Pattern analysis: predicting COVID-19 pandemic in India using AutoML

S. Gomathi
UK International Qualifications Ltd, Coimbatore, India
Rashi Kohli
Senior member IEEE, Pooler, USA
Mukesh Soni
Department of Computer Engineering, Smt Sr Patel Engineering College, Shihi, India
Gaurav Dhiman
Government Bikram College of Commerce, Patiala, India, and
Rajit Nair
Jagran Lakecity University, Bhopal, India

Abstract

Purpose – Since December 2019, global attention has been drawn to the rapid spread of COVID-19. Corona was discovered in India on 30 January 2020. To date, in India, 178,014 disease cases were reported with 14,011 deaths by the Indian Government. In the meantime, with an increasing spread speed, the COVID-19 epidemic occurred in other countries. The survival rate for COVID-19 patients who suffer from a critical illness is efficiently and precisely predicted as more fatal cases can be affected in advanced cases. However, over 400 laboratories and clinically relevant survival rates of all present critically ill COVID-19 patients are estimated manually. The manual diagnosis inevitably results in high misdiagnosis and missed diagnosis owing to a lack of experience and prior knowledge. The chapter presents an option for developing a machine-based prognostic model that exactly predicts the survival of individual severe patients with clinical data from different sources such as Kaggle data.gov and World Health Organization with greater than 95% accuracy. The data set and attributes are shown in detail. The reasonableness of such a mere three elements may depend, respectively, on their representativeness in the indices of tissue injury, immunity and inflammation. The purpose of this paper is to provide detailed study from the diagnostic aspect of COVID-19, the work updates the cost-effective and prompt criticality classification and prediction of survival before the targeted intervention and diagnosis, in particular the triage of the vast COVID-19 explosive epidemic.

Design/methodology/approach – Automated machine learning (ML) provides resources and platforms to render ML available to non-ML experts, to boost efficiency in ML and to accelerate research in machine learning. H2O AutoML is used to generate the results (Dulhare *et al.*, 2020). ML has achieved major milestones in recent years, and it is on which an increasing range of disciplines depend. But this performance is crucially dependent on specialists in human ML to perform the following tasks: preprocess the info and clean it; choose and create the appropriate apps; choose a family that fits the pattern; optimize hyperparameters for layout; and models of computer learning post processes. Review of the findings collected is important.

Findings – These days, the concept of automated ML techniques is being used in every field and domain, for example, in the stock market, education institutions, medical field, etc. ML tools play an important role in harnessing the massive amount of data. In this paper, the data set relatively holds a huge amount of data, and appropriate analysis and prediction are necessary to track as the numbers of COVID cases are increasing day by day. This prediction of COVID-19 will be able to track the cases particularly in India and might help researchers in the future to develop vaccines. Researchers across the world are testing different medications to cure COVID; however, it is still being tested in various labs. This paper highlights and deploys the concept of AutoML to analyze the data and to find the best algorithm to predict the disease. Appropriate tables, figures and explanations are provided.

Originality/value – As the difficulty of such activities frequently goes beyond non-ML-experts, the exponential growth of ML implementations has generated a market for off-the-shelf ML solutions that can be used quickly and without experience. We name the resulting work field which is oriented toward the radical automation of AutoML machine learning. The third class is that of the individuals who have illnesses such as diabetes, high BP, asthma, malignant growth, cardiovascular sickness and so forth. As their safe frameworks have been undermined effectively because of a common ailment, these individuals become obvious objectives. Diseases experienced by the third classification of individuals can be lethal (Shinde *et al.*, 2020). Examining information is fundamental in having the option to comprehend the spread and treatment adequacy. The world needs a lot more individuals investigating the information. The understanding from worldwide data on the spread of the infection and its conduct will be key in limiting the harm. The main contributions of this study are as follows: predicting COVID-19 pandemic in India using AutoML; analyzing the data set predicting the patterns of the virus; and comparative analysis of predictive algorithms. The organization of the paper is as follows,

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/1708-5284.htm**

World Journal of Engineering
19/1 (2022) 21–28
 [ISSN 1708-5284]
[DOI 10.1108/WJE-09-2020-0450]

Received 23 September 2020
Revised 29 October 2020
Accepted 3 November 2020

Sections I and II describe the introduction and the related work in the field of analyzing the COVID pandemic. Section III describes the workflow/framework for AutoML using the components with respect to the data set used to analyze the patterns of COVID-19 patients.

