SURVIVAL STUDY ON COVID DISEASE PREDICTION METHODS WITH PATIENT DATA

KALAISELVI S R
Ph.D. Research Scholar
Department of Computer Science
Avinashilingam Institute for
Home Science and Higher Education
srkalai2010@gmail.com

Assistant Professor
Department of Computer
Science, Dr.N.G.P. Arts
and Science College
Coimbatore

VIJAYABHANU R
Assistant Professor
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education Coimbatore
vijayabhanu_cs@avinuty.ac.in

***Abstract*— The healthcare domain is the prominent research field in existing scenarios with fast technology enhancement. COVID-19 causes respiratory disorder using manifestation such as cold, cough and fever. COVID-2019 was perceived as worldwide pandemic and used various numerical methods to predict the advancement. There are numerous for treatment of COVID-2019 diseases across the world. Machine Learning(ML) is an approach used for prediction and disease diagnosis. Many existing techniques were designed to carry outCOVID-2019 disease identication from input patient files. Classification is the process of classifying the ideas where the objects are identified and discriminated. But, it failed to enhance the prediction accuracy (PA) and not reduced the time complexity (TC)by existing data mining techniques. To overcome issues, conventional methods are reviewed for efficient disease diagnosis.**

Keywords — *Healthcare domain, coronavirus disease, respiratory ailment, manifestation, classification, prediction, disease diagnosis*

I. INTRODUCTION

Text mining employed the artificial intelligence technology to process data and created valuable insights for taking the data-driven decisions. At the present time, people get affected by diverse diseases like COVID-19 due to environmental condition and living habits. COVID-19 has been considered as the pandemic by World Health Organization (WHO). COVID-19 is the pandemic that affected over the 170 countries around world. The number of affected and dead patients is enhancing at alarming rate in infected nations. Machine learning processing is used for pattern recognition, prediction and classification.Prediction techniques act as an essential role for performing accurate prediction. The forecast of disease at starting phase is an important process. Prediction is significant tools for health care to identify COVID-19 disease patient at early stage.

The prediction is employed to discover COVID-19 patient's result based on patients features monitored at home, while in quarantine.

The prediction is applied for discovering disease people at risk of covid-19 disease or admitted to a hospitalin the general population. In addition to, maintain covid-19 using suspected infection. Regression analysis is applied for discovering the outcome of Covid-19 on different Indian states. Regression analysis employed to relationship among two variables. Regression is utilized to predict the both each day lively or latest confirmed COVID-19 cases.

Recognizing a pattern is a method of determining the regularities in data by using statistical or machine learning methods for performing further analysis such as clustering or classification. Pattern recognition plays a critical part for determining the risk factors and showing the trend of a disease from a large data set.To best appreciate and mitigate the broaden of disease, it is important for recognizing the pattern ofthe outbreak.

Classification is one of the most importanttasks in text mining with aid of dissimilar algorithms. Classification of COVID-19 is important for disease cure and manages. Classification is used for categorizing the COVID- 19 disease patients. In addition to, classification is employed for precisely categorize the patients whether they belong to disease class or not for improving the accurate classification performance.

The article is summarized by: Section 2 studies reconsider on different COVID prediction methods, Section 3 represents the study and analysis of existing COVID prediction methods, Section 4 illustrates the possible comparison of existing methods. In Section 5, discussion and limitations of the existing COVID prediction methods are studied with future direction. Section 6describes summary the paper.

II. LITERATURE REVIEW

Medical diagnosis is the method of recognizing which disease explains the individual symptoms and signs. "An intelligent framework was developed in [1] with disruptive method for COVID-19 analysis." The designed framework had many deficiencies to reduce huge stress on hospitals and improved patients to treat patients. But, error rate was not reduced-using designed intelligent framework.

"A novel hybrid feature representation method was designed in [2] with Hybrid feature representation method to forecast the antioxidant proteins". "However, the feature selection was not carried out in accurate manner. The forecasting methods were introduced in [3] to predict the COVID-19 dangering factors." Exponential Smoothing was designed for predicting current positive cases, decease rate. However, cost was not reduced.

"A new machine learning forecasting model was introduced in [4] to forecast COVID-2019 spread. "However, the designed model was unsuccessful for reducing time. Dimensionality reduction technique was designed in [5] to detect heart disease features. But prediction accuracy was not improved through dimensionality reduction technique.

