Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India

Niteesh Kumar, Harendra Kumar

PII: S0019-0578(21)00370-0
DOI: https://doi.org/10.1016/j.isatra.2021.07.003
Reference: ISATRA 4192

To appear in: ISA Transactions

Received date: 30 July 2020
Revised date: 1 July 2021
Accepted date: 1 July 2021

Please cite this article as: N. Kumar and H. Kumar, A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India. ISA Transactions (2021), doi: https://doi.org/10.1016/j.isatra.2021.07.003.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 ISA. Published by Elsevier Ltd. All rights reserved.
A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India

Niteesh Kumar\textsuperscript{1} & Harendra Kumar\textsuperscript{2}

\textsuperscript{1,2}Department of Mathematics and Statistics
Gurukula Kangri (Deemed to be University), Haridwar 249404, Uttarakhand, India.
Email: niteeshkumar963@gmail.com, balyan.kumar@gmail.com

Corresponding author: Harendra Kumar
Short Title of the Article

Highlights

The highlights of presented manuscript are summarized below:

- A hybrid fuzzy time series based model is proposed.
- FCM clustering technique is modified by using an exponential function to tolerate noisy data.
- The present model is applicable to predict approximate COVID-19 infected cases as well as deaths for trained and untrained data.
- The model presents prediction of COVID-19 infected cases and deaths for next 31 days till 31 August 2020.
A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India

ARTICLE INFO

Keywords
COVID-19
fuzzy C-means
fuzzy time series
clustering
pandemic

Abstract

World is facing stress due to unpredicted pandemic of novel COVID-19. Daily growing magnitude of confirmed cases of COVID-19 put the whole world humanity at high risk and it has made a pressure on health professionals to get rid of it as soon as possible. So, it becomes necessary to predict the number of upcoming cases in future for the preparation of future plan-of-action and medical set-ups. The present manuscript proposed a hybrid fuzzy time series model for the prediction of upcoming COVID-19 infected cases and deaths in India by using modified fuzzy C-means clustering technique. Proposed model has two phases. In phase-I, modified fuzzy C-means clustering technique is used to form basic intervals with the help of clusters centroid while in phase-II, these intervals are upgraded to form sub-intervals. The proposed model is tested against available COVID-19 data for the measurement of its performance based on mean square error, root mean square error and average forecasting error rate. The novelty of the proposed model lies in the prediction of COVID-19 infected cases and deaths for next coming 31 days. Beside of this, estimation for the approximate number of isolation beds and ICU required has been carried out. The projection of the present model is to provide a base for the decision makers for making protection plan during COVID-19 pandemic.

1. Introduction

Currently, a novel virus from a family of coronaviruses spread all over the world named as COVID-19 by world health organization (WHO). The number of infected COVID-19 cases has a rapid growth in most of the countries from December 2019 onward. That’s why WHO declared it as a pandemic. It is assumed that the spreading of this virus was originated from a city of China called Wuhan in mid-December 2019. On 30 January 2020, the first case of COVID-19 was reported in India. After the first case, COVID-19 shows rapid growth in almost all parts of India within a few months. As a result, the number of COVID-19 infected cases and deaths per month has increased. It shows that there is a drastic change in the growth of the novel coronavirus. Unfortunately, the vaccine for the preservation of COVID-19 is not invented yet in India and rest of the world. Thus, the mobility rate of COVID-19 at a peak in India. Hence, it becomes necessary to predict the number of upcoming COVID-19 infected cases and deaths in future for the preparation of future plan-of-action and medical set-ups in advance. For forecasting purposes, several mathematical models have been developed in the past few months.

