Prediction of COVID-19 Time Series – Case Studies of South Africa and Egypt using Interval Type-2 Fuzzy Logic System

Eyoh, Imo J. *[1], Udo, Edward N.[1], Umoeka, Ini J.[1] and Eyoh, Jeremiah E.[2]

[1]Department of Computer Science, University of Uyo, Uyo, Nigeria.

Email: *imoheyoh, iniumoeka, edwardudo@uniuyo.edu.ng

[2]School of Electrical, Electronics and Systems Engineering, AVRRC, Loughborough University, UK.

Email: j.e.eyoh@lboro.ac.uk

ABSTRACT

COVID-19 is a virus known to emanate from Wuhan, China in December 2019. COVID-19 spread widely to nearby countries like Japan and Korea, followed by Europe and America and later to Africa. Particularly, South Africa and Egypt have been worst hit by the virus. Generally, the COVID-19 data is highly uncertain and requires fuzzy logic approaches for the effective handling of these uncertainties. This study therefore presents the prediction of COVID-19 cases in South Africa and Egypt using interval type-2 fuzzy logic system with Takagi-Sugeno-Kang fuzzy inference and neural network learning. The parameters of the model are adapted using gradient descent backpropagation approach. The proposed model is found to outperform type-1 fuzzy logic system and artificial neural network in terms of the root mean squared error, mean absolute percentage error and mean absolute error.

Key words: COVID-19, gradient descent back propagation, interval type-2 fuzzy logic system, neural network, pandemic.

1. INTRODUCTION

The novel Coronavirus Disease 2019 (COVID-19) is an infectious disease caused by a virus; a member of the Beta coronavirus family called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is thought that the virus outbreak has animal origins [1]. The virus was first reported in Wuhan, Hubei Province, China on December 31, 2019 and was declared by World Health Organization (WHO) as a Public Health Emergency of International Concern on January 31, 2020 [2]. Its rapid spread across the world is in an exponential rate when compared to other outbreaks (Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)) and this caused WHO to recognize COVID-19 as a pandemic on March 11, 2020.

The cumulative incidence of COVID-19 causative virus (SARS-CoV-2) has affected 213 countries and territories around the world with significantly high confirmed and death rates. As at July 30, 2020, as reported by worldometers, there were 17,477,419 confirmed cases, 676,794 deaths and 10,940,377 recovered patients. A study in [3] reported that the spread of the virus put Africa in the first rank of high-risk region. As at July 30, 2020, there were 891,020 confirmed cases, 18,864 deaths and 540,397 recoveries. Moreover, various studies reported that the disease caused by CoV-2 is more dangerous for people with weak immune system. The elderly people and patients with life threatening diseases like cancer, diabetes, neurological conditions, coronary heart disease and HIV/AIDS are more vulnerable to severe effects of COVID-19 [4]. In the absence of proper vaccination and curative drug to arrest the spread and curtail the number of infected people, the best option, especially for developing and underdeveloped countries, to evade the effects of the virus and save the lives of their citizenry is to adhere to government and WHO guidelines regarding washing of hands, use of facemask, restrictions on public gatherings, observing physical distancing, travel restrictions and even lockdowns [5],[6]. Implementing some of these measures result in devastating economic crises, loses and damaging social impacts. These proffered preventive measures by policy makers are based on detailed analysis of the outbreak across the world and the results of some precise scientific predictions. Simple epidemiological and statistical models for COVID-19 have received more attention by authorities, and they are popular in the media (both online and print). With the unavailability of essential data and high level of uncertainty, application of standard models in predicting COVID-19 results in low accuracy, especially for long-term prediction [7]. In the literature, several methods and models are proposed to address this issue, but their robustness and preciseness are still not perfect and need improvement.

2. RELATED WORK

Different methods have been employed for the prediction of COVID-19 cases. For instance, [2] constructed a model based on time series data to predict the long-time period trend of COVID-19 globally. The COVID-19 epidemiological data was fed into logistic model to fit the cap of epidemic trend, and then fed the cap value into Fbprophet model to derive the epidemic curve and predict the trend of the pandemic. Reference [5] proposed a model for predicting the growth and trend of COVID-19 pandemic. The study used data collected from United States, Spain, Italy, United Kingdom, Germany, France and Russia. Data for the study was obtained from Our World in data website; updated daily from WHO situation reports from January 22, to May 4, 2020. Machine Learning (ML) models were built to make prediction of the number of new cases and the dates when the pandemic might end. These models were deployed in cloud data centers for hospitals and health centres to continuously send in their positive patients count. Population density, average and median age, weather conditions and health facilities were also integrated in the cloud model for enhanced prediction accuracy. FogBus framework for prediction, Generalized Inverse Weibull distribution for estimation, Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), Coefficient of determination (R^2) were the comparison metrics. The Weibull model based on iterative weighting, is able to make statistically better predictions than the baseline Gaussian model which shows an over-optimistic picture of the COVID-19 scenario. Reference [6] used deep learning-based models to predict the number of COVID-19 positive reported cases in India (28 states and 4 unions territories of India). Data was collected from the Ministry of Health and Family Welfare, India from March 14, to May 14, 2020. Indian states and unions were categorized into 3 zones (mild, moderate and severe) based on the spread of positive cases and daily growth rate. Long short-term memory (LSTM) models (Deep, Convolutional, Bi-directional) were trained, hyper parameter tuning was carried out, model was selected and future predictions were done. The proposed model yielded high accuracy for short term predictions. Deep LSTM, convolutional LSTM and Bi-directional LSTM gave better accuracy than simple LSTM models. Based on prediction errors, bi-directional LSTM gave the best results, and convolutional LSTM gave the worst.

