COVID-19 Pandemic analysis using SVM Classifier: Machine Learning in Health Domain

Swarn Avinash Kumar[1], Harsh Kumar[2], Vishal Dutt[3], Himanshu Swarnkar[4]

[1]*IIIT Allahabad, UP, India.*
[2]*Peoples'Friendship University of Russia, Moscow, Russia.*
[3]*Department of Computer Science, Arybhatta College, Ajmer, India.*
[4]*Department of Computer Science Engineering, Engineering College, Banswara, India.*

Abstract

Coronavirus has become very consistent in the medicinal field in recent months that has devastated many countries of the whole world in the last 5 -6 months. There is no exact cause of this disease, but research is still going on to find a proper treatment for this disease. Cases of coronavirus are increasing day by day, but there is not enough equipment available to identify the symptoms of the disease. This is why it takes a long time to identify the disease. Machine learning can help to reduce the time delay for the results of medical tests and to predict COVID-19 in patients. Detection of the corona (COVID-19) is now an important task for the physician. Corona spread among people so quickly and approaches 100,000 people worldwide. In this result, it is very important to identify the infected people then prevention of spread can be taken. This paper suggests an SVM-based methodology to detect coronavirus aimed at classification purposes. A set of procedures were recognized from works training including SVM (Support Vector Machine), RF (Random Forest), and ANN (Artificial Neural Network). Prediction of COVID-19 using machine learning may help grow the speed of virus detection as a result of decreased mortality. Analyzing the results obtained from the experiments, the Support Vector Machine (SVM) was identified to perform better than the other vectors.

Keywords: Severe Acute Respiratory Syndrome (SARS), MERS, COVID, Support Vector Machines, Regression, Deep Learning.

Introduction

COVID are a large group of infections that cause more common diseases than common viruses, such as the Middle East Respiratory Syndrome (MERS) and severe acute respiratory syndrome (SARS). These two transitions are spread by COVID named MERS-CoV and SARS-CoV. SARS was first seen in China in 2002 and MERS was first seen in 2012 in Saudi Arabia.The unexplained cause of pneumonia, identified in Wuhan, China, was first responded to by the

World Health Organization (WHO) country office in China on December 31, 2019. Since that time, cases of COVID have been expanding, along with higher losses.

These days, "COVID" is prefixed with "novel," as it is another strain of the infection. As demonstrated by the WHO, COVID has a spot with a wide family run from the typical cold to hazardous diseases (World Health Organization, 2020)[1].
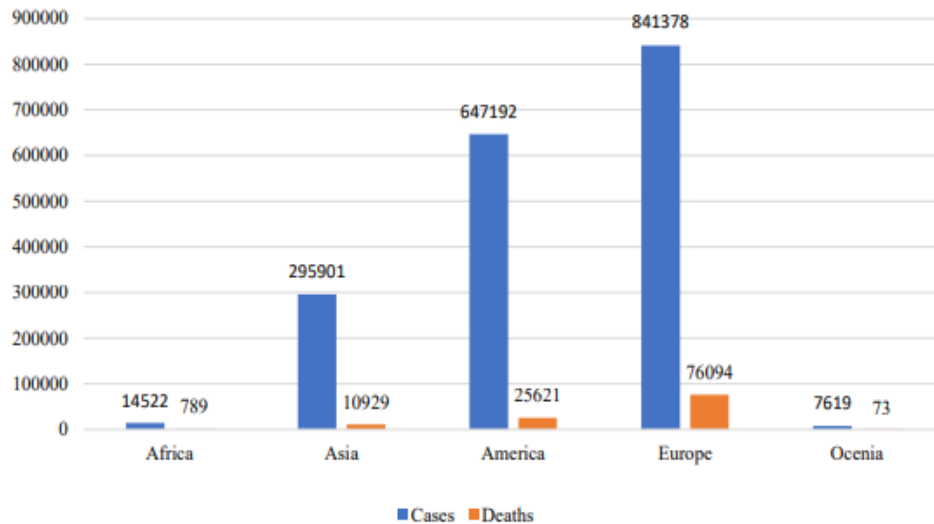

