Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 05 Issue 02
2023

Research Article
https://doi.org/10.56979/502/2023

Analysis and Evaluation of Coronavirus In-Patient Prediction Model Algorithm in Machine Learning

Amina Khalid[1], Rabia Afzaal[2*], Sheikh Luqman Ijaz[3], and Sadaf Hussain[4]
[1]Department of Software Engineering, The University of Lahore, Lahore, Pakistan.
[2]Department of Information Technology, Lahore Garrison University, Lahore, Pakistan.
[3]School of Software Engineering, Dalian University of Technology, China.
[4]Department of Computer Science, Lahore Garrison University,Lahore, Pakistan.
[*]Corresponding Author: Rabia Afzaal. Email: dr.rabiaafzaal@lgu.edu.pk

Received: July 28, 2023 **Accepted:** September 08, 2023 **Published:** September 17, 2023

Abstract: **C**oronavirus family belonged to the significant infection in both humans and animal's pathogens. The current universal of the Coronavirus respiring ailment (COVID-19) shows a state of worldwide public outbreak. Coronavirus affliction (COVID-19) is spread through large droplets produced during coughing and sneezing by symptomatic patients, as well as asymptomatic individuals before beginning of their symptoms. The bug beginning the affliction is named harsh severe respiring condition coronavirus (SARS-CoV-2) and has happened flowing during the whole of the experience, producing dispassionate exhibitions grazing from asymptomatic cases to harsh severe respiring condition and obliteration, particularly of things accompanying comorbidities [*hypertension, asthma, and diabetes*]. Identify a suitable algorithm model that can predict either a patient has COVID-19, the model was trained on data from infected SARS-CoV-2 inmates. Machine Learning a model identify is projected to label best choice invention; to judge the traits that influence the forecast model. To find highest in rank treasure for the forecast model an orderly composition review is administered. The number of algorithms recognized from experiments and deep study holds SVM and Logistic Regression namely ideal for prophecy. When the result of the identify was reasoning refine was found that SVM Model acted better than the alternative invention. Two classifier, Support Vector Machine and Logistic Regression, were selected; both have an accuracy of more than 90%. The SVM has 96.49% and LR 95.92% accuracy. These results can be used to take corrective measure by different governmental hospitals bodies. The infrastructure of methodology for forecasting infections disease can make it easier to fight COVID-19.

Keywords: Support Vector Machine & Logistic Regression Model Algorithm; COVID-19 Symptoms Prevention Disease; Performance; Accuracy, Comparison of Algorithm.

1. Introduction

The Coronavirus is a broad family of viruses that have been linked to illnesses ranging from the common cold to more dangerous infections, including Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) [1]. The 21st century has already witnessed significant epidemics caused by the SARS-CoV and MERS-CoV. However, the world population still suffers from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 [2]. These are three highly pathogenic viruses that belong to the order Nidovirales and the family Coronaviridae [3]. The first cases of SARS occurred in November 2002 in Guangdong, China [4,5] and had the identification of the causative agent SARS-CoV-2 in April 2003 [4]. From November 2002 to July 31, 2003, the World Health Organization (WHO) recorded 8096 cases of SARS in 27 countries, with 774 deaths attributed to the disease [6,7]. After this period, no new cases were detected, being then considered the end of the great epidemic of SARS [3].

Considering that emerging viruses are continually causing global public health threats, the of Artificial Intelligence (AI) approaches and open-source datasets, promoting earlier detection and fast governmental measures, can help to reduce the health and economic impacts of epidemics and pandemics [5,8]. AI has played a crucial role in contact tracing and predicting the spread of COVID-19. By analyzing large amounts of data, including demographic information, travel history, and social interaction, AI algorithms can identify potential infection sources and track the spread of the virus more effectively. These models help governmental authorities make informed decisions regarding quarantine measures, resource allocation, and targeted interventions[6]. By accurately predicting hotspots and potential outbreak areas, AI has helped control the spread of the virus and minimize its impact on communities [5]. AI has significantly expedited the vaccine development and drug discovery process for COVID-19[7]. Machine learning algorithms have been utilized to analyze vast amounts of genetic data, identify potential drug targets, and predict the efficacy of various compounds against the virus[8]. AI algorithms can also simulate the interaction between drugs and viral proteins, accelerating the process of finding effective treatments[9]. These AI-driven approaches have contributed to the rapid development and approval of several COVID-19 vaccines and therapeutics, aiding the global fight against the pandemic [10].

Globally, as of June 28, 2023, there have been 767,518,723 confirmed cases of COVID-19 had been registered, including 6,947,192 deaths, reported to WHO[11]. Despite not representing a pandemic anymore due to the rapid development of anti-SARS-CoV-2 vaccines, currently, there are 13,461,344,203 COVID-19 vaccine doses have been administered, and the virus still circulates throughout the world, affecting approximately 213 countries[12,13]. Continue Testing is not practicable in all countries, owing to time and expense restrictions, since the number of patients on screen for COVID Virus is continually spreading [10]. Meanwhile, the increase in worldwide electronic health data has represented an enormous potential to mitigate public health problems caused by future disease outbreaks.