In [7], a comparative analysis of machine learning and soft computing models for prediction of COVID-19 outbreak as alternative to Susceptible–Infectious–Recovered (SIR) and Susceptible–Exposed–Infectious–Recovered (SEIR) models using Italy, Germany, Iran, USA and China as case studies is presented. The parameters of several simple mathematical models (logistic, linear, logarithmic, quadratic, cubic, compound, power and exponential) were fitted using genetic algorithm (GA), particle swarm optimization (PSO), and Grey Wolf Optimizer (GWO). Machine learning models (MLP and ANFIS) were considered for two data scenarios.

Comparison between analytical and machine learning models was done. Extrapolation for long-term prediction of up to 150 days using the ML models was tested. The logistic model outperformed other models. The logistic GWO outperformed the PSO and GA and the computational cost for GWO was satisfactory. The performance of both ML models for the countries varied amongst the two different scenarios but the two scenarios were all suitable. The results of MLP and ANFIS had high generalization ability for long-term predictions. Reference [8] implemented different survival analysis models to measure the prediction accuracy of the discharge time of 1182 hospitalized COVID-19 patients. Open dataset collected by a group of researchers from different universities and research laboratories were utilized. Several filtering processes were applied to prepare the data for training and statistical analysis, the resulting data was then restructured to make it compatible with survival analysis methods and survival algorithms which were used to perform the time-to-event analysis. Statistical tools such as Kaplan-Meier estimator, Cox Proportional Hazard, Coxnet and Accelerated Time Failure and machine learning tools namely, Stagewise Gradient Boosting and Component-wise Gradient Boosting were adopted. Stagewise Gradient Boosting predicts discharge time compared to other algorithms while using only age and sex as model features. Results from Kaplan-Meier and Cox regression method suggest that sex and age of the hospitalized patients have a direct effect on their recovery time. Findings indicate that being male or being in older age groups is associated with lower hospital discharge probabilities.

Reference [9] provided statistical forecasts for the confirmed cases of COVID-19 globally using robust time series models. Data were retrieved from the Center for Systems Science and Engineering (CSSE) at John Hopkins University for confirmed cases, deaths and recoveries from January 22, to March 11, 2020. Forecasts were produced using models from the exponential smoothing family with limits to trended and non-seasonal models. Pragmatic approach was followed to give 10 days' interval ahead forecasts and predictions. Results show that there was continued increase in the confirmed COVID-19 cases with sizable associated uncertainty. Reference [10] built an adaptive neuro fuzzy inference system (ANFIS) forecasting model based on Flower Pollination Algorithm (FPA) and Salp Swarm Algorithm (SSA) (FPASSA-ANFIS) to estimate and forecast the number of confirmed cases of COVID-19 in China. COVID-19 datasets from January 21, to February 18, 2020 was collected and split into training and testing data. Fuzzy C-mean clustering was applied on the training data. The performance of FPASSA-ANFIS was evaluated and new data was forecasted. FPASSA-ANFIS performance measures were root mean squared error (RMSE), MAE, MAPE, root mean squared relative error (RMSRE), and R^2. The proposed FPASSA-ANFIS model provided a high ability to predict the number of confirmed cases within ten days. Besides, FPASSA-ANFIS outperforms other forecasting models in

terms of RMSE, MAE, MAPE, RMSRE, and R^2. Reference [11] proposed the modeling of the spread of COVID-19 using multilayer perceptron artificial neural network (MLP-ANN) to achieve accurate regression model. Data from John Hopkins University CSSE was collected from January 22, to March 12, 2020. Data collected was split into 3 groups (infected, recovered, deceased) and transformed into a regression dataset for training the MLP ANN. The training datasets used 15530 data points (75%) and the testing used 5176 data points (25%). Hyper parameters of the MLP were determined using grid search algorithm. Cross-validation was performed using K-fold algorithm with 5-folds. When cross-validation was performed, R^2 scores dropped to 0.94 for confirmed, 0.781 for recovered, and 0.986 for deceased patient models, showing high robustness of the deceased patient model, good robustness for confirmed, and low robustness for recovered patient model.

However, the COVID-19 pandemic data is highly non-linear and complex in nature, with high levels of uncertainty [9], [12] such that prediction using standard statistical models may not deliver accurate results because of these levels of uncertainty. According to [13], data collected for COVID-19 are highly uncertain. There is therefore a need for the adoption of fuzzy methodologies for the prediction of COVID-19 outbreak. In the literature, fuzzy logic is known to handle uncertainty well. Many studies therefore have integrated or use fuzzy logic concept in modelling COVID-19 outbreak. For example, [14] proposed an expert system to predict whether a patient is suffering from COVID-19 or not by adopting artificial neural network and fuzzy inference system of type-1 running in MATLAB. 11 symptoms were used as input values into the system while 15 rules were developed on the basis of guidelines given by WHO and other health agencies. The system predicts by comparing the patients' symptoms with COVID-19 symptoms and found to accurately predict at the early stages whether a patient was suffering from COVID-19 or not. Reference [15] presented a multiple ensemble neural network model with fuzzy response aggregation for the COVID-19 time series in Mexico. COVID-19 data of Mexico was collected. Nonlinear autoregressive NN was used to predict future values. Function fitting NN was used to train and learn from the data. MSE of the training and actual data were normalized and the normalized MSEs were used in fuzzy integrator to produce weights. The predictions from several predictors' modules were combined to obtain the final prediction. Result show very good predicted values in the validation data set. The prediction errors of the multiple ensemble neural networks with fuzzy response aggregation are significantly lower than using traditional monolithic neural networks. Reference [16] used SEIR to predict the peak of COVID-19 pandemic in Malaysia. ANFIS model was employed for the short-term prediction of number of infected cases. Analysis of results shows that ANFIS model shows a high forecasting accuracy with low normalized RMSE, low MAPE and high coefficient of determination.