Figure 1.Deaths and Cases due to COVID-19 on 13-April-2020

Machine Learning in COVID-19

Machine learning is a subject of artificial intelligence developed by pattern recognition where data is preserved for the user's understanding. Healthcare banking. A variety of applications have been developed using machine learning in the military equipment space, etc. Right now, Machine Learning is a quickly advancing and consistently creating field. It programs PCs utilizing information to streamline their exhibition. It learns the boundaries to enhance the PC programs utilizing the preparation information or its previous encounters. Utilizing the information can likewise foresee what's to come. AI likewise encourages us in building a numerical model utilizing the insights of the information.The main purpose of machine learning is to build mathematical models using data. This mathematical model learns from human experience without interference [2]. Machine learning is mainly classified into four types:

➢ Supervised (Administered) Machine Learning
➢ Unsupervised Machine Learning
➢ Re-Enforcement Machine Learning

Supervised Learning: Administered Learning is a Machine Learning model that is worked to give out forecasts. This calculation is performed by taking a marked arrangement of information as information and referred to reactions as yield to get familiar with the relapse/characterization model. It creates prescient models from order calculations and relapse strategies.[2]

- **Classification** predicts discrete responses. Type predicts discrete responses. Here, the set of rules labels with the aid of selecting two or more lessons for each instance. If it is performed between two instructions then it is known as binary category and if it's far done between or greater instructions then it's miles called multi-elegance category. Packages of class encompass handwriting reputation, medical imaging, etc.[3]

- **Regression** predicts continuous responses. Here, the algorithms return a statistical value. For example, a set of data is collected such that the people are happy when considered the amount of sleep. Here, sleep and happiness are both variables. Now, the analysis is done by making predictions. The types of popular regression techniques are: [3]
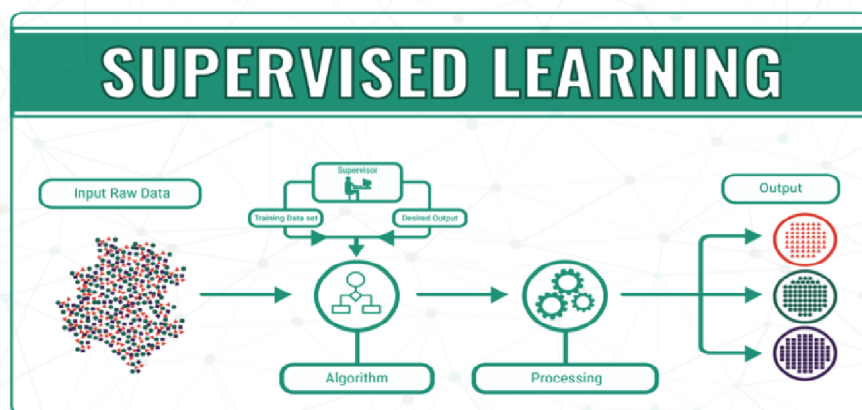 - Linear regression.
  - Logical regression.

Figure 2.Supervised Learning

Unsupervised Learning: Unlike supervised learning, there is no observer here and we only have input data.Unlike supervised learning, there is no observer here and we only have input data.Here, the basic objective is to find some patterns in the data to be compared to others. Here, assumptions are made such that bunches are discovered that match quite well with a taxonomy. For example, movies on Netflix.com are suggested based on the principle of clustering of movies where many similar films are based on a client's recently viewed movie list.[4]
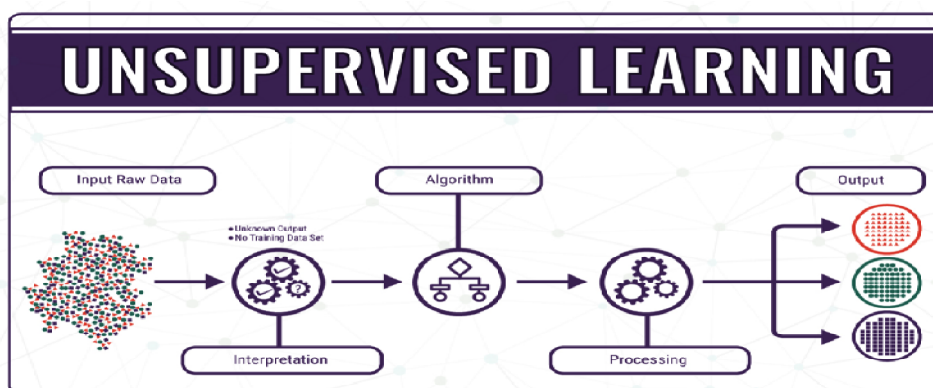