Therefore, using new machine learning approaches to predict COVID-19 in patients can help to reduce the time delay for the results of the medical tests and prepare the public health system to give proper medical treatment to control the disease. Here, the propose a machine-learning model that predicts a positive SARS-CoV-2 infection from common symptoms tests by asking basic questions [11]. To develop such a model, we performed a literature study alongside an experiment to identify a suitable algorithm and assess the features that affect the prediction model.

In this study are identify a Machine Learning model for predict COVID-19 in infected patient. It is also working discover characteristics form patients' clinic information that may affect the COVID-19 prediction outcome. This study does not include external factor such that other environmental variables that might affect the outcomes.

1.1 Interpretation of Support Vector Machine Over Logistic Regression Model Used for Solving

1.1.1 Suport Vector Machine (SVM) Algorithm

Support Vector Machine identify an N-dimensional classification system for a hyper data plan that divides it into two categories [3].

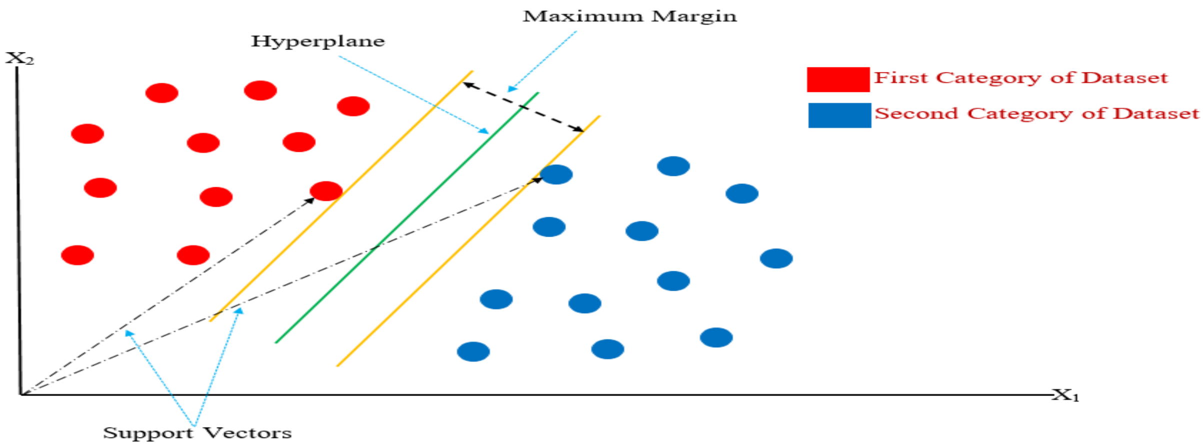

Figure 1. SVM

Figure-1 depicts a typical support vector machine example. Logistic Regression a predictive analytic approach that is based on the concepts of probability, logistic regression is one of the Machine Learning techniques used to address categorization issues.

In SVM, the predicted variable is referring to an attribute, and transformed points are referred to as features. Features selections refer to the selection of the process of the data that is most representative [4]. A vector is a collection of features that describe a single case. The final aim of SVM modelling is to determine the optimum hyperplane that divides the cluster, the target variable on one side of the plane and the other category on the other. Those support vectors that are near the plane are the strongest ones [24].

SVM are particularly useful when dealing with complex decision boundaries, non-linear separations, and high-dimensional data. They can also handle cases where the classes are not linearly separable by using kernel functions to implicitly transform the data into a higher-dimensional space where separation is possible.

1.1.2 Logistic Regression Model Algorithm

Logistic Regression is a simple yet effective algorithm used primarily for binary classification tasks. Despite its name, it is a classification algorithm, not a regression algorithm. Logistic Regression models the probability that a given input belongs to a certain class using the logistic function (sigmoid function). It estimates the log-odds of the probability, which can then be converted to a probability value using the sigmoid function. Logistic Regression is particularly suitable when the relationship between the features and the target variable is assumed to be linear on the log-odds scale. It is also commonly used when interpretability of the model is important, as the coefficients of the features can be directly interpreted as the impact of each feature on the log-odds of the outcome.

Model Assumption: Logistic Regression assumes that the log-odds of the probability of the positive class is a linear combination of the input features. Mathematically, this can be expressed as:

$$Log\text{-}Odds\ (Logit) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k {*} x_k$$

Where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of the features, $x_1, x_2, \ldots, x_k$, and k is the number of features.

Logistic Function (Sigmoid): The log-odds from the previous step are then passed through the logistic function (sigmoid function), which maps the log-odds to a probabaility value between 0 and 1. The sigmoid function is difined as:

$$P\left(\frac{y=1}{x}\right) = \frac{1}{1+e(-logit)}$$

Where $P\left(\frac{y=1}{x}\right)$ is the probability of the positive class given the features x, and logit is the value obtained from the linear combination of coefficients and features.

In summary, if you have a complex dataset with intricate decision boundaries, SVMs might be a better choice. However, if you prefer a simpler model, interpretability, or if your data exhibits a more linear separation, Logistic Regression could be more appropriate. Ultimately, the decision should be based on empirical evaluation, considering factors such as performance, interpretability, computational resources, and the specific characteristics of your data.