Ded and Majumdar [17] proposed model named as time series model for the analysis of incidence trends and also obtained reproductive number of COVID-19. Mandal et al. [31] presented an analysis of COVID-19 through restriction on travels from other countries and the impact of quarantine in India by using simple mathematical model. Bhardwaj [8] proposed regression based predictive model for the prediction of total infection that arises due to COVID-19 at the end of outbreak. Bhola et al. [9] discussed the predictive model for future projection of this pandemic in India. Chatterjee et al. [12] proposed a compartmental SEIR model for obtaining the impact on healthcare and number of infections due to the COVID-19 epidemic in India. Chakarborty and Ghosh [11] build a real time forecasting method for the prediction of future COVID-19 cases and obtained case fatality rates for different countries. Roy and Kar [40] presented a study of impact of natural causes like summer and humidity on COVID-19 by analysing the presently available data. Mondal and Ghosh [33] have focused on the analysis of exponential growth of COVID-19 cases in India with respect to other countries and made future sketching of it. Mandal et al. [50] have proposed a mathematical model by introducing quarantine class and other measures introduced by the government to control COVID-19 spreading and forecast trends of it in three states of India. Marimuthu et al. [32] have estimated the number of infected cases of COVID-19 in Delhi by using SEIR model. Scheiner et al. [47] have developed a mathematical model called death kinetic law by using SEIR model and compared with another approach infection-to-death delay rule.
El Koufi et al. [19] presented SIR epidemic model to stochastic from a deterministic frame with non-linear incidence and vertical rates. Al-qaess et al. [2] proposed a forecasting model to estimate confirmed cases of coronavirus in China based on previous cases. Labwani et al. [27] proposed 3-phase SIRD model for determining the lockdown period for highly affected regions and predict lockdown period for India and Italy. Patrikar et al. [36] have presented modified SEIR to forecast the estimated cases of novel coronavirus in India. Salgotra et al. [44] developed genetic programming-based prediction model to estimate confirmed cases and death of COVID-19 across three states and India as well. Pai et al. [34] predict the number of active cases of COVID-19 in India by knowing the impact of lockdown along with the inflation of active cases. Kuniya [26] has applied the SEIR compartmental model for predicting the of epidemic peak of COVID-19 in Japan. Tiwari et al. [53] built time series forecasting model to predict the epidemic of COVID-19 in India. Tomar and Gupta [54] proposed data driven model & predict the active cases in India for upcoming days and measure the impact of lockdown on COVID-19. Krishna and Prakash [24] presented a model to estimate the mobility of COVID-19. Sarkar et al. [46] presented SARIHqSq model to forecast the scenario of COVID-19 in India. Malavika et al. [29] presented short term prediction by using Logistic growth model and also used SIR model to forecast peak time, active cases and the effect of lockdown in India. Kaushik et al. [23] gave review on different terminology due to COVID-19 in India such as clinical presentation, treatment, virology etc.

Ambikapathy and Krishnamurthy [4] have proposed a model for knowing the impact of lockdown on COVID-19 spreading in India. Acharya and Porwal [1] present an ecological study by using a percentile ranking method for obtaining specific domain, overall vulnerability and present the number of active cases of coronavirus in 9 districts of India. Cooper et al. [15] proposed SIR model to examine the effectiveness of COVID-19 spreading in the community. Singhal et al. [50] have proposed two models, mathematical model with and without parameter, for investigating the trend of COVID-19 and gave some prediction for the upcoming days. Ranjan [39] analysed the outbreak of COVID-19 in India by using the epidemiological model for long-term and short-term predictions. Poddar et al. [37] proposed a model to study the spreading rate, death rate, recovery rate of COVID-19 and study the prediction of it in India. Vasantha and Patil [55] presented an overview of development of different models on COVID-19 in India. Pandey et al. [35] used regression and SEIR model to forecast the outbreak of COVID-19 in India. Alkahtani and Alraid [3] presented a model of COVID-19 based on fractional differential operator and numerical method by using Lagrange polynomial for solving the system equation. Arora et al. [5] presented a deep learning-based models for forecasting the number of positive cases of COVID-19 for union territories and 32 states of India. Sahoo and Sapra [43] have gave data driven epidemic model for the analysis of COVID-19 in India based on real data of COVID-19. Çakan [10] presented SEIR epidemic model by considering impact of health and analysed the global & local stability of this model. Giri et al. [21] have proposed neural network model by introducing the lockdown condition for showing the infection risk of COVID-19.

Clustering is an approach in which a grouped data are separated into a smaller data groups based on similarity measures. In the literature, many researchers used clustering technique with fuzzy time series (FTS) for forecasting purposes. Song and Chissom [51, 52] have proposed a time series model based on fuzzy set theory. Huang [22] proved that accuracy of the forecasting model can be improved by changing the length of intervals. Li et al. [28] have developed a forecasting algorithm for FTS based on fuzzy C-means (FCM) clustering. Qui et al. [38] proposed high order FTS model based on fuzzy logical relationship (FLR) and automatic clustering. Sang et al. [45] presented a forecasting FTS model based on IFCM clustering technique. Kumar et al. [25] proposed two distance metrics and developed two clustering algorithms AMFCM and EMFCM. Zhang et al. [60] developed FTS forecasting model based on time series clustering and multiple linear regressions. While the FCM clustering technique uses Euclidean distance (ED) which gets easily stuck in a noisy environment and doesn’t obtain good results. So, in the present manuscript basic FCM is modified by using an exponential function to tolerate noisy data before using with FTS technique. Along with this, most researchers generally are using mathematical modeling, like SEIR and SIR techniques, instead of using soft computing techniques for the prediction of COVID-19. Some of these models will forecast the effect of COVID-19 in upcoming weeks or days with more error rate. This gave us an encouragement to think out of the box and developed a novel hybrid fuzzy time series model (ANHTFS) which is based on modified FCM clustering technique for prediction of COVID-19 infected cases and deaths in India for coming 31 days. ANHTFS model has been used to forecast the infected cases and deaths in India. A primary reason for proposing this model is that this approach is more capable as compared to classical techniques and more durable in comparison to exist predicting models. Also, it can predict the COVID-19 infected cases and deaths for short-term and for long-term with small error. Proposed model has two phases. In phase-I, modified
fuzzy c-means (MFCM) clustering technique is used to form intervals with the help of centroid while in phase II, these intervals are upgraded to form sub-intervals and then predict the approximated cases of infection and deaths in India. The main contribution of the presented manuscript is:

1. Developed a hybrid model for forecasting by FTS technique.
2. FCM clustering technique is modified by using an exponential function to tolerate noisy data.
3. This model can predict the approximate COVID-19 infected cases and deaths for trained and untrained data in India.
4. Estimate the approximate number of isolation beds and ICU requirements till 31 August 2020 in India.
5. The proposed ANHFTS model is considered as an unsupervised learning process.