Keywords COVID-19, AutoML, Data processing, Machine learning, Data analysis, India, Laboratory, Pandemic WHO, Kaggle

Paper type Research paper

1. Introduction

It is broadly recognized nowadays that finding a well-performing machine learning (ML) pipeline (counting pre-preparing, artificial intelligence [AI] calculation choice, and hyperparameter enhancement) is a dreary and mistake inclined assignment for people. In the light of developing enthusiasm for AutoML techniques which are used in industry as well as they are used in academia (Dulhare *et al.*, 2020).

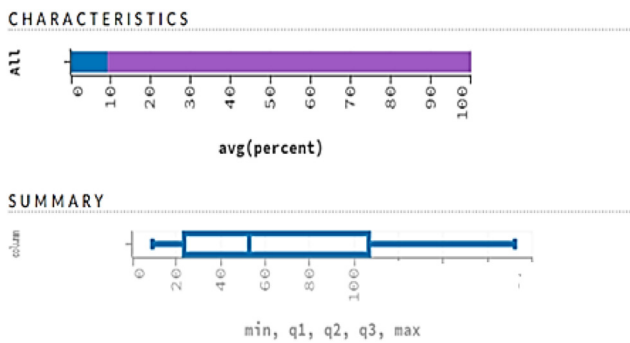
Our research focuses on analyzing the patterns of COVID-19, particularly in India region using automated machine learning (AutoML) techniques. One of the many unanswered scientific questions about COVID-19 is whether it is seasonal enough to flu-degrade during the warm summer months and regenerate in the fall and winter. Numerous nations forced a lockdown express that forestalls the development of the residents pointlessly (Elaziz *et al.*, 2020). Because of this social separating element and development limitations, the prosperity and economy of the different countries are being under peril. Scientists at the Lawrence Berkeley National Laboratory (Berkeley Lab) are now applying ML techniques to a large number of health and environmental data sets. Concept of AI can help us in dealing with the issues that should be tended to raised by the COVID-19 pandemic (Zhao *et al.*, 2020; Shinde *et al.*, 2020; Sidey-Gibbons and Sidey-Gibbons, 2019; Sujath *et al.*, 2020; Karras *et al.*, 2020). It is not just the advancement, be that as it may, that will influence yet rather the data and imaginativeness of the individuals who use it. Beyond question, the COVID-19 crisis will presumably reveal a bit of the key deficiencies of AI. AI (ML), the current kind of AI, works by perceiving structures in chronicled preparing data.

There is a dire need to evaluate and predict the clinical data to provide analysis and suggest the interventions based on the data. Driven by an expansion in computational force, stockpiling, memory and the age of stunning volumes of information, PCs are being used to play out a wide scope of complex errands with great precision. AI (ML) is the name given to both the scholarly order and assortment of procedures that permit PCs to attempt complex assignments. As a scholarly order, ML involves components of arithmetic, insights and software engineering. AI is the motor which is assisting with driving advances in the improvement of man-made consciousness. This paper focuses on the spread of virus-specific to India and provides an in-depth analysis to predict the extremely dangerous virus that has been spread across the nation. This paper will focus on the data set particularly from the India region. There is a requirement for innovative solutions for creating, oversee and dissect large information on the developing system of contaminated subjects, quiet subtleties, their locale developments and incorporate with clinical preliminaries and, pharmaceutical, genomic and general well-being information, there is a need to develop a strategy to predict and analyze the long-term effects of the highly spread coronavirus (Zhao *et al.*, 2020; Sidey-Gibbons and Sidey-Gibbons, 2019).

At the point when the COVID-19 began to spread at a phenomenal rate, preventive measures were worked out. These measures incorporated a total lockdown of the intensely contaminated territories, prohibition on universal voyages, suspending schools and other superfluous day-by-day exercises. The principal points of these measures were to constrain relational contact, thinking about the infectious idea of the malady. The time limitation was forced and carefully watched. As the brooding time of the infection is longer than different infections it is hard to dissect the ideal time required to watch a time limitation. On the off chance that the check-in time is lifted too early, the circumstance can get hazardous (Truong *et al.*, 2019). The individuals who get contaminated fall under three classifications. First in the classification are the old, who are profoundly vulnerable to the infection. Insights show that as a result of the feeble resistant framework the old surrender to the infection without any problem. The subsequent class is that of the youngsters. As the invulnerable frameworks of small kids are as yet a work in progress, the youngsters are at higher hazard (Chen and Gupta, 2015; He *et al.*, 2019; Zöller and Huber, 2019; Xu *et al.*, 2015).