1.2. Basic Methodology Infrastructure of Evaluation in COVID Infected Data Analysis.

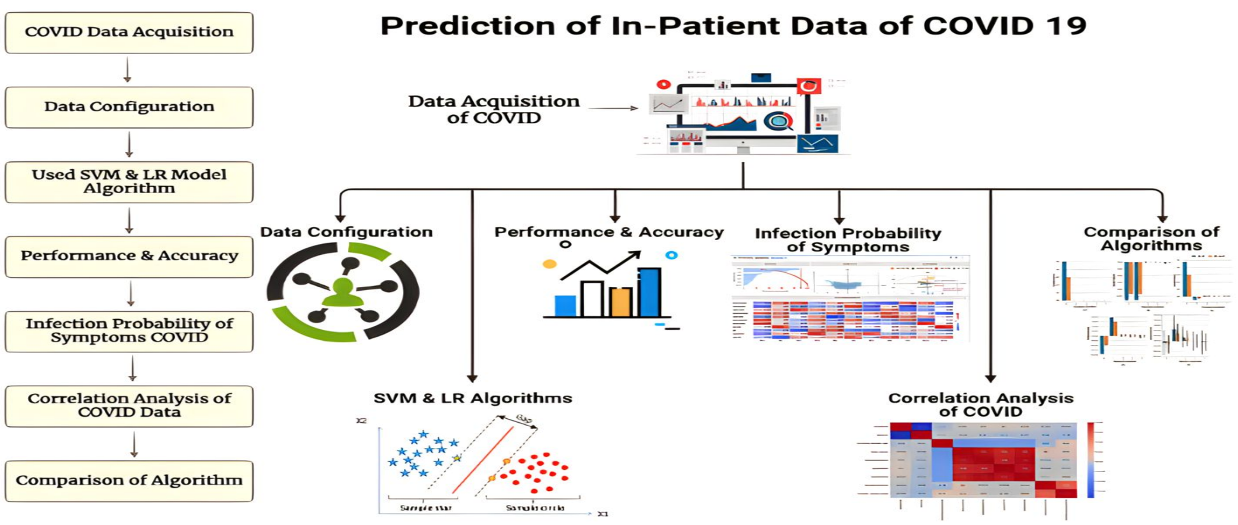

Figure 2. Prediction of In-Patient Data of COVID 19

2. Materials and Methods

2.1 Experiment Phase-I

Within the confines are carried out using the result of the (Systematic Literature Review) SRL to fulfil aims of RQ1 in which are determine better machine learning approach for COVID-19 predictions these work is then continuing to create prediction model using the chosen method to figure out RQ2 the elements influencing the prediction are discovered.

Jupyter Notebook – is a free & open-source online application that allows you to create and share documents that include live code, equation and graphics. Visualizations well as narrative pros Data cleaning, transformation, numerical simulation, and statistical modelling there are all example of machine learning are some of application[25]. For online quick tutorial is available at: https://nbviewer.org/github/pauleve/pint/blob/mast er/notebook/quick-tutorial.ipynb

2.1.1 Data Collected Instrument

Data Collection is an important and time-consuming procedure. Whatever the subject of study accuracy in data collection is essential for long-term success cohesiveness. Because patients' data gathering was a rigorous and time-consuming process since clinical information was not publically available procedure. Due to current scenario at hospital with a high intake of patient with COVID-19 many hospital and health institute in the world was consulted to obtain the most precise data but we could get data from different hospital [26,27]. The datasets consist of 10 million real patient data.

Saidu Hospital Swat provided the data set utilized to train the algorithm to predict COVID-19 [28]. The data collection includes information about COVID-19 hospitalized patients. It contained demographic information, symptoms and signs, prior medical data and laboratories values obtained from a computerized database [29]. The initial data set included specific on the medicines used by doctors to treat the illness. Because our model does not require such information, those fields have removed. The dataset consists of multidimensional data that has been merged [30].

Table 1. High Infected Patients

No of Features	Name of Features
1	Age
2	Body-Pain
3	Diff-Breath
4	Fever
5	Infection-Probability
6	Runny-Nose

2.1.2 Data Preparation & Implementation

Data Pre-processing is a crucial step in the creation of a machine-learning model. Data obtained is frequently uncontrolled with us of range numbers, missing values, and so on. Such data might indicate some unsatisfactory experiment outcomes.

- Missing Value Imputation – Our data missing value were handle by using a straightforward imputer from the Sklearn Python package. Its mean approach is used to replace missing values [31].
- Encoding Categorical Data – Since the DataFrame contains objects, we must transform them into a machine-readable format.

DataFrame.head(n=29998), return the first five and last rows data.

Table 2. Up and Down Machine Readable Checkpoint Solution of the Following Infected Patient's Records.

S.No	Fever	Body-Pain	Age	Runny Nose	Diff-Breath	Infection Probability
0	101.202600	1	26	0	1	0
1	99.371946	0	45	0	0	-1
2	100.637629	1	40	1	-1	0
3	99.927762	0	39	1	0	0
4	98.583098	0	40	0	1	1
29994	99.530594	0	18	1	1	-1
29995	101.315894	1	41	0	0	0

29996	98.866731	0	31	1	-1	1
29997	99.209840	-1	35	1	1	1
29998	100.530192	0	24	1	-1	0

The implementation is carried out in Python Jupyter Notebook the default integrated development and learning environment for the language the experiment was carried out in several stages, which are listed below:

Following records gathering, the patient's data is split into records set of 29998, recorded data

Because two columns' in the DataFrame are objects. We transform them into a machine readable format.