Figure 3.Unsupervised Learning

Reinforcement Machine Learning: Reinforcement learning learns its behavior from a trial and error method in a dynamic environment. Here, the problem is solved by taking appropriate action in a certain situation to maximize the output and to obtain the acquired results. In Reinforcement Learning, there is the presentation of the input or output data. Instead, when the desired action is chosen, the agent is immediately told the reward and the next state are not considering the long terms actions.[5]

Algorithm

At some stage in our studies, we've got investigated algorithms through which we have finished the supervised category, however, we worked on SVM.

Support Vector Machines (SVM): Support Vector Machines performs characterization by developing an N-dimensional hyperplane that isolates the information into two classifications. In SVM, the indicator variable is called a trait and the changed quality is known as an element. Choosing the most reasonable delegate information is called highlight choice.[4][5]

The objective of SVM design is to discover the ideal hyperplane that splits the clusters where on one side of the plane there is a target variable and on the extra adjacent of the plane another category. The vectors which might be immediate the hyperplane are the assist vectors. In parent four, an average example of an assist vector device is depicted
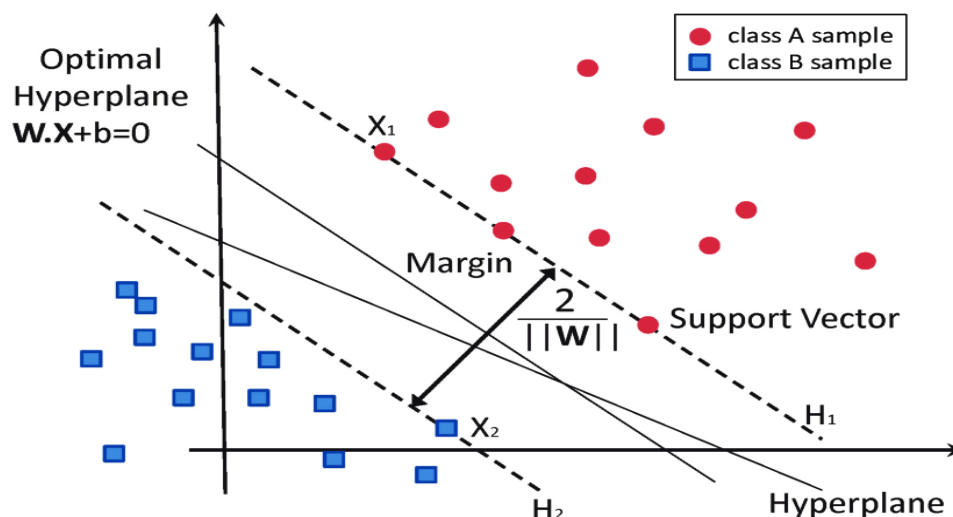

