Chapter 10
Machine Learning Approach Using KPCA-SVMs for Predicting COVID-19

Micheal Olaolu Arowolo, Roseline Oluwaseun Ogundokun, Sanjay Misra, Akeem Femi Kadri, and Tahir Olanrewaju Aduragba

10.1 Introduction

The world has met numerous epidemics in the past decades. Lately, a deadly sickness identified as COVID-19 has surfaced from China [1, 2]. Inimitable public health adversity is triggered by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [2–4]. The World Health Organization (WHO) termed the new epidemic as COVID-19. It is acknowledged as a Public Health Emergency of International Concern since the beginning of 2020. It was considered an epidemic around the first quarter of 2020 [5–7], as Americans joined forces with investigative institutions and scientific concerns for global artificial intelligence (AI) investigators' activities in evolving groundbreaking machine learning measures that will help tackle COVID-19 linked surveys [3]. COVID-19 is a novel solitary intelligent ribonucleic acid (RNA) germ comprising of a huge pathological genomic sequence. It alters advances fast with no specific limitation for evaluating or investigating techniques or suitable medications. Secluding infected individuals by quarantining is the utmost way of safeguarding the universe from more escalation of the deadly COVID-19 [8].

Currently, the utmost reliable analytical examination takes time as far as a week, hence, vulnerable to unevenly accurate, raising the risk of unproductive

M. O. Arowolo · R. O. Ogundokun
Department of Computer Science, Landmark University, Omu-Aran, Nigeria
e-mail: ogundokun.roseline@lmu.edu.ng

S. Misra (✉)
Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria
e-mail: sanjay.misra@covenantuniversity.edu.ng

A. F. Kadri · T. O. Aduragba
Department of Computer Science, Kwara State University, Malete, Nigeria

© Springer Nature Switzerland AG 2022

L. Garg et al. (eds.), *Healthcare Informatics for Fighting COVID-19 and Future Epidemics*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-030-72752-9_10

dissemination of medical kits and resources. There is a necessity for instantaneous findings for medical and precautionary actions [9].

Machine learning has proven to be an advanced method for discovering relevant information for the society of forthcoming immoderations and inconsistent discovery of outliers through developing data-driven health conditions, calculating risk scores, utilizing machine learning models for testing and training purposes, utilizing classification approaches, and developing application platforms, among others [10]. In a rapid period of COVID-19 occurrence, innovative machine learning procedures have been proposed in classifying COVID-19 genomes [11], subsistence prognosis of acute COVID-19 patients [12], and CRISPR-based COVID-19 discovery evaluation [13] and realizing probable medication candidates counter to COVID-19 [14]. Continuous efforts are ongoing to progress innovative investigative methods using machine learning procedures such as machine learning-built transmission of SARS-CoV-2 analytical strategies with a CRISPR-built infectious discovery structure [10]. Neural network classifiers approach important information about COVID-19-infected persons built on their diverse breathing design [15]. A deep learning analytical approach for thoracic CT images was carried out to perceive and control COVID-19 patients automatedly [16]. The readiness of COVID-19-associated medical data for processing detailed reachable records has been a crucial existing barrier [3].

This study proposes a machine learning method and exploits for prediction approach for COVID-19 occurrence in humans as artificial intelligence and machine learning are healthily incisive for prediction and diagnostics methods to fight COVID-19 pandemic. An enhanced kernel principal component analysis (KPCA) is used to fetch relevant kernel latent component features of genes that clinicians can efficiently use by clinicians to discover prescriptions, predictions, and diagnosis of patients suspected to be infected. Extracted features are classified by passing them into SVM-RBF classifier to classify the COVID-19 for prediction and the system performance is evaluated using the confusion matrix.

The remaining part of the paper is structured as thus: Sect. 10.2 addressed the overview of the investigation and of related investigations. Section 10.3 provides research materials and methods. Section 10.4 provides experimental research outcomes. Section 10.5 provides the conclusion of the investigation.

10.2 Background and Related Works

Machine learning is a statistical and mathematical method for predictions [17]. Prediction is a significant aspect of interest in healthcare, such as the kind of virus, which is probably predominant in the coming influenza period. What ampoules of virus vaccination to prepare ahead? Relatively, prediction is not the same as treatment properties. Numerous machine learning approaches can evaluate treatment properties, although there are discrepancies between treatment-effect and prediction assessment, which concern machine learning works.

Significant machine learning includes preprocessing, segmenting datasets into validation, and learning procedures to develop high developed classification accuracy algorithms on the proposed full dataset for prediction purposes. Machine learning approaches contain different systems, such as classification trees [17, 18].

Machine learning has increased and extracts general concepts from massive datasets in algorithms to predict outcomes, especially in healthcare. It is an emerging field promising for data preparation, preprocessing, normalization, evaluation, and parameter erudition. The purpose is to acquire an evaluation that corresponds participations to a result. Predicting impending hospitalization and pandemic cases as we witness them is crucial, and analysts gather evidences for successful measurements and forecasts. Machine learning uses a predictor and learner to perform its task by measuring how often the learned predictor appropriately labels patients' readmissions [19]. Machine learning methods are efficient to health facility investigators looking forward to improving the prediction of a healthcare result to train and improve the machine learning prediction algorithm. This approach can support comprehensive data driven by offering significant encounters for medical results that are of value. Computational enhancements in hardware and cloud computing knowledge have made machine learning approaches progressively available to medical investigators and medical industries. In a period of an epidemic as we presently observe, evolving machine learning procedures are possibly valuable for the healthcare practitioners if a forecast is a momentous and eloquent attempt. The prognosis could be joint with a fundamental investigation to advance consideration as machine learning approaches advance. Several machine learning algorithms have evolved and can be adopted for the novel COVID-19, such as the deep learning approach, SVM, and decision trees [18–23].

As shown in Fig. 10.1, the proposed workflow consists of a proposed machine learning model, which can be used to train a test several implementation processes that can enhance the prediction and detection of COVID-19 to help for proper medications and future occurrence.

Several research abstracts evolve from COVID-19 with several dataset repositories like Kaggle updating their outlets daily on the novel COVID-19. These documents help in extracting richer content for the application of machine learning [24]. Dimensionality reduction is a vital aspect of machine learning. It allows summary data from the curse of dimensionality and reduces computational complexity without halting the information's performance. Examples of efficient dimensionality reduction models include kernel principal component analysis (KPCA), ISOMAP, partial least square, and independent component analysis, among others [25].

