



# Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa

Zurki Ibrahim<sup>1</sup> · Pinar Tulay<sup>1</sup> · Jazuli Abdullahi<sup>2</sup>

Received: 9 May 2022 / Accepted: 30 July 2022 / Published online: 11 August 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Coronavirus disease 2019 (COVID-19) has produced a global pandemic, which has devastating effects on health, economy and social interactions. Despite the less contraction and spread of COVID-19 in Africa compared to some other continents in the world, Africa remains amongst the most vulnerable regions due to less technology and unequipped or poor health system. Recent happenings showed that COVID-19 may stay for years owing to the discoveries of new variants (such as Omicron) and new wave of infections in several countries. Therefore, accurate prediction of new cases is vital to make informed decisions and in evaluating the measures that should be implemented. Studies on COVID-19 prediction are limited in Africa despite the risks and dangers that the virus possessed. Hence, this study was performed to predict daily COVID-19 cases in 10 African countries spread across the north, south, east, west and central Africa considering countries with few and large number of daily COVID-19 cases. Machine learning (ML) models due to their nonlinearity and accurate prediction capabilities were employed for this purpose, including artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and conventional multiple linear regression (MLR) models. As any other natural process, the COVID-19 pandemic may contain both linear and nonlinear aspects. In such circumstances, neither nonlinear (ML) nor linear (MLR) models could be sufficient; hence, combining both ML and MLR models may produce better accuracy. Consequently, to improve the prediction efficiency of the ML models, novel ensemble approaches including ANN-E and SVM-E were employed. The advantage of using ensemble approaches is that they provide collective benefits of all the standalone models, thereby reducing their weaknesses and enhancing their prediction capabilities. The obtained results showed that ANFIS led to better prediction performance with  $MAD=0.0106$ ,  $MSE=0.0003$ ,  $RMSE=0.0185$  and  $R^2=0.9059$  in the validation step. The results of the proposed ensemble approaches demonstrated very high improvements in predicting the COVID-19 pandemic in Africa with  $MAD=0.0073$ ,  $MSE=0.0002$ ,  $RMSE=0.0155$  and  $R^2=0.9616$ . The ANN-E improved the standalone models performance in the validation step up to 10%, 14%, 42%, 6%, 83%, 11%, 7%, 5%, 7% and 31% for Morocco, Sudan, Namibia, South Africa, Uganda, Rwanda, Nigeria, Senegal, Gabon and Cameroon, respectively. This study results offer a solid foundation in the application of ensemble approaches for predicting COVID-19 pandemic across all regions and countries in the world.

**Keywords** COVID-19 · Africa · Modelling · Machine learning · Pandemic · Ensemble approaches

## Introduction

The outbreak of the novel Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV-2) has resulted in the worldwide pandemic, Coronavirus Disease 2019 (COVID-19 or 2019-nCoV) (Muhammad et al. 2020). The novel virus was first appeared in the City of Wuhan, Republic of China at the end of December 2019, with the fact that the highly transmissible virus was initially discovered in bats and transmitted through intermediate hosts like dog or raccoon as well as palm civets (Morens et al. 2020; Narin et al. 2021;

Responsible editor: Marcus Schulz

✉ Jazuli Abdullahi  
jazuli.abdullahi@bazeuniversity.edu.ng

<sup>1</sup> Department of Medical Genetics, Near East University, Mersin 10, Lefkosa, Turkey

<sup>2</sup> Department of Civil Engineering, Faculty of Engineering, Baze University, Abuja, Nigeria

Raphael and Stanley 2020). Middle East Respiratory Syndrome Coronavirus, SARS-CoV and newest 2019 Coronavirus (2019-nCoV) are the Coronaviruses that cause a range of diseases in birds and mammals from enteritis in cattle and pigs, cattle and chickens (Mahase 2020; Islam et al. 2020). After the novel virus was first identified, the Treatment and Diagnosis Protocol of the new Coronavirus Pneumonia by the Chinese noted that COVID-19 could be identified with no positive outcome of SARS-CoV-2 acid tests by means of the following strategies: (a) an affirmative chest Computerized Tomography scan; (b) major medical symptoms consist of pyrexia (cough), shortness of breath, cough, as well as other signs of infection in the lower respiratory tract; and (c) laboratory outcome presenting (optional) leucopenia and lymphopenia (CDTP 2020).

Major symptoms and manifestation of COVID-19 are cough (76%), fever (98%) as well as diarrhoea or watery stool (3%), which are repeatedly harsher amongst older people with chronic diseases (Huang et al. 2020), and several patients have experienced dumpiness of breath where in countless incidences appear like the symptom and manifestation of flu illness (Gralinski and Menachery 2020); since it was discovered in late December 2019, the novel virus (2019-nCoV) is exponentially spreading worldwide (Muhammad et al. 2020).

More than two hundred and nine countries and territories have been affected by the COVID-19 pandemic around the globe (Muhammad et al. 2021). Having been declared an Emergency Health problem of international concern, the novel virus is transmitted via direct and close contact with the body's fluids of the infected person whether through coughing and sneezing (WHO 2020). Furthermore, asymptomatic incidences and deficiencies of diagnosis equipment result in belated or even overlook diagnosis, rendering visitors and patients as well as healthcare personnel to the pathogenic virus (2019- nCoV) infection, and this causes a huge risk to the economic and healthcare sectors. COVID-19 is actually not the original or first coronavirus which has endangered the world in last 20 years (Zivkovic, et al. 2021). The initial virus epidemic was the Severe Acute Respiratory Syndrome (SARS) in the year 2003, and then Middle East Respiratory Syndrome (MERS) outbreak followed in the year 2012. There were numerous other disease epidemics around the globe in the last 2 decades such as swine flu, Ebola, H1N1 flu and the most recent Zika virus as well. Advanced and novel epidemiological models with high prediction performance were developed as a result of the virus outbreak.

However, the COVID-19 pandemic has also demonstrated a bunch of variations compared with the previous viral outbreaks, thereby putting doubt about the practical capacity of the on hand models to perform accurate forecasting and predictions. The outbreak of the COVID-19 still possessed

many unidentified variables that are influencing the novel virus spread; the varying behaviour and complex nature of the population within various nations and territories, different strategies adopted by officials and governments when applying the precautionary measures to curtail the spread of the virus, affirmed a state of urgent situation to mention a few. These uncertain indices have reduced the performance of the existing models drastically (Scarpino and Petri 2019). Some of the more recent models include the assessment of the influence of social distancing, quarantine and curfew into their outbreak prediction (Zhan et al. 2019; Rypdal and Sugihara (2019), evaluating whether social distance is enough to prevent COVID-19 (Mirza et al. 2022), assessment of the variation in air pollutants between Christmas and new year amidst COVID-19 pandemic (Praveen Kumar et al. 2022) and addressing the challenges of COVID-19 pandemic on human physical and psychological health, air quality, environment and climate (Thapliyal et al. 2022). Unfortunately, the COVID-19 pandemic has demonstrated a complex behaviour as implied by Ivanov (2020) study. Therefore, it is now clear that nonclinical techniques such as machine learning, data mining, expert system and other artificial intelligence techniques must play critical roles in diagnosis and containment of the COVID-19 pandemic. Using non-therapeutic approaches has the potential to reduce the huge burden on healthcare systems whilst providing the best diagnostic and predictable methods for COVID-19 (Muhammad et al. 2021).

Machine learning (ML) is one of the most advanced concepts of artificial intelligence (AI) techniques and provides a strategic approach to developing automated, complex and objective algorithmic techniques for multimodal and dimensional biomedical or mathematical data analysis (Sajda 2006). The ML algorithms are able to read and modify its structure based on a set of observed data with adaptation done by optimizing over a cost function or an objective (Jebara 2012). ML models including artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) have already shown prediction potentials in several field of studies including solar radiation (Nourani et al. 2019a) dew point temperature (Naganna et al. 2019), pan evaporation (Abdullahi and Tahsin 2020), reference evapotranspiration (Nourani et al. 2020), statistical downscaling (Elkiran et al. 2021), performance measurement of residential buildings (Mohammed et al. 2021), soils suitability in airfield applications (Sujatha et al. 2021), permeability prediction for hydrocarbon reservoirs (Talebkeikhah et al. 2021) to mention a few. However, in terms of outbreak prediction, the ML models have been considered as computing techniques with great potentials. Notable applications of ML models for diseases outbreak prediction include oyster norovirus (Chenar and Deng 2018), dengue fever (Anno et al. 2019) and H1N1 flu (Koike and Morimoto 2018), measles

(Uyar et al. 2019), hepatitis C virus epidemic (Khodaei-mehr et al. 2018) and tuberculosis (Mohammed et al. 2018).

With respect to ML model applications for COVID-19 prediction, many studies can be found in the literature. Pinter et al. (2020) applied hybrid ML method for the prediction of COVID-19 in Hungary. Zhavoronkov et al. (2020) used deep learning approaches to design potential COVID-19 3C-like protease inhibitors. Zivkovic et al. (2021) employed ML and nature-inspired algorithms in hybrid form to improve the time series prediction of COVID-19 in China. Muhammad et al. (2021) applied ANN, SVM and other ML models for the prediction of daily COVID-19 cases for Mexico. Kocadagli et al. (2022) used hybrid ML approach for clinical prognosis evaluation of COVID-19 patients at Koc University Hospital Istanbul, Turkey. Xiong et al. (2022) compared SVM, random forest (RF) and logistic regression (LR) models for predicting COVID-19 severity. Noy et al. (2022) employed ML model for deterioration of COVID-19 inpatients. Tiwari et al. (2022) applied SVM, MLR and Naïve Bayes models for COVID-19 pandemic prediction. Lucas et al. (2022) performed spatiotemporal COVID-19 incidence forecasting at the county level in the USA using ML approach.

One of the most serious and challenging issues in the application of ML model for tackling a specific problem is determination of near-optimal or optimal values of its parameters. Unfortunately, there is no universally accepted rule, and hence, different set of parameters' values are determined for a specific problem (Zivkovic et al. 2021). However, every natural process constitutes both linear and nonlinear aspects (Nourani et al. 2019b). The literature review presented above showed that all studies with application of ML models for COVID-19 prediction focused on application of standalone ML models or their combinations (hybrid models), which are nonlinear methods thereby neglecting the negative impact of the linear process of the system. Thus, errors induced by the linear aspect of COVID-19 may lead to inaccurate and less efficient prediction results. Consequently, combining the linear (multiple linear regression (MLR)) and nonlinear (ANN, ANFIS, SVM) models in form of ensemble approaches would better capture the complexity surrounding the COVID-19, thereby improving prediction. Moreover, every model has its strength and weakness; the advantage of ensemble approaches is to fill the weakness of one model with the strength of another and vice versa. Therefore, the motivation as well as the basic research question of this study can be stated as follows: using ensemble approaches, could it be possible to further improve the performance of ML models for COVID-19 prediction?

To accomplish this goal, initially, ML (ANN, ANFIS and SVM) and conventional MLR models were applied for the daily-confirmed COVID-19 prediction across ten selected countries from the Africa sub regions; Morocco and Sudan (Northern Africa), Uganda and Rwanda (Eastern Africa), Cameroon and Gabon (Middle Africa), South Africa and

Namibia (Southern Africa) as well as Nigeria and Senegal (Western Africa). Thereafter, two ensemble approaches including ANN-E and SVM-E were developed by replacing the variables of the input layers of ANN and SVM with outputs of the standalone models to improve performance. To the best knowledge of the authors and based on the present available literature, no similar study was performed for COVID-19 modelling using the considered models and countries in Africa. However, review of the literature also suggests that there were no studies carried out with the application of ensemble approaches for COVID-19 modelling in Africa.

The remainder of the work was organized as follows: The next section (Sect. 2) describes the study area, data, materials and methods employed for the study. Section 3 presents the results obtained and their discussions. Section 4 provides conclusion and recommendations for future works.