Figure 4.Support Vector Machine

Artificial Neural Networks (ANN): ANN is an attempt, in the only way, to mimic the neural gadget of the human brain. The basic unit of ANN is neurons. A neuron is stated to perform capabilities on input and bring an output. Neurons mixed are called neural networks. As soon as the neural networks are shaped, schooling of the records is started to limit the mistake. In the end, an optimizing algorithm is used to further reduce the mistakes. The layered structure of artificial neural networks (ANNs) is represented in figure 5.[5]
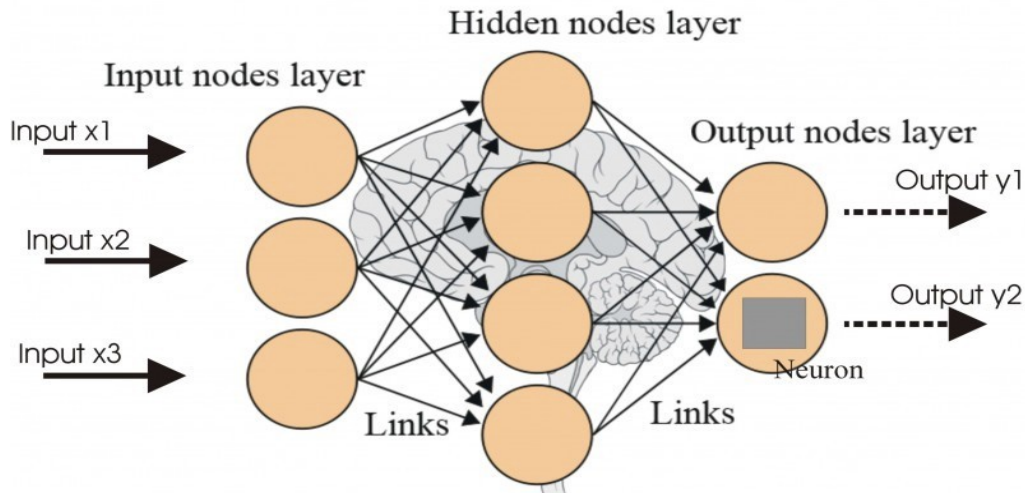
Figure 5.Artificial Neural Network

Random Forest: The random sampling and ensemble strategies utilized in RF enable it to acquire correct predictions as well as higher generalizations. The random forests encompass a big quantity of timber. The higher the variety of uncorrelated timber, the better the accuracy.

The expectation in random forests(RFS) is addressed in figure 7.An test is led with the outcomes accomplished from the SLR (Systematic Literature Review) to arrive at the objectives of RQ1 where we recognize the reasonable AI method for the forecast of COVID-19. The test is additionally kept on building a model of expectation with the chose calculation to decide RQ2 where the variables that impact the forecast are distinguished.[6]

Literature Survey

Li Q et. Al, COVID-19 was launched in Wuhan, China in December 2019, which has spread all over the world. According to the author's report, it does not affect public health.[7]

Liu K et.al., Clinical examinations on patients of various age gatherings of COVID-19 looked at clinical information and proposed that the ethical quality of older patients is bigger than that of young people and moderately aged patients. They likewise accentuated that more established patients have a higher extent of PSI score IV and V in contrast with juvenile and moderately aged patients.[8]

Cobb JS et.al, SIP recognized the impact of COVID-19 on the combined increment pattern of refuge set up. To dissect the pattern of the total development of COVID-19 they have utilized factual examination and an arbitrary approach. The reason that the expanding pattern of COVID-19 is viably decreased after the SIP request of the US government and nations with a higher thickness is getting more profit from the SIP request.[9]

S.J. Fong et.al, Characterize a hybrid system for time arrangement examination to predict the COVID-19 cases. The hybrid model is a mix of a composite.[10]

Tanujit Chakra body et. al, Performed hazard investigation and constant prediction on COVID-19 informational index of France, Canada, South Korea, India, and the UK. The examination has been led in two stages. The first stage covers half breed model of ARIMA and the Wavelet-based model for the on-going determination.[11]

Wei Yu et. al, SVM has been identified as the best classifier for uncertain data analysis by analyzing diabetes and pre-diabetes experimentally.[12]

Methodology

Collection of Data and Arithmetic Study

The overall information was gathered from January 22, 2020, to April 25, 2020, including normal instances, expired instances, and recuperated cases. [13]

Death Frequency = Total Deaths check-out / Total definite cases in the world

Projected Extrapolation with SVM Algorithm

One stroke can be used accurately to classify linearly divisible data as display in equation 1.

$$y = mx + c \tag{1}$$

The remaining x and y with Ψ_1 and Ψ_2 respectively, and finally, the equation is $m*\Psi_1-\Psi_2+C=0$.