KPCA is a kernel principal component analysis. It has played an essential role in dimensionality reduction by removing irrelevant information from a given unique dataset as a preprocessing phase in machine learning. It is executable and operative when working with mutual instances commonly encountered. Complex cases are unproductive using PCA. Recently KPCA is a modified PCA that has been proposed to confront the nonlinearity difficulties. As a nonlinear PCA method, KPCA could resourcefully calculate principal constituent in extreme dimensional prominence space using essential operators and nonlinear kernel purposes [26].

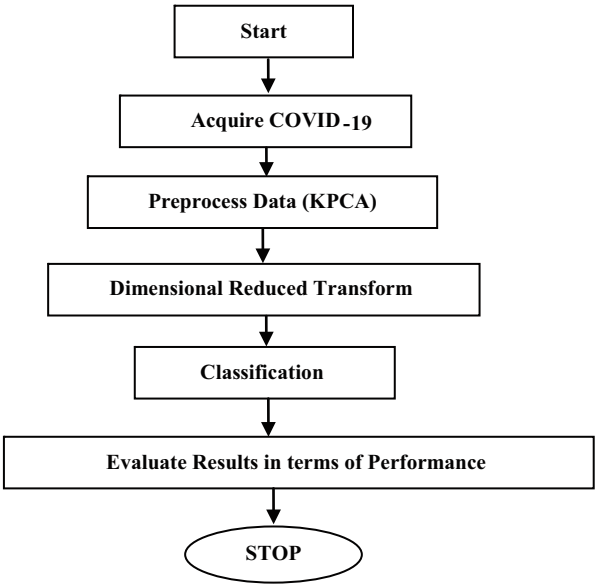


Fig. 10.1 Workflow approach for proposed methodology

PCA suffers from a probabilistic understanding, with no fundamental probability density model. PCA suffers from the predictable covariance matrix and turn out to be rank lacking. Case like this in the previous study is identified as small illustration delinquent in extreme dimensions. Due to the curse of dimensionality, classical multivariate approaches degenerate and break down. The data covariance matrix cannot be calculated and obtained ones classify poorly, due to high noise formed by irrelevance and redundancy of genetic factor (i.e., features) existing in the data.

This study addresses this limitation of PCA to develop and introduce a novel dimension reduction technique for gene expression analysis classification. This study introduces and uses probabilistic kernel principal component analysis (KPCA) as an alternative to the classic PCA.

KPCA methods enable the construction of different nonlinear versions of the algorithm and reduce dimensionality, thereby improving linear PCA.

Before applying the same optimization as PCA, kernel PCA (KPCA) maps the data into a higher dimensional feature space [46–48]. The mapping can be done implicitly by using a kernel function. The Gaussian kernel $K(xi, xj) = \exp(-\|xi - xj\|2\sigma 2)$ was applied in our study.

Clustering techniques are unsupervised learning methods for creating groups of similar samples from a population. Several clustering methods exist, such as the k-means and hierarchical agglomerative clustering, among others [23], intending to find the best clustering representation for a whole data sample space. Deep learning is a feature representation used to learn rich, informative illustrations of information

and extract highly representative features for overcoming overfitting, training, and augmentation of data and detecting inferences on clinical applications [27, 37].

The classification approach in machine learning helps discover class separations of samples for meaningful interpretations in substantive theory context. Several efficient classifiers exist, such as the support vector machine (SVM) [45], KNN, decision trees, and ensemble, among others, which can be proposed to help in the diagnosis and prediction of COVID-19 [28]. Machine learning techniques help the computer learn the use of training information processes to predict and make intelligent decisions. Several studies of its approaches and algorithms for modelling and performing tasks are of the essence.

An SVM is a robust classification algorithm with superior performance in several applications such as medical and educational, among other machine learning approaches. Due to its better understanding, it has been used for classification and analysis. Several machine learning methods have been proposed in predicting the diagnosis and threats of COVID-19, such as random forest and neural networks [9].

SVM move features to developed dimensions that can be distinguishable by a hyperplane by maximizing this hyperplane for more accurate classification. Consider a series of data points $D = [xi, di]$ i n where n is the size of data, di represents the target value, and xi represents the input space vector of the sample.

The SVM evaluates the function as given in the following two equations:

$$(x) = \omega\varphi(x) + b \tag{10.1}$$

where $\varphi(x)$ is the high-dimensional space feature, b is a scalar, and ω is a normal vector.

10.2.1 Related Work

Adegboye et al. [29] proposed a main advance detail of new COVID-19 in Nigeria by providing a preliminary epidemiological study of COVID-19 eruption in Nigeria by approximating the original contagious over time, varying imitation total constructed on the Bayesian method with uncertainty in the dispersion of serial discontinuity of indications in individuals and used for virus access. They test the transcription volume every day for the prevalence, and assessments suggested that Nigeria's COVID-19 occurrence is reduced than probable. The limitation of this study is the robustness of the study and analyzing the records.

Ohia et al. [30] suggested a COVID-19 and Nigeria certainties in the situation for critical need in positioning into perceptions, certainties strange globally by discovering joint procedures and conciliations for conveying the epidemic. The gaps in this study are coming up with relevant decision-making approach. Introducing statistical and computational learning is of the essence.

Ozili [31] suggested a COVID-19 epidemic, financial predicament knowledge and operational reasons. The globe is in an economic predicament by examining

overabundance to countries like Nigeria and operational faintness in organizations with the present disaster and predictions for transformation. The limitation of this study calls for the survey and statistical approaches for boosting the economy in post-COVID era.

Dutta and Bandyopadhyay [5] suggested a machine learning method for checking COVID-19 occurrence, with neural network deep learning procedure, efficient prediction, and faster achievable detailed virus recognition. Further insight is of the essence by consulting more efficient machine learning techniques.

Alimadadi et al. [3] discoursed on the necessity for artificial intelligence and machine learning to fight COVID-19, imperative developments of cyberinfrastructure in powering international associations and creations of links with creativities, and linking COVID-19-associated health information with predominant biorepositories that can utilize endeavors in the direction to a nearer and attainable technique for meaningful data mining by bioinformaticians and computational scientists, to progress predictive, diagnostic, and valuable methods against COVID-19 and similar future outbreaks.

Alom et al. [32] suggested an advanced determinate study method to control the amount of infested zones in x-ray and CT pictures on COVID-19 case discovery, with deep learning multi-task methods by using CNN with a transfer learning (TL) technique for COVID-19 detection; NABLA-N network was demonstrated to section areas COVID-19 has infested. Their template achieved an accuracy of 85% and 99%. This study requires further improvement by employing other data forms to the knowledge apart from the CT and x-ray images.

Pinter et al. [33] proposed a COVID-19 overall evaluation for Hungary with a combination machine learning technique for the prognosis of COVID-19 and signifying its concealed features using Hungarian data on an ANFIS-MLP-ICA integrated procedure to predict the period classification of ailing individuals and demise ratio. For future works, gaps in this study need to tackle computational time and accuracy enhancement.