This paper analyzes the different attributes which will be helpful in predicting the COVID impact. As the COVID-19 impact is still escalating at its peak, it is relatively important to analyze the impact it is still making and how the year 2021 will look like. These predictions can be made by analyzing the data based on the various ML and Auto ML predictive algorithms such as decision trees, Naïve Bayes, kNN, logistic regression and AutoML algorithm of Cox proportional hazards (CoxPH), various attributes were compared to predict the analysis of COVID. The attributes include age, gender, travel, contact with COVID patients, clinical manifestation and critically. Figure 1 categorizes the impact of age based on COVID-19 patients. The box plot describes the five-number summary including the minimum age which is averaged to be ten years, whereas the maximum age crosses 100. This data sample is

Figure 1 Age analysis of COVID-19 patients



geographically based out on India. The measure of variability with respect to age is calculated based on "Interquartile Range, i.e. IQR" which is in this case is

AutoML provides resources and platforms to render ML available to non-ML experts, to boost efficiency in ML and to accelerate research in ML. ML has achieved major milestones in recent years, and it is on which an increasing range of disciplines depend. But this performance is crucially dependent on specialists in human ML to perform the following tasks:

- preprocess the info, and clean it;
- choose and create the appropriate apps;
- choose a family that fits the pattern;
- optimize hyperparameters for layout;
- models of computer learning post processes; and
- review of the findings collected is important.

As the difficulty of such activities frequently goes beyond non-ML-experts, the exponential growth of ML implementations has generated a market for off-the-shelf ML solutions that can be used quickly and without experience. We name the resulting work field which is oriented toward the radical automation of AutoML ML.

The third class is that of the individuals who have illnesses like diabetes, high BP, asthma, malignant growth, cardiovascular sickness and so forth. As their safe frameworks have been undermined effectively because of a common ailment, these individuals become obvious objectives. Diseases experienced by the third classification of individuals can be lethal (Shinde *et al.*, 2020). Examining information is fundamental in having the option to comprehend the spread and treatment adequacy. The world needs a lot more individuals investigating the information. The understanding from worldwide data on the spread of the infection and its conduct will be key in limiting the harm.

The main contributions of this study are as follows:

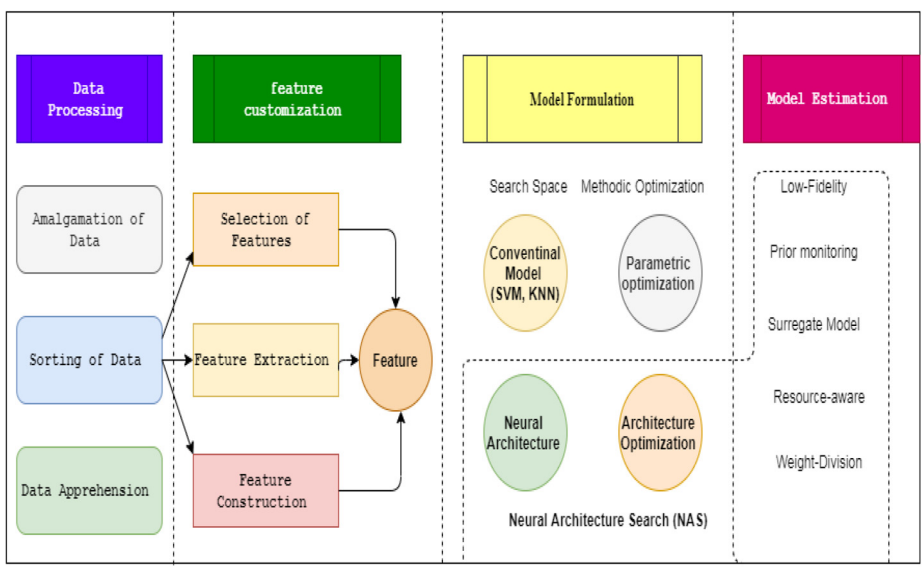
- predicting COVID-19 pandemic in India using AutoML;
- analyzing the data set predicting the patterns of the virus; and
- comparative analysis of predictive algorithms.

The organization of the paper is as follows: Sections 1 and 2 describe the introduction and the related work in the field of analyzing the COVID pandemic. Section 3 describes the workflow/Framework for AutoML using the components with respect to the data set used to analyze the patterns of COVID-19 patients.