If X= (Ψ_1, Ψ_2) and $\omega=(m, -1)$ then equation 2 of the hyperplane:

$$\omega*x*+C=0 \tag{2}$$

To make the assumption hyperplane is well-defined through premise function h_0 as per display below,

$$h_0(x_i)=\begin{cases} +1, if\ \omega*x+C \geq 0 \\ -1, otherwise \end{cases} \tag{3}$$

As described above the hyperplane will be categorized as group +1, as group -1.

$$\min\frac{1}{2}\left\|\frac{\omega^2}{1}\right\| subject\ to\ y(m*x+C)-1, i=1\dots z$$

This piece of content devours the idea of SVM, probably going to be a reasonable procedure aimed at a period arrangement set of data with the world's complete populace. An SVM model was given for design arrangement. SVM utilizes a strategy to as far as possible for mistakes by limiting the limit distance between the preparation information and hyperplane instead of conventional methods of decreasing observational testing blunders. [14] Train to SVM is equivalent to tackle quadratic programming issues with straight limitations while preparing

different organizations includes non-direct streamlining with the danger of stalling out in the nearby minima. In SVM, the critical thinking measure just relies upon a bunch of preparing information, which is known as a help vector [15]
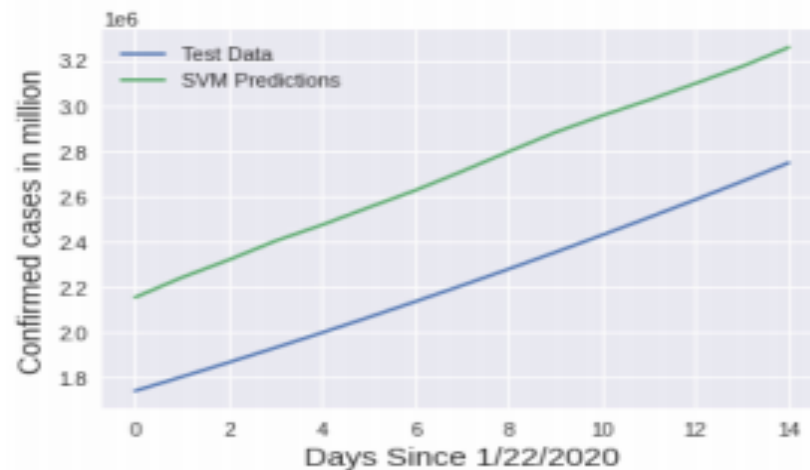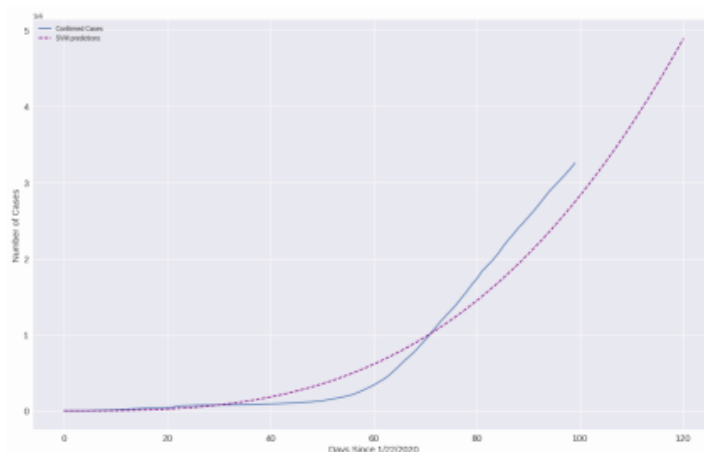

Figure 6.Record by SVM Model

Figure 7.Confirmed against Forecast cases by SVM prototype

Tests set of data by a trained model is utilized to compute errors. This shows parallel lines in figure 6.

The definite cases and SVM prediction cases are typically run concurrently and the SVM forecast is forward of the available data as of April 25, 2020, as shown in graph Figure 7.[16]

Result and Conclusion

The calculation of predicted sufferers of the coronavirus depends on the rankings of declared features with the aid of the usage of SVM calculations. Also, we discover an appropriate hyper aircraft for the usage of RBF. It facilitated us to evaluate the hyper aircraft constraints to better

study assist routes. At a similar period, completed numerical analysis over bar diagrams to differentiate agencies of topics.