Other studies adopting type-1 fuzzy concept for COVID-19 pandemic prediction include [17] - [21]. Unlike the type-1 fuzzy sets (T1FSs) with precise MF, type-2 fuzzy sets (T2FSs) [22] have MFs that are themselves type-1 FS (lower and upper T1FS). A T2FS is a 3-dimensional FS and may be general or interval T2 depending on the representations in the third dimension. Whilst general type-2 FSs have non-uniform weights in the third dimension which makes it complex to work with, the interval T2 FSs (IT2FSs) have uniform weights with value 1 and this makes it easier to manage and work with. Apart from these, the IT2FS can easily be represented in a 2-dimensional plane which makes it an appropriate choice for this analysis. The IT2FSs have found applications in diverse domains such as time series prediction and forecasting, classification, control, to mention but a few.

Although many studies have been conducted for the prediction of COVID-19 outbreak using the type-1 fuzzy logic systems (FLSs), however, there is a gap in the literature involving papers adopting IT2FLS for the prediction of COVID-19 cases. According to [12], the behaviour of COVID-19 in many countries exhibits high degree of uncertainty and complexity. This implies that using type-1 FLS may not be very appropriate and calls for the application of higher order fuzzy sets – the T2 FSs.

Hence, this paper proposes the use of IT2FLS with Takagi-Sugeno-Kang (TSK) inference for the prediction of COVID-19 outbreak in two African countries namely: South Africa and Egypt. These countries are chosen on the ground that they are the two most affected African countries with COVID-19 outbreak. The motivation is to improve the prediction accuracy of COVID-19 cases by applying IT2FLS-TSK which, unlike type-1 FS is furnished with fuzzy membership functions (lower and upper). For comparison purpose, experimental analysis using type-1 FLS and feedforward neural network (FFNN) are also carried out.

To the best knowledge of the authors, this is the first paper that utilizes IT2FLS for the prediction of COVID-19 pandemic cases. The IT2FLSs utilizing lower membership functions (LMFs) and upper membership functions (UMFs) capture abundance of information and handles uncertainty better. The intention is to encourage other researchers to cue into the narrative by using T2FL theory to address both the regional epidemic and global pandemic outbreaks such as COVID-19.

3. INTERVAL TYPE-2 FUZZY SET

Definition: A T2FS is defined using a LMFs $\underline{\mu}_{\tilde{A}}(x,u)$ and an UMFs, $\bar{\mu}_{\tilde{A}}(x,u)$.

$$\tilde{A} = \left\{ \left((x,u), \underline{\mu}_{\tilde{A}}(x,u), \bar{\mu}_{\tilde{A}}(x,u) \right) \mid \forall x \in X, \ \forall u \in J_x \subseteq [0,1] \right\}$$

(1)

where x is the primary variable in X and u is the secondary variable with domain J. The union of all primary memberships is known as the footprint of uncertainty (FOU) defined as $FOU_{\mu}(\tilde{A}^{*}) = \bigcup_{\forall x \in X} \left[\underline{\mu}_{\tilde{A}^{*}}(x), \overline{\mu}_{\tilde{A}^{*}}(x) \right]$. The amount of uncertainty in the T2FS is determined by the size of the FOU. The wider the FOU, the more uncertain about the primary memberships. The FOU together with the 3rd dimension of MFs provide additional degrees of freedom such that a T2FS can directly model and handle uncertainties. For IT2FS, the secondary memberships, $\underline{\mu}_{\tilde{A}}(x,u)$ and $\overline{\mu}_{\tilde{A}}(x,u) = 1$. A fuzzy logic system that uses IT2FS is known as interval type-2 fuzzy logic system (IT2FLS).

Figure 1 is the IT2FLS comprising inference engine, rule-base and output processing block. The output processing block comprises the type reducer that reduces the T2FS to type-1 FS and the defuzzifier that translates the generated type-1 FS to final crisp value as the output. For simplicity, singleton fuzzification is adopted.
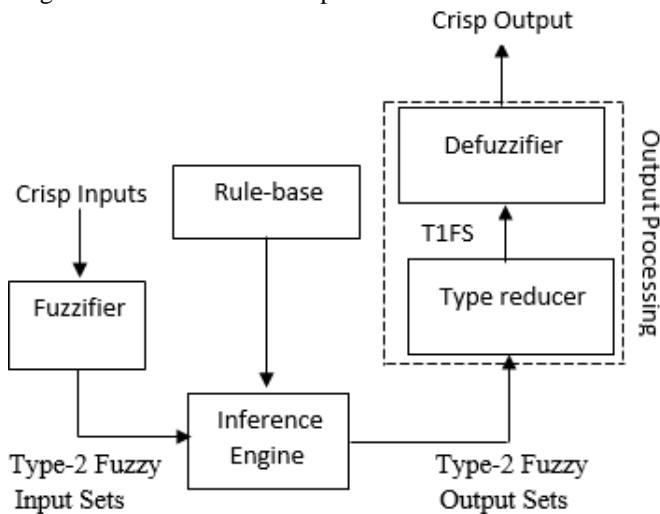

Figure 1: Interval type-2 fuzzy logic system (IT2FLS)

The crisp external inputs are propagated into the fuzzifier to generate the LMF and UMF of IT2FS. These IT2FS input values then activate the inference engine and the rule base to produce another IT2FSs as the outputs which are then reduced to type-1 FSs and further defuzzified to produce a crisp output. The rules of an IT2FLS are produced in the same way as the type-1 FLS. The only difference is the use of IT2FS in the rule base as against type-1 FS. The method of inferencing may be Mamdani or TSK. These two differ in their final output computation. While Mamdani entails type-reduction procedure which is computationally intensive, TSK inference uses weighted average to obtain the final output which is more cost effective and works well with optimization problems. Moreover, the TKS-based inference is known to accurately model an input-output relationship better [23]. For these reasons, the TSK inference is adopted in this study with the aim of obtaining an estimate that closely agrees

with the actual output of the system. The IT2FLS-TSK rule can be represented as in Eq. (2).