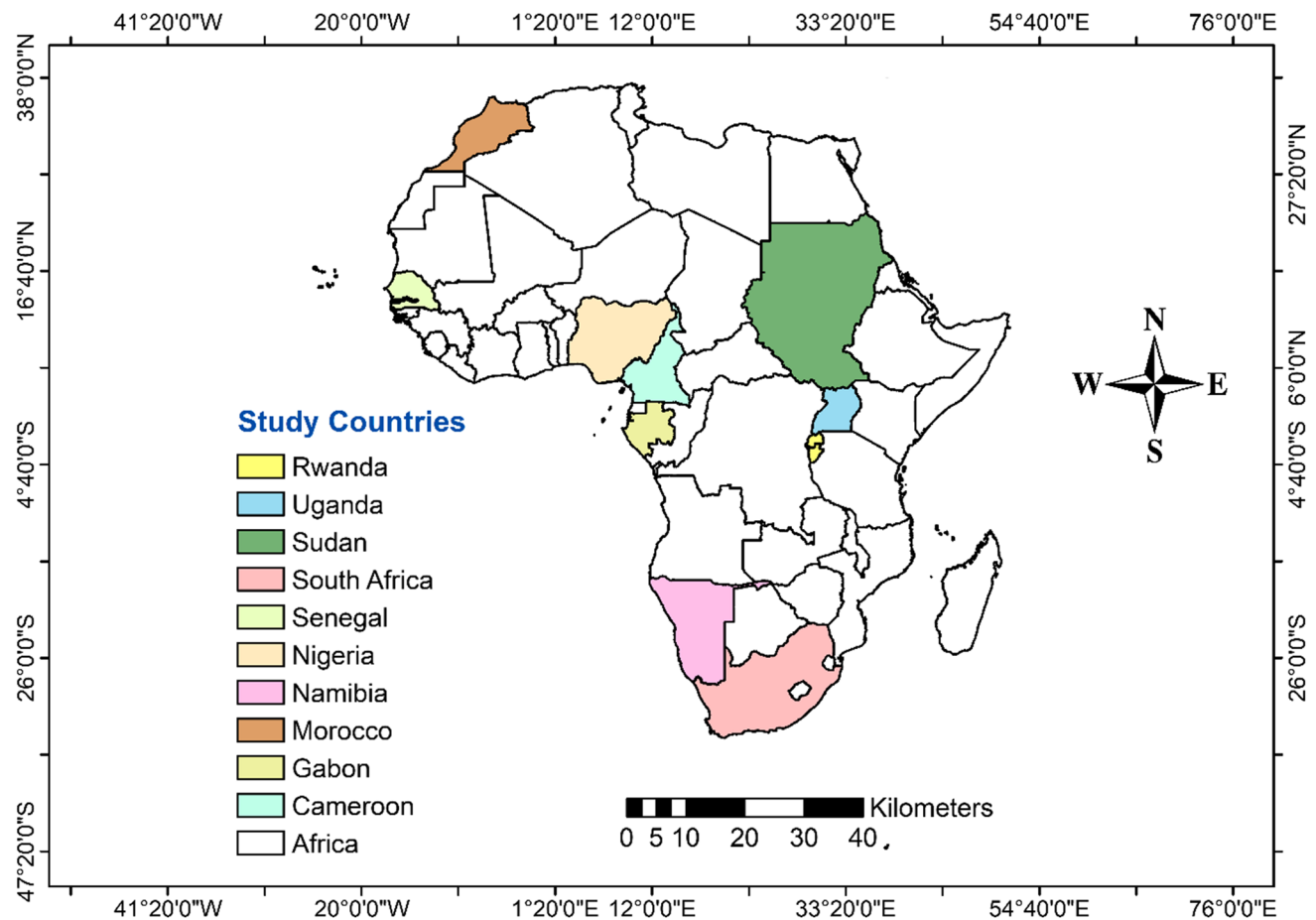
## Materials and methods

### Study area and data

This research predicted the daily-confirmed cases of COVID-19 using ML models including ANN, ANFIS, SVM and traditional MLR model, and their ensemble combinations (ANN-E and SVM-E) to improve performance. The total confirmed cases of COVID-19 in Morocco, Sudan, Uganda, Rwanda, Cameroon, Gabon, South Africa, Namibia, Nigeria and Senegal were used for the study purpose. These countries were chosen across different African regions to represent diversity. Furthermore, their figures of confirmed incidences are order of magnitudes variations, which provide enough chance to test the proposed models for the nations with both elevated and low numbers of confirmed cases. Moreover, a few of these nations have recorded the cases relatively longer period than many other countries, which is another reason for choosing them. Figure 1 shows the African map and the study countries.

The data used for the study were divided into two sections, comprising 75% and 25%. The former was used for training of the ML and MLR models, whilst the latter was employed for validation purposes. Thus, the predicted confirmed cases of the validation data were compared with those of observed ones. The sequential data of daily-confirmed COVID-19 cases were obtained from World Health Organization (WHO) database and can be extracted from <https://covid19.who.int/WHO-COVID-19-global-table-data.csv>. Table 1 shows the countries, duration of the data and data statistical description.

As COVID-19 cases in African nations started to be confirmed in March, 2020, this study considers 1<sup>st</sup> March, 2020 as the data collection period until 16<sup>th</sup> December, 2021. As



**Fig. 1** Location of the study countries in Africa

**Table 1** Statistical description of the daily-confirmed COVID-19 cases in some African countries

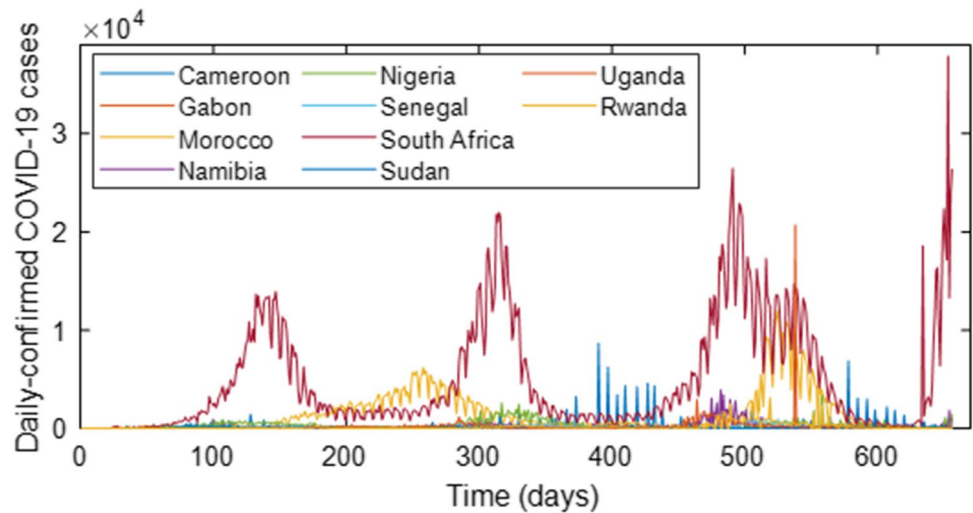
Country	Period	Minimum	Maximum	Mean	St. deviation
Morocco	1/3/2020–16/12/2021	0	12,039	1453	2095
Sudan	1/3/2020–16/12/2021	0	1215	69	101
Uganda	1/3/2020–16/12/2021	0	20,692	196	852
Rwanda	1/3/2020–16/12/2021	0	3141	154	325
Nigeria	1/3/2020–16/12/2021	0	3402	336	413
Senegal	1/3/2020–16/12/2021	0	1722	113	178
Namibia	1/3/2020–16/12/2021	0	3937	205	413
South Africa	1/3/2020–16/12/2021	0	37,875	4925	5677
Gabon	1/3/2020–16/12/2021	0	640	57	117
Cameroon	1/3/2020–16/12/2021	0	8681	164	689

seen from Table 1, all the countries have a minimum value of 0 case, which indicates that the period of the COVID-19 cases is appropriately covered by the study. It can also be seen from Table 1 that Morocco, Uganda and South Africa have the largest number of the daily-confirmed COVID-19 cases with 12,039, 20,692 and 37,875, respectively. Figure 2 shows the time series plots of the daily-confirmed COVID-19 cases across all countries considered in this study.

### Model validation

To ensure appropriate results are achieved in this study, k-fold cross validation was employed. For the 10 countries, the data samples were randomly divided into k folds (4-folds in this study) subsamples as can be seen from Fig. 3. In this way, 3-folds (k-1 or 4-1) were used for training, and the remaining fold subsample was used for validation (Table 2). The process

**Fig. 2** Time series plots for the daily-confirmed cases for all countries



**Fig. 3** The k-fold validation used in the study

continues up to k (4) times for different 4–1 training subsamples and single validation subsample. Thereafter, the final single results were obtained by taking average of the k results

from the folds. The advantage of using k-fold validation is that the entire observations are utilized for both training and validation (Sharma et al. 2018; Nourani et al. 2019b). Figure 3 shows the k-fold cross validation applied, whilst Table 2 illustrates the cumulative daily-confirmed COVID-19 cases of the study countries and the number of observations used for training and validation.

**Data normalization and performance criteria.**

To ensure all variables have equal attention and to eliminate their dimensional discrepancy, data normalization is usually applied for AI-based modelling (Abdullahi et al. 2017). For the normalization purpose in this study, the observations were scaled between 0 and 1. The normalization procedure is given by (Elkiran et al. 2021):

$$DC_n = \frac{DC_i - DC_{min}}{DC_{max} - DC_{min}} \tag{1}$$

**Table 2** Cumulative cases, validation and data partitioning

Country	Cumulative cases	No. of observation	Training sample (75%)	Validation sample (25%)	Validation type
Morocco	951,763	656	492	164	k-fold
Sudan	45,112	656	492	164	k-fold
Uganda	128,369	656	492	164	k-fold
Rwanda	100,978	656	492	164	k-fold
Nigeria	220,109	656	492	164	k-fold
Senegal	74,105	656	492	164	k-fold
Namibia	134,160	656	492	164	k-fold
South Africa	3,231,039	656	492	164	k-fold
Gabon	37,681	656	492	164	k-fold
Cameroon	107,662	656	492	164	k-fold

where  $DC_n$ ,  $DC_{\max}$ ,  $DC_{\min}$  and  $DC_i$  represent the normalized value, maximum value, minimum value and  $i$ th values of daily-confirmed COVID-19 cases, respectively.

To determine the accuracy and performance of the applied models for the modelling of COVID-19 pandemic across 10 African countries, 4 global statistical indices were used including mean absolute deviation (MAD) (Khatri et al. 2020), mean square error (MSE) (Hussain and Khan 2020), root mean square error (RMSE) (Abdullahi et al. 2019a) and determination coefficient ( $R^2$ ) (Abdullahi and Elkiran 2021) given by:

$$MAD = \frac{1}{N} \sum_{i=1}^n |p_i - a_i| \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^n (p_i - a_i)^2 \quad (3)$$

$$DC = 1 - \frac{\sum_{i=1}^N (a_i - p_i)^2}{\sum_{i=1}^N (a_i - \bar{a}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (a_i - p_i)^2}{N}} \quad (5)$$

where  $a_i$ ,  $p_i$ ,  $\bar{a}$  and  $N$  are the actual values, predicted values, mean of the actual values and number of observations, respectively.

## Research gap and study novelty

VOSviewer software was used to determine the research gap in this study. A search term “artificial intelligence applications for COVID-19 modelling” was entered into Scopus database for articles between December 2019 and February 2022. A total of 948 papers were downloaded and entered into the VOSviewer software with analysis type based on bibliographic coupling and unit of analysis based on countries. Figure 4 shows the results of the analysis carried out.

It can be seen from Fig. 4 that several COVID-19 studies were performed in several countries in the world based on AI-based applications, but the studies are very limited for African countries as only few countries may be seen including Morocco, Tunisia, Nigeria and South Africa. Therefore, there is need to have more research of COVID-19 pandemic in Africa for informed decisions and proper control measures to be applied.

Another analysis to determine the number of applications of ensemble approaches for COVID-19 modelling was also performed. For such purpose, co-occurrence is the type of analysis used, whilst author keywords were used as the

unit of analysis. Figure 5 presents the results of the second analysis.

As depicted by Fig. 5, so far, several keywords were used by authors for COVID-19 researches, but it can be seen that there is no single mention of ensemble approaches. This indicates that there are no/limited studies for COVID-19 with applications of ensemble approaches in the present literature and, thus, implies the novelty of this study.

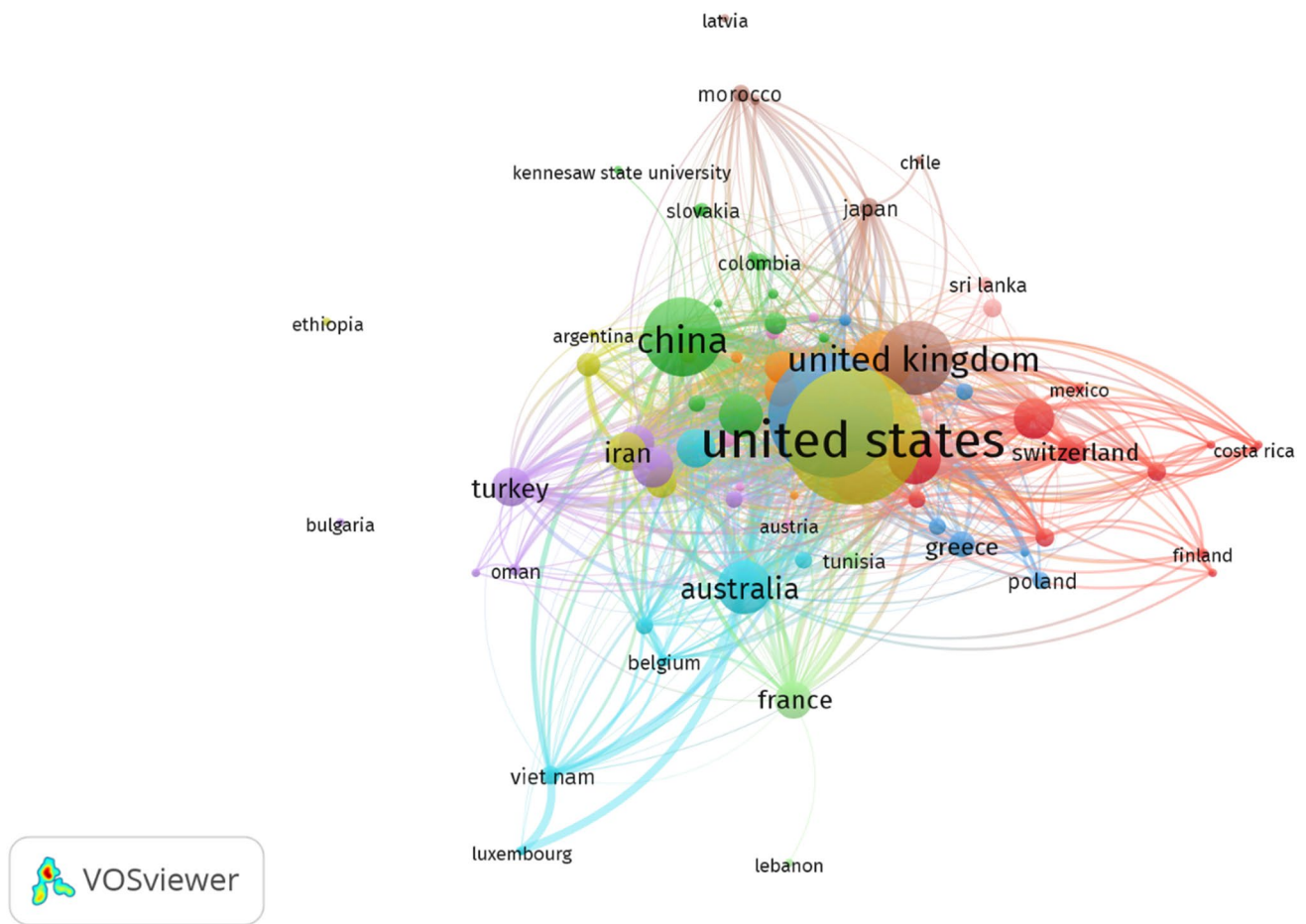
## Artificial neural network (ANN)

ANN is a well-established artificial intelligence model inspired by the structure of biological neurons of human (Nourani et al. 2019b). It has successfully been applied to many problems in various fields. In essence, it is an influential tool for exploring an association between input and output data. For accomplishing this task, it is necessary to be trained by utilizing a set of records consisting of input and the matching output data. The procedures for the training data are usually done by the soft architecture of ANN comprising 3 layers: (a) input layer, (b) hidden layer and (c) output layer (Ekhmaj 2012). The first and the third enclosed neurons were connected with both input and output vectors. Meanwhile, neurons enclosed in the hidden layer were linked with neurons of both hidden and output layers; they also basically lead to the turning of the input data into the matching output data. Moreover, the weighted summation of the input data was transferred via a transfer function. Usually, neurons enclosed in each layer of artificial neural network are normally allowed to have a link to the subsequent and previous layers, whilst inter-layered links are forbidden. The flow of the data via the network proceeds unless an association with needed precision is achieved; lastly, the better ANN is trained, the more desired outcomes may be achieved (Nourani and Fard 2016).

