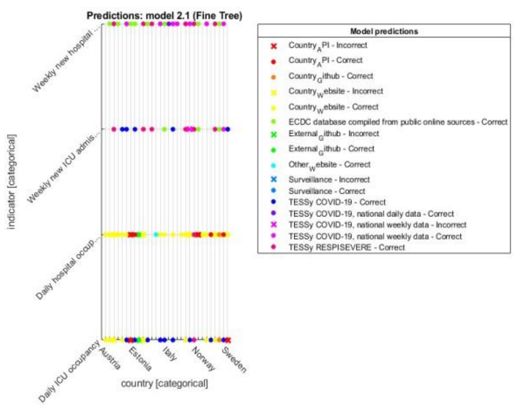

Figure 3(b) scatter plot 2

Figure 3(a) represents the scatter plot of the original data, where the x-axis denotes the country, which is categorized accordingly. The y-axis denotes the patients with covid in the weekly new hospital, weekly new ICU admissions, daily hospital occupied, and daily ICU occupied concerning the country. Figure 3(b) shows the prediction model on a scatter plot with a fine tree structure of covid in the aspect of the weekly new hospital, weekly new ICU admissions, daily hospital occupied, and daily ICU occupied concerning the country. In figure 3(a), color plots define the various datasets, and in figure 3(b), color plots show their prediction results, such as correct or incorrection prediction.

Parallel coordinates plot

High-dimensional datasets are frequently shown and analyzed using parallel coordinates. A backdrop made of n parallel lines, commonly vertical and evenly spaced, is drawn to represent a point collected in an n-dimensional space. A polyline with vertices on the parallel axes represents a point in n-dimensional space; the vertex position on the i-th axis correlates to the point's i-th coordinate. This representation closely resembles time series visualization, except that it is used with data that does not have a natural order because the axes do not relate to points in time. As a result, many axis configurations might be useful.

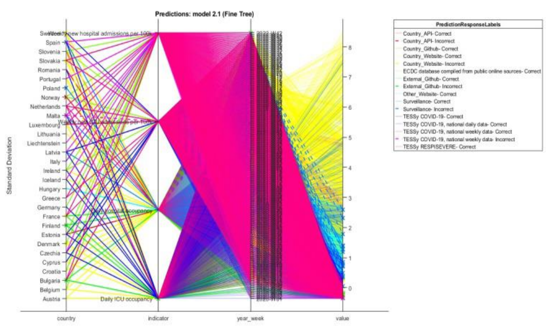

Figure 4(a) parallel coordinates plot 1

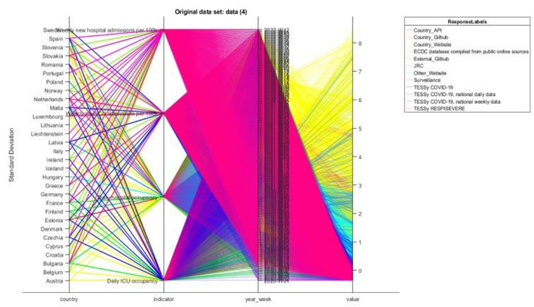

Figure 4(b) parallel coordinates plot 2

Figures 4(a) and 4(b) show the parallel coordinates concerning the input dataset. Figure 4(a) shows the parallel coordination of the correct and incorrect predictions concerning the country, indicator, year/week, and values. Figure 4(b) shows the parallel coordinates of the original data concerning the country, indicator, year/week, and values.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2024) 14(2), 1583-1590 | ISSN:2251-6727

Matlab Simulation Summary table

Accuracy (validation)	99.5%
Total cost (validation)	142
Predication speed	110000 obs/sec
Time	729 sec
Model	Hyperparameters
Maximum number of splits	100
Split criterion	Gine diversity index
Feature selection	4/4 individual features selected

Table 1: Simulation Summary table

In our research, MATLAB is used to determine the performance of the proposed SPRINT algorithm. The dataset is executed in the simulation environment, which contains the essential parameters such as accuracy in percentage, prediction speed, time, and model used for the execution. A details summary of the parameters used in the proposed execution is described the above table1.

Classifier	Accuracy
Proposed SPRINT	99.5 %
Ada boost with random forest	81%
SVM, KNN	65 % and 84 %

Table 2: Accuracy Comparison Table

Table 2 compares the proposed SPRINT algorithm with Ada boost with random forest and SVM, KNN. The comparison work is executed in MATLAB and then analyzed for accuracy. The proposed SPRINT algorithm has obtained maximum accuracy of 99.5 %, whereas Ada boosts with random forest with 81 %, SVM and KNN with 65 % and 84 %, respectively.

Conclusion:

COVID-19 is a worldwide pandemic endangering millions of humans' lives and livelihoods. It can be challenging to predict COVID-19 propagation, particularly in rural and local populations. It is difficult to predict the seriousness of the COVID-19 patients and Fake News detection from the text. The main reasons are its data volume is enormous and comes from various sources. The widespread COVID-19 pandemic and the critical situation worldwide have provoked responses from the research community in various sectors. The primary purpose of this research is to enhance the classification accuracy of machine learning algorithms. Even though existing classification models exist in the industry, accuracy, especially in predicting COVID-19, still needs to fulfill the needs. Hence, the SPRINT algorithm is proposed in this research. It contains the attribute table and Histogram, the two data structures the enhanced SPRINT method specifies. The proposed model's dataset is taken from ECDC, which contains the dataset of various European countries with the range of newly and weekly added normal and ICU-admitted Covid-19 cases. The proposed system undergoes the execution of a confusion matrix, scatter plot, and coordinates point on the prediction class. Using MATLAB simulation environment, a comparison work is carried out between the proposed SPRINT algorithm with AdaBoost algorithm with Random Forest model and with machine learning methods SVM and KNN in terms of accuracy. The proposed SPRINT algorithm has achieved an accuracy of about 99.5%, which is far better than the others.