Batista et al. [9] proposed a COVID-19 investigation evaluation in emergency convalescent care. Using the machine learning technique, by predicting COVID-19-positive investigation risks, they collected patient's information and established a helpful verdict of COVID-19 from RT-PCR tests. Random forests, neural networks, logistic regression, gradient boosting trees, and SVM algorithms were used to train patients' samples, and the presentation was verified on novel unobserved data. SVM outperforms other classification algorithms with an 85% AUC and found out that targeted choices for getting COVID-19 tests with data habitually composed are a capable, innovative part using machine learning procedures. Future works require the optimization of the classifier, as it should be further improved.

Rhadhawa et al. [11] proposed a machine learning method with intrinsic genomic signatures by merging supervised machine learning with digital signal processing (MLDSP) for gene calculations, enhanced by a decision tree and a Spearman's rank correlation persistent study for result evaluation with enormous limited epidemiologic gene classification dataset. Their method achieves 100% classification accuracy and understands relevant genetic genes' relations with raw

chromosome system data with no precise genetic data, training, protein sequence, or gene annotations. The alignment-based method limitations consist of a comparison of approaches to complement classification taxonomic is of the essence.

Onoja [10] worked on the machine learning process for treating persons medical extent on COVID-19 crisis, by using an imposing end-to-end data determined signifying technique for exploiting review privileges for information gatherings feeding Supervised Machine learning plans for determining person's medical status, by evolving an effective way determining, categorizing, and predicting their medical rank. They were supporting medical care personnel to identify cases and acknowledge suitable medical attention in their localities. There is a further necessity required for current state and need for minimizing COVID-19 spread.

Zhang et al. [1] used an in-depth learning approach for COVID-19 suppository profiling by gathering RNA-Seq disease with GISAID data by screening RNA-Seq records into protein sequences and then building a 3D protein model with homology modelling. COVID-19 primary protease is a critical, valuable goal and is occupied on medication screening grounded on the exhibited COVID-19 protease structure. DFCNN procedure was used to rank and recognize protein-ligand communications with high accuracy. The mechanism between the molecular dockings is a limitation and requires an important biological understanding of COVID-19.

Beck et al. [34] proposed a predicting antiviral drug that may be made for current or present COVID-19 pandemic with a deep-learning method with a proficient medication-target Fragment Transformer-Drug Target Interaction to discover saleable prevailing drugs that can accomplish pathological proteins of COVID-19. Antiretroviral medication for giving and averting HIV was suggested as the superlative biological composite, although there is no applied suggestion to back up the proposal. This study's limitation requires the development of useful control of spreading COVID infections by proposing prediction approached that will help experimental therapy identify the infection.

Ardabili et al. [35] worked on a machine learning comparative study for predicting COVID-19 prevalence as an auxiliary for SIR and SEIR illustrations. Two models obtained proficient results between the proposed machine learning designs identified as the MLP and ANFIS methods. Due to the complicated way of the COVID-19 existence and its interactive inconsistency worldwide, their study suggested machine learning as an effective method for demonstrating the prevalent by providing standards for indicating the likelihood of machine learning for impending investigations and originalities in general predictions can be realized with machine learning and SEIR methods. Modelling for the mortality rate is essential to plan for facilities, by integrating machine learning with the proposed model, to enhance existing study in terms of accuracy.

Xu et al. [36] operated on a deep learning method for COVID-19 and attained 87% accuracy for vigorous cases. This study presented a novel method for screening CT images of COVI-19 using deep learning. The model mechanism requires enhancement in terms of its accuracy, and it will be a promising, supplementary diagnostic approach by using robust data (Table 10.1).

Table 10.1 Comparison of existing works on COVID-19

S/No	Authors and work done	Findings	Deduction
1.	Adegboye et al. [29]	Assessed the replica amount of daily outbreak using a three weekly space through fine-tuning the cases	Requires relevant analysis
2.	Dutta and Bandyopadhyay [5]	Uses deep learning neural network with long short-term memory (LSTM) and gated recurrent unit (GRU) for training the data and the prediction results for COVID-19 cases	Call for accuracy enhancement
3.	Alom et al. [32]	An efficient COVID-19 patient identification approach using multi-task deep learning was proposed. CT scan and x-ray image dataset were evaluated. An inception recurrent residual convolutional neural network with transfer learning (TL) method for COVID-19 discovery and NABLA-N network model to segment infected regions were used	Call for further classification approaches with dimensionality reduction techniques
4.	Pinter et al. [33]	Proposed hybrid machine learning method for predicting COVID-19 using Hungary data. An adaptive network-based fuzzy inference system (ANFIS) and multi-layered perceptron-imperialist competitive algorithm (MLP-ICA) were utilized	Enhancement of the machine learning approach
5.	Beck et al. [34]	Proposed deep learning approach for drug-target interaction model for COVID-19	Prediction result enhancement
6	Raveendran et al. [38]	Random forest and support vector regression methods are used for the COVID-19 contagion dataset and compared the prediction accuracy. PCA fetched the relevant features classified	Call for enhanced dimensionality approaches for better classification
7.	Ardabili et al. [35]	Compared machine learning approaches using the susceptible, infected recover and susceptible exposed infectious removed model, with ANFIS using COVID-19 dataset	Enhancement of the accuracy
8.	Xu et al. [36]	Proposed a deep learning approach for screening COVID-19 from IAVP and pulmonary images	Accuracy enhancement

10.3 Materials and Methods

Recent helpful methods are essential for treating COVID-19 carriers globally. There is no definite, recognized medication. It is risky to establish inventive methods to repurpose tolerable health drugs or propose innovative drugs for SARS-CoV-2. Machine learning-built inversion and persistence context is proposed to rank present drugs for SARS-CoV-2 investigational examinations. Machine learning-built drug unearthing vessel has been projected for designing and producing advanced drug-like compounds for SARS-CoV-2. Probable protein structures associated with COVID-19 methods have been proposed to help as an indication for

the COVID-19 vaccine plan. Countless COVID-19 management data in hospitals all-inclusive desires innovative machine learning methods to examine improved beneficial properties for evaluating patient's prediction and improved care for patients and communications and suitable provisions. Diverse machine learning methods exist for various responsibilities such as detection and prediction [46–48] and segmentation models.