In this research, a feed forward back propagation network together with Levenberg Marquardt optimization algorithm was used to train the artificial neural network using MATLAB, and common features of artificial neural network were set in line with those utilized in the previous studies.

## Adaptive neuro-fuzzy inference system (ANFIS)

Neuro-fuzzy simulation points to the methods of using various learning algorithm to fuzzy modelling in the fuzzy inference system or neural network literature (Akrami et al. 2014). A unique way in the development of neuro-fuzzy is adaptive neuro-fuzzy inference system which was first articulated by Jang (1993) and employs the learning algorithm of neural network. As a general approximator, adaptive neuro-fuzzy inference system has the capability of compressing set of efficiency to any level for whichever



**Fig. 4** COVID-19 studies carried out for world countries based on artificial intelligence

real continuous function. Functionally, adaptive neuro-fuzzy inference system is correspondent to FIS according to a study by Jang et al. (1997). Precisely, the interest of the adaptive neuro-fuzzy inference system is equivalent functionally here to the primary order Sugeno fuzzy model. The adaptive neuro-fuzzy inference system common structure is presented in the following equation, and it considered that the inputs for the ANFIS are  $x$  and  $y$  as well as  $f$  as output (Aqil et al. 2007). The ideal rules sets for Sugeno first order which are 2 fuzzy-if then rules are written as:

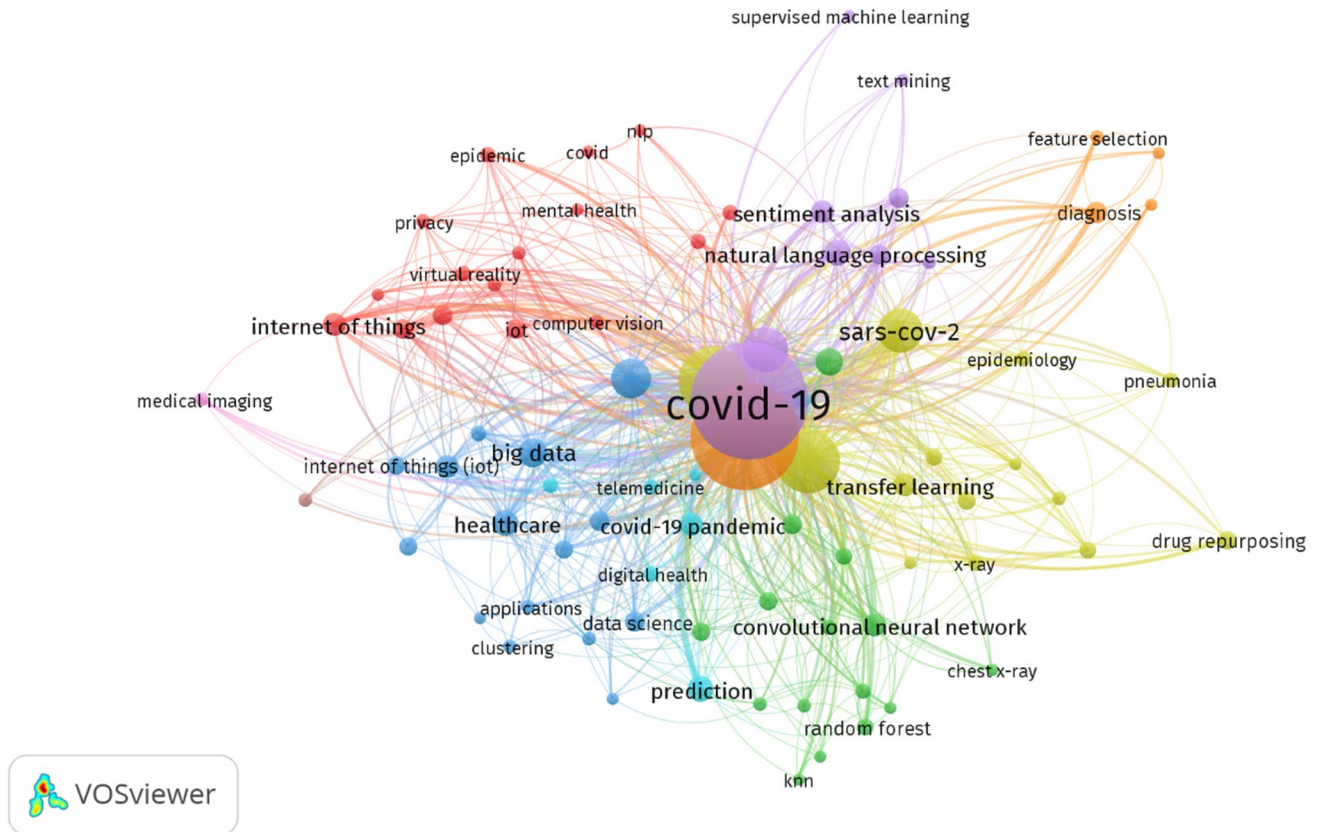
$$\text{Rule (1) : If } \mu(x) \text{ is } A1 \text{ and } \mu(y) \text{ is } B1; \text{ the } f1 = p1x + q1y + r1 \quad (6)$$

$$\text{Rule (2) : If } \mu(x) \text{ is } A2 \text{ and } \mu(y) \text{ is } B2; \text{ the } f2 = p2x + q2y + r2 \quad (7)$$

where,  $A1$  and  $A2$  stand for  $x$  inputs MFs,  $B1$  and  $B2$  are for the  $y$  inputs MFs, correspondingly. Moreover, the parameters for the output function are  $p1$ ,  $q1$ ,  $r1$ , and  $p2$ ,  $q2$ ,  $r2$ .

### Support vector machine (SVM)

Cortes and Vapnik (1995) proposed the concept of support vector machine. It applies mapping of nonlinear to an elevated dimensional hole or space based on the designed minimization rule which consists of regression model complexity and kernel function as well as regularization (Vapnik 1998). Several findings reported the accomplishment of support vector machine in forecasting stuff. Regarding the parameter selection, SVM lacks any theoretical direction. It utilizes quadratic-based programming to work out the support vector which results in its complexity (Li et al. 2019). In respect to quadratic-based programming, it requires huge memory, and it has elevated algorithmic complexity (Li et al. 2019). Furthermore, the suitable assortment of the kernel is awfully significant for the better model performance. However, it is often difficult to choose the appropriate kernel function. Detail information regarding SVM can be found from Vapnik (1998).



**Fig. 5** COVID-19 studies based on author keywords

## Multi-linear regression

Generally, in regard to multi-linear regression (MLR), the  $n$  regressor variables and the dependent variable  $y$  may be associated by (Elkiran et al. 2021):

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_ix_i + \xi \quad (8)$$

where  $b_0$  is the regression constant,  $x_i$  is the value of the  $i$ th forecaster and  $b_i$  stands as the coefficient of the  $i$ th predictor; likewise,  $\xi$  is the error term as well.

## Ensemble modelling

For a particular set of information or data, it is observable that the performance of one bright technology could outshine another; at the same time, if dissimilar sets of information are applied, the outcomes may totally be contrary (Nourani et al. 2019b). In order not to lose simplification and also to benefit from the significances of all procedures, an ensemble model is formed which makes use of the individual output of every technique with definite precedence level assigned to every one of them with the aid of a mediator to proffer the output (Kiran and Ravi 2008).

Weighted average ensemble, stack regression, simple average ensemble as well as nonlinear ensemble such as NN-based are some of the ensemble techniques applied. Two ensemble strategies have been reported by Kiran and Ravi (2008), which are: (i) Nonlinear ensemble procedure; for example, an artificial neural network is usually trained to achieve an ensemble output; (ii) Linear ensemble procedure; which comprises linear ensemble by means of weighted averaging, linear ensemble through simple averaging and linear ensemble by means of weighted median.

In this research, the ensemble modelling was done through 2 nonlinear ensemble procedures including ANN ensemble (ANN-E) and SVM ensemble (SVM-E). Although, other algorithms such as ANFIS could be used for the nonlinear ensemble modelling, the choice of the mentioned models is based on the following: (i) ANN-E is the most widely nonlinear ensemble model applied; it is simple to use and leads to efficient performance (Nourani et al. 2019b), whilst (ii) SVM-E has never been tested before in any field of study. The general procedure of the ensemble modelling is given in Fig. 6.



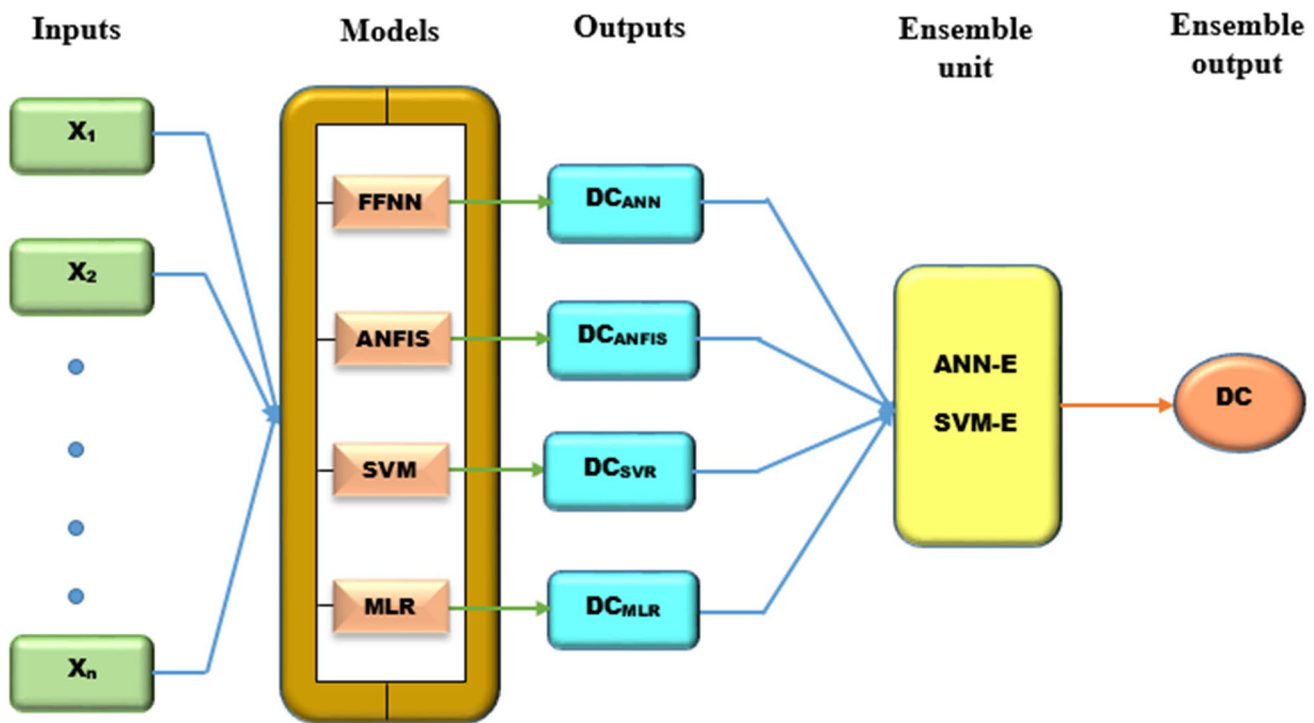


Fig. 6 The general ensemble procedure applied

**Proposed methodology**

In this study, the feasibility of employing ensemble concept to further improve COVID-19 prediction accuracy was investigated. Firstly, ML models including ANN, ANFIS and SVM and conventional MLR model were applied for daily-confirmed COVID-19 cases prediction across 10 African countries including Morocco, Sudan, Namibia, South Africa, Uganda, Rwanda, Nigeria, Senegal, Gabon and Cameroon. Thereafter, two ensemble approaches were applied to improve the COVID-19 prediction.