"A case-based reasoning (CBR) model was introduced in [6] with convincing ideas for case- specific knowledge exploitation and concrete problem circumstances for resolving new cases." However, time consumption was not minimized through case-based reasoning (CBR) model. "Machine learning technique called one-class approach was designed in [7] to assist research community for future examination regarding COVID-19 prevention and treatment. "But, the computational cost was not minimized through machine learning techniques.

"A H5N1 virus caused infection with clinical similarity in [8] for COVID-19. But, time consumption was not reduced during COVID-19 virus prediction. "A novel feature reduction (NFR) technique was designed in [9] with machinelearning and data mining to reduce the error rate." The designed approach was introduced depending on heuristic process through minimizing the features regarding AUC enhancement.

"Deep learning disease prediction model was designed in [10] with stepwise approach during feature extraction and categorization." However, prediction model failed to minimize the error rate.

"Two-stage method was introduced in [11] to integrate the structured EHR data for early assessment of Kawasaki Disease (KD)." The missing data with imputation techniques were designed with higher accuracy. However, computational complexity was not minimized."A smart healthcare system was designed in [12] to forecast heart disease." Though computational complexity was minimized, the time consumption was not reduced through designed smart healthcare system.

III. COVID DISEASE PREDICTION METHODS

COVID-19 is a worldwide health emergency through different infected cases and deaths reportedover 180 countries. Early recognition of COVID-19 is an essential one for patient isolation and rapid clinical involvement.The whole world is experiencing the critical pandemic salvation (COVID-19). COVID-19 was caused with SARS- CoV-2 and detected iin China COVID-19 is essential one to patient isolation as well as rapid clinical involvements.

A. *"AN INTELLIGENT FRAMEWORK USING DISRUPTIVE TECHNOLOGIES FOR COVID-19 ANALYSIS"*

An intelligent framework was introduced to limit the COVID-19 epidemics. It employed the disruptive concept for COVID-19 prediction. The designed framework guaranteed the healthcare teams safety and maintained the patient corporal and mental healthcare conditions. Designed approach managed the cruel shortage of PPE for medical panel, and reduced the huge stress for indulgence COVID-19. Designed framework provided to minimize the unprecedented out-breaks impacts for COVID-19. The empirical case study was carried out for COVID-19 patient analysis. The safe environment was provided to preserve the patient physical and psychological healthcare for medical group and medical caregivers. Designed framework handled PPE shortage, minimized the huge stress follow improved patients for indulgence the COVID-19.

B. *"PREDICTION OF ANTIOXIDANT PROTEINS USING HYBRID FEATURE REPRESENTATION METHOD AND RANDOM FOREST"*

Antioxidant proteins are identified for era dicating the extreme radicals for preserving cells over injury, protect as well as indulgencediseases. An exact recognition of antioxidant proteins was essential to the improvement of Novel studies. Antioxidant proteins interrelate with free radicals for preserving cells. A novel Hybrid feature representation method was designed to forecast the antioxidant proteins exactly. The single feature extraction technique was employed to carry out the feature extraction. It is employedto recognize the best features. Hybrid feature was utilized with excellent predictors of antioxidant proteins.

C. *COVID-19 FUTURE FORECASTING USING SUPERVISED MACHINE LEARNING MODELS*"

ML-based forecasting mechanism establish result for predicting with preoperative results for increasing the decision-making performances. ML models were employed in different application areas for identification and prioritization of difficult factors. Different prediction techniques were employed to handle the forecasting issues.ML techniques were introduced to predict patients infected through COVID-19. Benchmark predictingtechniques namely LR, LASSO, SVM, and ES were employed for predicting features of COVID-19.An early forecast model termed SARS-CoV-2 was employed for spread of corona virus by WHO. COVID-19 was serious threat for human life all over world. "The virus was identified in city of China called Wuhan." It has different reaction on human body. Thousands of people were infected with epidemic all over the globe using many deaths. It was statements positive every day.

IV. PERFORMANCE ANALYSIS OF CORONA VIRUS DISEASE PREDICTION TECHNIQUES

In order to compare the coronavirus disease prediction methods, number of patient data taken tocarry out an experiment. Experimental evaluation of three methods namely intelligent framework,random forest and hybrid feature and Machinelearning (ML) based forecasting mechanism aredeveloped using Java. With the aim of predictcorona virus disease, Corona Virus 2019 Dataset is occupied over Kaggle. Datasetishttps://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset. It contains everyday details on number of inclined cases, deaths andimprovement over 2019 coronavirus. It contains 44attributes and 13174 instances. Attributes are:identity, age, sex, city, country, province, etc.Through the features, relevant attributes are chosen to carry out categorization of COVID. Analysis of conventional techniques is evaluated with metricssuch as accuracy, time, and Error rate.