Rest of the presented manuscript is organised as follows. Some basic preliminaries of FCM and FTS are introduced in section 2. In section 3, some notation and MFCM with their necessary conditions are presented here. Section 4, described the description of the proposed model. Section 5 contains some performance measures which conclude the performance of proposed model. Section 6 reveals the implementation of proposed algorithm on two examples with the prediction of COVID-19 infected cases and deaths in India for next 31 days till 31 August 2020. Section 7 concludes this work.

2. Background information

In this section, we have discussed basic about FCM and FTS techniques which are used in our proposed model.

2.1. Fuzzy C-means clustering technique

Well known soft clustering technique, FCM partition the historical data in such a way that it can exist in more than one cluster with distinct membership value. Ruspini [41, 42] form clusters by using the concept of fuzzy set theory. Later on, Bezdek [7] improved the process of clustering after FCM formulation by Dunn [18]. The main objective of FCM is to minimize the objective function with their necessary conditions as follows.

\[
J(v_i, C_k) = \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik}^m ||v_i - C_k||^2, \quad m \in (1, \infty),
\]  

(1)

where, \(v_i\) is the \(i^{th}\) data point, \(\mu_{ik}\) represent its membership value in \(k^{th}\) cluster, \(||.||\) denotes Euclidean distance, \(c\) is number of cluster, \(C_k\) is \(k^{th}\) cluster centroid.

\[
\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{||v_i - C_k||^2}{||v_i - C_j||^2}\right)^{\frac{1}{m-1}}},
\]  

(2)

\[
C_k = \frac{\sum_{i=1}^{n} \mu_{ik}^m v_i}{\sum_{i=1}^{n} \mu_{ik}^m},
\]  

(3)

Equation (2) and (3) are necessary conditions that will minimize equation (1).

Here, \(\mu_{ik} \in [0, 1] \) & \(\sum_{k=1}^{c} \mu_{ik} = 1; \quad i = 1, 2, ..., n.\)

2.2. Some basic definitions

This section contains important definitions which are used in throughout the present manuscript. Song and Chissom defined the FTS first time. The basic definitions of FTS, time-variant & time-invariant, FLR and FLR group (FLRG) are briefly reviewed as follows.

**Definition 1 Fuzzy time series:** Let \(f_i(t)\) be a fuzzy set defined on a universe of discourse \(X(t)\), a subset of \(R\). Then the set of \(f_1(t), f_2(t), ...\) is denoted by \(F(t)\) and it is called FTS defined on \(X(t)\). [51]
Definition 2 First order model: Suppose $F(t)$ is formed by $F(t−1)\circ R(t,t−1)$, then, fuzzy relation can be expressed as

$$F(t) = F(t−1)\circ R(t,t−1),$$

where $\circ$ represents max-min composition and $R(t,t−1)$ is fuzzy relationship between $F(t−1)$ & $F(t)$.

Then, $F(t)$ is called first order model. [52]

Definition 3 Time-variant and time-invariant: If relation $R(t,t−1)$ of $F(t)$ is not depend on $t$, $R(t,t−1) = R(t_1,t_2−1)$, then $F(t)$ is known as time-invariant FTS otherwise it is called time-variant. [14]

Definition 4 Fuzzy logical relationship: If $F(t−1) = I_t$ and $F(t) = I_t$, then the relationship between $F(t−1)$ & $F(t)$ is known as FLR and can be expressed as $I_k \rightarrow I_j$, where $I_k$ and $I_j$ are previous and current state of FLR. [13]

Definition 5 Fuzzy logical relationship group: let assume that $I_k \rightarrow I_{k1}$, $I_k \rightarrow I_{k2}$, ..., $I_k \rightarrow I_{km}$ are FLR’s. Then these FLR’s can be grouped to form FLRG as $I_k \rightarrow I_{k1}, I_{k2}, ..., I_{km}$. [22]