The prediction accuracy of each algorithm is compared and assessed for each record set in order to pick the optimal method for the datasets.

2.2 Experiment Phase-II

2.2.1. Configuration of Algorithms

The configuration of the algorithm is discussed in this section. Change to the algorithm's setup can have an impact on the results [32].

- Support Vector Machine – SVM

SVC (Kernel = 'linear', random state = 0)

- Logistic Regression:

It is a method of predicting the future based on past data,

clf = Logistic Regression () clf.fit (**X** train, **Y** train) clf.fit (**X** train, **Y** train)

clf.fit (**X** train, **Y** train)

2.2.2. Performance & Accuracy

Measuring the performance of a machine-learning model is a crucial task, because our approach necessitates prediction. We chose accuracy as the prediction outcomes. The metric utilized to evaluate the algorithms in this research is accuracy [33]. It is the most often used performance statistic for assessing classification methods. This metric tells us whose model is best to finding patterns in the training dataset in order to make in the unknown text, better prediction dataset [34].

Table 3. Collected Datasets [35].

Date	**P-Name**	**Age**	**Sex**	**Origin**	**P-Phone**	**No-**
30/01/2020	Yaser Khan	26	M	Odigram	3.02E+09	Quar-
20/03/2020	Dr. Usman	45	M	Saidu Sharif	Nil	Quar-
23/03/2020	Gouhar Ali	28	M	Saidu Sharif	Nil	Quar-
23/03/2020	Usman	29	M	Saidu Sharif	Nil	Quar-
24/03/2020	Dr. Sajad	40	M	Chakisar	3.33E+09	Isola-
24/03/2020	Khalid	60	M	Rahim	0346	Hospi-
25/03/2020	Jawad	29	M	Manyar	3.43E+09	Hospi-
24/03/2020	Fozia Am-	35	F	Cham	3.43E+09	Quar-
27/03/2020	Syed Am-	40	M	Cham	3.43E+09	Quar-
27/03/2020	Ishal Khan	5	F	Cham	3.43E+09	Quar-
27/03/2020	Manal	3	F	Maho Patail	3.43E+09	Quar-
27/03/2020	Rokhana	12	F	Odigram	Nil	Quar-
27/03/2020	Amjad	25	M	Balakot	0346	Isola-
26/03/2020	Junid Khan	30	M	Amankot	000 000	Hospi-
30/03/2020	Fazal	70	M	Balakot	0300	Hospi-
31/03/2020	M. Anwar	48	M	Balakot	0345	Isola-
31/03/2020	Bawar	70	M	Balakot	0313	Isola-
31/03/2020	Alam Zab	60	M	Beha	000 000	Isola-
31/03/2020	Baghi Sha-	60	M	Barthana	000 000	Hospi-
1/4/2020	Imran	25	M	Matta	0.00E+00	Isola-
1/4/2020	Ali Sartaj	60	F	Kanju	4.00E+00	Isola-
1/4/2020	Bacha	55	M	Mingora	000 000	Isola-

Patient ID	Report-	Report-	Reported
COVID19/PAK/	Peshawar	Administra-	Dr. M. Isaq
COVID19/PAK/	Peshawar	Administra-	Dr. Akram
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Shangla	DHO	Dr. Sajad
COVID19/PAK/	Swat	DHO Swat	Dr. Najib
COVID19/PAK/	Swat	Secretary	Dr. Riaz
COVID19/PAK/	Swat	Secretary	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Peshawar	Administra-	Dr. Ahmad
COVID19/PAK/	Swat	DHO Swat	Dr. Naeeb Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Najib Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Najib Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Najib Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Najib Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Najib Ul-
COVID19/PAK/	Swat	DHO Swat	Dr. Akram
COVID19/PAK/	Swat	DHO Swat	Dr. Akram
COVID19/PAK/	Swat	DHO Swat	Dr. M. Akram

3. Results

3.1. Infection Probability Symptoms & Count Plot

According to Fig Count paragraphs shows that the patients suffered from COVID-19 and 2575 total patients. In which there are almost 375 patients that had fever and in the figure that show the 95% have normal fever and the other patients have high fever Similarly, the 94% patients have no body pain symptoms but that 1.0% have the body pain symptoms of that in COVID-19 [36]. The symptoms of COVID-19 shows on the different age of the people like according to age 1 to 25 and 40 to 65 have less range but according to the age 25 to 40 and 85 to 100 have high symptoms of COVID-19. Also there are another symptoms occurs in COVID-19 like Runny Nose and Infection Prob both have the similar % of patients. But in the Breath there is much more difficulties, if there are same number of patients that are suffering from this and almost equal people that are save from this disease [37]. According to the checking of the COVID patient the machine learning model is trained for that and after that we predict patient very easily and in future it is very helpful for us. The Figure 3, shows all the patients % in which how much patients are under this disease and how many are save from this. The machine learning trained model give very easiness to checking the patient COVID symptoms. The figure is shown below.

