

---

# COVID PATIENT DATA ANALYSIS SYSTEM USING SVM ALGORITHM IN MACHINE LEARNING

---

<sup>1</sup>Mrs Ramya S, <sup>2</sup>Dhanalakshmi M, <sup>2</sup>Sri Shakthi M R, <sup>2</sup>Swathi J

**1 Assistant Professor, 2 Undergraduate Students**

*1, 2 Department Of Information Technology*

*1, 2 M. Kumarasamy College of Engineering*

## **Abstract.**

COVID-19, is also called as 2019-nCoV, is just to know a pandemic but an endemic illness that has killed over 651,247 people over the world.COVID-19 seems to have no specified treatment at this time, hence it is unavoidable to live with the disorder and its effects. This fact has put enormous demand on the world's in sufficient health care systems, especially in developing countries. Even though there is no efficient, medically tested anti - viral agents' tactic or a certified vaccine to destroy the COVID-19 disease. The consequences of this epidemic, particularly the lockdown tactics, are multifaceted. There are non-clinical methods such as algorithms, data science, machine learning based, and other machine intelligence that may reduce the excessive toll on not only restricted healthcare systems as well as the economy. The spread of the COVID-19 virus is now posing a significant threat to worldwide healthcare. As an effect, the proposed model based on support vector machine learning algorithms offer a talented solution to this challenge that may be accurately predict the infectivity of the disease and identify earlier the affected people.

Keywords: COVID- Corona virus Disease, SVM-Support Vector Machine, NB- Naive Bayes, k-nn - k -Nearest Neighbors, CAD-Coronary Artery Disease

## **1. INTRODUCTION**

The outbreak of developing coronal virus diseases 2019(COVID-19), initiated by SARS-CoV-2019, continue to pose a grim and imminent threat to global health. The endemic began in the democracy of China's Hubei province in beginning Dec 2019

and since then has increase across the world. As from October 2020, there were more than 39,500,000 reported cases of the infection in more over 180 nations, with the total number of persons afflicted likely much higher. COVID-19 has claimed the lives of about 1,110,000 people. Increased demand for hospital beds and crucial shortages of medical equipment, as well as the infection of several healthcare workers, continue to be issues for medical systems in the world. As a result, physicians' able to produce timely clinical choices and efficiently utilize healthcare resources has grown. A study of alternative involvement tactics[5] towards restrict the extend of the virus in the city is ended utilising such a set-up and the robust outbreak simulation stage.

## 2. LITERATURE REVIEW

Romney B Duffey [2] created the technologies of classification and data mining are excellent for classifying data. They have been used in medicine to make diagnosis and analysis so that decisions can be made. On the Wisconsin Breast Cancer (original) datasets, a comparison of results is undertaken between k Nearest Neighbor (k-NN), Naive Bayes classifier (NB)), SVMs (SVM) and Logistic Regression are examples of machine learning methods. The major goal is to classify data according to the accuracy, sharpness, sensitivity, and specificity of the each algorithm in terms of efficiency and efficacy. By looking at their results of the experiment, The SVM algorithm does have uppermost precision (97.13%) as well as the simple error rate. The Issue of Covid-19 is discussed in paper [2]. In this study, the recovery time after infectious disease outbreaks is predicted. By attempting to impose preventive measures such as medical services, isolation, took a step back, and so on, and in particular by adjusting the virus increase control by lowering illness rates when this is effective, the rate of infection declines after reaching a maximum, following what is recognized as the Worldwide Healing bend. When the transmission rate slows down, a phenomenon known as 'flattening the curve,' the danger of communicable diseases decreases. Once the rate reaches a high, successful countermeasures should cause the rate to fall.

PetarRadanliev [3] proposed removal techniques of scientific papers data as from Web of Science Collection was used to gather Covid-19 data, immunity information, and vaccination. The results are then compared against data from all Covid-19 explore areas shared, and each record is examined separately. According to a historical review of scientific information records on virus, pandemics, and death, Chinese colleges have not yet historically been leaders in these fields. We found a few bunch with allusions to exercising, irritation, smoke, fatness, and a variety of other topics.