The main advantages of using ensemble approaches are: (i) Understanding whether the underlying process for a particular problem is induced by linear or nonlinear aspect is difficult task to accomplish in practical situations or the most preferable method to be chosen between others. Therefore, for a unique issue, choosing a befitting method has become a difficult task before predictors. Thus, problem of selecting the most appropriate models could be handled by ensemble approaches (Nourani et al. 2019a). (ii) The real-world process may involve both linear and nonlinear characteristics. Hence, for such a circumstance, the nonlinear ML models (ANN, ANFIS and SVM) or the linear MLR will neither be sufficient for the time series prediction since MLR could not cope with the nonlinear relationship and ML models could magnify errors of a linear pattern. Consequently, by combining the ML and MLR models, the system’s complex

manner could be captured more accurately (Nourani et al. 2020). (iii) There is no unique method that can perfectly detect the distinct patterns of time series due to the complex nature of the real-world problem (Sharghi et al. 2018). The applied ensemble models are:

**(i) ANN-E**

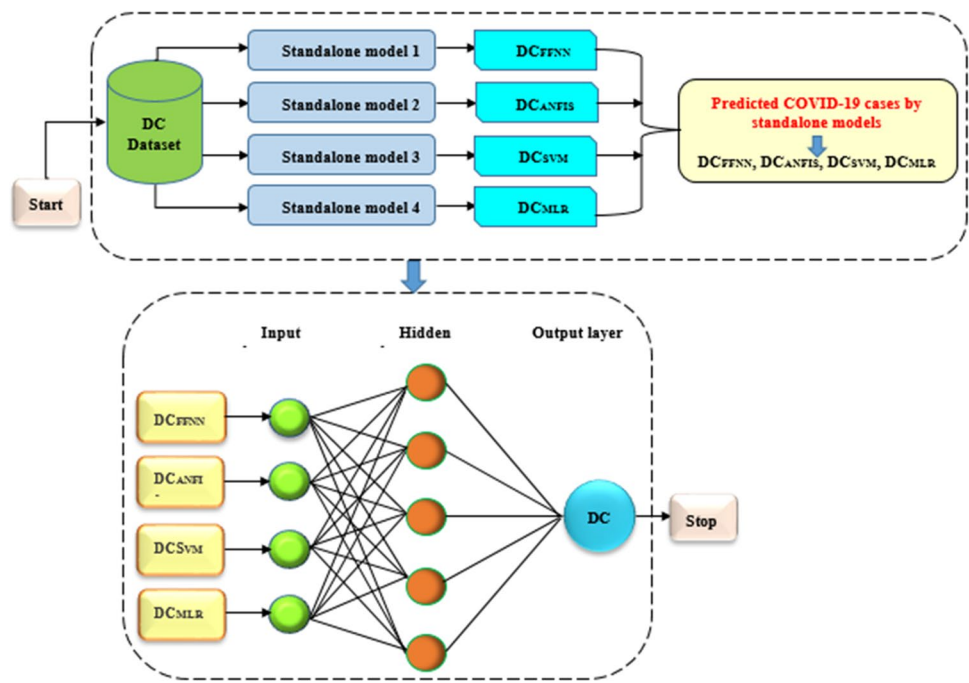
For ANN-E, the daily-confirmed COVID-19 cases were simulated as a function of the outputs of the single models based on ANN model, given as

$$DC_{ANN-E} = f(DC_{ANN}, DC_{ANFIS}, DC_{SVM}, DC_{MLR}) \tag{9}$$

where  $DC_{ANN-E}$  represents the daily-confirmed values by ANN-E, and  $DC_{ANN}, DC_{ANFIS}, DC_{ANFIS}, DC_{SVM}$  and  $DC_{MLR}$  are the outputs of the daily-confirmed cases of the individual countries produced by ANN, ANFIS, SVM and MLR, respectively. Figure 7 shows the proposed nonlinear ensemble approach based on ANN model (ANN-E).

As seen in Fig. 7, the COVID-19 data obtained after passing through data preprocessing, ANN, ANFIS, SVM and MLR models were then applied as standalone models. The ANN-E prediction of the COVID-19 was then performed using ANN as the ensemble kernel. In this way, the outputs of the standalone models were used to replace the input

**Fig. 7** The proposed ANN-E approach applied



layer neurons, which comprised input, hidden and output layers structure. With its ability to checkmate the minimum required error, feed forward neural network (FFNN) with back propagation algorithm was employed. Levenberg Marquardt (LM) was used as the training algorithm, whilst the adaptation learning function utilized was LEARNGDM and mean square error (MSE) was used as the performance function. Trial and error method was applied to determine the optimum number of hidden layer neurons. In order to have sufficient iterations for improve performance, the epoch number was set by trial and error to fall between 100 and 200.

## (ii) SVM-E

The SVM-based ensemble modelling was performed using the SVM kernel to combine the outputs of the single models, given as

$$DC_{SVM-E} = f(DC_{ANN}, DC_{ANFIS}, DC_{SVM}, DC_{MLR}) \quad (10)$$

where  $DC_{SVM-E}$  implies daily-confirmed COVID-19 values by SVM-E for the each country. Figure 8 shows the proposed nonlinear ensemble approach based on SVM model (SVM-E).

SVM-based ensemble prediction (SVM-E) of the daily-confirmed COVID-19 cases was performed using the outputs of the standalone models (ANN, ANFIS, SVM and MLR). The outputs were used to replace the input layer variables as shown in Fig. 8. For a complicated nonlinear process

(such as COVID-19), the Gaussian kernel function is more suitable (Ghorbani et al. 2016). Therefore, Gaussian kernel function was chosen for the SVM-based ensemble prediction to take care of the uncertain and complex nature of COVID-19 pandemic.

The general methodology proposed by this study is given in Fig. 9.

## Results and discussion

In this study, the proposed methodology contains: (i) Prediction of daily cases of COVID-19 in 10 African countries using AI-based and linear models including ANN, ANFIS, SVM and MLR. (ii) To ensure higher predictions are achieved, nonlinear ensemble models including ANN-E and SVM-E were developed. Therefore, the results in this section are presented accordingly.

### Results of the standalone models

Although may not be practically proven, many hydro-climatological variables (such as temperature, precipitation wind speed, solar radiation etc.) may have an impact on COVID-19 spread. However, the cumulative cases, number of deaths and cumulative number of deaths may be sensitive to the daily-confirmed cases of COVID-19. These variables have been taken into account in this study. But, previous studies including Ardabili et al. (2020) and Niazkar and Niazkar (2020) have shown that a successful prediction of daily cases of COVID-19 can

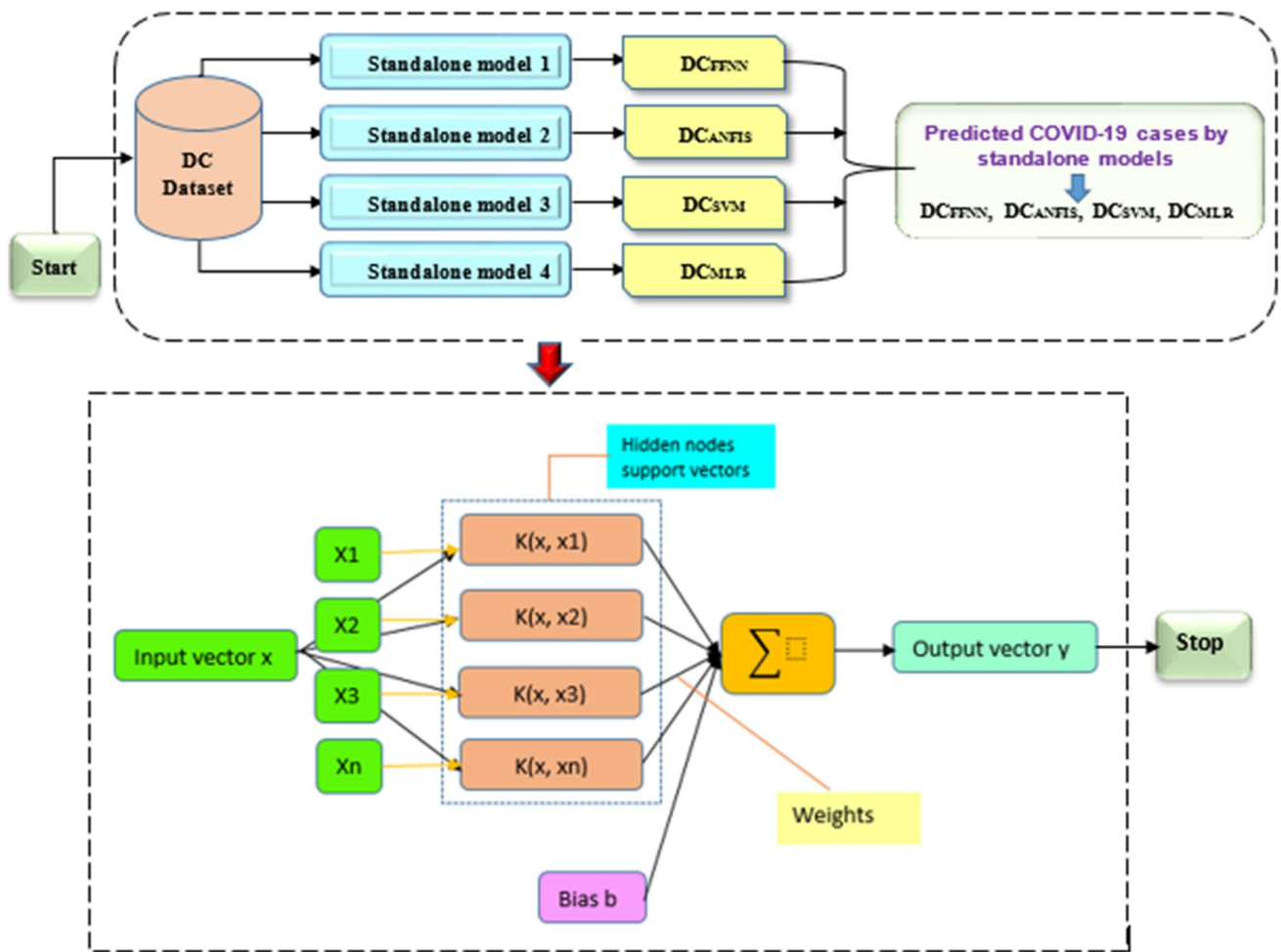


Fig. 8 The proposed SVM-E approach applied

be accomplished using the COVID-19 outbreak data at previous time step ( $t_n$ ). Therefore, several time lags were considered in order to meet the Markov strength of the previous cases with respect to the current case. It was found that up to seven-time lag ( $t - 7$ ), strong relationship exists between current and previous cases. In other words, previous cases up to 7-day period are sensitive to the current case of COVID-19. Hence, for the prediction of COVID-19 outbreak in Africa, the following were used as inputs:

$$DC^i = f(DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}, DC^i_{(t-7)}) \tag{11}$$

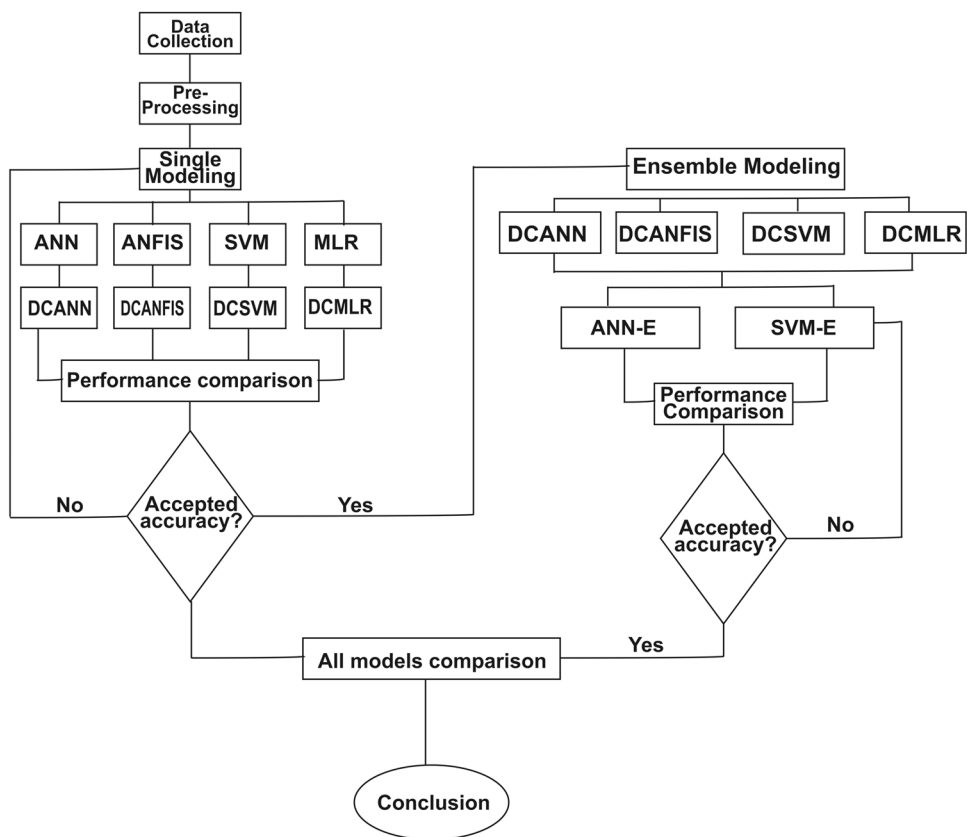
where  $i$  represents the African country under consideration, DC implies daily cases of the virus,  $DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}, DC^i_{(t-7)}$  are the  $i$ th country outbreak data at previous time steps  $t - 1, t - 2, t - 3, t - 4, t - 5, t - 6$  and  $t - 7$  (or 1, 2, 3, 4, 5, 6 and 7 days ago).

One of the most significant aspects of any ML-based prediction is the selection of the most dominant inputs; failure to do that may lead to errors and inaccuracy in results (Abdullahi and Elkiran 2021; Elkiran et al. 2021). However, with difference in the rate of infections per day, population density and mitigating measures put in place by the African countries, variation in performance based on the 7-input variables is observed. Therefore, by trial and error, the best input variables representing the most sensitive inputs to COVID-19 output were selected for every country as shown

in Table 3.

For ANN models, three-layered FFNN method was adopted in the study that consists of input, hidden and output layers. The ANN models were trained using LM algorithm, whilst the adaptation learning

**Fig. 9** The applied methodological approach of the study



**Table 3** Input variables selected for the study countries

	Country	Inputs
North Africa	Morocco	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}$
	Sudan	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}$
South Africa	Namibia	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}$
	South Africa	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}$
East Africa	Uganda	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}$
	Rwanda	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}$
West Africa	Nigeria	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}$
	Senegal	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}$
Central Africa	Gabon	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}$
	Cameroon	$DC^i_{(t-1)}, DC^i_{(t-2)}, DC^i_{(t-3)}, DC^i_{(t-4)}, DC^i_{(t-5)}, DC^i_{(t-6)}$

function utilized was LEARNGDM and mean square error (MSE) was used as the performance function. To ensure accuracy in the ANN predictions, several number of neurons in the hidden layer were tried, and through trial and error, the maximum performance was achieved. According to a suggestion by Fletcher and Goss (1993), the most appropriate number of hidden layer neurons falls between  $2n^{1/2} + m$  and  $2^{n+1}$ , where  $m$  signifies the number of output nodes and  $n$  represents the number of input nodes. Apart from the number of hidden layer neurons, Emamgholizadeh et al. (2014)

emphasized that the transfer function between nodes adversely affects prediction precision of ANN models. This study examined several transfer functions in order to achieve the best results including hyperbolic tangent ( $f(x) = \tanh(x)$ ), sigmoid ( $f(x) = \frac{1}{1+\exp(-x)}$ ), hyperbolic secant ( $f(x) = \text{sech}(x)$ ) and Gaussian ( $f(x) = e^{-x \cdot x}$ ). The learning rate used was 0.01, and the epoch number varied between 100 and 300.