A. *"IMPACT ON PT"*

PT is described as sum of time used forforecasting the number of patient data. PT is product of number of data as well as sum of time taken to forecast the data. Consequently, PT „Pre_{Time}" is computed below,

$$Pre_{Time} = Number\ of\ patient\ data * \ umed\ for\ \text{forecasting}\ one\ data \quad -(1)$$

From (1), PT is calculated. PT is calculated by milliseconds (ms).

Table.1. Tabulated for PT

No. of Patient Data (Number)	PT (ms)		
	Intelligent Framework	Hybrid feature representation method	ML based forecasting mechanism
50	25	31	44
100	28	34	47
150	31	37	50
200	33	40	53
250	36	43	56
300	39	46	59
350	41	49	62
400	43	51	65
450	45	54	68
500	48	57	71

Tab1 illustrate PT using number of patientdata. PT performed on existing Intelligent Framework, Hybrid feature representation method and ML based forecasting mechanism. The graphical illustration of PT is demonstrated in the figure 1

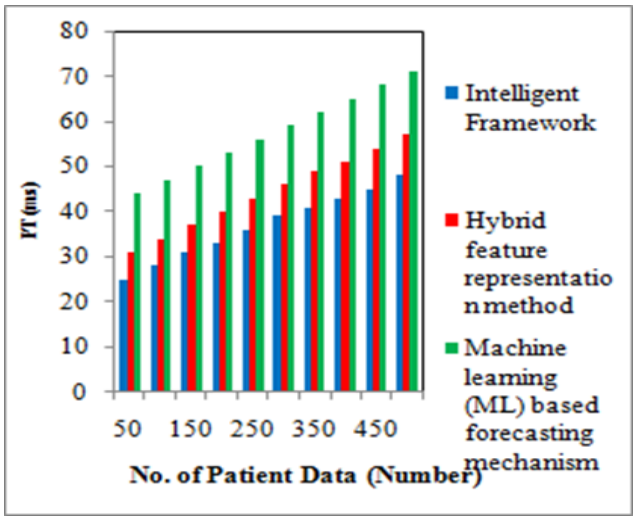

Fig.1. Measurement of PT

From figure 1, PT depend on various number of patient data is demonstrated. In fig 1 blue colour bar indicates PT of Intelligent Framework. Redcolor bar and green colorbar represent PT of Hybrid feature representation method and ML-based forecasting mechanism correspondingly. It is examined that PT using Intelligent Framework is minimum compared to the Hybrid feature representation method and Machine learning (ML) based forecasting mechanism. This is due to the application of disruptive technology for COVID-19 analysis. The designed framework assured the healthcare group safety and maintained patient corporal and psychological healthcare conditions. Therefore, prediction time of IntelligentFramework is reduced by 17% when compared to Hybrid feature representation method and 36% when compared to the ML based forecasting mechanism.

B. " IMPACT ON PA"

PA is computed as proportion of the number of patient data which are accurately forecasted the disease to entire number of patient data taken. „Pre_{Acc} " determined as below

$$Pre_{Acc} = \left(\frac{Number\ of\ patient\ data\ that\ are\ correctly\ forecasted\ disease}{Number\ of\ patient\ data}\right) * 100 \quad (2)$$

From (2), the PA is determined.