Jooyeon Park [4] proposed with the advancement of transportation technologies, the worldwide transmission of illnesses continue to speed up. This paper uses GIS to investigate the relationship among transportation and infection, as well as statistical analysis of highway transportation and domestic flights. By employing logic in the SEIR imitation, the local infection could be forecasted when outbreaks occur, allowing for timely responses to disease risks. It may be determined through simulation that places with a huge population also have had the most illness. In conclusion, places with big inflows rather than flows have a high number of infected people, while infectivity is relatively low in some areas with poor transportation.

Vasilis Z. Marmarelis [6] proposed SIR is a framework that separates a population into "Susceptible," "Infectious," and "Recovered/ Removed" fractions and specifies their dynamic interrelationships using first-order differential equations. Estimated parameters result in a progressive decrease in infectivity rate, albeit the most recent wave is likely to be the greatest. The effects of various mitigation strategies can be quantified by analyzing data from particular states or countries.

Petropoulos and Spyros Makridakis [7] divided by predicting necessitates a large amount of past data. There is no such thing as a sure bet when it comes to the future. The forecast will be influenced by the data's dependability and the variables being anticipated. Predictions in medicine are rarely wrong, but their ambiguity is a severe problem. Estimating the future of pandemic and spate is difficult due to the large numeral of instances to be studied. For a basic time series, the prediction of confirmed COVID-19 instances is used. Algorithms from the simple exponential family are used to create the forecasts. Among several forecasting techniques, this provides good forecast accuracy, especially for short series. Exponential smoothing models feature patterns that are either multiplicative or additive.

Hiba Asri, Hajar Mousannif [8] created classifier and data gathering technologies are excellent ways to classify data. They are used for the medical field to make diagnoses and analyses in order to make judgments. The major goal is to classify data according to the efficiency and efficacy of each algorithm on the basis of precision, as well as precision, sensitivity, and specificity. By observing the trial findings, SVM provides the highest efficiency (97.13 percent) with the lowest error rate.

Shaobo He. Yuexi Peng [10] proposed for common data and various manage techniques like as quarantine, peripheral input, and sickbay, COVID-19 is studied using the SEIR epidemic model. Optimization algorithm (PSO) method is used to estimate the parameters using data from Hubei province. New limitation, such as seasonally and stochastic disease, are used in this model. The system's dynamics are influenced by the parameters, and the effects vary depending on the parameter.

### 3. PROPOSED METHODOLOGY

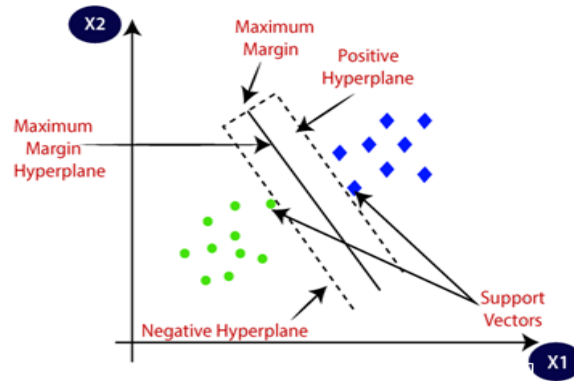


Fig 1 Support vector machine

The purpose of our project is to enable user to correctly predicted COVID-19 illness diagnoses in the proposed system. Age, gender, fever observation, travel history, and clinical data such as cough strength and lung infection incidence were also evaluated. We used our data to test a variety of machine learning techniques and found that the SVM model had the higher precision (>85%) in detecting and picking variables that correctly reflect COVID-19 condition in people of all ages. Support Vector Machines divide data into two groups to categorize it using an N-dimensional hyper plane. In SVM, the predicting variable is considered a feature, and the modified feature is called a converted attribute. Fig 1 describes mapping the data and on the Hyper plane, based on this divisible of data takes place. After a COVID diagnosis, doctors provide recommendations and prescribe medication to patients. The Take Entire is a fully - featured patient treatment system that collects all pertinent patient information. It's utilized to create new users who can interact with the interface. [9] To review the efficiency of varying time series in huge datasets, machine learning (ML)[11-12] approaches are applied. For a large data set, the assumptions are formed by separating the data into many time series. Also provided was a method for retrieve judgment from the improved SVM. The performance of proposed approach was then evaluated by comparing the behaviour[14-15] of the data sets with that of other algorithms. Finally, improved SVM was utilised to estimate the life expectation of lung cancer patients after surgery.