For SVM technique, the kernel function selected was Gaussian. The advantage of using the Gaussian kernel function for SVM model is that it makes the modelling

and analysis easier in complicated nonlinear problems (Abunama et al. 2019). Cortes and Vapnik (1995) give full details of SVM and its equations.

The MLR models find the linear relationship between input and output variables and are also utilized to compare their performance with the ML techniques. Tables 4, 5, 6, 7 and 8 give the results of all the developed models for the daily COVID-19 cases across the African continent based on 5 African regions.

It is worthy to mention that four global statistical indices were used in this study to determine the performance of the applied models for the prediction of the daily cases of COVID-19 in African. The error measures including MAD, MSE and RMSE have no units since the data were

normalized, whereas the goodness of fit measure of  $R^2$  is dimensionless.

As can be seen from Table 4 for North African countries, different models lead to different outcomes for both Morocco and Sudan in the training and validation steps, respectively. Considering the validation step, it can be seen that for Morocco, all the applied models have  $R^2$  value greater than 0.7, which is an indication of the models accuracy. Despite the promising results of the applied models, SVM shows better efficiency having minimum errors and stronger fitting with  $MAD=0.0185$ ,  $MSE=0.0008$ ,  $RMSE=0.0287$  and  $R^2=0.9185$ . This is followed closely by ANFIS model with  $MAD=0.0204$ ,  $MSE=0.0011$ ,  $RMSE=0.0326$  and  $R^2=0.9154$ . For Sudan, it can be seen that the models

**Table 4** Results of the applied models for North Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
Morocco	ANN	0.0699	0.0122	0.1106	0.8486	0.0336	0.0019	0.0436	0.8302
	ANFIS	0.0321	0.0035	0.0591	0.9515	0.0204	0.0011	0.0326	0.9154
	SVM	0.0423	0.0059	0.0767	0.9347	0.0185	0.0008	0.0287	0.9185
	MLR	0.0604	0.0115	0.0107	0.9078	0.0208	0.0001	0.0341	0.8405
Sudan	ANN	0.0369	0.0032	0.0564	0.6000	0.0353	0.0028	0.0564	0.4854
	ANFIS	0.0306	0.0029	0.0536	0.8315	0.0213	0.0012	0.0345	0.5343
	SVM	0.0299	0.0041	0.0642	0.5929	0.0242	0.0029	0.0537	0.3330
	MLR	0.0374	0.0048	0.0691	0.3248	0.0443	0.0055	0.0740	0.1135

**Table 5** Results of the applied models for East Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
Uganda	ANN	0.0078	0.0002	0.0126	0.3552	0.0141	0.0060	0.0777	-0.0025
	ANFIS	0.0064	0.0001	0.0113	0.4807	0.0181	0.0056	0.0750	0.0650
	SVM	0.0036	0.0001	0.0090	0.6682	0.0080	0.0060	0.0774	0.0048
	MLR	0.0081	0.0002	0.0134	0.2712	0.0144	0.0060	0.0775	0.0015
Rwanda	ANN	0.0319	0.0072	0.0851	0.9232	0.0112	0.0003	0.0182	0.7012
	ANFIS	0.0233	0.0023	0.0478	0.9205	0.0106	0.0003	0.0185	0.9059
	SVM	0.0432	0.0110	0.1006	0.5824	0.0150	0.0020	0.0446	0.5366
	MLR	0.0485	0.0094	0.0970	0.6417	0.0166	0.0015	0.0392	0.6119

**Table 6** Results of the applied models for West Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
Nigeria	ANN	0.0374	0.0035	0.0592	0.7964	0.0295	0.0031	0.0560	0.7206
	ANFIS	0.0330	0.0029	0.0537	0.8958	0.0247	0.0016	0.0400	0.7699
	SVM	0.0387	0.0071	0.0844	0.7947	0.0288	0.0032	0.0562	0.4316
	MLR	0.0514	0.0088	0.0936	0.7312	0.0351	0.0041	0.0643	0.3015
Senegal	ANN	0.0284	0.0055	0.0737	0.8334	0.0195	0.0009	0.0308	0.6285
	ANFIS	0.0189	0.0017	0.0408	0.9492	0.0187	0.0008	0.0287	0.6765
	SVM	0.0335	0.0063	0.0791	0.8089	0.0185	0.0009	0.0292	0.6563
	MLR	0.0363	0.0061	0.0783	0.8128	0.0187	0.0009	0.0292	0.6647

**Table 7** Results of the applied models for South Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
Namibia	ANN	0.0505	0.0076	0.0870	0.5786	0.0394	0.0044	0.0667	0.3827
	ANFIS	0.0255	0.0024	0.0488	0.8884	0.0183	0.0012	0.0343	0.8059
	SVM	0.0406	0.0093	0.0965	0.5297	0.0255	0.0050	0.0705	0.2400
	MLR	0.0477	0.0091	0.0955	0.5467	0.0259	0.0048	0.0692	0.2556
South Africa	ANN	0.0351	0.0050	0.0706	0.8924	0.0224	0.0018	0.0427	0.8553
	ANFIS	0.0364	0.0040	0.0630	0.9355	0.0195	0.0011	0.0331	0.8846
	SVM	0.0375	0.0063	0.0796	0.9110	0.0209	0.0015	0.0388	0.8160
	MLR	0.0523	0.0096	0.0977	0.8689	0.0240	0.0022	0.0471	0.7225

**Table 8** Results of the applied models for Central Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
Gabon	ANN	0.0567	0.0106	0.1006	0.8575	0.0447	0.0079	0.0888	0.5866
	ANFIS	0.0490	0.0077	0.0878	0.9007	0.0411	0.0055	0.0741	0.6983
	SVM	0.0490	0.0120	0.1097	0.8142	0.0441	0.0103	0.1014	0.5289
	MLR	0.0835	0.0209	0.1193	0.6228	0.0700	0.0142	0.1445	0.4429
Cameroon	ANN	0.0037	0.0006	0.0249	0.9042	0.0085	0.0022	0.0465	0.6531
	ANFIS	0.0097	0.0005	0.0220	0.9225	0.0080	0.0012	0.0341	0.8200
	SVM	0.0089	0.0011	0.0339	0.8230	0.0104	0.0013	0.0354	0.7987
	MLR	0.0091	0.0011	0.0338	0.8231	0.0114	0.0012	0.0349	0.8041

with the best performance is ANFIS with  $MAD=0.0213$ ,  $MSE=0.0012$ ,  $RMSE=0.0345$  and  $R^2=0.5343$ .

Comparing the results of Table 4 in the validation step for Morocco and Sudan, it can be deduced that performances of the models are higher for Morocco. This can be attributed to the fact that Morocco has the highest number of confirmed daily COVID-19 cases with maximum value up to 12,039, whereas Sudan has the maximum daily value of 1215. The predictive models were developed to provide accurate prediction based on previous experience, and the absence of cases at a particular day and presence of cases in another day (as in the case of Sudan) make it difficult for the predictive models to perform at the highest level.

Based on Table 5 results for East Africa, the results show a weak performance by the models for Uganda, and the model with highest performance in the validation step is ANFIS with  $MAD=0.0181$ ,  $MSE=0.0056$ ,  $MSE=0.0750$  and  $R^2=0.0650$ . The poor performance of the models may be due to the nature of confirmed daily cases of COVID-19 in the country with sudden increase and decrease. For Rwanda in the validation step, the models in comparison to Uganda achieve relatively better performances. Nevertheless, high disparity can be seen between the models, which demonstrate uncertainty of confirmed daily COVID-19 cases in Eastern African countries. Despite the drawback in the prediction efficiency, it can be observed that ANN and ANFIS have appreciable

performance above 0.7  $R^2$  value and ANFIS led to most efficient results with  $MAD=0.0106$ ,  $MSE=0.0003$ ,  $RMSE=0.0185$  and  $R^2=0.9059$ .

The results for the West African countries are presented by Table 6. The performance of the applied models shows that AI models are capable of predicting the confirmed cases of COVID-19 in Nigeria, whereas MLR model can also be employed. The better prediction capability of the AI models could be due to their ability of dealing with the nonlinear, stochastic and uncertain phenomena associated with COVID-19. Despite the prediction capability of the AI-based models, it is observed that ANN and ANFIS model led to better performance. This emphasized the wide adaptation and general application of ANN due to some of its advantages including easy application, good generalization and above all efficient and accurate prediction. ANFIS on the other hand is a hybrid model that combines the learning capability of fuzzy system and prediction capability of ANN. This makes ANFIS unique with high precision.

For Senegal results shown in Table 6, it can be observed in the validation step that the results are comparable with that of Nigeria. This is because Nigeria and Senegal share same region in Africa. The culture, behaviour and social mingling are similar between the two countries; COVID-19 is mostly contracted through these means, and thereby led to similarity in daily-confirmed cases as well as predictive performance of the models.

Table 7 presented the results of daily-confirmed cases of COVID-19 prediction by the four applied models for Southern Africa. It can be deduced that for Namibia in the validation step, less accurate and less appreciable predictions are achieved by all models with exception of ANFIS, which has  $MAD=0.0183$ ,  $MSE=0.0012$ ,  $RMSE=0.0343$  and  $R^2=0.8059$ . The inefficiency of the results by ANN, SVM and MLR models might be due to the uncertain nature of the cases which makes the prediction tedious.

For the results of South Africa given by Table 7 in the validation step, it can be seen that all models archive high performance accuracy. This is because South Africa has the highest number of daily-confirmed COVID-19 cases (37,875 in this study period); the steady flow of the cases helps the models to have precise trend of COVID-19 in the country thereby improving prediction accuracy. ANFIS has the best performance with  $MAD=0.0195$ ,  $MSE=0.0011$ ,  $RMSE=0.0331$  and  $R^2=0.8846$ . The second most efficient model is ANN, followed by SVM model, and MLR model is the least in performance owing to its linear approach and its inability of solving nonlinear aspects.

The results for Central Africa countries including Gabon and Cameroon are presented by Table 8. For Gabon based on the results in the validation step, ANFIS provided the highest accuracy with  $MAD=0.0411$ ,  $MSE=0.0055$ ,  $RMSE=0.0741$  and  $R^2=0.6983$ , followed by ANN with  $MAD=0.0447$ ,  $MSE=0.0079$ ,  $RMSE=0.0888$  and  $R^2=0.5866$ , then SVM with  $MAD=0.0441$ ,  $MSE=0.0103$ ,  $RMSE=0.1014$  and  $R^2=0.5289$  and lastly, MLR model with  $MAD=0.0700$ ,  $MSE=0.0142$ ,  $RMSE=0.1445$  and  $R^2=0.4429$ .

For Cameroon from Table 8, ANFIS also shows better prediction skills with  $MAD=0.0080$ ,  $MSE=0.0012$ ,  $RMSE=0.0341$  and  $R^2=0.8200$ . Despite linearity of MLR model, it still produced reliable performance in comparison to ANN and SVM. The MLR model predictive capability is actually not baffling as it is a nonlinear system identification evolving tool and it showed more predictive ability in several studies (Kouadri et al. 2021).

The performance of the individual models can be compared and assessed graphically by Fig. 10 using a radar chart. The radar chart has the ability to assemble several models into one chart for easy comparison. In terms of  $R^2$ , the wider the internal lines are, the higher the precision of the models and vice versa.

As depicted by Fig. 10, depending on the number of daily-confirmed cases of the COVID-19 and the frequency of their occurrence, the performance of the models is different. For ANN model (Fig. 10a-d), with positive results for everyday COVID-19 test, ANN is able to produce the best performance for South Africa and Morocco. However, it is of paramount significance to understand the fact that not in all situations the large number of daily-confirmed

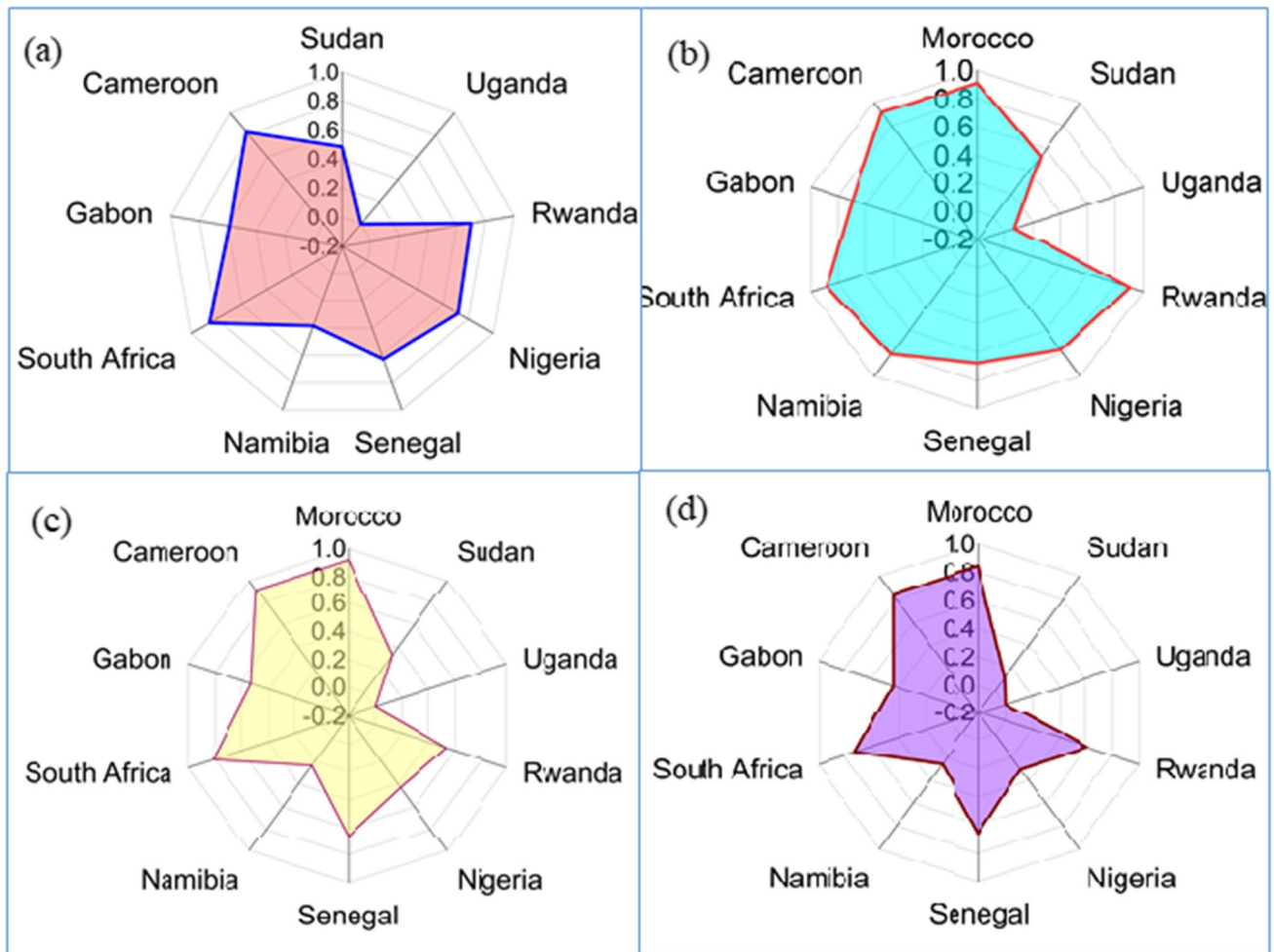
COVID-19 cases matters with regard to the accuracy and efficiency of the predictive models. Stringent protective measures taken by authorities such as lockdown, social distancing, use of sanitizers play a major role in the identification of cases and efficient prediction. For example, Cameroon has less number of cases compared to Nigeria and several other countries. Nevertheless, the measures taken by the Cameroonian authority to curve the effect and spread of COVID-19 make it easier to unravel the uncertainties surrounding the COVID-19 and thereby making the models feasible to have reliable and accurate prediction.