2. Related work

In India, the episode of coronavirus has upset the working of life in general. All were pushed to remain back to protect from the shocking transmission. In the underlying stages, the affirmed cases are those that come back from regulates followed by transmission through the nearby transmission. More alert is given to the older and the insusceptibility of fewer individuals. The segment of the contaminated individuals in India demonstrates that 39 years is the middle. Nearly, individuals somewhere in the range of 21 and 40 years are being influenced more. Presently, the entire world is seeing COVID-19 pandemic. More than 100 or more nations to date are significantly affected by COVID-19. This tally is expanding as each spending day. Since the commencement of these epidemics, one thing was watched, that is, with the advancement in time, these scourges swelled into pandemics or commonly alluded to as the episode of the infection/ailment. The COVID-19 crisis, subsequently, will include something that has reliably been legitimate about AI: it is a gadget, and the estimation of its use in any condition is directed by the individuals who structure it and use it. In the current crisis, human movement and improvement will be particularly fundamental in using the force of what AI can do. One approach to managing the novel situation issue is to amass new preparing data under current conditions. For both human pioneers and AI structures the equivalent, each new bit of information about our current situation is particularly critical in exhorting our decisions proceeding. The more practical we are at sharing information, the more quickly our situation isn't,

Figure 2 AutoML pipeline framework analysis



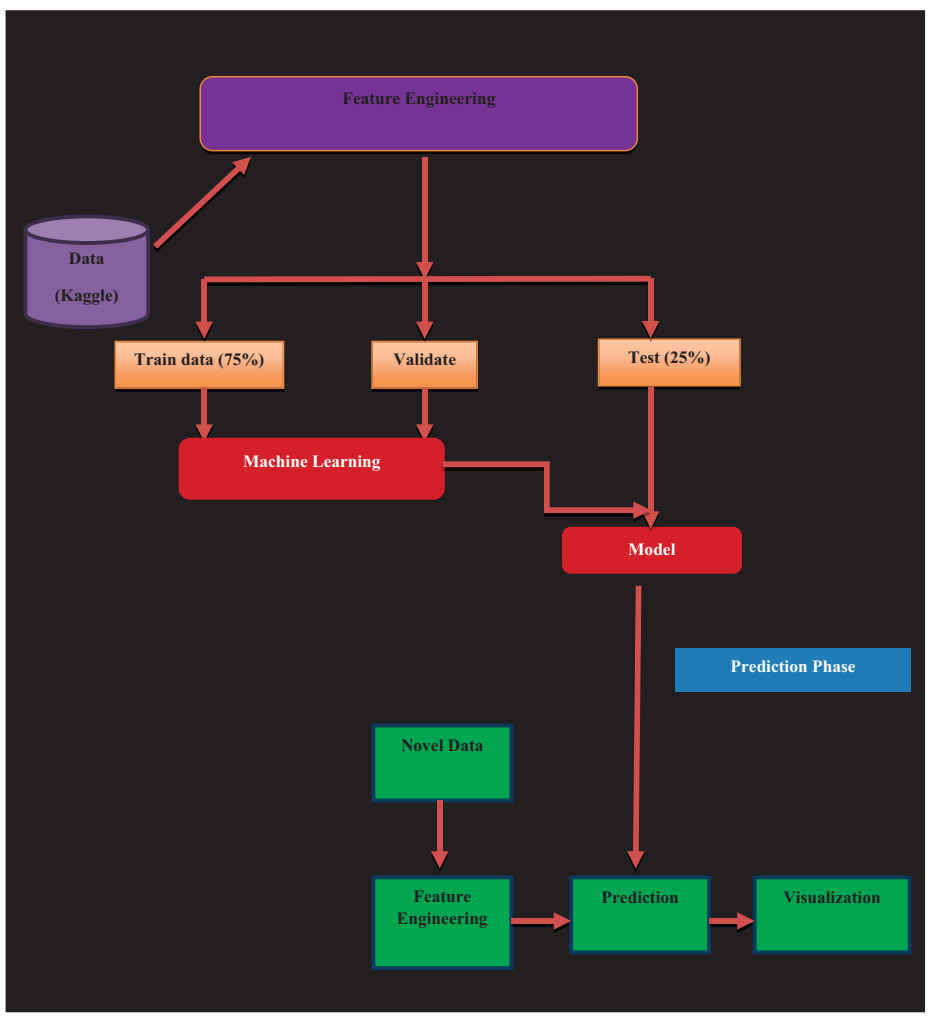
now novel and we can begin to see a path ahead. Simulated intelligence can help us in taking care of the issues that should be tended to raise by the COVID-19 pandemic (Sujath *et al.*, 2020).