#### 3.1. Architecture Proposed System

Fig 2 describes the architecture of the system in that the first phase is data cleansing. The dataset had a few null values in a few columns, which were handled by removing the rows matching to those columns because these values couldn't be substituted with

anything else relevant. The second phase is feature selection. Because the dataset had several columns, this step involved picking the most important ones to employ for learning algorithm. The data was randomly partitioned into training data and real data in the third stage. The percentage of test data used for testing and validation in the current study ranged from 15% to 40%. The fourth step is to train the model, the best parameters for a given classification method were chosen and utilised to train a model. Step five is to evaluate the prototype that was implemented to the testing of data, and the model's accuracy was assessed by analysing the confusion matrix that was generated. The model was assessed in the sixth stage by purposes of calculating error metrics. Model prediction was matched to sample data in step seven, and the projected value was validated.

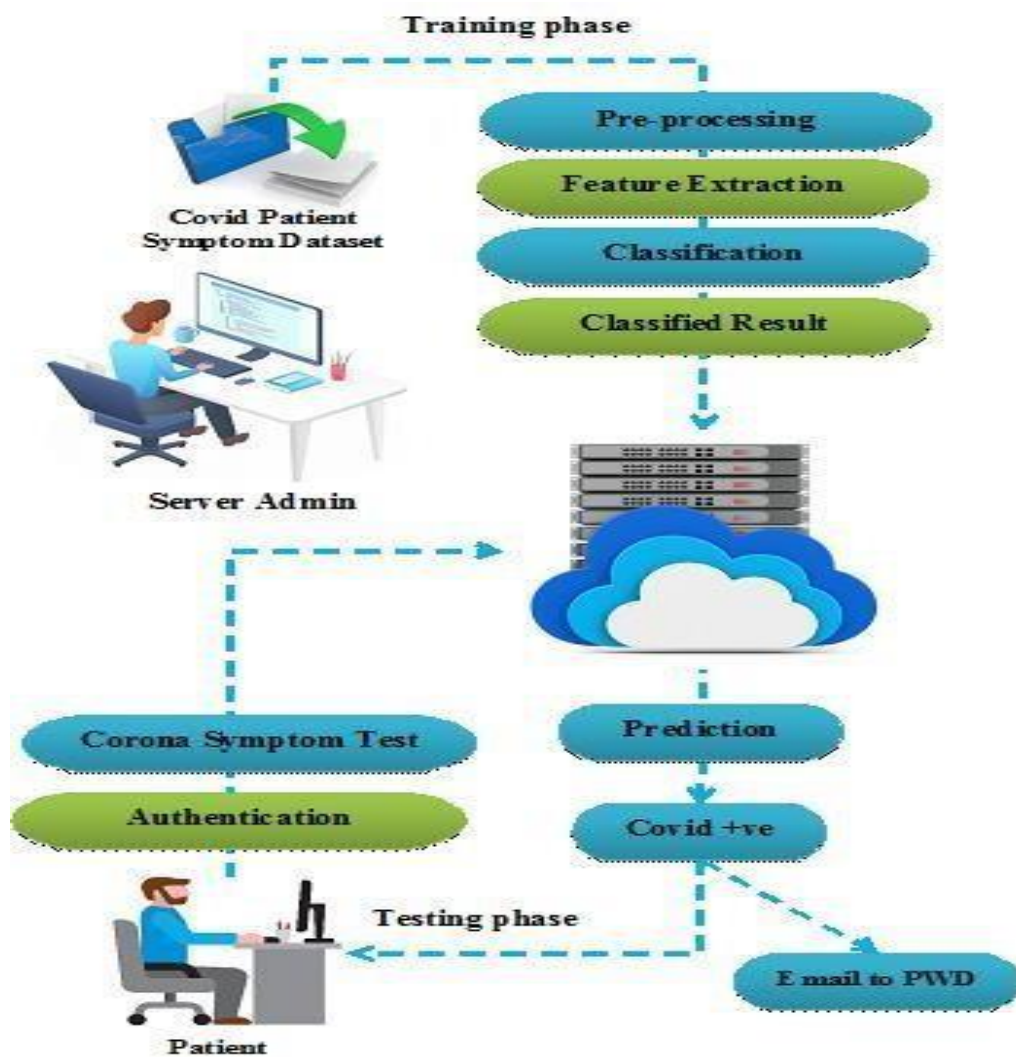


Fig 2 System Architecture

6

#### 4. ADVANTAGES

- SVM can provide an a prior probability of disease, which can then be used to designate patients for angiogram. Other patients will save money and time as a result of this. They are also free of the angiography's negative effects.
- It could lead to earlier discovery and, as a result of lower mortality rate.
- Using SVM, latent patterns in the data collected can be extracted. This could lead to the development of new approaches for early diagnosis of ailments such as coronary artery disease.
- This is a novel method that builds on current research to derive quick and precise diagnostics.
- The method significantly out performs other published research in this area due to its superior accuracy.