By inspection of the models performance from Fig. 10a-d, it can be realized that the behaviour of the models in terms of performance is similar with respect to countries. The models have the highest accuracy for South Africa, Morocco, followed by Cameroon, Nigeria, Rwanda, Senegal, Gabon, Sudan, Namibia and Uganda. Moreover, by visual observation of the Fig. 10a-d, it can be seen that ANFIS provided better performance in almost all countries, which is due to its combined efficiencies of fuzzy logic and neural network. Comparing the models performance between Tables 4, 5, 6, 7 and 8 and Fig. 10, it can be said that the country with the best models accuracy is Morocco. Therefore, time series plots in order to see the trend between predicted and observed daily-confirmed COVID-19 cases in the validation step (from 05/07/2021 to 12/12/2021) for Morocco are given by Fig. 11.

## Results of the ensemble models

Figure 11a demonstrates the performance of all models for Morocco. It can be seen from the figure that all the models generally follow the trend of the observed data. However, close observation of predicted values cannot be clearly seen due to the fluctuations of large values. Consequently, Fig. 11b is plotted which zoomed the values in order to have precision in observing the predicted against the actual data values. In spite of the fact that the models performed better for Morocco than any other country, they still show room for improvement as closer look shows wide margins between predicted and observed values. Therefore, ensemble models based on ANN-E and SVM-E are employed to improve the modelling accuracy. The results of the ensemble models are presented in Tables 9 and 10.

The results of ensemble models applied show high improvement with minimum errors and high  $R^2$  values of mostly more than 0.9. Comparing the ANN-E (Table 9) results with single models (Tables 4, 5, 6, 7 and 8), it can be seen that a highly significant enhancement in performance is achieved. The ANN-E improved prediction accuracy of ANN models in the validation step up to 10%, 14%, 42%, 6%, 83%, 11%, 7%, 5%, 7% and 31% for Morocco, Sudan, Namibia, South Africa, Uganda, Rwanda, Nigeria, Senegal,



**Fig. 10** Performance comparison of the individual models based on  $R^2$  for (a) ANN, (b) ANFIS, (c) SVM and (d) MLR

Gabon and Cameroon, respectively. In view of the achieved results, it can be said that with less performance of the single models the ensemble models performed better (Nourani et al. 2019b; 2020). For instance, the highest increment in models performance was achieved for Uganda by 83%, which is the country with least performance of the single models. On the other hand, countries with highest single modelling accuracy were found to have the least improvement in their efficiency by ensemble models. For example, Morocco and South Africa have enjoyed the most successful prediction of the daily-confirmed COVID-19 cases by single models but were found to have efficiency improvement by 10% and 6% only. This indicates that with weak or poor performance single models, huge space would be left to enhance prediction, whereas for single models that performed excellently, little gap would be left behind to improve the prediction performance.

Nonetheless, the results of this study show that the advent of several variants of COVID-19 ensure that the ensemble

model does not only have large improvement over weak performance single models for daily-confirmed COVID-19 cases in Africa, but huge improvement can be achieved even for high performance single models. For instance, Cameroon is amongst the countries with highest single models performance, and hence, the ensemble models have improved its performance by 31%.

Comparing Tables 9 and 10, it can be deduced that ensemble models have a comparable performance, which perhaps could be due to similar methodology they shared of combining the single models. It can be vividly seen from Tables 9 and 10 that there is no much superiority in performance between ANN-E and SVM-E. For some countries including Morocco and Sudan, ANN-E edged a little bit higher, whereas for countries including Rwanda and Gabon, a superior accuracy is demonstrated by SVM-E. Based on this, it can be stated that there is no better algorithm in ensemble prediction and any ensemble kernel could lead to high performance improvements. The results of all the single



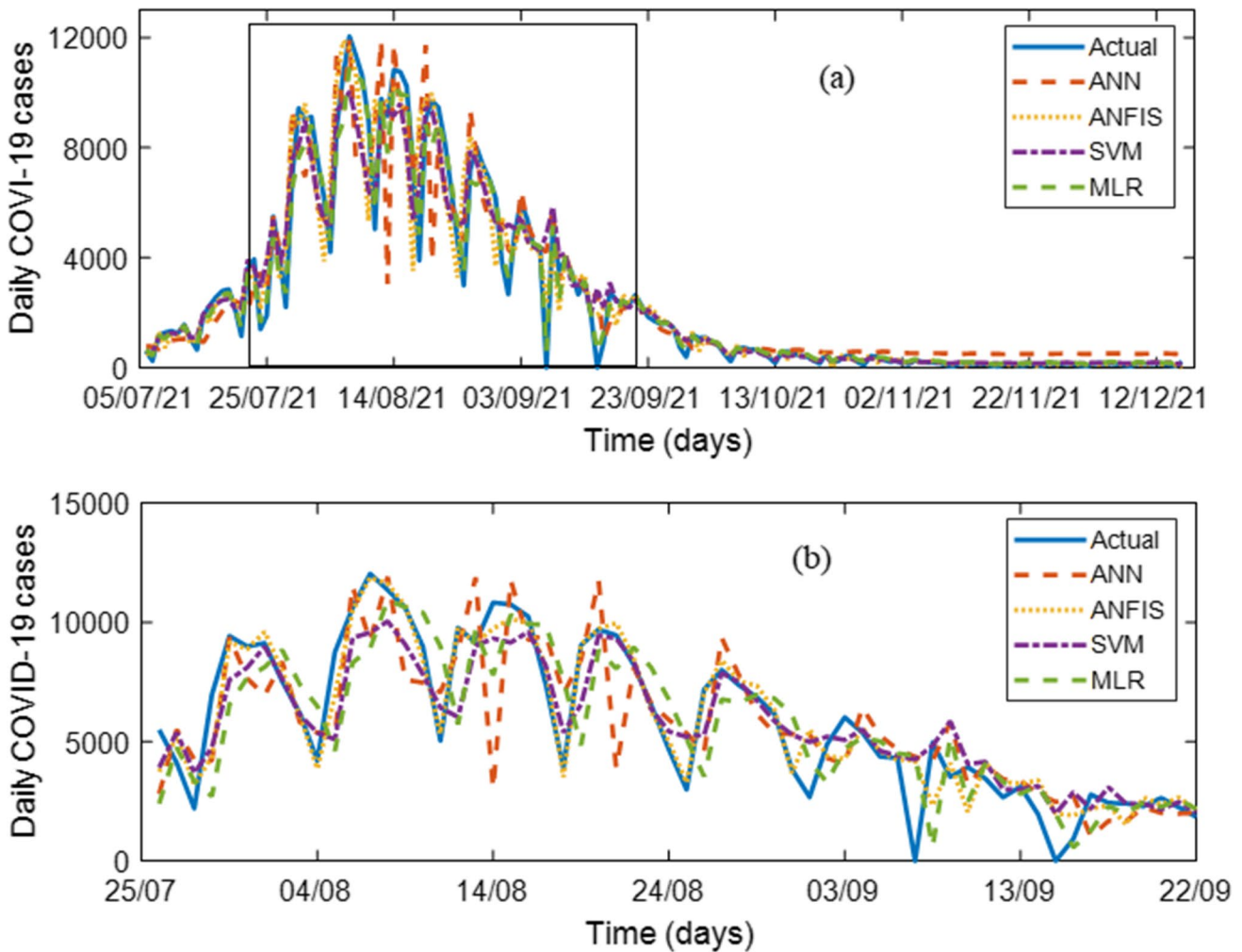


Fig. 11 Observed versus predicted daily-confirmed COVID-19 cases in the validation step for (a) complete dataset. (b) Zoom view

Table 9 Results of the applied ensemble models based on ANN-E

Region	Country	Training				Validation			
		MAD	MSE	RMSE	R <sup>2</sup>	MAD	MSE	RMSE	R <sup>2</sup>
North Africa	Morocco	0.0289	0.0025	0.0500	0.9653	0.0206	0.0009	0.0300	0.9282
	Sudan	0.0210	0.0012	0.0343	0.8338	0.0265	0.0023	0.0478	0.6299
South Africa	Namibia	0.0277	0.0025	0.0497	0.8968	0.0204	0.0011	0.0330	0.7980
	South Africa	0.0276	0.0027	0.0519	0.9366	0.0203	0.0011	0.0328	0.9219
East Africa	Uganda	0.0032	0.0000	0.0050	0.9958	0.0031	0.0000	0.0064	0.8314
	Rwanda	0.0265	0.0047	0.0686	0.9207	0.0111	0.0003	0.0185	0.8059
West Africa	Nigeria	0.0248	0.0014	0.0368	0.9118	0.0315	0.0026	0.0510	0.7926
	Senegal	0.0168	0.0016	0.0394	0.9526	0.0185	0.0008	0.0288	0.6756
Central Africa	Gabon	0.0485	0.0088	0.0939	0.8954	0.0404	0.0058	0.0761	0.6550
	Cameroon	0.0042	0.0002	0.0145	0.9674	0.0073	0.0002	0.0155	0.9616

and ensemble models are compared by Taylor diagrams and presented in Fig. 12.

A Taylor diagram takes in to account the RMSE between prediction by the models and observed data as well as pattern

correlations and variability, which summarizes the overall performance of the models (Abdullahi et al. 2019b). In the graph, correlation coefficients (CC), RMSE and standard deviation (SD) are used to determine the similarity between

**Table 10** Results of the applied ensemble models based on SVM-E

Region	Country	Training				Validation			
		MAD	MSE	RMSE	$R^2$	MAD	MSE	RMSE	$R^2$
North Africa	Morocco	0.0299	0.0028	0.0527	0.9615	0.0202	0.0009	0.0302	0.9276
	Sudan	0.0251	0.0024	0.0488	0.8408	0.0208	0.0011	0.0336	0.6142
South Africa	Namibia	0.0277	0.0025	0.0498	0.8978	0.0181	0.0011	0.0328	0.7980
	South Africa	0.0289	0.0028	0.0531	0.9383	0.0206	0.0010	0.0323	0.9180
East Africa	Uganda	0.0058	0.0002	0.0146	0.9645	0.0040	0.0001	0.0087	0.6943
	Rwanda	0.0295	0.0020	0.0452	0.9157	0.0204	0.0008	0.0283	0.8139
West Africa	Nigeria	0.0331	0.0027	0.0520	0.8930	0.0286	0.0016	0.0406	0.7845
	Senegal	0.0205	0.0022	0.0466	0.9337	0.0193	0.0009	0.0293	0.6629
Central Africa	Gabon	0.0506	0.0072	0.7175	0.8999	0.0429	0.0055	0.0849	0.7175
	Cameroon	0.0083	0.0002	0.0133	0.9728	0.0097	0.0004	0.0189	0.9429

predictive models and observed records. The observed dataset is positioned along the abscissa of the circle from which the performance of the predictive models is assessed (Al-Sultani et al. 2021). In general, if the predicted SD values are surpassed by the observed values, then an underestimation occurs. Meanwhile, overestimation occurs if on the other hand, the predicted values surpass the SD of the observed values (Abdullahi et al. 2019b).

As seen from Fig. 12, based on CC values, the ANN-E and SVM-E have lower values (close to 1) which signify the most reliable and efficient daily-confirmed COVID-19 cases prediction across all countries. This is an indication that besides the tabulated superiority of ensemble models over single models, which are based on global statistical indices applied, graphically, ensemble models also outperformed other models. In terms of RMSE, it can be seen that ensemble models have lower error values and, hence, led to more accurate prediction. With respect to SD, the values that are more close to the actual line signify more reliability. It can be observed that the ensemble values expressed better prediction skill.

In general, the results obtained in this study demonstrated the capability of ensemble models in improving the modelling efficiency of standalone models. Even though, the number of cases as well as precautionary measures adopted by each country may have an impact on the prediction efficiency of the single models. The stochastic and uncertain nature of daily-confirmed COVID-19 cases in African countries can be greatly described and ascertained by using ensemble models.

## Conclusion

In this study, novel ensemble machine learning (ML) approaches called ANN-E and SVM-E were applied to predict COVID-19 pandemic across 10 African countries including Morocco, Sudan, Namibia, South Africa, Uganda,

Rwanda, Nigeria, Senegal, Gabon and Cameroon. The advantage of these methods over others is that they take into cognizance both the linear and nonlinear aspects of COVID-19 in their predictions. To achieve the study aim, three ML models including artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) were used initially as standalone models for the COVID-19 prediction. Multiple linear regression (MLR) model was also applied for comparison. Thereafter, the input kernels of ANN and SVM were replaced with the outputs of the standalone models for performance improvement.