R_k : IF x_i is $\tilde{A}_{ik}$ and ... and x_n is $\tilde{A}_{nk}$ THEN $y_k = \sum_{i=1}^{n} w_{ik} \; x_i + b_k$ (2)

where $\tilde{A}_{ik}$'s are IT2FSs, $x_i's$ (i = 1, ..., n) are inputs, y_k represents the output for each of the rule, w and b are the weight and bias respectively. As earlier stated, the outputs of a TSK FLS are computed directly as linear combination of the inputs and so does not require type-reduction which is computationally intensive. For the final output of IT2FLS-TSK, the Begian-Melek-Mendel (BMM) [24] approach is utilized which is computed as in Eq. (3):

$$y = (1-\beta) \frac{\sum_{k=1}^{M} \underline{f}_k y_k}{\sum_{k=1}^{M} \underline{f}_k} + \beta \frac{\sum_{k=1}^{M} \overline{f}_k y_k}{\sum_{k=1}^{M} \overline{f}_k} \qquad (3)$$

where M represents the number of rules, $\underline{f}_k$ and $\overline{f}_k$ are the firing strength for the lower and upper MFs and are computed as in Eq. (4) and (5) using product t-norm.

$$\underline{f}_k = \prod_{i=1}^{n} \underline{\mu}_{\tilde{A}_{ik}}(x_i) \qquad (4)$$

$$\overline{f}_k = \prod_{i=1}^{n} \overline{\mu}_{\tilde{A}_{ik}}(x_i) \qquad (5)$$

and β is a user defined parameter that weighs the contributions of the LMF and UMF to the final output. Although there are many MFs in the literature such as triangular, trapezoidal, Gaussian, pi, and more, the MF adopted in this work is the Gaussian MF. Apart from the smoothness of the Gaussian MF, it is also continuous at all points. Furthermore, obtaining the partial derivatives of the MF parameters is easier and straightforward especially in GD-based optimization problems [25] undertaken in this study. In other words, Gaussian function is very useful for FLS's optimization problems. The Gaussian MF is defined as follows:

$$\mu_{ik}(x_i) = \left[exp \left(-\frac{(x_i + m_{ik})^2}{2\sigma_{ik}^2} \right) \right] \quad \sigma_{ik} \in [\sigma_{i1}, \sigma_{i2}] \qquad (6)$$

where x_i, (i = 1, ..., n) is the input, m is the mean of Gaussian, σ is the standard deviation.

The uncertainty about a Gaussian MF may be associated with either the mean or the standard deviation of the MF. Thus, a Gaussian MF comes in two flavours namely: Figure 2a which shows Gaussian with uncertain mean and Figure 2b which shows Gaussian with uncertain standard deviation. Here, and for this optimization problem, the Gaussian MF with uncertain standard deviation is very appropriate because it is one MF that is differentiable at all points [26].
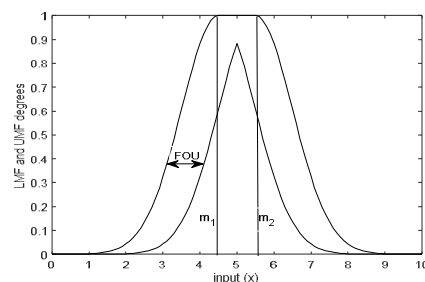

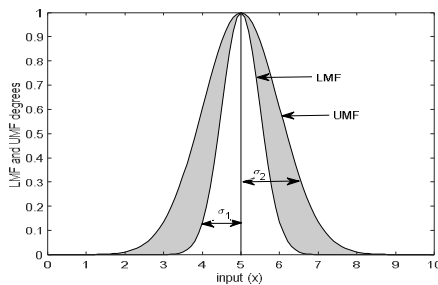**Figure 2 (a):** IT2FS with uncertain mean

Figure 2. (b): IT2FS with uncertain standard deviation

4. PARAMETER UPDATE RULE

The parameters of the IT2FLS-TSK are updated using the gradient descent (GD) back propagation, the most popular optimization algorithm for FLSs [26]. Optimization involves adjusting the MF parameters of the IT2FLS-TSK in order to obtain as accurate an estimate as possible under uncertain conditions.

The cost function for a single output is defined as:

$$E = \frac{1}{2}(y^r - y)^2 \tag{7}$$

where y^r is the real output and y is the predicted output. For a TSK-based IT2FLS, the parameters to be updated include the mean, m, standard deviation, σ, weight, w, bias, b and the user defined parameter, β. The parameters of the comparative models (T1FLS and ANN) are also updated using GD-backpropagation algorithm.

The generic GD update rule for any parameter, θ, is as follows:

$$\theta_{ik}(t + 1) = \theta_{ik}(t) - \gamma \frac{\delta E}{\delta \theta_{ik}} \tag{8}$$

where γ is the learning rate and θ represents m, σ, w, b, and β.

The update for each parameter is given as:

$$m_{ik}(t + 1) = m_{ik}(t) - \gamma \frac{\delta E}{\delta m_{ik}} \tag{9}$$

$$\sigma_{ik}(t + 1) = \sigma_{ik}(t) - \gamma \frac{\delta E}{\delta \sigma_{ik}} \tag{10}$$

$$w_{ik}(t + 1) = w_{ik}(t) - \gamma \frac{\delta E}{\delta w_{ik}} \tag{11}$$

$$b_k(t + 1) = b_k(t) - \gamma \frac{\delta E}{\delta b_k} \tag{12}$$

The learning rate must be carefully chosen to avoid slow learning or instability in the learning process. The derivatives in Eq. (9) to (12) are computed as follows:

$$\frac{\delta E}{\delta m_{ik}} = \sum_1^M \frac{\delta E}{\delta y} \left(\frac{\delta y}{\delta \underline{f}_k} \frac{\underline{f}_k}{\underline{\mu}_{ik}} \frac{\underline{\mu}_{ik}}{m_{ik}} + \frac{\delta y}{\delta \overline{f}_k} \frac{\overline{f}_k}{\overline{\mu}_{ik}} \frac{\overline{\mu}_{ik}}{m_{ik}} \right) \tag{13}$$