#### 5. RESULTS AND EXPERIMENTS



Fig 3 Home Page

# COVID19



User Login

Username

Username

Password

Password

Login

[Register here](#)

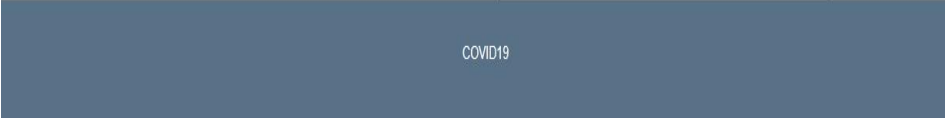


Fig 4 User Login

# COVID19



Doctor Login

Username

Username

Password

Password

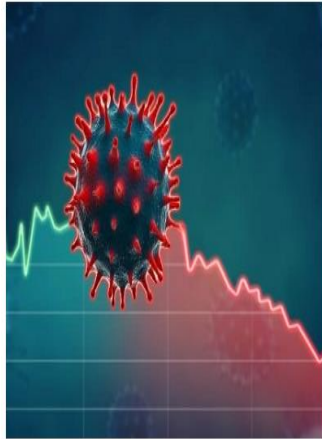
Login

[Register here](#)



Fig 5 Doctors Login

# COVID19



Admin Login
Username
Username
Password
Password
<input type="button" value="Login"/>

Fig 6 Admin Login

# COVID19

Testing
<p>Covid Positive, Call your health care provider or 911 for immediate guidance.</p> <p>Accuracy: 69.91%</p>
Height (Cm)
Height
Weight (Kg)
Weight
Diabetics
Yes

Fig 7 COVID Test



## COVID19

[Home](#) [Suggestion](#) [Test Report](#) [Logout](#)

Test Report				
Sno	Date	BMI	Symptoms	Status
1	31-03-2022	BMI: 20.4518, You are healthy	Fever: Yes Headache: Yes Vomiting: Yes Tiredness: Yes Dry-Cough: No Difficulty-in-Breathing: No Sore-Throat: Yes Pains: Yes Nasal-Congestion: No Runny-Nose: Yes Diarrhea: Yes Known contact: Yes	Covid Positive
2	31-03-2022	BMI: 20.4518, You are healthy	Fever: Yes Headache: Yes Vomiting: No Tiredness: Yes Dry-Cough: Yes Difficulty-in-Breathing: No Sore-Throat: Yes	Covid Negative

Fig 8 Test Report

## COVID19

[Home](#) [Suggestion](#) [Test Report](#) [Logout](#)

Suggestions and Prescription			
Sno	Suggestions	Prescriptions	Date
1	take rest for 14 days	eat healthy foods	31-03-2022
2	serdx	etdr	31-03-2022

COVID19

Fig 9 Suggestion and Prescriptions

Fig 3, is the home page where the patient doctor and admin have their login page and the registration page. Fig 4,5,6 are the login pages where the patients, doctors and admin can login with their username and the password or giving new user they can register and login. Fig 7 is the COVID test page in that the doctor tests the patients by giving the parameters given in that page and submits it after submitting the prediction of COVID disease is done. Fig 8 is the test reports of the patients and Fig 9 is suggestion and prescriptions given by the doctors are able to view in the user login.

## 6. CONCLUSION AND FUTURE ENHANCEMENT

Global health is currently being threatened by the outbreak of a COVID-19 pandemic. The develop of ways to detect infected people as soon as feasible is critical to stop this from happening. Given the lag in symptom manifestation, this may be problematic; nevertheless, machine in the event of a pandemic, training methods offer a viable solution that can be adopted rapidly and cheaply. The classification methods to look at information criterion and analyze performance using machine learning techniques on variety of clinical choices of sick people with COVID-19 diseases in a data set from India. The doctors' can improve ability to recognize patients early on by forecasting the probability of COVID-19 infection using predictive clinical signs. Several of the classifiers, on the other hand, did not produce reliable findings, perhaps because, despite their unquestionable exactitude, they produce done-sided answers for these datasets. Imaging new features were not used in this investigation to maximize speed and convenience. Imaging features can be compared to laboratories and pro features in future studies to see how effective they are at predicting outcomes.