The proposed ANN-E and SVM-E were tested on COVID-19 because it is amongst the major challenges currently facing the entire humanity. The proposed methods also can be generalized and applied for any time series prediction. The results of the simulation and comparative analysis carried out showed that the proposed ANN-E and SVM-E approaches can be useful tools for time series prediction performance improvement and outperformed all the other standalone methods tested using the same datasets. The results demonstrated very high improvements in predicting the COVID-19 pandemic in Africa with MAD = 0.0073, MSE = 0.0002, RMSE = 0.0155 and  $R^2 = 0.9616$ . The ANN-E improved the prediction accuracy of ANN models in the validation step up to 10%, 14%, 42%, 6%, 83%, 11%, 7%, 5%, 7% and 31% for Morocco, Sudan, Namibia, South Africa, Uganda, Rwanda, Nigeria, Senegal, Gabon and Cameroon, respectively.

The two main contributions of this research are: (i) The prediction accuracy of the ML models has been improved and enhanced by the proposed approaches for daily-confirmed COVID-19 prediction in Africa. Despite the complex nature of the COVID-19 pandemic, promising improvements in results were achieved by the proposed ensemble approaches. These can serve as alternative methods for disease outbreak predictions, which can assist the policy makers as well as the authorities to make decisions on measures to apply and the time of their implementation. (ii) The

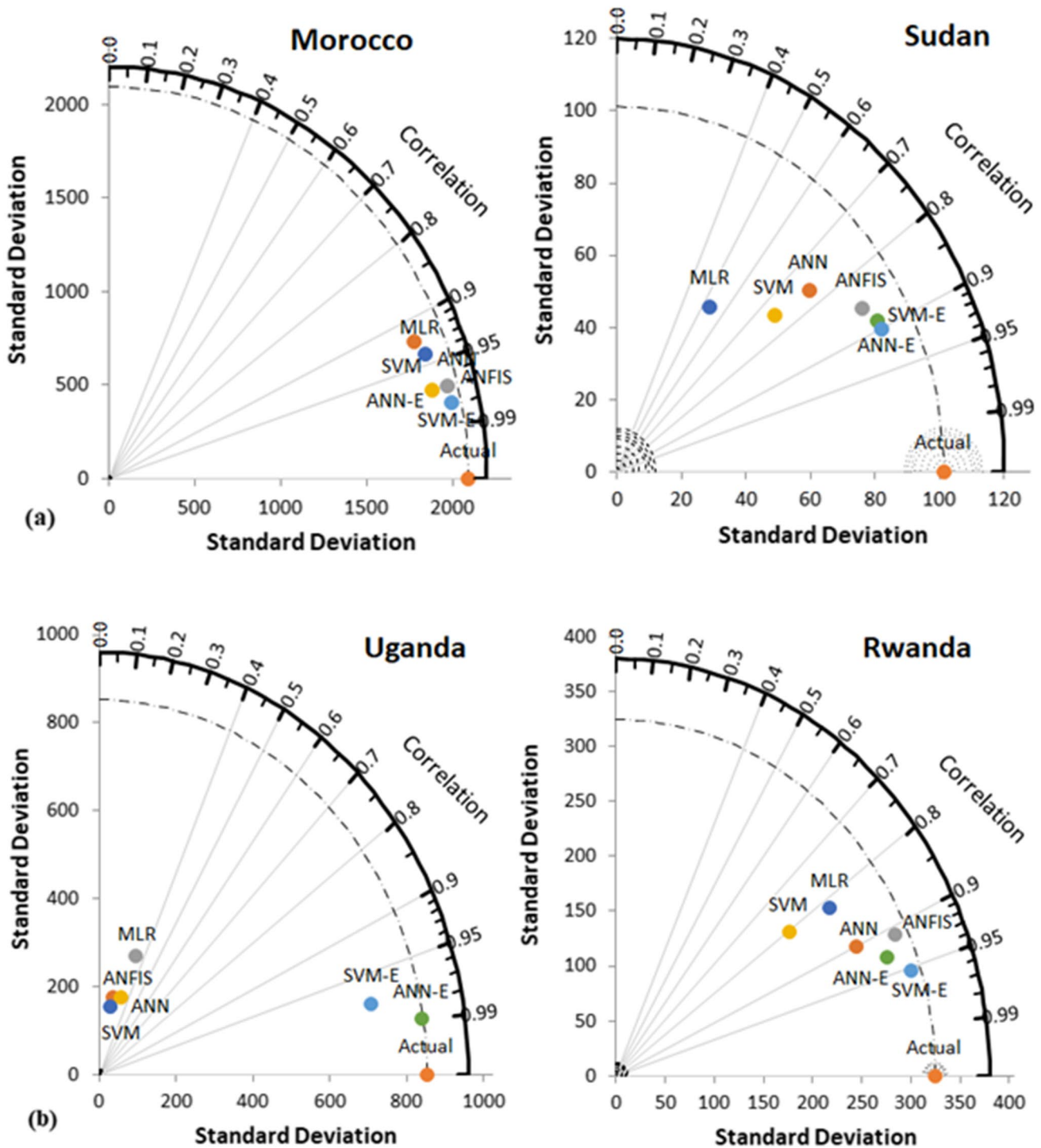


Fig. 12 Performance comparison of all models for (a) North Africa, (b) East Africa, (c) West Africa, (d) South Africa and (e) Central Africa

proposed approaches also implied that in case of an outbreak of disease, the traditional epidemiological models together with the ML-based ensemble approaches could be employed for new cases prediction.

The major challenge in the application ANN-E and SVM-E is that despite combining both linear and nonlinear

models, which successfully helped in capturing both the linear and nonlinear complex nature of COVID-19, their kernel functions are still nonlinear (i.e. only nonlinear kernels were utilized). Therefore, to have an efficient performance comparison of ensemble approaches, for future work, linear ensemble approaches including simple linear

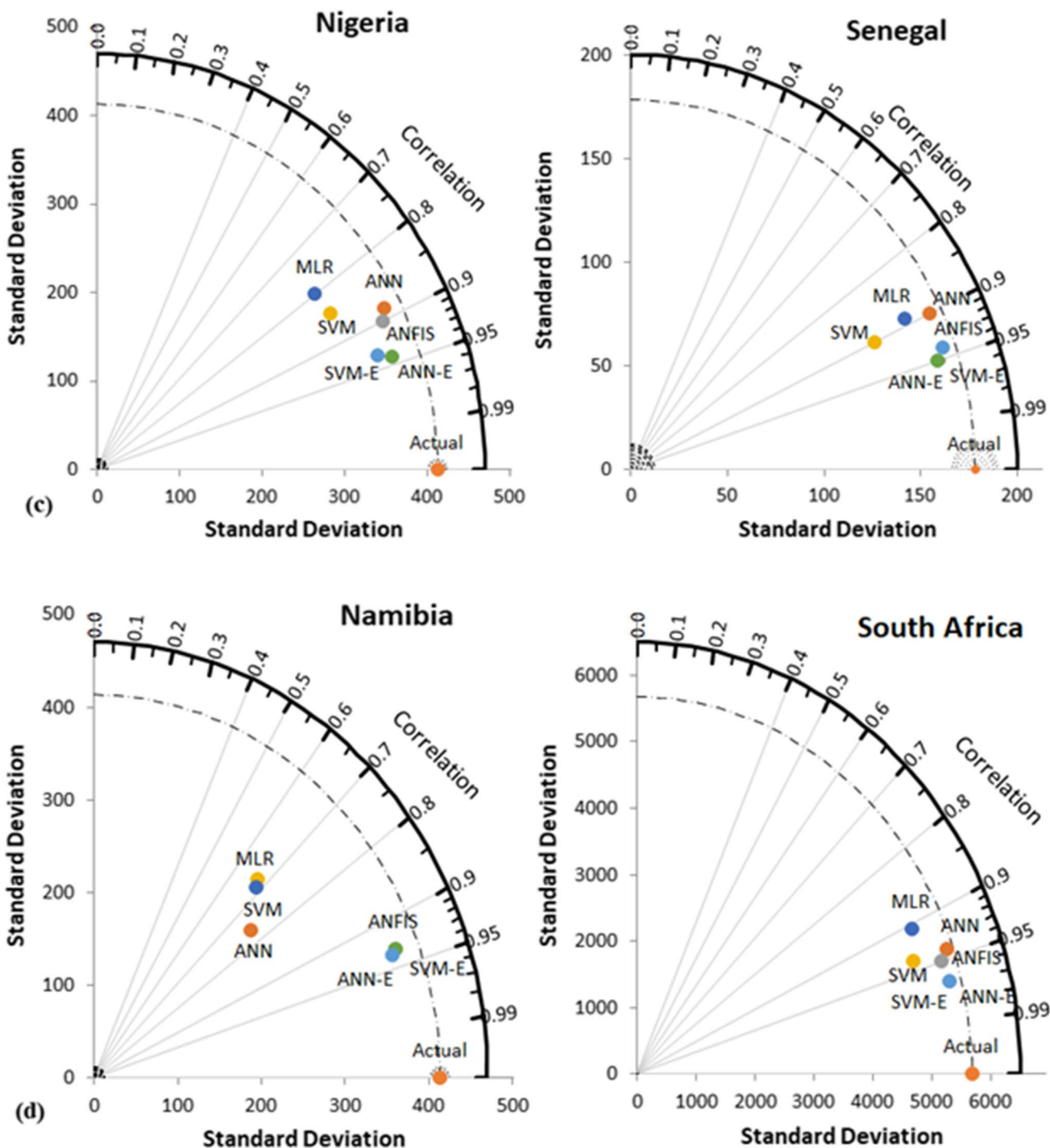


Fig. 12 (continued)

average ensemble (SLAE) and weighted linear average ensemble (WLAE) should be applied in order to determine the most efficient ensemble approaches for COVID-19 prediction. Further studies should also consider application of the ensemble models for modelling cumulative cases and

mortality rate of COVID-19 in Africa. Other types of ML models as well as other ensemble kernels such as genetic algorithms, etc. could be employed for further studies to assess their performance.

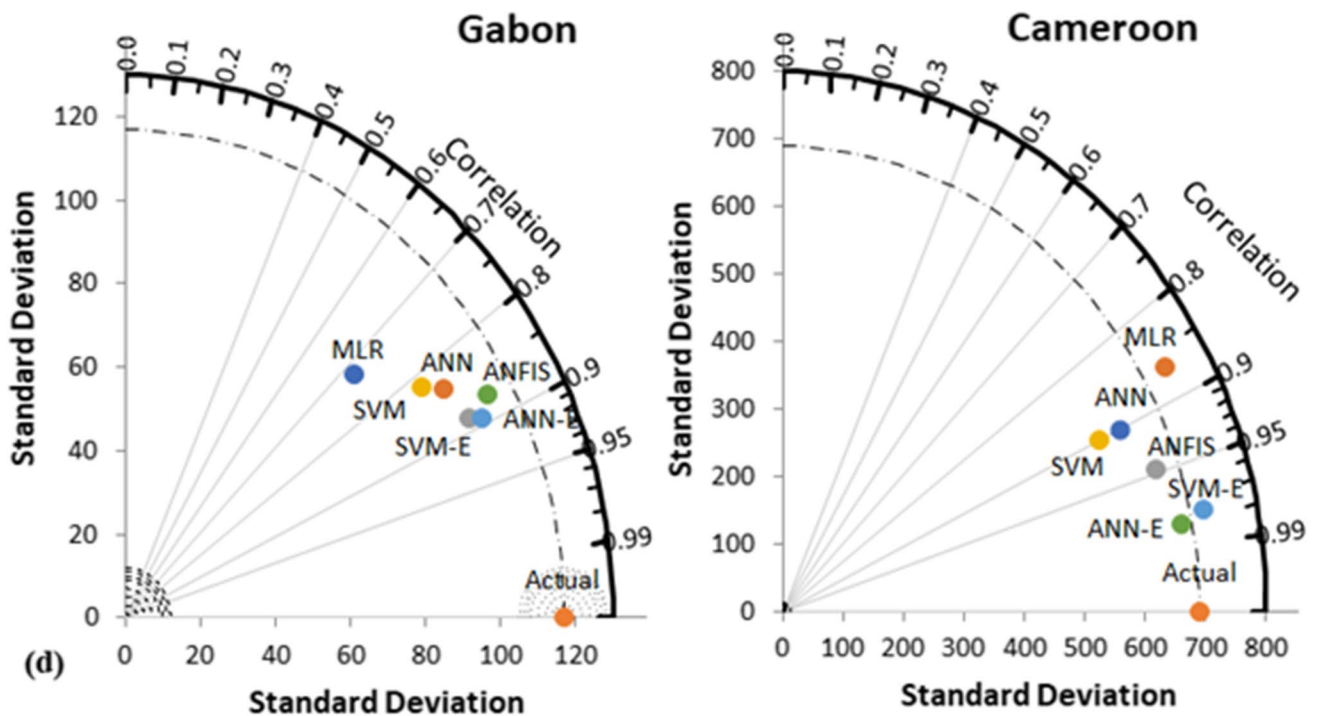


Fig. 12 (continued)

**Author contribution** Sourced the daily COVID-19 data and provided the general study concept/idea (ZI). Introduction and assembly of all sections of the study (PT). Methodology, results and discussion (JA).

**Data availability** The data and materials used in this study would be available on request.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** All authors gave their consent to participate in writing of the manuscript.