These days AI methods are used worldwide for forecasts because of its exactness. Be that as it may, to use AI (ML) methods, there are a couple of difficulties as almost no information is accessible. For example, the difficulties associated with preparing a model are the fitting choice of boundaries and the determination of the best ML model for expectation (Ilyas, 2016).

One of the biggest challenges in predicting the virus using AutoML (Gijsbers *et al.*, 2019; Yang *et al.*, 2018; Feurer *et al.*, 2018; Mahdavi *et al.*, 2019) is that the model performs really well on unseen data, there is still not enough data. This is very much a proof of concept, and there is a dire need of labeled data to continue to improve the models further and provide accurate data. To diminish the major improvement costs, a novel thought of computerizing the whole pipeline of ML has evolved, which is refer to as AutoML. It is intended to decrease the interest for information researchers and empower space specialists to automatically manufacture ML applications absent a lot of prerequisite for measurable and ML information (El Shawi *et al.*, 2019). Also some common algorithms such as linear regression, multilayer perceptron and vector auto regression are used to predict COVID-19. (Sujath *et al.*, 2020) also used day-level forecasting for COVID-19 spread (Elmousalami and Hassanien, 2020), a hybrid deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization algorithm (Ezzat and Ella, 2020) and used some CT scan images for detecting COVID-19 (Barstugan *et al.*, 2020).

Figure 2 describes the overview and pipeline framework of AutoML which comprises data processing, feature customization, model formulation and model estimation. Each section of AutoML works in a pipeline structure and it has its own components. There are different methods of optimization, architecture optimization is part of Neural Architecture Search that communicates with the model estimation. On the other hand, hyperparametric optimization forms the important component of model formulation (El Shawi *et al.*, 2019). This process clearly explains the AutoML pipeline framework. This

Figure 3 Proposed AutoML architecture



paper uses this model to analyze the data set of COVID 19 patients to analyze the patterns and formulate the hypothesis based on the analysis.

3. Methodology

3.1 AutoML

AutoML provides resources and platforms to render ML available to non-ML experts, to boost efficiency in ML and to accelerate research in ML. H2O AutoML is used to generate the results (Dulhare *et al.*, 2020).

ML has achieved major milestones in recent years, and it is on which an increasing range of disciplines depend. But this performance is crucially dependent on specialists in human ML to perform the following tasks:

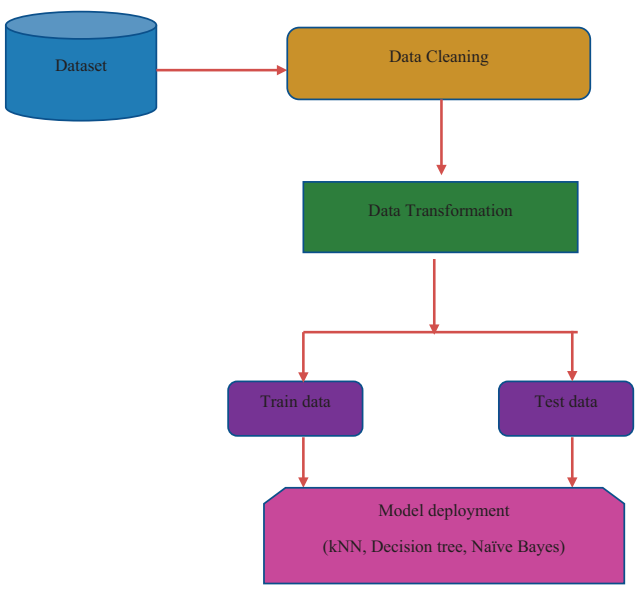
- preprocess the info, and clean it;
- choose and create the appropriate apps;
- choose a family that fits the pattern;
- optimize hyper parameters for layout;
- models of computer learning post processes; and
- review of the findings collected is important.

As the difficulty of such activities frequently goes beyond non-ML-experts, the exponential growth of ML implementations has generated a market for off-the-shelf ML solutions that can be used quickly and without experience. We name the resulting work field which is oriented toward the radical automation of AutoML (Figure 3).

3.2 Phase 1: training phase

This phase is the initial phase. The accuracy of the training data dramatically affects the reliability of the model you are making. By definition, the consistency of the forecasts that have come back from that process. Once you have submitted it, you cannot change the details. When you need to modify the training data after importing it, you may need to delete and re-import the source data.