$$\frac{\delta E}{\delta \sigma_{ik}} = \sum_1^M \frac{\delta E}{\delta y} \left(\frac{\delta y}{\delta \underline{f}_k} \frac{\underline{f}_k}{\underline{\mu}_{ik}} \frac{\underline{\mu}_{ik}}{\sigma_{1,ik}} + \frac{\delta y}{\delta \overline{f}_k} \frac{\overline{f}_k}{\overline{\mu}_{ik}} \frac{\overline{\mu}_{ik}}{\sigma_{2,ik}} \right) \tag{14}$$

$$\frac{\delta E}{\delta w_{ik}} = \frac{\delta E}{\delta y} \frac{\delta y}{\delta y_k} \frac{\delta y_k}{\delta w_{ik}} \tag{15}$$

$$\frac{\delta E}{\delta b_k} = \frac{\delta E}{\delta y} \frac{\delta y}{\delta y_k} \frac{\delta y_k}{\delta b_k} \tag{16}$$

The user-defined parameter β is updated the same way as in Eq. (17).

$$\beta(t + 1) = \beta(t) - \gamma \frac{\delta E}{\delta \beta} \tag{17}$$

5. DATASET DESCRIPTION

The COVID-19 cases datasets are obtained from Kaggle (https://www.kaggle.com/datasets), a publicly available data repository. The Kaggle website contains COVID-19 cases for all African countries collected from February 15, to June 24, 2020. In this analysis, two African countries namely: South Africa and Egypt are considered, being the most affected countries in the region. Figure 3 shows the trends of COVID-19 cases in South Africa while Figure 4 shows the trends of COVID-19 cases in Egypt respectively. There are five recorded cases of COVID-19 namely: daily cases (confirmed cases), daily deaths, active cases, total cases and total deaths. The two countries exhibit exponential growth in the total and active cases of COVID-19 while the other cases are very insignificant. The different COVID-19 cases are modelled as time series using the current value and four previous values for each case to predict the next case (one-step ahead). The input-output generating vector is given as: $\{x(t), x(t - 4), x(t - 3), x(t - 2), x(t - 1), x(t + 1)\}$, where $x(t + 1)$ is the output. Each COVID-19 case data is normalized to a range of 0 and 1 so that larger values do not overwhelm the smaller values. For any given COVID-19 case, $X = \{x_1, x_2, x_3, ..., x_n\}$, the data elements are normalized as in Eq. (18).

$$x_{i\ new} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{18}$$

Normalization helps to improve the convergence speed of the prediction model. Table 1 shows the statistical characteristics of the different COVID-19 cases for South Africa and Egypt regions. Data from February 15, to May 16, 2020 are used for training while data from May 17, to June 24, 2020 are used for testing. For each case, 10 simulation runs are carried out with 100 epochs in each run. The IT2FLS for COVID-19 outbreak prediction is implemented using MATLAB© 2020a, running on Windows 8.

Table 1: COVID-19 Cases Characteristics for South Africa and Egypt

	South Africa		Egypt	
	mean	std	mean	std
Daily cases	859.969	1316.982	458.154	554.14
Daily deaths	16.954	25.847	18.846	24.657
Active cases	8009.2	11875.623	8109.608	11404.191
Total cases	16990.223	27004.076	11641.862	16378.535
Total deaths	343.423	558.961	493.785	614.499

6. EXPERIMENTAL ANALYSIS

The performance indexes adopted in this study are the root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

$$RMSE = \sqrt{\frac{1}{T}\sum_{1}^{T}(y^r - y)^2} \qquad (19)$$

$$MAE = \frac{1}{T}\sum_{1}^{T}|y^r - y| \qquad (20)$$

$$MAPE = \frac{1}{T}\sum_{1}^{T}\frac{|y^r - y|}{y^r}*100 \qquad (21)$$

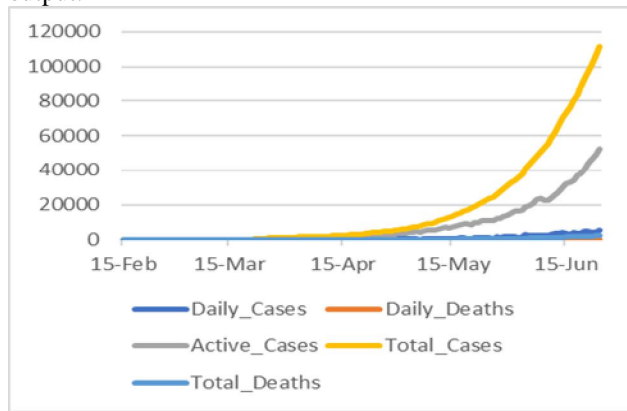where y^r is the real output and y is the model predicted output.

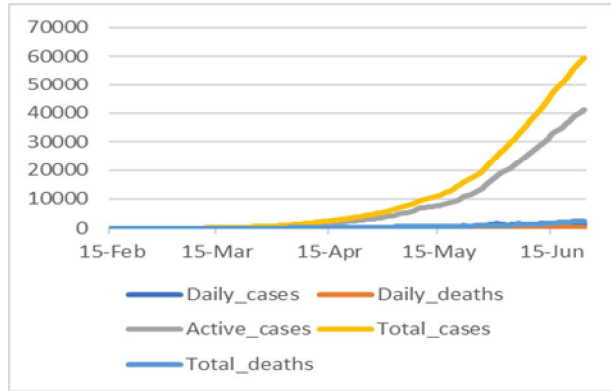**Figure 3:** COVID-19 cases in South Africa