10.3.1 Database

Since the arrival of COVID-19, scholars have been gathering associated data for the prevalence. Several datasets for engaging machine learning methods have been proposed, such as the pneumonia detection models from openly accessible datasets [3, 9, 20, 21, 34, 39], to support the fast and extensive distribution of study for the COVID-19 epidemic. Numerous bodies like the World Health Organization have been fetching investigators and health consultants together to speed exploring procedures and growths to advance values that can hold the spread of this deadly epidemic and support care for the affected people. The dataset from the COVID-19 medical record worldwide sample of the available data online given by WHO is concerned about patients diagnosed positively with COVID-19 [40].

The upper airway gene expression differentiates COVID-19 from other acute respiratory diseases in this analysis. It shows the suppression of innate immune responses by the SARS-CoV-2 dataset from the University of California San Francisco after August 2020.

The data involves high-throughput sequencing expression profiling with GPL24676 234 samples of *Homo sapiens* species. Out of 234 patients with COVID-19 ($n = 93$), other acute respiratory diseases (ARIs) are infectious ($n = 100$) or non-viral ($n = 41$). Compared to other viral ARIs, COVID-19 patients displayed substantially decreased amounts of neutrophils, macrophages, and increased amounts of goblet, dendritic, and B cells [41].

S/No	COVID-19 data archives
1.	https://github.com/ieee8023/covid-chestXray-dataset
2.	www.worldometers.info/coronavirus/country/hungary/
3.	www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset
4.	data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases
5.	dataverse.harvard.edu/
6.	www.who.int/tb/country/data/download/en/
7.	github.com/datasets/covid-19
8.	ourworldindata.org/coronavirus-testing
9.	springernature.com/gp/researchers/campaigns/coronavirus/research-data
10.	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE156063
11.	https://github.com/owid/covid-19-data/tree/master/public/data

10.3.2 Economy

Since the global economic tragedy, the world has been brought into the middle of the nastiest recession. The economic downturn of countries like Nigeria has been impelled by the incorporation of worsening oil value and spillovers from the COVID-19 widespread and engaged to a radical lineage in the cry for oil products and still economic events from taking place when communal hostility strategies were required.

Anxiety thrives concerning the new epidemic and its consequences as cumulative records losses surge with costs owing to loss of production, and comorbidities are of concern [42]. The financial results are unfavorable to employment, health, travel, sport, trading, agriculture, market world, and merchandise, among others [43]. Some of the recommended strategic solutions are positive administration methods, health strategy basis to address societal causes of health, education, administration, knowledge, nationwide and global modifications in investments, private and public corporations, and World Technical Council establishments on COVID-19. Operative application of policy keys will necessitate the total backing of all investors, such as governments, the mass media, NGOs, health specialists, public, and persons. An extensive socioeconomic progress strategy in sector diplomacies and an environment encouraging enterprise are required to enhance commercial productions. It is wise for administrations and economic institutes to always evaluate the situation and ensure a better living [17, 44].

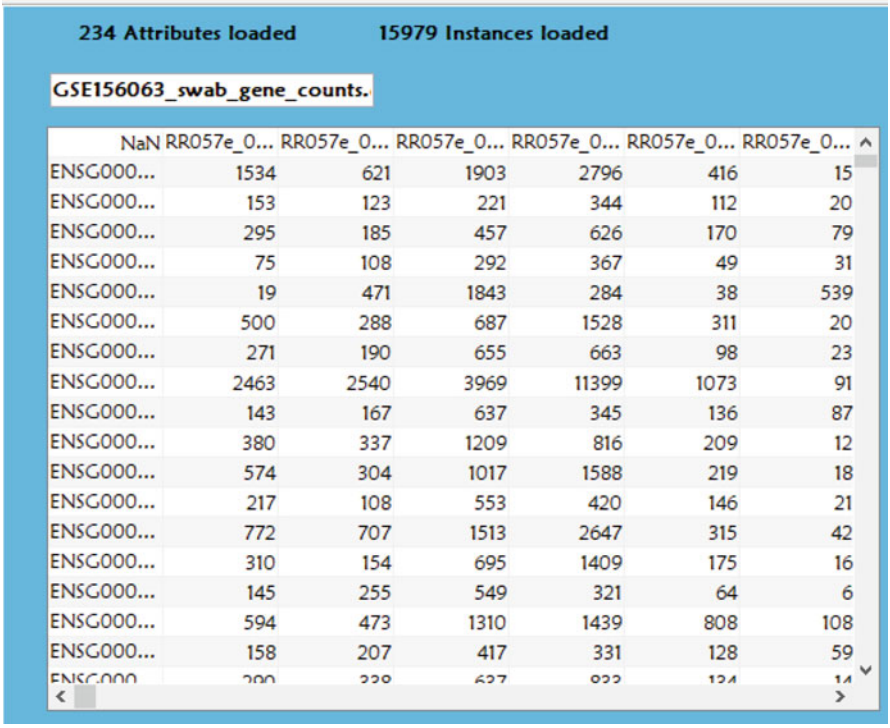
10.4 Experimental Results

In this study, KPCA was used to fetch out relevant information from the high-dimensional data comprising of 234 attributes and 15,979 instances of COVID-19 sample genes. SVM classification algorithms were used as a model to test the experiment's prediction; Fig. 10.2 shows the dataset information.

KPCA fetched relevant information from the data with 33 latent component attributes of the relevant genes. Figure 10.3 shows the extracted features. The extracted features projects relevancy with the irrelevant information been eliminated. The KPCA aids nonlinearity. It helps classify improvements by representing and integrating the input variables to reveal interpretability.

Figure 10.4 shows the computational time it takes for KPCA to execute and 15.0318 s was recorded as the time taken. The extracted features are then passed to the classifier for evaluation.

The findings discovered from this study can help clinicians in decision-making. Figure 10.5 shows the classification scattered plot; it represents the individual values of the data point with their variable relationships. KPCA output result is also shown in Fig. 10.5.



234 Attributes loaded 15979 Instances loaded

GSE156063_swab_gene_counts.