A test is performed on python software with the outcomes done from the SLR (systematic literature evaluation) to attain the dreams of RQ1 wherein we perceive the proper gadget learning technique for the prediction of COVID-19. The experiment is further persisted to construct a model of prediction with the selected algorithm to determine rq2 where the elements that impact the prediction are recognized.

One of the greatest qualities of the SVM model is AI procedures utilized for the forecast given its effortlessness.In this examination, the current COVID-19 pandemic circumstance all around the planet is featured, and the Continuing example and seriousness of the episode must determine through the SVM technique. The planned research is an endeavor to acquaint SVM models with gauge COVID-19 cases everywhere in the world

References

1. S. J. Fong, G. Li, N. Dey, R. G. Crespo, E. Herrera-Viedma, Composite monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction, Applied Soft Computing (2020) 106282

2. S. Sasubilli, K. Attanguddi and A. Kumar "APSO based feature extraction approach and SVM based classification for sentimental analysis", Global Journal on Innovation, Opportunities, and Challenges in Applied Artificial Intelligence and Machine Learning Vol. 2, Issue 1-2018.

3. K. Liu, Y. Chen, R. Lin, K. Han, Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients, Journal of Infection (2020).

4. S. Chandrasekaran and A. Kumar Implementing Medical Data Processing with Ann with Hybrid Approach of Implementation Journal of Advanced Research in Dynamical and Control Systems-JARDCS issue 10, vol.10, page 45-52, ISSN-1943-023X. 2018/09/15.

5. D. B. Percival, A. T. Walden, Wavelet methods for time series analysis, Vol. 4, Cambridge university press, 2000

6. S. Sasubilli, K. Attanguddi and A. Kumar "APSO based feature extraction approach and SVM based classification for sentimental analysis", Global Journal on Innovation, Opportunities, and Challenges in Applied Artificial Intelligence and Machine Learning Vol. 2, Issue 1-2018.

7. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia, New England Journal of Medicine (2020).

8. R. Raturi and A.Kumar "An Analytical Approach for Health Data Analysis and finding the Correlations of attributes using Decision Tree and W-Logistic Modal Process", 2019, "IJIRCCE", Vol 7, Issue 6, ISSN: 2320-9801

9. M. A. Al-cases, A. A. Ewers, H. Fan, M. Abd El Aziz, an Optimization method for forecasting con_rmed cases of COVID-19 in China, Journal of Clinical Medicine 9 (3) (2020) 674.

10. S. J. Fong, G. Li, N. Dey, R. G. Crespo, E. Herrera-Viedma, Composite monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction, Applied Soft Computing (2020) 106282.

11. W. Yu, T. Liu, R. Valdez, M. Gwinn, M. J. Khoury, Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes, BMC medical informatics and decision making 10 (1) (2010) 16.

12. T. Chakraborty, I. Ghosh, Real-time forecasts and risk assessment of novel coronavirus (covid-19) cases: A data-driven analysis, Chaos, Solitons& Fractals (2020) 109850.

13. A. Dubey, S. Narang, A. Kumar, S. Sasubilli, V.Díaz (2020), "Performance estimation of machine learning algorithms in the factor analysis of COVID-19 dataset" Computers, Materials & Continua, ISSN: 1546-2226.

14. E. Gorbalenya, Severe acute respiratory syndrome-related coronavirus–the species and its viruses, a statement of the coronavirus study group, BioRxiv (2020).

15. A. Kumar, Surbhi Sharma and Sammah Rasheed "A Shape Mechanism along Text-Based Advanced Face Detection Methodology Using EFM Model", IJIRCCE Vol 7, Issue 2-2019, 23209798.

16. R. Singh, R. Adhikari, Age-structured impact of social distancing on the COVID-19 epidemic in India, arXiv preprint arXiv: 2003.12055 (2020).

17. A.Kumar, D.Hung, D. Le, T.Sairam "Smart Cities Management Using Deep Learning and Iot with Cloud Computing", "Global Journal on Innovation, Opportunities, and Challenges in Applied Artificial Intelligence and Machine Learning Vol. 3, Issue 1-2019.