Figure 4: COVID-19 cases in Egypt

Figures 5 and 6 show the actual and predicted daily cases and daily deaths of COVID-19 outbreak in South Africa using type-1 FLS, ANN and IT2FLS while Figures 7 and 8 show the actual and predicted daily cases and daily deaths in Egypt. Tables 2 and 3 show a one-day ahead prediction of daily cases in South Africa and Egypt for 10 days using type-1 FLS, ANN and IT2FLS and their corresponding absolute prediction errors. As shown in the tables 2 and 3, IT2FLS has the lowest absolute prediction errors and thus exhibits good performance compared to type-1 FLS and ANN.
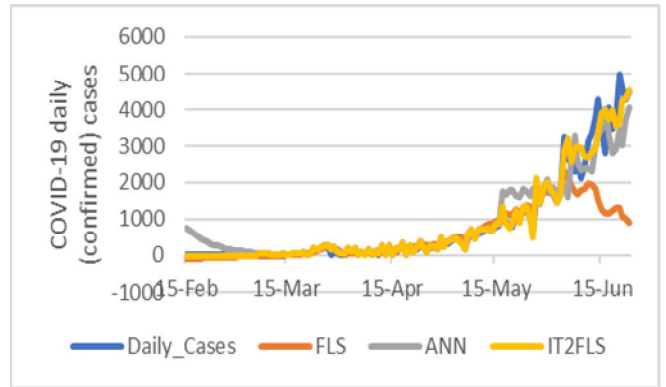

Figure 5: Actual and predicted South Africa COVID-19 daily cases using FLS, ANN and IT2FLS
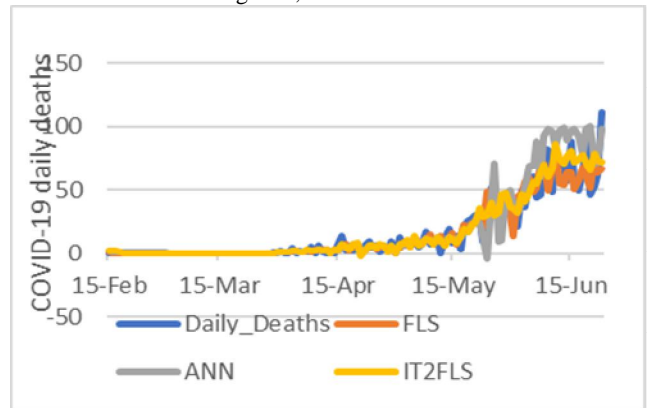

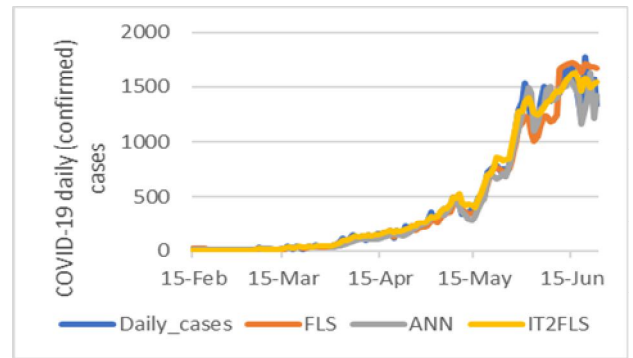**Figure 6:** Actual and predicted daily deaths cases in South Africa using FLS, ANN and IT2FLS

Figure 7: Actual and predicted COVID-19 daily cases in Egypt using FLS, ANN and IT2FLS
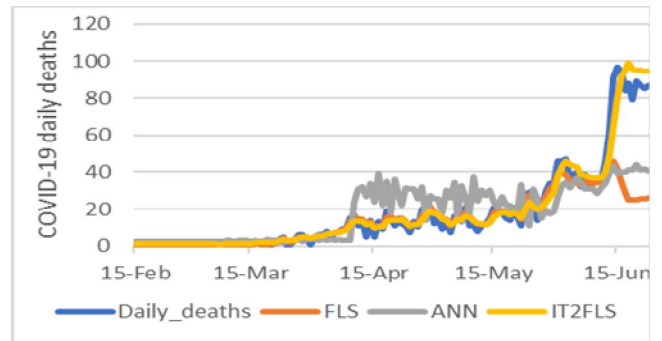

Figure 8: Actual and predicted daily deaths cases in Egypt using FLS, ANN and IT2FLS

Table 2: Performance Comparison of Actual and Predicted Values of Daily Cases of COVID-19 in South Africa using FLS, ANN and IT2FLS Models with their Corresponding Absolute Prediction Errors.

Day	Actual daily case	FLS predicted case	ANN predicted case	IT2FLS predicted case	FLS absolute error	ANN absolute error	IT2FLS absolute error
1	4302	1531.7	3290.1	3355.5	2770.3	1011.9	946.5
2	3495	1255.4	3475.1	3907.7	2239.6	19.9	412.7
3	2801	1144.7	4015.3	3991	1656.3	1214.3	1190
4	4078	1167	3313.5	3739	2911	764.5	339
5	3478	1242.8	2809.1	3987.4	2235.2	668.9	509.4
6	3825	1304.2	2935.3	3546.2	2520.8	889.7	278.8
7	4966	1282.7	3483	3665.3	3683.3	1483	1300.7
8	4621	1078	3028.3	4232.5	3543	1592.7	388.5
9	4288	993.7	3746.6	4304.2	3294.3	541.4	16.2
10	4518	878.2	4066.5	4544.9	3639.8	451.5	26.9
Average absolute error					2849.36	868.78	540.87

Table 3: Comparison of Actual and Predicted Values of Daily Cases in Egypt using FLS, ANN and IT2FLS with their Corresponding Absolute Prediction Errors.

Day	Actual daily case	FLS predicted case	ANN predicted case	IT2FLS predicted case	FLS absolute error	ANN absolute error	IT2FLS absolute error
14/06/2020	1618	1712.5	1582	1585.4	94.5	36	32.6
15/06/2020	1691	1724.1	1528.1	1624.7	33.1	162.9	66.3
16/06/2020	1567	1716.4	1488.8	1630.5	149.4	78.2	63.5
17/06/2020	1363	1681	1421.5	1564.6	318	58.5	201.6
18/06/2020	1218	1638.8	1164.8	1460.5	420.8	53.2	242.5
19/06/2020	1774	1715.2	1303.3	1582.5	58.8	470.7	191.5
20/06/2020	1547	1707	1411.2	1547.7	160	135.8	0.7
21/06/2020	1475	1687.2	1633.2	1495.6	212.2	158.2	20.6
22/06/2020	1576	1689.9	1218.3	1521.5	113.9	357.7	54.5
23/06/2020	1332	1673.8	1422.4	1537.6	341.8	90.4	205.6
Average absolute error					190.25	160.16	107.94

Table 4: Comparison of Prediction Accuracy of COVID-19 Cases using FLS, ANN and IT2FLS in South Africa and Egypt.