NaN	RR057e_0...	RR057e_0...	RR057e_0...	RR057e_0...	RR057e_0...	RR057e_0...
ENSG000...	1534	621	1903	2796	416	15
ENSG000...	153	123	221	344	112	20
ENSG000...	295	185	457	626	170	79
ENSG000...	75	108	292	367	49	31
ENSG000...	19	471	1843	284	38	539
ENSG000...	500	288	687	1528	311	20
ENSG000...	271	190	655	663	98	23
ENSG000...	2463	2540	3969	11399	1073	91
ENSG000...	143	167	637	345	136	87
ENSG000...	380	337	1209	816	209	12
ENSG000...	574	304	1017	1588	219	18
ENSG000...	217	108	553	420	146	21
ENSG000...	772	707	1513	2647	315	42
ENSG000...	310	154	695	1409	175	16
ENSG000...	145	255	549	321	64	6
ENSG000...	594	473	1310	1439	808	108
ENSG000...	158	207	417	331	128	59
ENSG000...	200	338	637	833	134	14

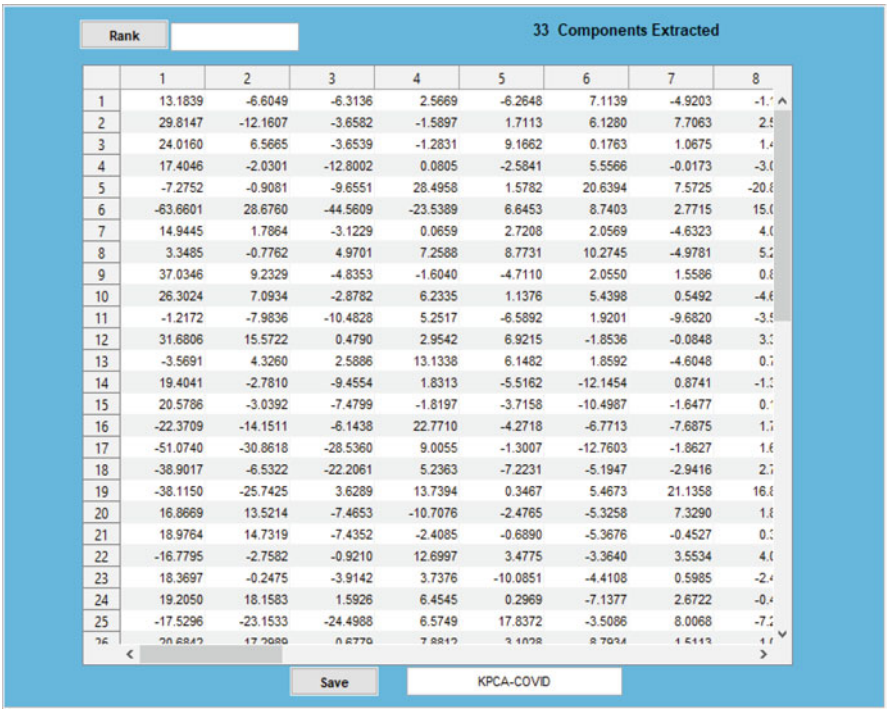
Fig. 10.2 The COVID-19 dataset

SVM-kernels using radial basis function and linear SVM were employed for classification with tenfold cross-validation using MATLAB. The confusion matrix is further used for the evaluation procedure, and their results were compared. Figure 10.6 shows the KPCA-L-SVM classification confusion matrix, used for evaluating the performance metrics for the experiment.

Figure 10.7 depicts the KPCA-SVM-RBF classification confusion matrix; it is used for evaluating the performance metrics for the experiment.

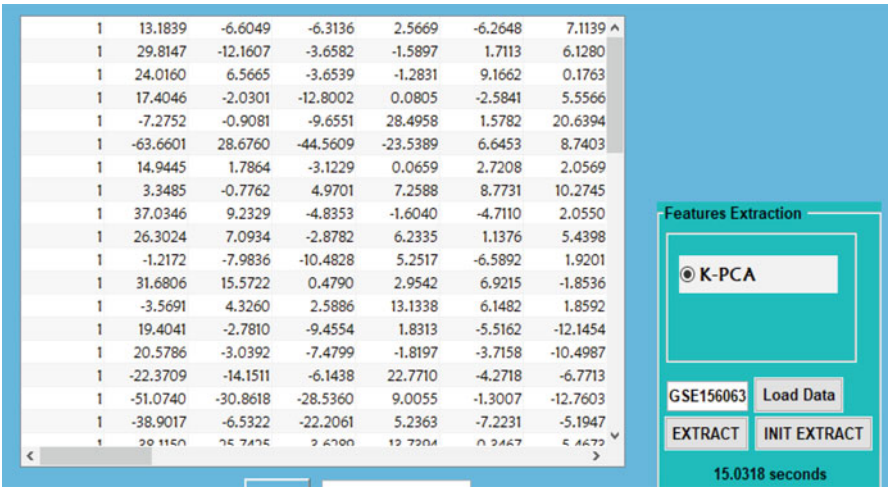
In this study, the reduction and classification of high-dimensional COVID-19 data were developed. The experiment used the KPCA to reduce the data and classify the experiment using the L-SVM and SVM-RBF classifiers. The experimental results are tabulated in Table 10.2.

In this study, KPCA and SVM were used for COVID-19 clinical data provided online to predict the pandemic. The result achieved 93% and 87% accuracy, respectively. With this prediction, this study can help clinicians' decision-making process by translating the data analysis using the machine learning perspectives. Presently with the data available, there are cumulated challenging issues and needs



	1	2	3	4	5	6	7	8
1	13.1839	-6.6049	-6.3136	2.5669	-6.2648	7.1139	-4.9203	-1.1
2	29.8147	-12.1607	-3.6582	-1.5897	1.7113	6.1280	7.7063	2.5
3	24.0160	6.5665	-3.6539	-1.2831	9.1662	0.1763	1.0675	1.4
4	17.4046	-2.0301	-12.8002	0.0805	-2.5841	5.5566	-0.0173	-3.0
5	-7.2752	-0.9081	-9.6551	28.4958	1.5782	20.6394	7.5725	-20.8
6	-63.6601	28.6760	-44.5609	-23.5389	6.6453	8.7403	2.7715	15.0
7	14.9445	1.7864	-3.1229	0.0659	2.7208	2.0569	-4.6323	4.0
8	3.3485	-0.7762	4.9701	7.2588	8.7731	10.2745	-4.9781	5.2
9	37.0346	9.2329	-4.8353	-1.6040	-4.7110	2.0550	1.5586	0.8
10	26.3024	7.0934	-2.8782	6.2335	1.1376	5.4398	0.5492	-4.8
11	-1.2172	-7.9836	-10.4828	5.2517	-6.5892	1.9201	-9.6820	-3.5
12	31.6806	15.5722	0.4790	2.9542	6.9215	-1.8536	-0.0848	3.3
13	-3.5691	4.3260	2.5886	13.1338	6.1482	1.8592	-4.6048	0.7
14	19.4041	-2.7810	-9.4554	1.8313	-5.5162	-12.1454	0.8741	-1.1
15	20.5786	-3.0392	-7.4799	-1.8197	-3.7158	-10.4987	-1.6477	0.1
16	-22.3709	-14.1511	-6.1438	22.7710	-4.2718	-6.7713	-7.6875	1.7
17	-51.0740	-30.8618	-28.5360	9.0055	-1.3007	-12.7603	-1.8627	1.8
18	-38.9017	-6.5322	-22.2061	5.2363	-7.2231	-5.1947	-2.9416	2.7
19	-38.1150	-25.7425	3.6289	13.7394	0.3467	5.4673	21.1358	16.8
20	16.8669	13.5214	-7.4653	-10.7076	-2.4765	-5.3258	7.3290	1.8
21	18.9764	14.7319	-7.4352	-2.4085	-0.6890	-5.3676	-0.4527	0.1
22	-16.7795	-2.7582	-0.9210	12.6997	3.4775	-3.3640	3.5534	4.0
23	18.3697	-0.2475	-3.9142	3.7376	-10.0851	-4.4108	0.5985	-2.4
24	19.2050	18.1583	1.5926	6.4545	0.2969	-7.1377	2.6722	-0.4
25	-17.5296	-23.1533	-24.4988	6.5749	17.8372	-3.5086	8.0068	-7.2
26	20.6842	17.2989	0.6779	7.8812	3.1028	8.7934	1.5113	1.1