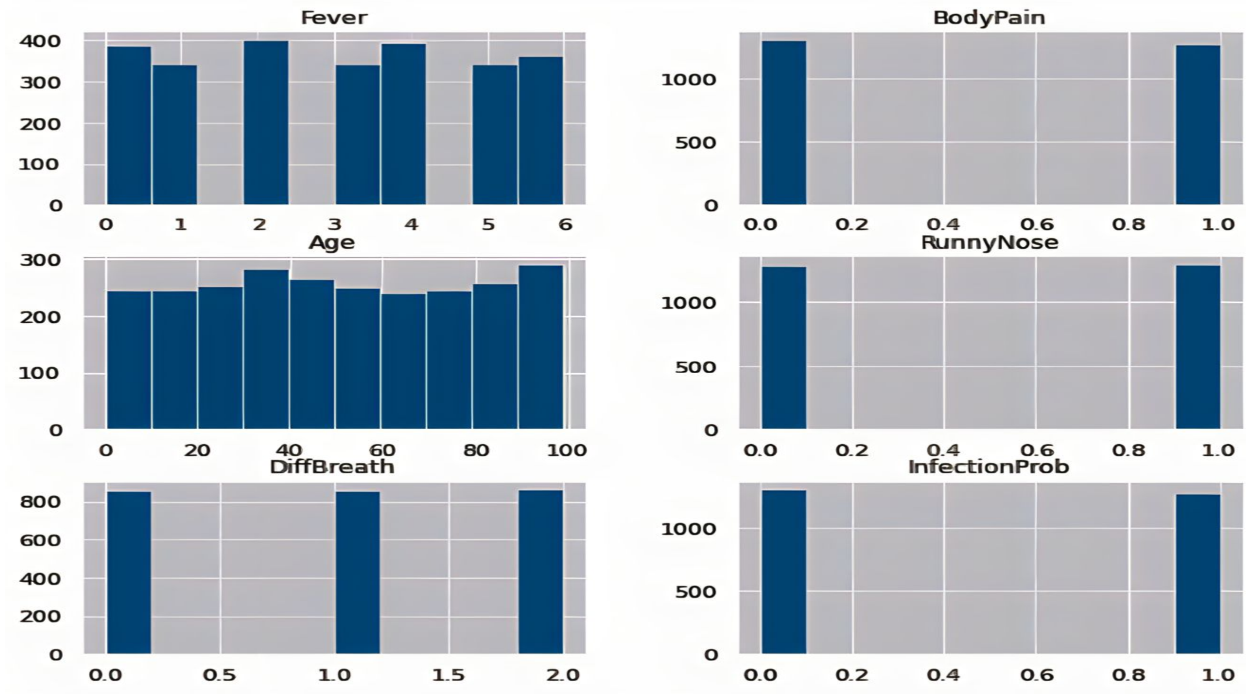

Figure 3. Gather data that includes information about individuals, their symptoms, infection probabilities, and symptoms counts Plot represent the probability of patients suffering from COVID-19 with relevant symptoms.

3.2. COVID-19 Correlation Coefficients & Matrix Datasets Cleaning Operation

Table-4 lists the Pearson, Spearman, and Kendallau correlation coefficients. Features like mask-wearing and market sanitization are not taken into account because they had null values [38,39]. These characteristics are eliminated because they are highly associated with chronic lung, cardiac, and digestive conditions, as well as running noses [40].

Table 4. Different correlation coefficient of the given dataset. Correlation coefficients are used to assess the strength and direction of the linear relationships between pairs of variables.

Stats	Pearson	Spearman	KendallTau
	Pearson	Spearman	KendallTau
Highest Positive Correlation	0.048	0.048	0.045
Highest Negative Correlation	-0.036	-0.036	-0.036
Lowest Correlation	0.0	0.0	0.0
Mean Correlation	-0.006	-0.006	-0.005

	Fever	BodyPain	Age	RunnyNose	DiffBreath	InfectionProb
Fever	1.000000	-0.013039	-0.021883	0.011719	0.000281	-0.007187
BodyPain	-0.013039	1.000000	-0.012199	-0.036059	0.005747	-0.031646
Age	-0.021883	-0.012199	1.000000	0.029824	0.006668	-0.035496
RunnyNose	0.011719	-0.036059	0.029824	1.000000	0.047533	-0.017423
DiffBreath	0.000281	0.005747	0.006668	0.047533	1.000000	-0.027542
InfectionProb	-0.007187	-0.031646	-0.035496	-0.017423	-0.027542	1.000000

Figure 4. Obtained correlation matrix for the given dataset after data cleaning operation.

3.3. Number of Cases of the COVID-19 Diagnoses

Within the confiense to the graph there is a 2000 count down in COVID-19 of the people in which there are the range of 0 to 450 is not COVID pattient but the total number that the COVID symptoms show are almost 95% that's show in the graph. We use the machine learning model to predict the COVID-19 symptoms easily to save the people lives that's why we use the SVM and other algorithm technologies are used. In the graph its clear that there are a number of patient who are dignosed from this disease. After apply the algorithm through machine learning we defined the number of patients through this [39]. Number of Casese circle diagram there are showing the % of the dignoses people and the people who are safe from this. There are the number of clases and show the total amount of the people that are dignose from coronavirus. There are the 14.5% people that are not the Covid patients but the other 85.5% that are the dignoses from this the symptoms like body pain, diff-breath, fever, infection-probability and runny nose.