Figure 4 Machine learning architecture



The data set has been obtained from the Kaggle [1]. The first step in generating useful training data is to ensure the question is well defined and can produce the outcomes you expect to anticipate. If you are new to ML, you can study the types of problems that are solved by AutoML tables and determine which sort of model you want to build.

Determine the details that will be used in the preparation. In certain situations, you might be able to use a table already present in your company records. For other instances, to build the most efficient training data, you must combine data from various places, probably execute certain data transformations or remove specific columns.

3.3 Phase 2: prediction phase

3.3.1 Cox proportional hazards

The most commonly used approach to estimating period to incident details is the Cox proportional hazards simulations. As the name implies, the danger function measures the instantaneous probability of an incident occurrence.

This mixture of a non-parametric baseline danger feature and a parametric risk score results in a semi-parametric definition of the Cox proportional hazard models. Furthermore, a quick rearrangement of terms reveals that unlike generalized linear models, the presence of a baseline hazard function and intercept (constant) term in risk score contribute little benefit to the model fit.

Table 1 Comparison metrics of GramSVD and PowerSVD

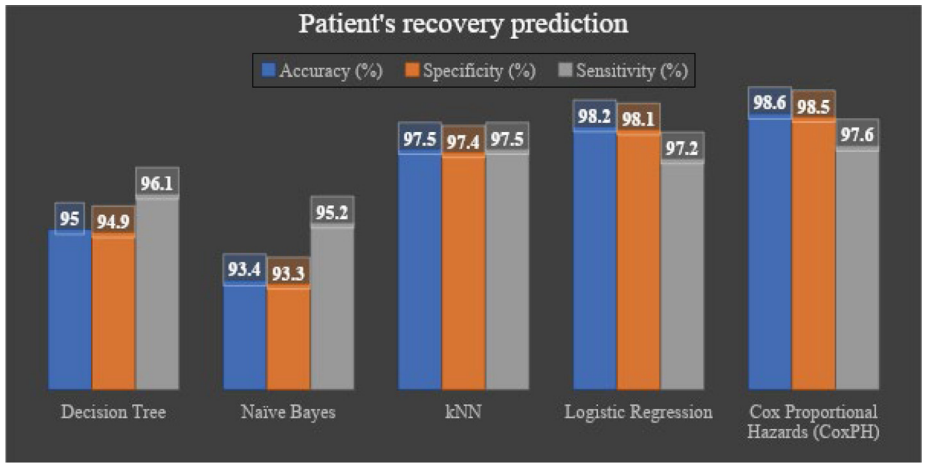
Statistical value	Gram SVD	Power SVD
Standard deviation	10,330.6604	10,330.6604
Proportion of variance	0.9664	0.9664
Cumulative proportion	0.9664	0.9664
Throughput time	3.540 s	3.000 s

Table 2 Data set attributes

S.No.	Attributes	Value
1.	Age	05–100
2.	Gender	Male: 0 Female: 1
3.	Visited any foreign countries	Yes: 1 No: 0
4.	Contact with COVID-19 patients	Yes: 1 No: 0
5.	Clinical manifestation	Fever: 1 Headache: 2 Chest pain: 3 Dry cough: 4 Tiredness: 5 Sore throat: 6 Fatigue: 7 Breathing difficulty: 8
6	Criticality	Normal: 1 Critical: 2 Severe: 3

Note: No. of data set acquired: 150,000

Figure 5 Patients recovery prediction



3.3.2 Principal component analysis

Principal components analysis (PCA) connected to the regression of principal components. The calculation is actualized on a lot of likely collinear qualities and executes a change to produce another assortment of uncorrelated characters (Aït-Sahalia and Xiu, 2019).

PCA generally used for reenactment without regularization or for rising dimensionality. It can likewise be helpful to lead as a pre-preparing stage before separation based calculations, for example, K-Means as PCA guarantees that all elements of a complex are symmetrical. The goal of PCA is to boost change while limiting covariance (Cui *et al.*, 2019).