Fig. 10.3 The extracted features using KPCA



1	13.1839	-6.6049	-6.3136	2.5669	-6.2648	7.1139
1	29.8147	-12.1607	-3.6582	-1.5897	1.7113	6.1280
1	24.0160	6.5665	-3.6539	-1.2831	9.1662	0.1763
1	17.4046	-2.0301	-12.8002	0.0805	-2.5841	5.5566
1	-7.2752	-0.9081	-9.6551	28.4958	1.5782	20.6394
1	-63.6601	28.6760	-44.5609	-23.5389	6.6453	8.7403
1	14.9445	1.7864	-3.1229	0.0659	2.7208	2.0569
1	3.3485	-0.7762	4.9701	7.2588	8.7731	10.2745
1	37.0346	9.2329	-4.8353	-1.6040	-4.7110	2.0550
1	26.3024	7.0934	-2.8782	6.2335	1.1376	5.4398
1	-1.2172	-7.9836	-10.4828	5.2517	-6.5892	1.9201
1	31.6806	15.5722	0.4790	2.9542	6.9215	-1.8536
1	-3.5691	4.3260	2.5886	13.1338	6.1482	1.8592
1	19.4041	-2.7810	-9.4554	1.8313	-5.5162	-12.1454
1	20.5786	-3.0392	-7.4799	-1.8197	-3.7158	-10.4987
1	-22.3709	-14.1511	-6.1438	22.7710	-4.2718	-6.7713
1	-51.0740	-30.8618	-28.5360	9.0055	-1.3007	-12.7603
1	-38.9017	-6.5322	-22.2061	5.2363	-7.2231	-5.1947
1	-38.1150	-25.7425	3.6289	13.7394	0.3467	5.4673

Fig. 10.4 KPCA computational time

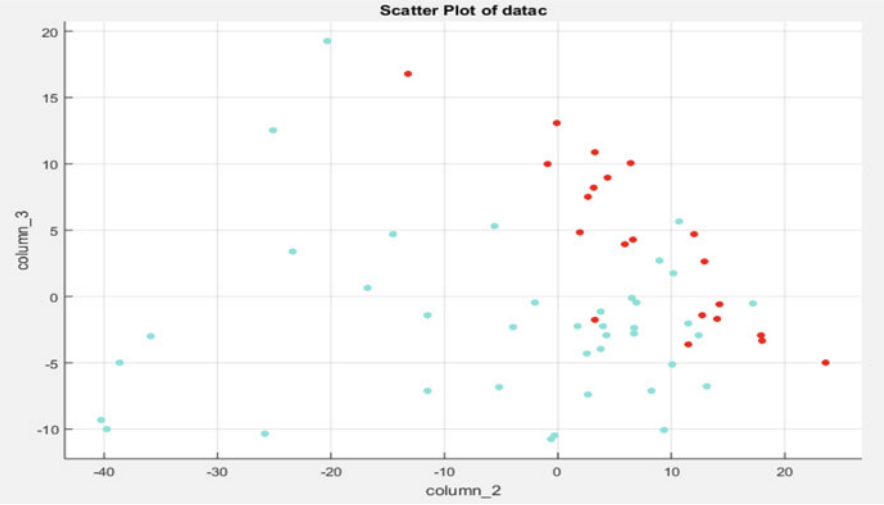


Fig. 10.5 Classification scattered plot

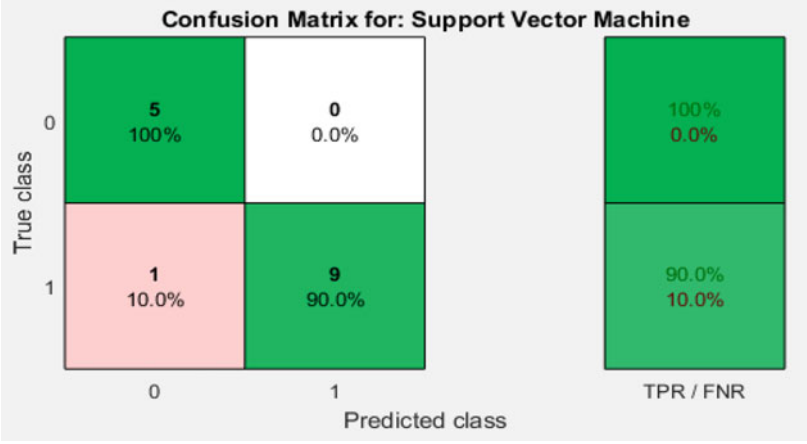


Fig. 10.6 Confusion matrix for the KPCA-L-SVM

relevant to genetic information to help predict and detect the epidemic. Prediction correctness of models is essential for hospitals and immune systems of infected persons. KPCA-L-SVM has proven to be an efficient approach for predicting the COVID-19 outbreak; the comparative chart for the experiment is depicted in Fig. 10.8.