3. Problem formulation and modified fuzzy c-means

3.1. Notation

The various notations used throughout this article are as:

| Notation | Description |
|----------|-------------|
| $v_i$ | $i^{th}$ data point |
| $\mu_{k_i}$ | membership value of $i^{th}$ data point in $k^{th}$ cluster |
| $C_k$ | $k^{th}$ cluster centroid |
| $c$ | number of clusters ($2 \leq c \leq n − 1$) |
| $m$ | fuzzy index |
| $J(v_i, C_k)$ | objective function |
| $F(t)$ | forecasted value at any time $t$ |
| $I_p$ | $p^{th}$ linguistic variable |
| $AV_p$ | $p^{th}$ actual value |
| $r$ | rate of increment or decrement |
| $C(t)$ | number of confirmed cases at any time $t$ |
| $A(t)$ | number of active cases at any time $t$ |
| $R(t)$ | number of recovered cases at any time $t$ |
| $D(t)$ | number of deaths at any time $t$ |
| MFCM | modified fuzzy C-means |
| ANHFTS | a novel hybrid fuzzy time series |

3.2. Problem statement

Now-a-days, the whole India faced a pandemic problem in the form of COVID-19. Due to this, the number of confirmed cases shows a rapid growth in India day-by-day. Therefore, it is crucial to predict the number of upcoming infected cases and deaths in India. So, health professional are prepared for that situation and able to control the pandemic situation of COVID-19. By considering this problem, a novel hybrid predictive model has been developed by using the fuzzy time series based on MFCM.

3.3. Modified fuzzy C-means

Among the clustering technique, basic FCM [7] technique is frequently used by the researcher because of its easy implementation. But in a noisy environment, it gets easily stuck and doesn’t form the desired output. So, to overcome this problem, the basic FCM has been modified by introducing a negative exponential variable for better robustness and the objective function of MFCM is given below.

$$J(v_i, C_k) = \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{k_i}^m \|v_i − C_k\|^2 \exp \left(−\frac{u}{a}\right).$$  (4)
where, \( a = \left| \sum_{i=1}^{n} v_i / n \right| \) and \( u_i = \left| \sum_{j=1}^{d} m_{ij} \right| \), \( d \) is the dimension of data and \( m_{ij} \) denote the \((ij)^{th}\) entry of \( j^{th}\) data point. The necessary conditions for the minimization of equation (4) are derived in the upcoming sub-section.

3.3.1. Derivation for cluster centroid

The minimization of objective function has necessary condition in the form of cluster centroid, it has been derived below for \( c \) clusters. Let derive it for first cluster centroid i.e. \( C_1 \).

\[
J(v, C_1) = \sum_{i=1}^{n} \mu_{1i}^m \| v_i - C_1 \|^2 \exp \left\{ -u_{1i}/a \right\}.
\]

(5)

Differentiate the equation (5) with respect to \( C_1 \) by considering other variables as constant.

\[
\sum_{i=1}^{n} \mu_{1i}^m (v_i - C_1) \exp \left\{ -u_{1i}/a \right\} = 0,
\]

(6)

\[
C_1 = \frac{\sum_{i=1}^{n} \mu_{1i}^m \exp \left\{ -u_{1i}/a \right\} (v_i)}{\sum_{i=1}^{n} \mu_{1i}^m \exp \left\{ -u_{1i}/a \right\}}.
\]

In the same way, we assumed that it can be easily obtained for \( k^{th} \) cluster. Hence, the general form of \( k^{th} \) cluster centroid for MFCM is as follows:

\[
C_k = \frac{\sum_{i=1}^{n} \mu_{ki}^m \exp \left\{ -u_{ki}/a \right\} (v_i)}{\sum_{i=1}^{n} \mu_{ki}^m \exp \left\{ -u_{ki}/a \right\}}.
\]

(7)

3.3.2. Derivation for membership value

Another necessary condition for the minimization of equation (4) is determined with respect to membership. So, the Lagrangian function for equation (4) is

\[
J_k(\lambda, \mu_k) = \sum_{i=1}^{n} \mu_{ki}^m \| v_i - C_k \|^2 \exp \left\{ -u_{ki}/a \right\} - \lambda \left( \sum_{i=1}^{c} \mu_{ki} - 1 \right).
\]

(8)

For obtaining necessary condition, differentiates the equation (7) with respect to \( \lambda \) by considering other variables as constant.

\[
\frac{dJ_k}{d\lambda} = \left( \sum_{i=1}^{c} \mu_{ki} - 1 \right) = 0.
\]

(9)

Again, differentiate the equation (8), but this time with respect to \( \mu \) and obtain the following form.

\[
\mu_{ki} = \left( \frac{\lambda}{m} \right)^{1/(m-1)} \frac{1}{\left[ \| v_i - C_k \|^2 \exp \left\{ -u_{ki}/a \right\} \right]^{1/(m-1)}}.
\]

(10)

From equation (8) and (10), we get

\[
\left( \frac{\lambda}{m} \right)^{1/(m-1)} = \left( \sum_{j=1}^{c} \frac{1}{\left[ \| v_j - C_i \|^2 \exp \left\{ -u_{ki}/a \right\} \right]^{1/(m-1)}} \right)^{(-1)}.
\]
By putting the value of \( \left( \frac{1}{m} \right) \frac{1}{2} (m-1) \) in equation (9), the general form of membership value is obtained.