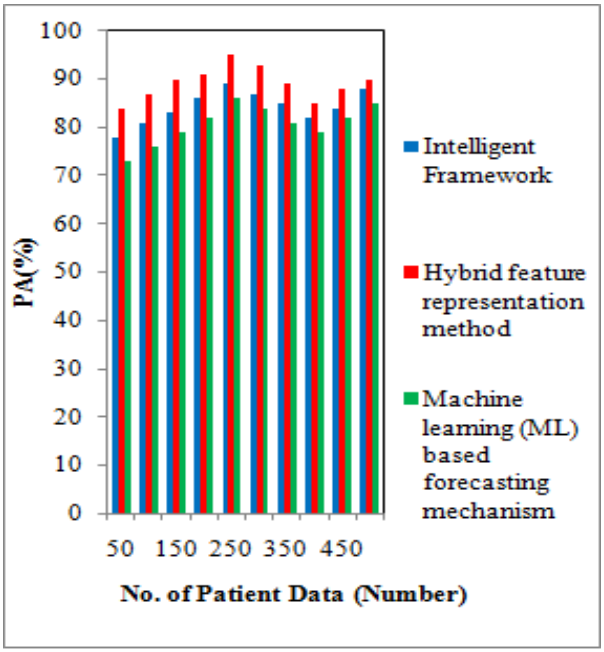

Fig.2. Measurement of PA

Table.2. Tabulated for PA

No. of Patient Data (Number)	PA (%)		
	Intelligent Framework	Hybrid feature representation method	ML based forecasting mechanism
50	78	84	73
100	81	87	76
150	83	90	79
200	86	91	82
250	89	95	86
300	87	93	84
350	85	89	81
400	82	85	79
450	84	88	82
500	88	90	85

Table 2 demonstrates the PA with number of patient data. PA is compared with existing Intelligent Framework, Hybrid feature representation method and ML based forecasting mechanism. The graphical illustration of PA is depicted in figure 2.

The above figure 2, PA depend on various number of patient data is explained. Blue colour bar indicates PA of Intelligent Framework. . Red color bar and green color bar indicates PA of Hybrid feature representation method and ML based forecasting mechanism correspondingly.

It is examined that PA using Hybrid feature representation method is higher when compared to Intelligent Framework and Machine learning (ML) based forecasting mechanism. This is due to the application of random forest classifier to forecast the antioxidant proteins. Hybrid feature was employed with excellent predictors of antioxidant proteins for forecasting the antioxidant proteins. Consequently, PA of Hybrid feature representation method improved by 6% than the Intelligent Framework and 11% than the ML based forecasting mechanism.

C. "IMPACT ON ER"

ER is described as proportion of number of patient data which are wrongly forecasted to entire number of data. „Err_{Rate} " is computed by,

$$Err_{Rate} = \left(\frac{Number\ of\ patient\ data\ that\ are\ incorrectly\ forecasted}{Number\ of\ patient\ data}\right) * 100 \quad (3)$$

From (3), ER is calculated. ER is determined in percentage (%).

Table.3. Tabulated for ER

No. of Patient Data (Number)	ER (%)		
	Intelligent Framework	Hybrid feature representation method	ML based forecasting mechanism
50	18	15	12
100	14	11	10
150	12	8	9
200	10	7	7
250	8	6	5
300	11	8	6
350	14	10	8
400	16	12	9
450	18	14	11
500	21	17	13

Table 3 demonstrates ER with number of patient data Error rate is compared with existing Intelligent Framework, Hybrid featurerepresentation method and ML based forecasting mechanism. The graphical illustration of ER is described in fig 3.

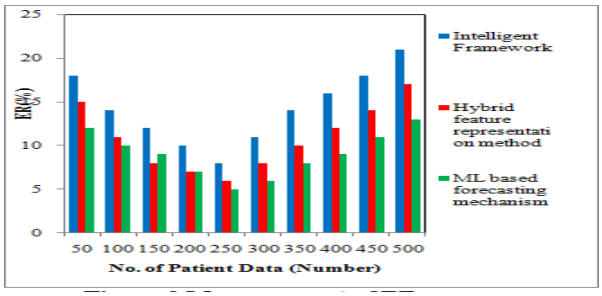

Fig.3. Measurement of ER

The above figure 3, error rate depend on various number of patient data is explained. Blue colour bar indicates ER of Intelligent Framework. Red color bar and green color bar indicatesER of Hybrid feature representation method and ML based forecasting mechanism correspondingly. It is examined that ER using ML based forecasting mechanism is minimum compared to Intelligent Framework and Hybrid feature representation method. This is because of application of ML techniques to predict no. of future patients. ML methods analysed patient information to reduce the error rate. Consequently, error rate of ML based forecasting mechanism reduced by 36 than the Intelligent Framework and 15% than the Hybrid feature representation method.

V.DISCUSSION ON COVID DISEASE PREDICTION TECHNIQUES

Intelligent Framework limited the COVID-19 spread outbreak, assured thehealthcare team safety and conserved the patient physical as well as psychological healthcare conditions. The designed frameworkmanaged lack of PPE for medical group, reduced largestress on hospitals and followed the healthier patients to indulgence the COVID-19 patients with plasma. But, failed to reduce the ER by designed intelligent framework.