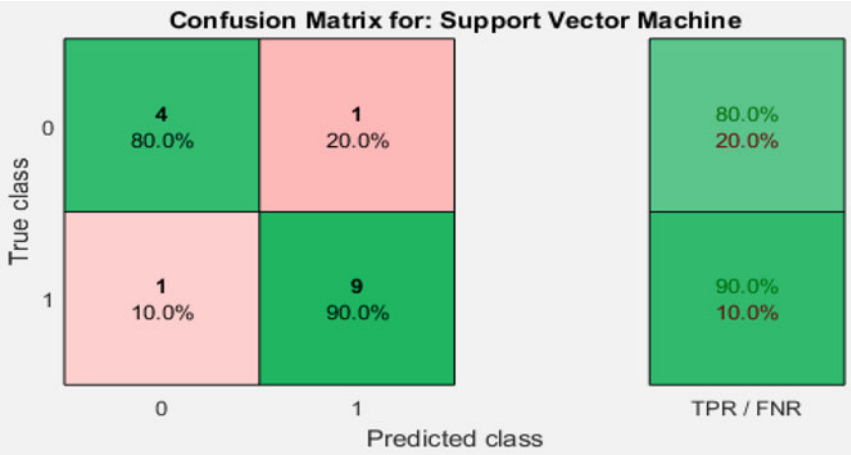


Fig. 10.7 Confusion matrix for the KPCA-SVM-RBF

Table 10.2 Experimental results

Performance metrics	KPCA-L-SVM	KPCA-RBF-SVM
Accuracy (%)	93.3	86.67
Sensitivity (%)	90.0	90.0
Specificity (%)	100	80.0
F-Score (%)	94.74	90
Matthews correlation coefficient	86.6	70
Precision	100	90
Negative predictive value	83.33	80

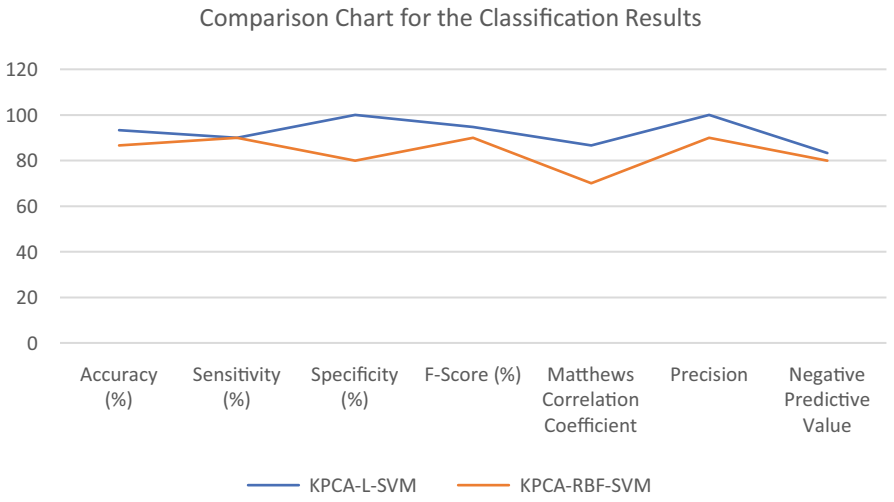


Fig. 10.8 Result comparison chart

10.5 Conclusion

Machine learning is a popular approach for predictive modelling, yet it does not significantly combat COVID-19. Machine learning can help define treatment effects, pharmaceutical views, and diagnostic prediction and improve several facets such as the economy the recent worldwide pandemic can bring. Improving proposed approaches with its risk factors should concern and present efficient algorithms for rapid prediction, and improvement should be considered. In this study, KPCA-SVM was used for predicting the novel COVID-19, and an accuracy of about 93% and 87% was attained. Frameworks have assisted healthcare to evaluate the strength of suggestions over the years. Several algorithms for assessing and understanding machine learning have been proposed, and significant public categories of learners can help predict and analyze COVID-19. In the future, several Machine learning algorithms can be used to enhance the performance and compared with the state-of-art. Using other approaches such as probabilistic PCA, Probabilistic Partial Least Square with classifiers such as SVM an KNN could also be employed.

References

1. H. Zhang, K.M. Saravanan, Y. Yang, M.T. Hossain, J. Li, X. Ren, Y. Wei, Deep learning based drug screening for novel coronavirus 2019-nCov. Preprints **2020**, 2020020061 (2020). https://doi.org/10.20944/preprints202002.0061.v1
2. R.O. Ogundokun, A.F. Lukman, G.B. Kibria, J.B. Awotunde, B.B. Aladeitan, Predictive modelling of COVID-19 confirmed cases in Nigeria. Infect. Dis. Model. **5**, 543–548 (2020)
3. L. Garg, E. Chukwu, N. Nasser, C. Chakraborty, G. Garg, Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model. IEEE Access **8**, 159402–159414 (2020)
4. E.A. Adeniyi, J.B. Awotunde, R.O. Ogundokun, P.O. Kolawole, M.K. Abiodun, A.A. Adeniyi, Mobile health application and covid-19: Opportunities and challenges. J. Crit. Rev. **7**(15), 3481–3488 (2020)
5. S. Dutta, S.L. Bandyopadhyay, Machine learning approach for confirmation of COVID-19 cases: Positive, negative, death and release. medRxiv Preprint (2020). https://doi.org/10.1101/2020.03.25.20043505
6. R.O. Ogundokun, J.B. Awotunde, Machine learning prediction for COVID 19 pandemic in India. medRxiv (2020)
7. A.F. Lukman, R.I. Rauf, O. Abiodun, O. Oludoun, K. Ayinde, R.O. Ogundokun, COVID-19 prevalence estimation: Four most affected African countries. Infect. Dis. Model. **5**, 827–838 (2020)
8. W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, D.S.C. Hui, Clinical characteristics of 2019 novel coronavirus infection in China. MedRxiv (2020)
9. A.F.M. Batista, J.L. Miraglia, T.H.R. Donato, C.A.D.P. Filho, COVID-19 diagnosis prediction in emergency care patients: A machine learning approach. medRxiv. https://doi.org/10.1101/2020.04.04.20052092
10. A.A. Onoja, A Proposed Machine Learning Approach for Monitoring Individual's Health Status on Corona Virus, researchgate.net/publication/339827100