\[
\mu_{x} = \frac{1}{\sum_{j=1}^{c} \left[ \frac{\left\| u_j - c \right\|^2 \exp\left(-u_j / u_\lambda \right)}{\left\| u_j - c \right\|^2 \exp\left(-u_j / u_\lambda \right)} \right]^{1/(m-1)}}.
\]

(10)

4. Description of the proposed ANHFTS model MFCM based approach

The cases of COVID-19 are growing rapidly in India, due to this obtaining the information regarding approximate number of infected cases and deaths in India become difficult. The main issues of government are how to control this disease and predict the upcoming confirmed cases of COVID-19 to prepare in advance for public health and economic decision on the basis of mathematical model.

The present manuscript addresses a novel model for the prediction of COVID-19 infected cases and deaths by using FTS technique based on MFCM clustering to prepare in advance during this ongoing pandemic problem. The present model predicts the cases in two phases:

Phase I: Form basic intervals by using cluster centroid obtained from MFCM clustering technique.

Phase II: Upgrade the basic intervals into sub-intervals to forecast more accurately and predict the upcoming infected cases and deaths in India.

Phase I

In this phase, basic intervals are formed with help of obtained cluster centroid during MFCM clustering technique. To start the process of MFCM, first we have to know about how many clusters have to be formed. To obtain the number of clusters make partition the universe of discourse \( X \) as \( X = [D_{\text{min}} - S_1, D_{\text{max}} + S_2] \), where \( S_1 \) and \( S_2 \) are randomly chosen positive numbers and \( D_{\text{min}} \) and \( D_{\text{max}} \) are the minimum and maximum value of the collected historical data set, into equal length intervals. Instead of applied basic FCM to obtain centroid, we modified the FCM to get better results. By using MFCM, the centroids are obtained by assigning membership value randomly. Now, the basic intervals \( a_1, a_2, ..., a_c \) are formed with help of obtained centroid \( C_k \); \( k = 1, 2, ..., c \).

Phase I is elaborated in the form of algorithm A with involved steps.
Algorithm A  
To make basic interval by using MFCM clustering technique

Step 1: Partition the universe of discourse $X$ into several equal length intervals for obtaining the numbers of clusters $c$.

Step 2: Input $m$ and fuzzy stopping criteria $\varepsilon$, here $\varepsilon = 0.0001$; $m = 2$.

Step 3: Randomly initialize membership value for each historical data s.t. 
\[ \sum_{i=1}^{c} \mu_{ij}^{(0)} = 1 \text{, where } i \text{ is number of iterations}. \]

Step 4: Calculate the value of $a$ and $u_i$ by following formula 
\[ a = \left( \sum_{i=1}^{n} y_i / n \right) \text{ and } u_i = \left| \sum_{j=1}^{m} m_{ij} \right|. \]

Step 5: Calculate the cluster centroid by 
\[ c_k = \frac{\sum_{i=1}^{n} \mu_{ij}^{(0)} \exp\left(-\frac{u_i}{a}\right)}{\sum_{i=1}^{n} \mu_{ij}^{(0)} \exp\left(-\frac{u_i}{a}\right)} . \]

Step 6: Update the membership value by 
\[ \mu_{ij}^{(t+1)} = \frac{\| x_j - c_k \| \exp\left(-\frac{u_i}{a}\right)}{\sum_{i=1}^{n} \| x_j - c_k \| \exp\left(-\frac{u_i}{a}\right)} \].

Step 7: If $\| \mu_{ij}^{(t+1)} - \mu_{ij}^{(t)} \| \leq \varepsilon$, then go to next step, otherwise go to step 5.

Step 8: Calculate the basic intervals with the help of centroid by following steps:

Step 8.1: 
\[ UB_k = \frac{C_k + C_{k-1}}{2}; \]

Step 8.2: 
\[ LB_k = UB_k; \] where UB and LB are upper bound and lower bound respectively.

Step 9: 
\[ UB = 2C_1 - LB; \]

Phase II

In this phase, the basic intervals are upgraded by FTS technique to forecast more accurate value. Initially, the total number of elements $y_1, y_2, ..., y_n$ is obtained from the set of historical data which belong to their respective basic intervals i.e. $a_1, a_2, ..., a_c$. Now, partition the interval $a^c$ into $y_1$ sub-interval with equal length. Repeat this process until all basic intervals partitioned into sub-intervals. Then, select only those sub-intervals in which historical data belong and referred them as $b_i, i = 1, 2, ..., n$. Now, define the linguistic variable for each intervals and allocate them to the historical data to fuzzified it as $I_1, I_2, ..., I_n; n \in N$. The fuzzified time series data set can be defuzzified by using the mid points of upgraded interval for those who’s FLR are non-empty is given below