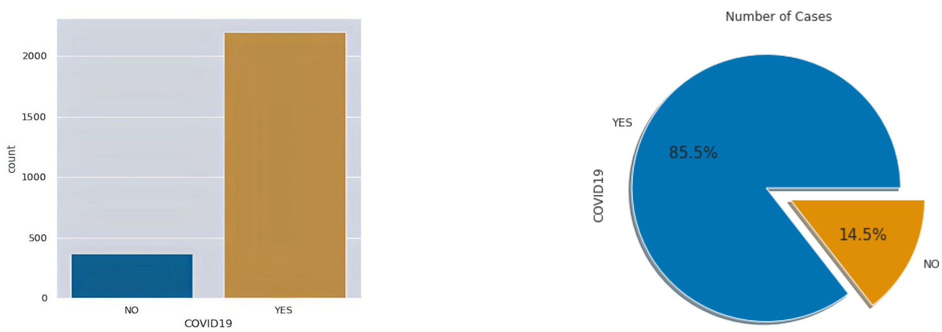

Figure 5. The number of cases has been represented the count people of diagnosis from coronavirus. Found the result of plot and circle graph. Yes, represent the range of infected patient's and No, represent the healthy states.

3.4. Experiment Overview

To address RQ1, Systematic Literature Review (SLR) is performed. Which machine learning approach is best for predicting COVID-19? The SLR's objective is to discover the best appropriate algorithms that will aid in the accurate prediction of COVID-19. I used two algorithms here one is SVM and another Regression in these algorithms regression model is the best model for prediction of COVID-19. The results are the following!

InputFeatures = [104, 1, 90, 1, 1]

InfProb = clf.predict_proba([InputFeatures])[0][1]

InfProb

When the program run the infection probability on the basis of the given symptoms value is 0.5492305552082519 and the result of the algorithm is approximately equal to **99%.** In the Systematic Experiments, several works utilization machine learning algorithms in the healthcare area were found (SLR). The majority of the papers compared machine learning approaches. According to [14, 35, 18] and [23], clinical datasets necessitate a comparison of algorithms to choose the best one. When accuracy is the performance criterion, the most commonly used algorithm are Regression model.

3.5. Experiment Algorithm

The findings of experiment are presented in this chapter. The performance of algorithm describes in section 4.2.5 are used to assess the effectiveness of algorithm chosen. The following two algorithms were found as the best suited for classification of predicting COVID-19.

- SVM- Support Vector Machine
- Logistic Regression

Each method listed above was train using the dataset that was gathered and the results were interpret. Every algorithm performance was assessed at various phases of the training set. Every algorithm was train using records sets of 1,000 records, 15,000 records, 17,000 records, and 20,000 records, in that order. This experiment is carried out to determine which COVID-19 predictions would be best served by this method. Also because the data has been divided into smaller groups and larger records was determine which method will perform best with the various dataset available.

3.6. Support Vector Machine Result

The Support Vector Machine (SVM) algorithm was trained on each record set to determine its correctness at each step. Using k-fold cross validation, the data was split into training and test data at all phases (5-folds). SVM has a 96.00 percent accuracy rate. The accuracy for each batch of data attained by the Support Vector Machine (SVM) method is shown in table.

Table 5. The number of patient records in identified the accuracy result with SVM algorithm.

Number of Patient Records	Accuracy
60000	94.73%
10,000	96%
12,000	97.36%
15,000	97.18%
18,000	97.71%
20,000	98.33%

3.7. Logistic Regression Model Result

Let's apply Logistic Regression on it

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
#Fit the model
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
#Score/Accuracy
acc_logreg=model.score(x_test, y_test)*100
acc_logreg