Country	COVID-19 cases	FLS			ANN			IT2FLS		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
South Africa	Daily cases	965.45	781.50	19.89	804.54	640.80	15.46	723.79	556.65	14.68
	Daily death	22.92	21.26	34.38	25.80	19.15	33.06	20.98	18.53	29.12
	Active cases	9558.20	7939.58	20.66	9511.4	8226.3	25.08	6668.03	5682.13	17.57
	Total cases	17625.54	14921.4	17.32	24539	19940	21.14	14581.6	12153.43	16.40
	Total death	274.05	234.62	13.53	276.23	228.19	12.58	219.20	184.73	10.81
Egypt	Daily cases	229.87	188.33	13.34	197.33	154.52	10.94	196.82	151.72	10.86
	Daily death	23.97	23.07	27.96	27.80	23.88	29.02	21.26	19.63	23.69
	Active cases	4374.69	3535.30	10.99	6137.3	5068.7	13.67	4174.19	3547.26	10.71
	Total cases	6324.50	5137.59	11.09	7952.7	6516.7	12.24	6022.89	4841.49	10.43
	Total death	316.57	274.13	15.17	524.5	432.65	20.70	307.64	258.47	14.77

Table 4 summarizes the performances of the different models (FLS, ANN and IT2FLS) in terms of RMSE, MAE and MAPE for the five COVID-19 cases in South Africa and Egypt respectively. As shown in the table 4, IT2FLS also exhibits good performance overall with reduced RMSE, MAE and MAPE.

Also shown in the table 4 is that ANN outperforms type-1 FLS prediction in the daily cases prediction with accuracy very close to IT2FLS. In the overall, IT2FLS achieves greater prediction accuracy with reduced RMSE, MAE and MAPE compared to type-1 FLS and ANN. Generally, the FL approaches outperform the ANN in most of the COVID-19 prediction cases.
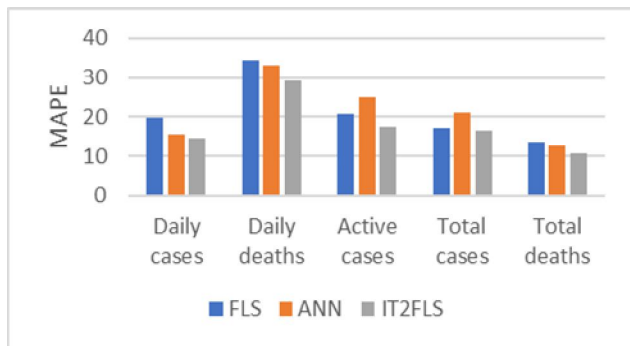

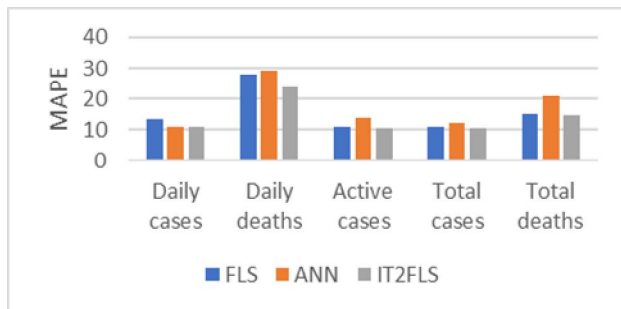**Figure 9**: Comparison of MAPE using FLS, ANN and IT2FLS for South Africa COVID-19 cases.

Figure 10: Comparison of MAPE using FLS, ANN and IT2FLS for Egypt COVID-19 cases.

Figures 9 and 10 show the MAPE of FLS, ANN and IT2FLS for the different COVID-19 cases (daily cases, daily deaths, active cases, total cases and total deaths) in South Africa and Egypt respectively. As shown in the figures 9 and 10, IT2FLS provides good performance overall in terms of MAPE compared to FLS and ANN.

7. CONCLUSION

This paper presents an IT2FLS-TSK approach for the prediction of COVID-19 outbreak in two African countries namely South Africa and Egypt. The COVID-19 outbreak has been shown to possess high levels of uncertainty that standard methods and classical type-1 FLS may struggle with. IT2FLS is known to handle uncertainty better and has been used in

this study to predict COVID-19 pandemic cases in South Africa and Egypt well. As shown from the experimental analysis, IT2FLSs with LMFs and UMFs perform better than the type-1 FLS with single (precise) MFs. The IT2FLS also outperforms the feedforward neural network in the prediction of COVID-19 cases. In the future, we intend to adopt other fuzzy approaches such as intuitionistic fuzzy logic of type-2 and hybrid learning methods such as particle swarm optimization (PSO) plus salp swarm algorithm (SSA), flower pollination algorithm (FPA) plus ant colony optimization for the parameter tuning of the IT2FLS model. Moreover, we intend to explore the causative parameters of the COVID-19 outbreak.

REFERENCES

1. P. Zhou, X. Yang, Wang et al., **A pneumonia outbreak associated with a new coronavirus of probable bat origin.** *Nature,* vol. 579, no. 7798, 2020, pp. 270–273.