3.3.3 Singular value decomposition

Singular value decomposition (SVD) is among the most significant lattice factorizations of the computational period, giving an establishment to almost the entirety of the information techniques in this book. The SVD gives a mathematically steady lattice deterioration that can be used for an assortment of purposes and is ensured to exist. We will use the SVD to get low-position approximations to networks and perform pseudo-inverses of non-square frameworks to find the arrangement of an arrangement of conditions Ax = b. Another significant utilization of the SVD is the hidden calculation of head part examination (PCA), where high-dimensional information is disintegrated into its most factually spellbinding variables. SVD/PCA has been applied to a wide assortment of issues in science and building (Figure 4; Tables 1 and 2).

4. Results and discussions

Figure 5 shows the comparison chart of patient's recovery (Tables 3 and 4). Figure 6 shows the algorithm of patient's comparison to predict the patients may affect with the disease in future. Cox proportional hazards algorithm shows the best accuracy when compared with the other four algorithms. Different colors show different parameters like blue color represent the accuracy parameter, orange color shows the specificity parameter and gray color shows the sensitivity parameter. By doing comparison, it has found that the Cox proportional hazards (CoxPH) has better performance in all the mention parameters among all the available methods.

5. Conclusion and future work

These days, the concept of AutoML techniques is being used in every field and domain for example, in the stock market, education institutions, medical field, etc. The ML tools play an important role in harnessing the massive amount of data. In this paper, the data set relatively holds a huge amount of data and appropriate analysis and prediction are necessary to

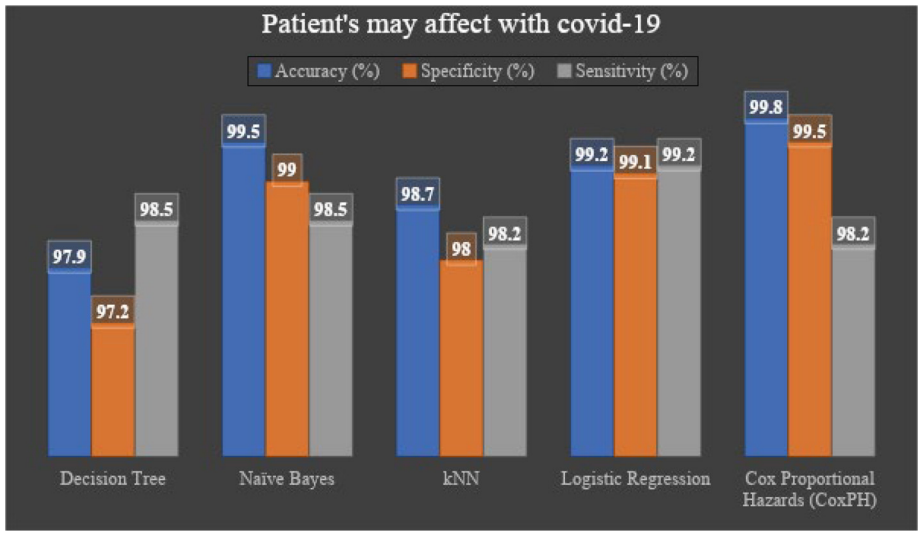
Table 3 Performance evaluation of machine learning and AutoML algorithm to predict patient's recovery

Predictive algorithms	Accuracy (%)	Specificity (%)	Sensitivity (%)
Decision tree	95	94.9	96.1
Naïve Bayes	93.4	93.3	95.2
kNN	97.5	97.4	97.5
Logistic regression	98.2	98.1	97.2
Cox Proportional Hazards (CoxPH) [AutoML algorithm]	*98.6*	*98.5*	*97.6*

Table 4 Performance evaluation of machine learning and AutoML algorithm to predict patient's may affect with COVID

Predictive algorithms	Accuracy (%)	Specificity (%)	Sensitivity (%)
Decision tree	97.9	97.2	98.5
Naïve Bayes	99.5	99	98.5
kNN	98.7	98	98.2
Logistic regression	99.2	99.1	99.2
Cox Proportional Hazards (CoxPH) [AutoML algorithm]	99.8	99.5	98.2

Figure 6 Algorithm comparison of patients may affect with COVID-19



track as the numbers of COVID-19 cases are increasing day by day. This prediction of COVID-19 will be able to track the cases particularly in India and might help researchers in the future to develop vaccines. Researchers across the world are testing different medications to cure COVID-19; however, it is still being tested in various labs. This paper highlights and deploys the concept of AutoML to analyze the data and to find the best algorithm to predict the disease. Appropriate tables, figures and explanations are provided.

Note

1 www.kaggle.com/sudalairajkumar/covid19-in-india

References

Aït-Sahalia, Y. and Xiu, D. (2019), "Principal component analysis of high-frequency data", *Journal of the American Statistical Association*, Vol. 114 No. 525, pp. 287-303.