95.92233009708738
```

Figure 6. Accuracy of regression model. The chart in clearly display the regression model classification accuracy for each record set.

3.8. Comparison of SVM & LR Algorithm Result

The total accuracy results from the tests are summarized for comparison in Figure 5 and 6, in these models regression model give us 96% accuracy.

Table 6. Summarized comparison algorithm and accuracy.

	MODEL	SCORE
0	Support Vector Machine	96.494090
1	Losgistic Regression	95.922330

3.9. Analysis of Experiment

Many studies have purposefully undertaken a clear comparison of various machine learning algorithms, but the collection could not be reached, all proposed a comparison model SVM & Regression Model were chosen to undertake an experimental assessment to determine the best method to predict COVID-19. The experiment is divided into two stages: Assessment of machine learning methods chosen from the Literature Review and Experimental work to address RQ1. Importance Feature generation for detemining the influence of a certain feature on the prediction of COVID-19 which is used to answer RQ2.

3.10. Analysis of Evaluation Phase I and II

Quantitative data compared to estimated accuracy for each machine learning method in order to determine the best algorithm for COVID-19 predications. When compared to Support Vector Machine and other methods, the regression model performs significantly better in terms of accuracy and predication. When compared to other algorithm, the Regression Model produced superior results with less and bigger training data records. When the quantity of records was doubled, there was no discernible difference in prediction accuracy.

Phase_2 is being carried in order to responds RQ2. The goal of this experiment is determine which feature in the dataset have an impact on the predicted outcome. It was determined that nor single algorithm can have designated as the best appropriate algorithm. A set of algorithm was chosen, which included: To do a comparison study, Support Vector Machine and Regression Model were used. The correctness of the chosen collection of algorithms is analyzed and assessed at several stages, based on the findings of the experiment, Regression is identified as the best machine learning approach for predicting COVID-19.

Table 7. Represent the name of feature and values in different several stages.

Name of Features	Values of Features
Age	1, 120
Body Pain	0, 1
Runny Nose	0, 1
Diff Breath	0, -1, 1

4. Conclusions

A thorough literature review was conducted in this study and experimental Algorithms have conduct find the accurate algorithm for COVID-19 forecast in patient. There were no pure arguments discovered an appropriate method to summarize methods for predictions. Hence Support Vector Machine is one of a group of algorithms (SVM) and Regression Models. The chosen algorithm is trended with recorded sets of varying numbers of patients to determine the accuracy of machine learning models. After result analysis, Regression Model show better prediction in comparison Support Vector Machine (SVM) and KNN. The trained algorithms were further evaluated in order to identify the characteristics that influence COVID-19 prediction in patients. Machine Learning has a lot of potential in Medical Department. It is suggested that further effort be focused on prediction of COVID-19 on AI (Artificial Intelligence) and Machine Learning techniques that might solve issues compared to current algorithms. It is quicker and produces superior results. An artificial intelligence (AI) application that makes advantage of different sensors and characteristics to find and assist in the diagnosis illnesses can also be built.

References

1. E. de Wit, N. van Doremalen, D. Falzarano, and V. J. Munster, "SARS and MERS: recent insights into emerging coronaviruses," Nat. Rev. Microbiol., vol. 14, no. 8, pp. 523–534, Aug. 2016, doi: 10.1038/nrmicro.2016.81.
2. K. Ratnasiri, A. J. Wilk, M. J. Lee, P. Khatri, and C. A. Blish, "Single-cell RNA-seq methods to interrogate virus-host interactions," Semin. Immunopathol., Nov. 2022, doi: 10.1007/s00281-022-00972-2.
3. E. Osuna, R. Freund, and F. Girosi, "An improved training algorithm for support vector machines," in Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop, pp. 276–285. doi: 10.1109/NNSP.1997.622408.
4. A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha, "Machine learning for soil moisture assessment," in Deep Learning for Sustainable Agriculture, Elsevier, 2022, pp. 143–168. doi: 10.1016/B978-0-323-85214-2.00001-X.
5. C. R. MacIntyre, S. Lim, and A. Quigley, "Preventing the next pandemic: Use of artificial intelligence for epidemic monitoring and alerts," Cell Reports Med., vol. 3, no. 12, p. 100867, Dec. 2022, doi: 10.1016/j.xcrm.2022.100867.
6. M. Senthilraja, "Application of Artificial Intelligence to Address Issues Related to the COVID-19 Virus," SLAS Technol., vol. 26, no. 2, pp. 123–126, Apr. 2021, doi: 10.1177/2472630320983813.
7. A. Bali and N. Bali, "Role of artificial intelligence in fast-track drug discovery and vaccine development for COVID-19," in Novel AI and Data Science Advancements for Sustainability in the Era of COVID-19, Elsevier, 2022, pp. 201–229. doi: 10.1016/B978-0-323-90054-6.00006-4.
8. S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, and M. J. Ahsan, "Machine Learning in Drug Discovery: A Review," Artif. Intell. Rev., vol. 55, no. 3, pp. 1947–1999, Mar. 2022, doi: 10.1007/s10462-021-10058-4.
9. D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, "Artificial intelligence in drug discovery and development," Drug Discov. Today, vol. 26, no. 1, pp. 80–93, Jan. 2021, doi: 10.1016/j.drudis.2020.10.010.
10. M. Khan et al., "Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review," Expert Syst. Appl., vol. 185, p. 115695, Dec. 2021, doi: 10.1016/j.eswa.2021.115695.