2. P. Wang, X. Zheng, J. Li, and R. Zhu. **Prediction of Epidemic Trends in COVID-19 with Logistic Model and Machine Learning Technigues**. *Chaos, Solitons and Fractals*, 2020, doi: https://doi.org/10.1016/j.chaos.2020.110058

3. M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, et al., **Preparedness and Vulnerability of African Countries against Importations of COVID-19: A Modelling Study**. *The Lancet*, Volume 395, Issue 10227, 2020, pp. 871 - 877. DOI: 10.1016/s0140-6736(20)30411-6.

4. K. Liu, Y. Chen, R. Linand, K. Han. **Clinical Features of covid-19 in Elderly Patients: A Comparison with Young and Middle-Aged Patients.** *Journal of Infection*, ;80(6):e14-e18. doi:10.1016/j.jinf.2020.03.0052020.

5. S. Tuli, S. Tuli, r. Tuli, and S. Gill. **Predicting the Growth and Trend of COVID-19 Pandemic using Machine Learning and Cloud Computing**. *Internet of Things,* 2020, doi: https://doi.org/10.1016/j.iot.2020.100222.

6. P. Arora, H. Kumar, and B. Panigrahi. **Prediction and Analysis of COVID-19 Positive Cases using Deep Learning Models: A Descriptive Case Study of India.** *Chaos, Spolitons and Fractals,* 139, 2020, pp. 1-9.

7. S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, P. M. Atkinson. **COVID-19 Outbreak Prediction with Machine Learning**. Preprints 2020, 2020040311 (doi: 10.20944/preprints202004.0311.v1).

8. M. Nemati, J. Ansang, and N. Nemati. **Machine Learning Approaches in COVID-19 Survival Analysis and Discharge Time Likelihood Prediction using Clinical Data Pattern**, 2020, doi:https://doi.org/10:1016/j.patter.2020.100074.

9. F. Petropoulos, and S. Makridakis, **Forcasting the Novel Coronavirus COVID-19.** *Plus One*, 15(3):eo231236, 2020, https://doi.org/101371/journal.pone.0231236.

10. A. Al-qaness, A. Ewees, H. Fan and M. El Aziz. **Optimization Method for Forecasting Confirmed Cases of COVID-19 in China.** *Journal of Classical Medicine*, 9, 674, 2020, doi:10.3390/jcm9030674.

11. Z. Car, S. Segota, N. Andelic, I. Lorencin and V. Mrzljak. **Modeling the Spread of COVID-19 Infection using Multilayer Perceptron.** *Computational and mathematical Methods in Medicine*, 2020, doi:10.1155/2020/5714714

12. **L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu.** Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model**. *IEEE Access***, **8***,* **2020, 51761-51769, doi:10.1109/ACCESS.2020.2979599.**

13. C. Anastassopoulou, L. Russo, A. Tsakris and C. Siettos. **Data-based analysis, modelling and forecasting of the COVID-19 outbreak**. *PLoS ONE,* 15(3): 2020, e0230405. https://doi.org/10.1371/journal.pone.0230405

14. D. Painuli, D. Mishra, S. Bhardwaj and M. Aggarwal. **Fuzzy Rule Based System to Predict COVID-19: A Deadly Virus**. *International Journal of Management and Humanities.* 4(8), 2020, 78 – 82.

15. P. Melin, J. C. Monica, D. Sanchez and O. Castillo. **Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: The case of Mexico,** *In Healthcare*, Vol. 8, No. 2, p. 181, Multidisciplinary Digital Publishing Institute, 2020. Vol. 29, No. 12s, 2020, pp. 2179-2183.

16. A. Alsayed, H. Sadir, R. Kamil, and H. Sari, **Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia**, *International Journal of Environmenta. Research and. Public Health,* 17, 4076; doi:10.3390/ijerph17114076

17. N. Dhiman and M. K. Sharma. **Fuzzy logic inference system for identification and prevention of coronavirus (COVID-19),** *International Journal of Innovative Technology and Exploring Engineering* (IJITEE), ISSN: Volume-9 Issue-6, April 2020, 1575-1580, *DOI:10.35940/ijitee.F4642.049620.*

18. S. J. Fong, G. Li, N. Dey, R. G. Crespo and E. Herrera-Viedma. **Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction**, *Applied Soft Computing Journal.* 93, 2020, 106282.

19. S. A. Fatima, N. Hussain, A. Balouch, I. Rustam, M. Saleem, M. Asif. **IoT enabled smart monitoring of coronavirus empowered with fuzzy inference system**, *International Journal of Advance Research, Ideas and Innovations in Technology,* Vol. 6, issue 1, 2020, 188-194.

20. P. Verma, M. Khetan, S. Dwivedi and S. Dixit. **Forecasting the Covid-19 outbreak: An application of arima and fuzzy time series models,** 2020.

21. N. V. Tinh. **Forecasting of COVID-19 confirmed cases in Vietnam using fuzzy time series model combined with particle swarm optimization**, *Computational Research Progress in Applied Science & Engineering* (CRPASE), Vol. 06(02), 114-120, June 2020

22. L. A. Zadeh. **The concept of a linguistic variable and its application approximate reasoning-i,"** *Information Sciences*, vol. 8, pp.199–249, 1975.

23. J. Casillas, O. Cord´on, F. H. Triguero and L. Magdalena. **Accuracy improvements in linguistic fuzzy modeling**, 129. Springer, 2013.

24. M. B. Begian, W. W. Melek and J. M. Mendel. **Parametric design of stable type-2 tsk fuzzy systems**, in IEEE Annual Meeting of the North American Fuzzy Information Processing Society, (NAFIPS), 2008, pp. 1-6.

25. J. M. Mendel. **Explaining the performance potential of rule-based fuzzy systems as a greater sculpting of the state space**, IEEE Transactions on Fuzzy Systems, Vol. 26, Issue 4, 2017, 2362-2373. DOI:10.1109/TFUZZ.2017.2774190.

26. E. Kayacan, and M. Khanesar. **Fuzzy Neural Networks for Real Time Control Applications: Concepts, Modeling and Algorithms for Fast Learning**, 2015, (pp. 89), Elsevier Science.