## 7. REFERENCES

- [1] 2020 COVID-19 Future Forecasting Using Supervised Machine Learning Models” IEEE ACCESS pp 101489-101499. Furqan Rustam , Aijaz Ahmad Reshi , Arif Mehmood , Saleem Ullah , Byung-Won On, Waqar Aslam and Gyu Sang Choi may.
- [2] The Case of CoVid-19 vol.8 IEEE Engineering in Medicine and Biology Society Section pp 110789-110795. Romney B Duffey and Enrico Zio may. 2020 Analysing Recovery From Pandemics by Learning Theory:
- [3] 2020 Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development - In the first wave of the Covid-19 pandemic vol.3 Science direct pp 1121-1132, .Petar Radanliev, David De Roure and Rob Walton jun.
- [4] Jooyeon Park, Jinhwa Jang and Insung Ahn Oct. 2017 Epidemic Simulation of H1N1 Influenza Virus using GIS in South Korea IEEE, 2017 International Conference on Information and Communication Technology Convergence (ICTC) pp 58-60
- [5] Evaluating strategies for pandemic response in Delhi using realistic social networks IEEE International Conference on Healthcare Informatics pp 121-130 Huadong Xia, Kalyani Nagaraj, Jiangzhuo Chen and Madhav Marathe sep. 2013

- [6] Vasilis Z. Marmarelis 2020 Predictive modeling of Covid-19 data in the US: Adaptive phasespace approach IEEE Open Journal of Engineering in Medicine and Biology, pp 207-211
- [7] Fotios Petropoulos and Spyros Makridakis Mar.2020 Forecasting the novel coronavirus COVID-19 ( ARGENTINA Lidia Adriana Braunstein, Universidad, Nacional de Mar delPlata ) pp 1-8
- [8] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime and Thomas Noel 2016 Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis
- [9] Statistical and Machine Learning forecasting methods: Concerns and ways forward Plos one (Alejandro Raul Hernandez Montoya,UniversidadVeracruzana MEXICO) 1-26 Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos Mar.2018
- [10] Shaobo He. Yuexi Peng and Kehui Sun 2020 SEIR modeling of the COVID-19 and its dynamics Springer Nature B.V
- [11] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [12] Zięba, M., Tomczak, J. M., Lubicz, M., & Świątek, J. (2014). Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied soft computing*, 14, 99-108.
- [13] Hemanth Reddy, K., & Saranya, G. (2021). Prediction of cardiovascular diseases in diabetic patients using machine learning techniques. In *Artificial Intelligence Techniques for Advanced Computing Applications* (pp. 299-305). Springer, Singapore.
- [14] Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: an experimental analysis. In *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies* (pp. 51-56).
- [15] Shaikh, F. J., & Rao, D. S. (2022). Prediction of Cancer Disease using Machine learning Approach. *Materials Today: Proceedings*, 50, 40-47.

## BIOGRAPHIES



Ms. S. Ramya is currently working as Assistant Professor, Department of Information Technology in M. Kumarasamy College of Engineering, Karur, Tamilnadu, India. She has completed Master of Engineering in Computer Science and Engineering in Anna University in the year 2011. She possesses nearly 12 years of experience in teaching and published several papers in international journals. Her areas of interest are Cloud Computing, Big data, Cluster Computing and Blockchain.



Ms. M.Dhanalakshmi is currently pursuing 4<sup>th</sup> year B.Tech, Department of Information Technology in M. Kumarasamy College of Engineering, Karur, Tamilnadu, India. Her areas of interest are Cloud Computing.



Ms. M.R.Sri Shakthi is currently pursuing 4<sup>th</sup> year B.Tech, Department of Information Technology in M. Kumarasamy College of Engineering, Karur, Tamilnadu, India. Her areas of interest are Web Development.



Ms. J.Swathi is currently pursuing 4<sup>th</sup> year B.Tech, Department of Information Technology in M. Kumarasamy College of Engineering, Karur, Tamilnadu, India.. Her areas of interest are Machine Learning.