11. G.S. Randhawa, M.P.M. Soltysiak, H.E. Roz, C.P.E.D. Souza, K.A. Hill, L. Kari, Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS One **15**(4) (2020). https://doi.org/10.1371/journal.pone.0232391
12. L. Yan, H.T. Zhang, Y. Xiao, M. Wang, C. Sun, J. Liang, S. Li, M. Zhang, Y. Guo, Y. Xiao, Prediction of survival for severe Covid-19 patients with three clinical features: Development of a machine learning-based prognostic model with clinical data in Wuhan. medRxiv (2020). https://doi.org/10.1101/2020.02.27.20028027
13. H.C. Metsky, C.A. Freije, T.-S.F. Kosoko-Thoroddsen, P.C. Sabeti, C. Myhrvold, CRISPR-based COVID-19 surveillance using a genomically comprehensive machine learning approach. bioRxiv (2020). https://doi.org/10.1101/2020.02.26.967026
14. Y. Ge, T. Tian, S. Huang, F. Wan, J. Li, S. Li, H. Yang, L. Hong, N. Wu, E. Yuan, L. Cheng, Y. Lei, H. Shu, X. Feng, Z. Jiang, Y. Chi, X. Guo, L. Cui, L. Xiao, Z. Li, C. Yang, Z. Miao, H. Tang, L. Chen, H. Zeng, D. Zhao, F. Zhu, X. Shen, J. Zeng, A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. bioRxiv (2020). https://doi.org/10.1101/2020.03.11.986836
15. Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, N. Yao, Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. Arxiv2002.05534 (2020)
16. O. Gozes, M. Frid-Adar, H. Greenspan, P.D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv2003.05037 (2020)
17. W.H. Crown, Potential application of machine learning in health outcomes research and some statistical cautions. Value Health **18**(2), 137–140 (2015)
18. T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, G. Yang, Y. Chen, A machine learning-based framework to identify type 2 diabetes through electronic health records. Int. J. Med. Inform. **97**, 120–127 (2017)
19. P. Doupe, J. Faghmous, S. Basu, Machine learning for health services researchers. Value Health **22**(7), 808–815 (2019). https://doi.org/10.1016/j.jval.2019.02.012
20. S. Basu, J.H. Faghmous, P. Doupe, Machine learning methods for precision medicine research designed to reduce health disparities: A structured tutorial. Ethn. Dis. **30**(1), 217–228 (2020). https://doi.org/10.18865/ed.30.s1.217
21. Y. Chen, V.V. Chirikov, X.L. Marston, J. Yang, H. Qiu, J. Xie, N. Sung, C. Gu, P. Dong, X. Gao, Machine learning for precision of health economics and outcomes research. Methodol. Health Care Policy **7**(1), 1–10 (2020)
22. H. Storm, K. Baylis, T. Heckelei, Machine learning in agricultural and applied economics. Eur. Rev. Agric. Econ.. https://doi.org/10.1093/erae/jbz033
23. S. Vollmer, B.A. Mateen, G. Bohner, Machine learning and artificial intelligence research for patient benefit. BMJ Res. Methods Report. **368** (2020). https://doi.org/10.1136/bmj.l6927
24. A. Doanvo, X. Qian, D. Ramjee, H. Piontkivska, A. Desai, M. Majumder, Machine learning maps research need in Covid-19 literature. bioRxiv preprint (2020) https://doi.org/10.1101/2020.06.11.145425
25. S.K. Sonbhadra, S. Agarwal, P. Nagabhushan, Target specific mining of COVID-19 scholarly articles using one-class approach. arXiv:2004.11706v1
26. Y. Zhou, B. Sun, F. Li, W. Song, NC machine tools fault diagnosis based on kernel PCA and K-nearest neighbor using vibration signals. Shock Vibrat. **2015**, 139217 (2015)
27. J.B. Awotunde, R.O. Ogundokun, F.E. Ayo, O.E. Matiluko, Speech segregation in background noise based on deep learning. IEEE Access **8**, 169568–169575 (2020)
28. S. Anto, S. Chandramathi, Supervised machine learning approaches for medical dataset classification. A review. IJCST **2**(4), 234–240 (2011)
29. O.A. Adegboye, A.I. Adekunle, E. Gayawan, Early transmission dynamics of novel coronavirus (COVID-19) in Nigeria. Int. J. Environ. Res. Public Health **17**(9) (2020). https://doi.org/10.3390/ijerph17093054

30. C. Ohia, A.S. Bakarey, T. Ahmad, COVID-19 and Nigeria: Putting the realities in context. Int. J. Infect. Dis. **95**, 279–281 (2020)
31. P.K. Ozili, COVID-19 pandemic and economic crisis: The Nigerian experience and structural causes (2020). https://doi.org/10.2139/ssrn.3567419
32. M.Z. Alom, M.M.S. Rahman, M.S. Nasrin, T.M. Taha, V.K. Asari, COVID_MTNet: Covid-19 detection with multi-task deep learning approaches. Electr. Eng. Syst. Sci.. arXiv:2004.03747
33. G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, R. Gloaguen, COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. https://doi.org/10.20944/preprints202005.0031.v1
34. B.R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. bioRxiv preprint (2020). https://doi.org/10.1101/2020.01.31.929547
35. S.F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A.R. V-Koczy, U. Reuter, T. Rabczuk, P.M. Atkinsion, COVID-19 outbreak prediction with machine learning, Health Economics and Outcomes Research Artificial Intelligence and Machine Learning (2020) https://doi.org/10.21203/rs.3.rs-27130/v1
36. X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, K. Xu, L. Ruan, W. Wu, Deep learning system to screen coronavirus disease 2019 Pneumonia. Appl. Intelligence **22**, 1–7 (2020)
37. A.A. Adeyinka, M.O. Adebiyi, N.O. Akande, R.O. Ogundokun, A.A. Kayode, T.O. Oladele, A deep convolutional encoder-decoder architecture for retinal blood vessels segmentation, in *International Conference on Computational Science and Its Applications*, (Springer, Cham, 2019), pp. 180–189
38. S. Raveendran, P. N. Indi, S. Agrahari, S. Menon, D. A. Sathia Seelan, Machine Learning Based Prognostic Model and Mobile Application Software Platform for Predicting Infection Susceptibility of COVID-19 Using Health Care Data (2020)
39. Coronavirus-Dataset, Version 1, https://www.kaggle.com/kimjihoo/coronavirusdataset-old. Accessed 17 Mar 2020
40. E.H. Aboul, S. Aya, D. Ashraf, Artificial intelligence approach to predict the COVID-19 patient's recovery
41. https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE156063
42. O. Evans, Socio-economic impacts of novel coronavirus: The policy solutions. BizEcons Q. **7**, 3–12 (2020)
43. M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A.A. Jabir, C. Iosifidis, M. Agha, R. Aghaf, The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int. J. Surg. **78**, 185–193 (2020)
44. https://github.com/owid/covid-19-data/tree/master/public/data
45. R.O. Abolade, S.O. Famakinde, S.I. Popoola, O.F. Oseni, A.A. Atayero, S. Misra, Support vector machine for path loss predictions in urban environment, in *International Conference on Computational Science and Its Applications*, (Springer, Cham, 2020), pp. 995–1006
46. S.I. Popoola, S. Misra, A.A. Atayero, Outdoor path loss predictions based on extreme learning machine. Wirel. Pers. Commun. **99**(1), 441–460 (2018)
47. R.K. Behera, S.K. Rath, S. Misra, M. Leon, A. Adewumi, Machine learning approach for reliability assessment of open source software, in *International Conference on Computational Science and Its Applications*, (Springer, Cham, 2019), pp. 472–482
48. G. Blessing, A. Azeta, S. Misra, F. Chigozie, R. Ahuja, A machine learning prediction of automatic text based assessment for open and distance learning: A review, in *International Conference on Innovations in Bio-Inspired Computing and Applications*, (Springer, Cham, 2019), pp. 369–380