**Consent for publication** All authors gave their consent to publish the manuscript.

**Conflict of interest** The authors declare no competing interests.

## References

- Abdullahi J, Elkiran G, Nourani V (2017) Application of artificial neural network to predict reference evapotranspiration in Famagusta, North Cyprus. In 11th International Scientific Conference on Production Engineering Development And Modernization of Production (pp. 549–554)
- Abdullahi J, Elkiran G, Nourani V (2019a) Artificial intelligence based and linear conventional techniques for reference evapotranspiration modeling. In International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions (pp. 197–204). Springer, Cham. [https://doi.org/10.1007/978-3-030-35249-3\\_25](https://doi.org/10.1007/978-3-030-35249-3_25)
- Abdullahi J, Iravani A, Nourani V, Elkiran G (2019b) Application of artificial intelligence based and multiple regression techniques for monthly precipitation modeling in coastal and inland stations. *Desalin Water Treat* 177:338–349. <https://doi.org/10.5004/dwt.2020.24954>
- Abdullahi J, Elkiran G (2021) Monthly prediction of reference evapotranspiration in northcentral Nigeria using artificial intelligence tools: a comparative study. In International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions (pp. 165–172). Springer, Cham. [https://doi.org/10.1007/978-3-030-92127-9\\_25](https://doi.org/10.1007/978-3-030-92127-9_25)
- Abunama T, Othman F, Ansari M, El-Shafie A (2019) Leachate generation rate modeling using artificial intelligence algorithms aided by input optimization method for an MSW landfill. *Environ Sci Pollut Res* 26(4):3368–3381. <https://doi.org/10.1007/s11356-018-3749-5>
- Akrami SA, Nourani V, Hakim SJS (2014) Development of nonlinear model based on wavelet-ANFIS for rainfall forecasting at Klang Gates Dam. *Water Resour Manag* 28(10):2999–3018. <https://doi.org/10.1007/s11269-014-0651-x>
- Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, Tadono T (2019) Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. *Geospat Health* 14(2). <https://doi.org/10.4081/gh.2019.771>
- Al-Sultani AO, Al-Mukhtar M, Roomi AB, Farooque AA, Khedher KM, Yaseen ZM (2021) Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access* 9:108527–108541. <https://doi.org/10.1109/ACCESS.2021.3100490>
- Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Atkinson PM (2020) Covid-19 outbreak

- prediction with machine learning. *Algorithms* 13(10):249. <https://doi.org/10.3390/a13100249>
- Aqil M, Kita I, Yano A, Nishiyama S (2007) Analysis and prediction of flow from local source in a river basin using a neuro-fuzzy modeling tool. *J Environ Manage* 85(1):215–223. <https://doi.org/10.1016/j.jenvman.2006.09.009>
- Chenar SS, Deng Z (2018) Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ Int* 111:212–223. <https://doi.org/10.1016/j.envint.2017.11.032>
- Chinese Diagnosis and Treatment Plant (CDTP) of COVID-19 patients (The fifth edition). <http://www.nhc.gov.cn/yzygj/s7653/p/202002/3b09b894ac9b4204a79d6b5b8912d4440.shtml>. 2020. Accessed 5 Jun 2020.
- Cortes C, Vapnik V (1995) Support-Vector Networks. *Machine Learning* 20(3):273–297. <https://doi.org/10.1007/BF00994018>
- Elkiran G, Nourani V, Elvis O, Abdullahi J (2021) Impact of climate change on hydro-climatological parameters in North Cyprus: application of artificial intelligence-based statistical downscaling models. *J Hydroinf* 23(6):1395–1415. <https://doi.org/10.2166/hydro.2021.091>
- Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E (2014) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int J Environ Sci Technol* 11(3):645–656. <https://doi.org/10.1007/s13762-013-0378-x>
- Ekhmaj AI (2012) Prediction of evapotranspiration using artificial neural networks model. Malaysia In: International Annual Symposium on Sustainability Science and Management. Terengganu pp. 937–943
- Fletcher D, Goss E (1993) Forecasting with neural networks: an application using bankruptcy data. *Information & Management* 24(3):159–167. [https://doi.org/10.1016/0378-7206\(93\)90064-Z](https://doi.org/10.1016/0378-7206(93)90064-Z)
- Ghorbani MA, Zadeh HA, Isazadeh M, Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ Earth Sci* 75(6):1–14. <https://doi.org/10.1007/s12665-015-5096-x>
- Gralinski LE, Menachery VD (2020) Return of the coronavirus: 2019-nCoV. *Viruses* 12(2):135. <https://doi.org/10.3390/v12020135>
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *China the Lancet* 395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Hussain D, Khan AA (2020) Machine learning techniques for monthly river flow forecasting of Hunza River. *Pak Earth Sci Inf* 13(3):939–949. <https://doi.org/10.1007/s12145-020-00450-z>
- Islam M, Mahmud S, Muhammad LJ, Nooruddin S, Ayon SI (2020) Wearable technology to assist the patients infected with novel coronavirus (COVID-19). *SN Computer Science* 1(6):1–9. <https://doi.org/10.1007/s42979-020-00335-4>
- Ivanov D (2020) Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transp Res Part e: Logist and Transp Rev* 136:101922. <https://doi.org/10.1016/j.tre.2020.101922>
- Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence. Prentice Hall, New Jersey
- Jebara T (2012) Machine learning: discriminative and generative (Vol. 755). Springer Science & Business Media
- Khatri N, Khatri KK, Sharma A (2020) Artificial neural network modelling of faecal coliform removal in an intermittent cycle extended aeration system—sequential batch reactor based wastewater treatment plant. *J Water Process Eng* 37:101477. <https://doi.org/10.1016/j.jwpe.2020.101477>
- Kiran NR, Ravi V (2008) Software reliability prediction by soft computing techniques. *J Syst Softw* 81(4):576–583. <https://doi.org/10.1016/j.jss.2007.05.005>
- Kouadri S, Elbeltagi A, Islam ARM, Kateb S (2021) Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Appl Water Sci* 11(12):1–20. <https://doi.org/10.1007/s13201-021-01528-9>
- Khodaei-mehr J, Tangestanizadeh S, Vatankeh R, Sharifi M (2018) ANFIS-based optimal control of hepatitis C virus epidemic. *IFAC-PapersOnLine* 51(15):539–544. <https://doi.org/10.1016/j.ifacol.2018.09.211>
- Kocadagli O, Baygul A, Gokmen N, Incir S, Aktan C (2022) Clinical prognosis evaluation of COVID-19 patients: an interpretable hybrid machine learning approach. *Curr Res Transl Med* 70(1):103319. <https://doi.org/10.1016/j.rettram.2021.103319>
- Koike F, Morimoto N (2018) Supervised forecasting of the range expansion of novel non-indigenous organisms: alien pest organisms and the 2009 H1N1 flu pandemic. *Glob Ecol Biogeogr* 27(8):991–1000. <https://doi.org/10.1111/geb.12754>
- Li M, Wei D, Liu T, Liu Y, Yan L, Wei Q, Du B, Xu W (2019) EDTA functionalized magnetic biochar for Pb (II) removal: adsorption performance, mechanism and SVM model Prediction. *Sep Purif Technol*. <https://doi.org/10.1016/j.seppur.2019.115696>
- Lucas B, Vahedi B, Karimzadeh M (2022) A spatiotemporal machine learning approach to forecasting COVID-19 incidence at the county level in the USA. *Int J Data Sci Anal p*. 1–20. <https://doi.org/10.1007/s41060-021-00295-9>
- Mahase E (2020) China coronavirus: what do we know so far? <https://doi.org/10.1136/bmj.m308>
- Mirza S, Niwalkar A, Gupta A, Gautam S, Anshul A, Bherwani H, Kumar R (2022) Is safe distance enough to prevent COVID-19? Dispersion and tracking of aerosols in various artificial ventilation conditions using OpenFOAM. *Gondwana Research*. <https://doi.org/10.1016/j.gr.2022.03.013>
- Mohammed SH, Ahmed MM, Al-Mousawi AM, Azeez A (2018) Seasonal behavior and forecasting trends of tuberculosis incidence in Holy Kerbala Iraq. *Int J Mycobacteriology* 7(4):361. [https://doi.org/10.4103/ijmy.ijmy\\_109\\_18](https://doi.org/10.4103/ijmy.ijmy_109_18)
- Mohammed SJ, Abdel-khalek HA, Hafez SM (2021) Predicting performance measurement of residential buildings using an artificial neural network. *Civ Eng J* 7(3):461–476
- Morens DM, Daszak P, Taubenberger JK (2020) Escaping Pandora's box—another novel coronavirus. *N Engl J Med* 382(14):1293–1295. <https://doi.org/10.1056/NEJMp2002106>
- Muhammad LJ, Islam M, Usman SS, Ayon SI (2020) Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. *SN Computer Science* 1(4):1–7. <https://doi.org/10.1007/s42979-020-00216-w>
- Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA (2021) Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN Computer Science* 2(1):1–13. <https://doi.org/10.1007/s42979-020-00394-7>
- Naganna S, Deka P, Ghorbani M, Biazar S, Al-Ansari N, Yaseen Z (2019) Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water*. <https://doi.org/10.3390/w11040742>
- Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Anal Appl* 24(3):1207–1220. <https://doi.org/10.1007/s10044-021-00984-y>
- Niazkar HR, Niazkar M (2020) Application of artificial neural networks to predict the COVID-19 outbreak. *Glob Health Res Policy* 5(1):1–11. <https://doi.org/10.1186/s41256-020-00175-y>

- Nourani V, Elkiran G, Abdullahi J, Tahsin A (2019a) Multi-region modeling of daily global solar radiation with artificial intelligence ensemble. *Nat Resour Res* 28(4):1217–1238. <https://doi.org/10.1007/s11053-018-09450-9>
- Nourani V, Elkiran G, Abdullahi J (2019b) Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J Hydrol* 577:123958. <https://doi.org/10.1016/j.jhydrol.2019.123958>
- Nourani V, Elkiran G, Abdullahi J (2020) Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. *J Hydrol* 581:124434. <https://doi.org/10.1016/j.jhydrol.2019.124434>
- Nourani V, Fard MS (2016) Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv Eng Softw* 47(1):127–146. <https://doi.org/10.1016/j.advengsoft.2011.12.014>
- Noy O, Coster D, Metzger M, Atar I, Shenhar-Tsarfaty S, Berliner S, Shamir R (2022) A machine learning model for predicting deterioration of COVID-19 inpatients. *Sci Rep* 12(1):1–9. <https://doi.org/10.1038/s41598-022-05822-7>
- Praveen Kumar R, Samuel C, Raju SR, Gautam S (2022) Air pollution in five Indian megacities during the Christmas and New Year celebration amidst COVID-19 pandemic. *Stoch Environ Res Risk Assess* p. 1–31. <https://doi.org/10.1007/s00477-022-02214-1>
- Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R (2020) COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics* 8(6):890. <https://doi.org/10.3390/math8060890>
- Raphael D, Stanley P. Novel coronavirus from Wuhan China, 2019–2020, Chapter 155, Mandell, Douglas, and Bennett's Principles and practice of infectious diseases, ninth edition (Elsevier, 2020)
- Rypdal M, Sugihara G (2019) Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics. *Nat Commun* 10(1):1–8. <https://doi.org/10.1038/s41467-019-10099-y>
- Sajda P (2006) Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng* 8(1):537–565
- Sharghi E, Nourani V, Behfar N (2018) Earthfill dam seepage analysis using ensemble artificial intelligence based modeling. *J Hydroinf* 20(5):1071–1084. <https://doi.org/10.2166/hydro.2018.151>
- Sharma A, Vijay R, Bodhe GL, Malik LG (2018) An adaptive neuro-fuzzy interface system model for traffic classification and noise prediction. *Soft Comput* 22(6):1891–1902. <https://doi.org/10.1007/s00500-016-2444-z>
- Scarpino SV, Petri G (2019) On the predictability of infectious disease outbreaks. *Nat Commun* 10(1):1–8. <https://doi.org/10.1038/s41467-019-08616-0>
- Sujatha A, Govindaraju L, Shivakumar N, Devaraj V (2021) Fuzzy knowledge based system for suitability of soils in airfield applications. *Civil Engineering Journal* 7(1):140–152
- Talebkeikhah M, Sadeghtabaghi Z, Shabani M (2021) A comparison of machine learning approaches for prediction of permeability using well log data in the hydrocarbon reservoirs. *Journal of Human, Earth, and Future* 2(2):82–99
- Thapliyal J, Bhattacharyya M, Prakash S, Patni B, Gautam S, Gautam AS (2022) Addressing the relevance of COVID–19 pandemic in nature and human socio-economic fate. *Stoch Environ Res Risk Assess* p. 1–15. <https://doi.org/10.1007/s00477-022-02191-5>
- Tiwari D, Bhati BS, Al-Turjman F, Nagpal B (2022) Pandemic coronavirus disease (Covid-19): world effects analysis and prediction using machine-learning techniques. *Expert Syst* 39(3):e12714. <https://doi.org/10.1111/exsy.12714>
- Uyar K, Ilhan U, Iseri EI, Ilhan A (2019) Forecasting measles cases in Ethiopia using neuro-fuzzy systems. In 2019 3rd international symposium on multidisciplinary studies and innovative technologies (ISMSIT) (pp. 1–5). IEEE
- Vapnik V (1998) The support vector method of function estimation, in: *Nonlinear modeling*. Springer, pp. 55–85. [https://doi.org/10.1007/978-1-4615-5703-6\\_3](https://doi.org/10.1007/978-1-4615-5703-6_3)
- WHO (2020), Key messages and actions for COVID-19 prevention and control in schools
- Xiong Y, Ma Y, Ruan L, Li D, Lu C, Huang L (2022) Comparing different machine learning techniques for predicting COVID-19 severity. *Infect Dis Poverty* 11(1):1–9. <https://doi.org/10.1186/s40249-022-00946-4>
- Zhan Z, Dong W, Lu Y, Yang P, Wang Q, Jia P (2019) Real-time forecasting of hand-foot-and-mouth disease outbreaks using the integrating compartment model and assimilation filtering. *Sci Rep* 9(1):1–9. <https://doi.org/10.1038/s41598-019-38930-y>
- Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov DS, Polykovskiy D, Shayakhmetov R, Filimonov A, Orekhov P. (2020). Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *Insilico Med Hong Kong Ltd A* 307: E1. <https://doi.org/10.26434/chemrxiv.12301457>
- Zivkovic M, Bacanin N, Venkatachalam K, Nayyar A, Djordjevic A, Strumberger I, Al-Turjman F (2021) COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustain Cities Soc* 66:102669. <https://doi.org/10.1016/j.scs.2020.102669>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.