\[ F(t) = \begin{cases} 
\frac{1.5m_i + m_j}{m_j + 0.5m_i} & \text{if } j = 1 \\
\frac{2m_i + m_j + m_{j+1}}{2m_i + m_j + m_{j+1} + 0.5m_i + 0.5m_j} & \text{if } 2 \leq j \leq n - 1 \\
\frac{1.5m_i + 0.5m_j}{m_{j+1} + 0.5m_i} & \text{if } j = n \end{cases} \]  \tag{11}

where $n$ is the total number of sub-intervals, $t$ represent time and $m_j$ is the mid-point of $j^{th}$ interval. If the FLR is empty (untrained data), i.e. $I_1 \rightarrow \phi$, $\phi \rightarrow \phi$; where $\phi$ represent empty FLR, then the predicted values are obtained by equation (12)

\[ F(t) = F(t - 1) + \frac{F(t - 2) + F(t - 3) + F(t - 4)}{3} \times (r - h) . \]  \tag{12}

where $F(t - 1), F(t - 2), F(t - 3)$ & $F(t - 4)$ are previous predicted values, $F(t)$ is current predicted value at any time $t$, $r$ is the average rate of increment and decrement in all forecasted values obtained by equation (11) and $h$
is any small number lying between 0 to 0.5 to overcome the effect of high increment and decrement in the value of $r$.

Phase II is also elaborated in the form of algorithm B with summarised steps.

**Algorithm B**

To make forecasting by using FTS

| Step | Description |
|------|-------------|
| Step 1: | Basic interval obtained by previous algorithm A are partitioned into sub-intervals according to the number of elements $y_i$ belong to them. |
| Step 2: | Select those sub-intervals $b_j$ in which historical data belong and calculate the mid-points $m_i$ of each sub-interval. |
| Step 3: | Linguistic variables are defined for each selected sub-interval obtained in step 2 as $I_j = \sum_{i=1}^{n} \frac{b_i}{a_{ij}}$; $i \in \mathbb{N}$, where $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } j = j - 1 \text{ or } j = j + 1 \\ 0 & \text{otherwise} \end{cases}$ |
| Step 4: | Allocate the linguistic variable to all historical data according to the belonging of data to their respective sub-interval. |
| Step 5: | Create first order FLR and FLRG from step 4. |
| Step 6: | Defuzzify the historical data i.e., calculate the forecasted value by using equation (11), if FLR is non-empty. |
| Step 7: | Calculate the average rate of increment or decrement $r$ of defuzzified value obtained in previous step; $0 \leq r \leq 1$. |
| Step 8: | Determine the predicted value by equation (12), if FLR is empty. |
| Step 9: | End. |

The Fig. 1 depicts the flow chart for proposed ANHFTS model containing algorithm A and algorithm B.

### 5. Performance measure

The measurement of the performance of proposed ANHFTS model is evaluated with three different parameters which are mean square error (MSE), root mean square error (RMSE) and average forecasting error rate (AFER).

#### 5.1. Mean square error

The average of squared difference between forecasted $F_i$ and actual value $AV_i$ is estimated by mean square error [49]. The value of MSE can be formulated in equation (13), lower its value, the better forecasted value.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (AV_i - F_i)^2,$$  \hspace{1cm} (13)

where $n$ is the total number of data points.

#### 5.2. Root mean square error

Root mean square error is used to calculate that how much the forecasted value differs with actual value [48]. The value of RMSE should be small for better forecasting and calculated by using equation (14).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (AV_i - F_i)^2},$$  \hspace{1cm} (14)

where $n$ is the total number of data points.
5.3. Average forecasting error rate

Average forecasting error rate [20] is the percentage of error that reflects the absolute difference between the actual value and forecasted value at any point of time and it is defined in equation (15).

\[
\text{AFER} \% = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \times 100, 
\]

where \( n \) is the total number of data points.

6. Experimental results and analysis

In this section, first we show that the MFCM is free from noisy data then, the proposed ANHFTS model is implemented on two examples to figure out the performance of it and prediction for the upcoming COVID-19 infected cases and deaths in India at the end of month August 2020 has been carried out.

6.1. Noisy environment effect

For clustering purpose, FCM is well-known technique. However, noisy environment easily affects the result of FCM. While clustering should be free from it. So, FCM is modified by introducing a variable in negative exponential form to conquer this problem.

Let \( X = \{x_1, x_2, \ldots, x_n\} \) be a universe of discourse with \( m \) data points defined on \( \mathbb{R}^n \). By minimizing

\[
\sum_{j=1}^{m} \left| y_j - C \right|^2
data points with respect to \( C \), the estimated value of \( C \) is obtained by equation (16).

\[
C = \frac{\sum_{j=1}^{m} y_j}{m}. 
\]

Let \( \{5, 3, 4.7, 6, 5.6, 5.1, 4.4, 5.3, 7, 4\} \) be an artificial data set [59], has to be tested by the procedure of least-square method. 5 and 7.0833 are the estimated value of \( C \) before and after adding the noisy value 30, in artificial data set, by equation (16), it shows noisy data highly affects the minimizer. The same data is applied to MFCM for obtaining minimizer result by equation (6) are 4.9931 and 4.9259 before adding noisy data and after respectively. Hence, it shows that MFCM tolerate noisy environment.