Hybrid feature representation methodcomprised mixture of hybrid feature to forecast the antioxidant proteins. Feature selection methods helped the optimal features. The designed method minimized feature dimension and redundant feature vectors. t-SNE minimized the high-dimensional features and imagined features in 2D space. However,feature selection was not performed in more accurate way.

ML based forecasting mechanism addressed the forecasting issues. It predicts number of patients concerned with COVID-19. Exponential Smoothing carried out well in predicting novel confirmed cases, death rate as well as recovery rate. ML modelsare used to compute the R2, R2 adjusted, MSE, MAE, and RMSE.R2 score is applied to determine scatteredness of information represented as the coefficient of determination. R2 is adapted by R2 such as R2.MSE is employed for computing performance of regression models. RMSE is represented as standard deviation of forecasting errors. This is helps to reduce the error rate. But, the real-time liveforecasting was not carried out by designed model and also failed to reduce computational cost byexisting forecasting techniques.

A. Future Scope

The future scope of COVID disease prediction can be performed using ML and deep learning techniques to enhance disease diagnosis performance by improved accuracy with minimumtime consumption.

VI. SUMMARY

A comparison of various existing COVID disease prediction methods was explained. ML based forecasting mechanism techniques failed to minimize the computational cost. The survival examine demonstrates that feature selection was notperformed

more accurate way by hybrid feature representation method. In addition, error rate was not minimized by designed intelligent framework. The broad-spectrum of experiments on many existing COVID disease prediction method establish performance through its limitations. Research method can be performed using MLtechniques and deep learning methods for enhancing the COVID disease prediction performance.

References

[1] Mohamed Abdel-Basset, Victor Chang and Nada A. Nabeeh, "An intelligent framework using disruptive technologies for COVID-19 analysis", Technological Forecasting and Social Change,October 2020, Pages 1-14

[2] ChunyanAo, Wenyang Zhou, Lin Gaoa, Benzhi Dong, Liang Yu, "Prediction of antioxidant proteins using hybrid feature representation method and random forest", Genomics, Elsevier, Volume 112, 2020, Pages 4666-4674

[3] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam and Gyu Sang Choi, "COVID-19 Future Forecasting using Supervised Machine Learning Models", IEEE Access, Volume 8, May 2020, Pages 101489 – 101499

[4] R. Sujath, Jyotir Moy Chatterjee and Aboul Ella Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India", Stochastic Environmental Research and Risk Assessment,Springer, Volume 34, 2020, Pages 959 – 972

[5] Anna Karen Garate-Escamila, Amir Hajjam El Hassani, Emmanuel Andres, "Classification models for heart disease prediction using feature selection and PCA", Informatics in Medicine Unlocked,Elsevier, Volume 19, 2020, Pages 1-11

[6] Olaide N. Oyelade and Absalom E. Ezugwu, "A case-based reasoning framework for early detection and diagnosis of novel coronavirus", Informatics in Medicine Unlocked, Elsevier, Volume 20, 2020,Pages 1-22

[7] Sanjay Kumar Sonbhadra, Sonali Agarwal and P. Nagabhushan, "Target specific mining of COVID-19 scholarly articles using one-class approach", Chaos, Solitons & Fractals, Elsevier,Volume 140, November 2020, Pages 1-11

[8] NopharGeifman and Anthony D. Whetton, "A consideration of publication-derived immune- related associations in Coronavirus and related lung damaging diseases", Journal of Translational Medicine, Volume 18, Issue 297, 2020, Pages 1-11

[9] Syed Javeed Pasha and E. Syed Mohamed, "Novel Feature Reduction (NFR) Model with Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction", IEEE Access,Volume 8, Pages 184087 – 184108

[10] Srinivas Koppu, Praveen Kumar Reddy Maddikunta and Gautam Srivastava, "Deep learning disease prediction model for use with intelligent robots", Computers & Electrical Engineering, Elsevier, Volume 87, October 2020, Pages 1-13

[11] Haolin Wang, Xuhai Tan, Zhilin Huang, Bo Pan and Jie Tian, "Mining incomplete clinical data for the early assessment of Kawasaki disease based on feature clustering and convolutional neural networks", Artificial Intelligence In Medicine, Elsevier, Volume 105, 2020, Pages 1-7