Barstugan, M. Ozkaya, U. and Ozturk, S. (2020), "Coronavirus (covid-19) classification using ct images by machine learning methods", arXiv preprint arXiv:2003.09424.

Chen, X. and Gupta, A. (2015), "Webly supervised learning of convolutional networks", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1431-1439.

Cui, Z., Li, F. and Zhang, W. (2019), "Bat algorithm with principal component analysis", *International Journal of Machine Learning and Cybernetics*, Vol. 10 No. 3, pp. 603-622.

Dulhare, U.N., Mubeen, A. and Ahmad, K. (2020), "Hands-on H2O machine learning tool", *Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications*, Vol. 1, pp. 423-453.

El Shawi, R. Maher, M. and Sakr, S. (2019), "Automated machine learning: state-of-the-art and open challenges".

Elaziz, M.A., Hosny, K.M., Salah, A., Darwish, M.M., Lu, S. and Sahlol, A.T. (2020), "New machine learning method for image-based diagnosis of COVID-19", *PLoS One*, Vol. 15 No. 6, available at: https://doi.org/10.1371/journal.pone.0235187

Elmousalami, H.H. and Hassanien, A.E. (2020), "Day level forecasting for coronavirus disease (COVID-19) spread: analysis, modeling and recommendations", arXiv preprint arXiv:2003.07778.

Ezzat, D. and Ella, H.A. (2020), "GSA-DenseNet121-COVID-19: a hybrid deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization algorithm", arXiv preprint arXiv:2004.05084.

Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M. and Hutter, F. (2018), "Practical automated machine learning for the automl challenge 2018", *Proc. Int. Workshop Autom. Mach. Learn. (ICML)*, pp. 1189-1232.

Gijsbers, P. LeDell, E. Thomas, J. Poirier, S. Bischl, B. and Vanschoren, J. (2019), "An open source AutoML benchmark", arXiv preprint arXiv:1907.00909.

He, X. Zhao, K. and Chu, X. (2019), "Automl: a survey of the state-of-the-art", arXiv preprint arXiv:1908.00709.

Ilyas, F. (2016), "Effective data cleaning with continuous evaluation", *IEEE Data Eng. Bull*, Vol. 39 No. 2, pp. 38-46.

Karras, T. Laine, S. and Aila, T. (2020), "A style-based generator architecture for generative adversarial networks", arXiv preprint arXiv:1812.04948.

Mahdavi, M., Neutatz, F., Visengeriyeva, L. and Abedjan, Z. (2019), "Toward automated data cleaning workflows", *Machine Learning*, Vol. 15.

Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J. and Hassanien, A.E. (2020), "Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art", *SN Computer Science*, Vol. 1 No. 4, p. 197, available at: https://doi.org/10.1007/s42979-020-00209-9

Sidey-Gibbons, J. and Sidey-Gibbons, C. (2019), "Machine learning in medicine: a practical introduction", *BMC Medical Research Methodology*, Vol. 19 No. 1, p. 64, available at: https://doi.org/10.1186/s12874-019-0681-4

Sujath, R., Chatterjee, J.M. and Hassanien, A.E. (2020), "A machine learning forecasting model for COVID-19 pandemic in India", *Stochastic Environmental Research and Risk Assessment*, Vol. 34 No. 7, pp. 959-972, available at: https://doi.org/10.1007/s00477-020-01827-8

Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C.B. and Farivar, R. (2019), "Toward automated machine learning: evaluation and comparison of automl approaches and tools", *IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE.

Xu, Z., Huang, S., Zhang, Y. and Tao, D. (2015), "Augmenting strong supervision using web data for fine-grained categorization", *Proceedings of the IEEE international conference on computer vision*, pp. 2524-2532.

Yang, J., Sun, X., Lai, Y.-K., Zheng, L. and Cheng, M.-M. (2018), "Recog- nition from web data: a progressive filtering approach", *IEEE Transactions on Image Processing*, Vol. 27 No. 11, pp. 5303-5315.

Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L. and He, D. (2020), "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak", *International Journal of Infectious Diseases*, Vol. 92, pp. 214-217.

Zöller, M.-A. and Huber, M.F. (2019), "Benchmark and survey of automated machine learning frameworks", arXiv preprint arXiv:1904.12054.

Corresponding author

Mukesh Soni can be contacted at: soni.mukesh15@gmail.com