6.2. Implementation of proposed ANHFTS model

For the future prediction of COVID-19 infected cases and deaths in India, the proposed ANHFTS model has been implemented. But before the prediction, the proposed model is tested against available data of infected cases and deaths due to COVID-19 form April to July 2020, for checking the efficiency of the proposed model.

Example 6.2.1: For predicting the number of infected cases of COVID-19 in India, the data form 1st April to 30th July 2020 are considered. This epidemic data of COVID-19 has been taken from Government authorized portal [16]. For simplicity, the whole data is divided into four groups i.e., April, May, June and July months group.

Before the prediction of infected cases of COVID-19 for August 2020, the proposed model is tested against the actual data of infected cases by this virus in India. For simplicity, available four months data of COVID-19 is tested by proposed ANHFTS model month-wise. First apply the ANHFTS model for April 2020 COVID-19 data. Let \( X \) be the universe of discourse which contain the COVID-19 data of infected person for April 2020. The minimum and maximum values of data set are 2059 and 34866 respectively, which are denoted by \( D_{\text{min}} \) and \( D_{\text{max}} \) respectively.

For the simplicity, the values of \( S_1 \) and \( S_2 \) are taken randomly 59 and 2134 respectively. Then, the universe of discourse will be \( X = [D_{\text{min}} - S_1, D_{\text{max}} + S_2] = [2000, 37000] \).

Partition \( X \) into randomly chosen 7 equal length intervals such as \( x_1 = [2000, 7000] \), \( x_2 = [7000, 12000] \), \( x_3 = [12000, 17000] \), \( x_4 = [17000, 22000] \), \( x_5 = [22000, 27000] \), \( x_6 = [27000, 32000] \), \( x_7 = [32000, 37000] \). So, the 7 number of clusters will be formed according to the present model. Now, initialize the membership value and obtain the average of data as \( a = 14975.1 \). Apply the step 5 and 6 of algorithm A until the termination condition is not obtained. The iterative process of MFCM upgrades the membership value along with centroid successively. After
satisfying the termination condition, the centroid will be $\mathbf{c}_1 = 2800.2066$, $\mathbf{c}_2 = 13029.9076$, $\mathbf{c}_3 = 5337.9829$, $\mathbf{c}_4 = 8712.6469$, $\mathbf{c}_5 = 18441.7603$, $\mathbf{c}_6 = 24450.5756$ and $\mathbf{c}_7 = 31798.9668$ in increasing order. After applying the step 8 of algorithm A, the following basic intervals are obtained:

$$a_1 = [1531.3185, 4069.0947], a_2 = [4069.0947, 7025.3149], a_3 = [7025.3149, 10871.2772],$$

$$a_4 = [10871.2772, 15735.8339], a_5 = [15735.8339, 21446.1679], a_6 = [21446.1679, 28124.7712],$$

$$a_7 = [28124.7712, 35473.1624].$$

Now, obtain the number of elements belong to the interval $a_i$ are 2059, 2545, 3105 and 3684. Partition the interval $a_i$ into 4 sub-intervals by using the step 1 of the algorithm B. Repeat the same procedure with other remaining intervals and select those sub-intervals $b_i; i \in \mathbb{N}$, which contain the given data. The obtained results are shown in Table I.

After applying the step 3 of algorithm B, the linguistic variables $I_i; i = 1, 2, ..., 28$ for each sub-intervals $b_i; i = 1, 2, ..., 28$ are as:

$$I_1 = \frac{b_1}{0 } + \frac{b_2}{0.5} + \frac{b_3}{0} + ... + \frac{b_6}{0} + \frac{b_7}{0} + \frac{b_8}{0},$$

$$I_2 = \frac{b_1}{0.5} + \frac{b_2}{0} + \frac{b_3}{0} + ... + \frac{b_6}{0} + \frac{b_7}{0} + \frac{b_8}{0},$$

$$I_3 = \frac{b_1}{0} + \frac{b_2}{0} + \frac{b_3}{0.5} + ... + \frac{b_6}{0} + \frac{b_7}{0} + \frac{b_8}{0},$$

$$I_{28} = \frac{b_1}{0.5} + \frac{b_2}{0} + \frac{b_3}{0} + ... + \frac{b_6}{0} + \frac{b_7}{0} + \frac{b_8}{0},$$

where, the denominators of $b_i$ denotes the membership value of each sub-intervals to their respective linguistic variables $I_i$. Now, allocate these linguistic variables to the historical data according to their belongingness to the sub-intervals $b_i$. Then, the first order FLR’s and FLRG’s will be in the following form

$$FLR(x) = \{(1, I_1 \rightarrow I_1), (2, I_2 \rightarrow I_1), ... , (29, I_{28} \rightarrow I_1)\},$$

$$FLRG(i) = \{(G_1, I_1 \rightarrow I_1), (G_2, I_2 \rightarrow I_1), ... , (G_{29}, I_{28} \rightarrow I_1)\},$$

where $G_i$ represent $i^{th}$ group and FLR $(x) = (x, \text{Linguistic variable}) : \forall x = \text{dates}$. After applying the remaining steps involved in the proposed ANHFTS model, the forecasted COVID-19 infected cases for April-2020 month with their linguistic variables are find out, which are shown in Table II.