11. M. Khan et al., "COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far," Molecules, vol. 26, no. 1, p. 39, Dec. 2020, doi: 10.3390/molecules26010039.
12. M. Cascella, M. Rajnik, A. Aleem, S. C. Dulebohn, and R. Di Napoli, Features, Evaluation, and Treatment of Coronavirus (COVID-19). 2023.
13. T. L. Dao, T. D. Nguyen, and V. T. Hoang, "Controlling the COVID-19 pandemic: Useful lessons from Vietnam," Travel Med. Infect. Dis., vol. 37, p. 101822, Sep. 2020, doi: 10.1016/j.tmaid.2020.101822.
14. A. Al-Hazmi, "Challenges presented by MERS corona virus, and SARS corona virus to global health," Saudi J. Biol. Sci., vol. 23, no. 4, pp. 507–511, Jul. 2016, doi: 10.1016/j.sjbs.2016.02.019.
15. J. Guarner, "Three Emerging Coronaviruses in Two Decades," Am. J. Clin. Pathol., vol. 153, no. 4, pp. 420–421, Mar. 2020, doi: 10.1093/ajcp/aqaa029.
16. N. Zhong et al., "Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003," Lancet, vol. 362, no. 9393, pp. 1353–1358, Oct. 2003, doi: 10.1016/S0140-6736(03)14630-2.
17. W. Sungnak et al., "SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes," Nat. Med., vol. 26, no. 5, pp. 681–687, May 2020, doi: 10.1038/s41591-020-0868-6.
18. K. Roosa et al., "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," Infect. Dis. Model., vol. 5, pp. 256–263, 2020, doi: 10.1016/j.idm.2020.02.002.
19. Z. Wu and J. M. McGoogan, "Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China," JAMA, vol. 323, no. 13, p. 1239, Apr. 2020, doi: 10.1001/jama.2020.2648.
20. N. Shaukat, D. M. Ali, and J. Razzak, "Physical and mental health impacts of COVID-19 on healthcare workers: a scoping review," Int. J. Emerg. Med., vol. 13, no. 1, p. 40, Dec. 2020, doi: 10.1186/s12245-020-00299-5.
21. M. Pal et al., "Symptom-Based COVID-19 Prognosis through AI-Based IoT: A Bioinformatics Approach," Biomed Res. Int., vol. 2022, 2022, doi: 10.1155/2022/3113119.
22. H. S. Maghdid and K. Z. Ghafoor, "A Smartphone Enabled Approach to Manage COVID-19 Lockdown and Economic Crisis," SN Comput. Sci., vol. 1, no. 5, p. 271, Sep. 2020, doi: 10.1007/s42979-020-00290-0.
23. M. C. Robinson, "Clinical Features," 1955.
24. E. García-Gonzalo, Z. Fernández-Muñiz, P. García Nieto, A. Bernardo Sánchez, and M. Menéndez Fernández, "Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers," Materials (Basel)., vol. 9, no. 7, p. 531, Jun. 2016, doi: 10.3390/ma9070531.
25. B. M. Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, "Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study," in 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Jun. 2017, pp. 1–2. doi: 10.1109/JCDL.2017.7991618.
26. Hamed Taherdoost, "Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects," (IJARM), vol. 10, pp. 10–38, 2021.
27. J. Sutton and Z. Austin, "Qualitative Research: Data Collection, Analysis, and Management," Can. J. Hosp. Pharm., vol. 68, no. 3, Jun. 2015, doi: 10.4212/cjhp.v68i3.1456.

28. "Saidu Teaching Hospital, abbreviated as STH, is the fifth biggest teaching hospital and instruction in Khyber Pakhtunkhwa, Pakistan. It is located in Swat, KPK.", Accessed: Mar. 28, 2023. [Online]. Available: https://en.widipedia.org/wiki/Saidu_Teaching_Hospital
29. A. A. Khan et al., "Assessment of COVID-19 Management at Saidu Group of Teaching Hospital Swat: Analytical Cross-Sectional Study," Pakistan J. Public Heal., vol. 11, no. 2, pp. 102–106, Jul. 2021, doi: 10.32413/pjph.v11i2.726.
30. J. Yang et al., "Brief introduction of medical database and data mining technology in big data era," J. Evid. Based. Med., vol. 13, no. 1, pp. 57–69, Feb. 2020, doi: 10.1111/jebm.12373.
31. Scikit-Learn, "Imputation of missing values techniques for missing data imputation," Google Summer Code, 2007, Accessed: Mar. 28, 2023. [Online]. Available: https://scikit-learn.org/stable/modules/impute.html
32. NIRAJ VERMA, "Support Vector Machine detail analysis," Kaggle, 2018, Accessed: Mar. 28, 2023. [Online]. Available: https://www.kaggle.com/code/nirajvermafcb/support-vector-machine-detail-analysis/notebook
33. A. Alonso Robisco and J. M. Carbó Martínez, "Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction," Financ. Innov., vol. 8, no. 1, p. 70, Dec. 2022, doi: 10.1186/s40854-022-00366-1.
34. I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
35. Data.World, "Pakistan COVID-19 Dataset of Infected Patient's," Nov. Coronavirus cases Pakistan., 2020, Accessed: Mar. 28, 2023. [Online]. Available: https://data.world/dekhpakistan/pakistan-covid-19-dataset
36. O. A. Gbinigie, J. M. Ordóñez-Mena, T. Fanshawe, A. Plüddemann, and C. J. Heneghan, "Limited evidence for diagnosing bacterial skin infections in older adults in primary care: systematic review," BMC Geriatr., vol. 19, no. 1, p. 45, Dec. 2019, doi: 10.1186/s12877-019-1061-y.
37. M. T. Amin, M. Hasan, and N. M. M. A. Bhuiya, "Prevalence of Covid-19 Associated Symptoms, Their Onset and Duration, and Variations Among Different Groups of Patients in Bangladesh," Front. Public Heal., vol. 9, Sep. 2021, doi: 10.3389/fpubh.2021.738352.
38. J. Algina and H. J. Keselman, "Comparing squared multiple correlation coefficients: Examination of a confidence interval and a test significance.," Psychol. Methods, vol. 4, no. 1, pp. 76–83, Mar. 1999, doi: 10.1037/1082-989X.4.1.76.
39. M. M. Mukaka, "Statistics corner: A guide to appropriate use of correlation coefficient in medical research.," Malawi Med. J., vol. 24, no. 3, pp. 69–71, Sep. 2012.
40. V. Kakulapati, S. M. Reddy, and N. Kumar, "Lexical modeling and weighted matrices for analyses of COVID-19 outbreak," in Lessons from COVID-19, Elsevier, 2022, pp. 313–340. doi: 10.1016/B978-0-323-99878-9.00005-4.
41. A. Khan, S. Ahmad, B. Tahira, and A. Khalid, "Static Analysis of Dengue Biological Regulatory Network's", biolsciences, vol. 3, no. 2, pp. 406–416, May 2023