Reference:

1) Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and coronavirus disease-2019 (covid-19): The epidemic and the challenges. Int J Antimicrob Agents. 2020;55(3): 105924
2) Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. MedRxiv;2020.
3) Jiang X, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Comput Mater Continua. 2020;63(1):537–51
4) Zhou Y, Wang L, Zhang L, Shi L, Yang K, He J, et al. A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States. Harvard Data Science Review. 2020;. pmid:32607504
5) Tang L, Zhou Y, Wang L, Purkayastha S, Zhang L, He J, et al. A review of multi-compartment infectious

disease models. International Statistical Review. 2020;88(2):462–513. pmid:32834402
6) Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the United States: An ecological analysis as of May 2020. Journal of Rural Health. 2020;36(3):433–445. pmid:32543763
7) Al-Ahmad, B.; Al-Zoubi, A.M.; Abu Khurma, R.; Aljarah, I. An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information. Symmetry 2021, 13, 1091. [CrossRef]
8) COVID-19 Pandemic—Wikipedia. Available online: https://en.wikipedia.org/wiki/COVID-19_pandemic (accessed on 20 December 2021)
9) Gupta, A.; Sukumaran, R.; John, K.; Teki, S. Hostility detection and COVID-19 fake news detection in social media. arXiv 2021, arXiv:2101.05953
10) Kaliyar, R.K. Fake news detection using a deep neural network. In Proceedings of the 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 14–15 December 2018; pp. 1–7.
11) Gupta, A.; Sukumaran, R.; John, K.; Teki, S. Hostility detection and COVID-19 fake news detection in social media. arXiv 2021, arXiv:2101.05953.
12) Kaliyar, R.K.; Goswami, A.; Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed. Tools Appl. 2021, 80, 11765–11788.
13) Ali, R.; Lee, S.; Chung, T.C. Accurate multi-criteria decision-making methodology for recommending machine learning algorithm. Expert Syst. Appl. 2017, 71, 257–278.
14) Chowdhury, N.K.; Kabir, M.A.; Rahman, M. An Ensemble-based Multi-Criteria Decision-Making Method for COVID-19 Cough Classification. arXiv 2021, arXiv:2110.00508.
15) Pirouz, B.; Ferrante, A.P.; Pirouz, B.; Piro, P. Machine Learning and Geo-Based Multi-Criteria Decision Support Systems in Analysis of Complex Problems. ISPRS Int. J. Geo-Inf. 2021, 10, 424.
16) Celestine Iwendi, Ali Kashif Bashir, Atharva Peshkar, R Sujatha, Jyotir Moy Chatterjee, Swetha Pasupuleti, Rishita Mishra, Sofa Pillai, and Ohyun Jo. Covid-19 patient health prediction using boosted random forest algorithm. Frontiers in public health, 8:357, 2020.
17) Brinati D, Campagner A, Ferrari D, Locatelli M, Banf G, Cabitza F. Detection of covid-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst. 2020;44(8):1–12
18) Yassine Meraihi, Asma Benmessaoud Gabis, Seyedali Mirjalili, Amar Ramdane-Cherif and Fawaz E. Alsaadi,"Machine Learning-Based Research for COVID-19 Detection, Diagnosis, and Prediction: A Survey",SN Computer Science (2022) 3:286 https://doi.org/10.1007/s42979-022-01184-z
19) Mariam Laatif, Samira Douzi, Abdelaziz Bouklouz, Hind Ezzine , Jaafar Jaafari , Younes Zaid, Bouabid El Ouahidi and Mariam Naciri, "Machine learning approaches in Covid-19 severity risk prediction in Morocco", Journal of Big Data (2022) 9:5 https://doi.org/10.1186/s40537-021-00557-0
20) Nicholson C, Beattie L, Beattie M, Razzaghi T, Chen S. A machine learning and clustering-based approach for county-level COVID-19 analysis. PLoS One. 2022 Apr 27;17(4):e0267558. DOI: 10.1371/journal.pone.0267558. PMID: 35476849; PMCID: PMC9045668
21) Tashtoush, Y.; Alrababah, B.; Darwish, O.; Maabreh, M.; Alsaedi, N. A Deep Learning Framework for Detection of COVID-19 Fake News on Social Media Platforms. Data 2022, 7, 65. https://doi.org/10.3390/data7050065
22) Jiang, T.; Li, J.P.; Haq, A.U.; Saboor, A. Fake News Detection using Deep Recurrent Neural Networks. In Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 18–20 December 2020; pp. 205–208.
23) Umer, M.; Imtiaz, Z.; Ullah, S.; Mehmood, A.; Choi, G.S.; On, B.W. Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access 2020, 8, 156695–156706. [CrossRef]
24) Wang K, et al. Clinical and laboratory predictors of in-hospital mortality in patients with coronavirus disease-2019: a cohort study in Wuhan, China. Clin Infect Dis. 2020;71(16):2079–88. https://doi.org/10.1093/cid/ ciaa538.
25) Banoei MM, Dinparastisaleh R, Zadeh AV, et al. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. Crit Care. 2021;25:328. https://doi.org/10.1186/ s13054-021-03749-5.