Again, repeat the whole process for forecasting the COVID-19 infected cases for May, June and July 2020 months. The number of clusters formed for these months is 10, 8 and 10 respectively. After applying the steps involved in the proposed model, the forecasted results for the months May, June and July 2020 with their linguistic variables are evaluated and shown in Table II.

By the analysis of Table II, it can be concluded that the calculated value of AFER ($\%$), 2.0093, 0.6061, 0.4802 and 0.5561 for April, May, June and July 2020 months respectively are very small. Therefore, the forecasted values of the infected persons obtained by the proposed model are very close to the actual values. Hence, the proposed model is well trained and suitable for the future prediction of novel corona virus. Fig. 2 shows the graphical representation of Table II. In which Fig 2(a), 2(b), 2(c) and 2(d) show the graphical representation of forecasted and actual COVID-19 infected cases for the month of April, May, June and July 2020 respectively. In this figure, the forecasted infected data (red line) is the data used for training purposes and actual values (blue line) indicate the official data of infected cases till the end of July 2020 in India. From this graph, it is observed that the forecasted COVID-19 infected cases closely match the available official data.

From Table II, it can be concluded that the spreading of COVID-19 is increasing day-by-day in India. Now, apply step 7 of algorithm $B$ to calculate the average rate of increment of COVID-19 for the month of July 2020. The average rate of increment for July is 0.03451 and the randomly generated value of $h$ is 0.007. Now-a-days, the health ministry of India facing a lot of stress due to the COVID-19 virus. So, it becomes necessary to predict the number of COVID-19 cases in upcoming days to take protective measures in a worst situation. Therefore, the proposed ANHFTS model can also predict the number of upcoming new infected cases approximately. The
predicted newly infected cases for the upcoming August-2020 month is shown in Table III. The results of Table III shows that there could be approximately 3659185 COVID-19 infected cases upto the end of August 2020. The predicted results of Table III are also depicted graphically in Fig. 3.

Example 6.2.2: In this example, the number of predicted approximate deaths due to COVID-19 is obtained for
the month of August 2020. The available data of deaths due to COVID-19 have been collected from Government
authorised portal [16] during the period of April-July 2020.

The prediction of COVID-19 deaths in India for August 2020 month will be determined, but before this pre-
pdiction, the proposed model has been tested against the official data of deaths from COVID-19 in India. The
proposed ANHFTS model is applied to the official data month-wise. The universe of discourse for April 2020 will
be [50, 1200]. According to step 1 of algorithm A, the estimated number of clusters for April 2020 will be 10.
After applying the remaining steps involved in the proposed ANHFTS model, the forecasted COVID-19 deaths for
the month of April 2020 with their linguistic variable are shown in Table IV. Similarly, the universe of discourse
for May, June and July 2020 will be [1200, 5500], [5500, 17500] and [17800, 35800] respectively. By step 1 of
algorithm A, the estimated number of clusters for these months will be 10, 8 and 10 respectively. The forecasted
value of COVID-19 deaths in India with their linguistic variables for the month of May, June and July 2020 are
shown in Table IV.

It can be observed by the analysis of Table IV that the value of performance measure AFER (%), 2.2989,
1.0395, 1.0560 and 0.4196 for April, May, June and July 2020 months respectively are very close to 0. Therefore,
the forecasted values of deaths due to COVID-19 obtained by proposed model have a minor difference from
the actual value. The forecasted and actual values of deaths in India versus date for April, May, June and July 2020
months are also depicted in Fig. 4(a), 4(b), 4(c) and 4(d) respectively. From this graph, it is observed that the
forecasted COVID-19 deaths obtained by proposed model are closely match with the available official data.

By using forecasted COVID-19 deaths data by proposed model present in Table IV, the average rate of incre-
ment in COVID-19 death cases for July is 0.02441 and randomly generated value of $h$ is 0.005. The prediction
of the deaths of COVID-19 is also necessary for taking protective measures against it. The predicted deaths due
to novel virus for August 2020 month is shown in Table III. The predicted COVID-19 deaths in August 2020 are
also represented in Fig. 5. The results of Table III shows that there could be approximately 64410 deaths due
to COVID-19 upto the end of August 2020. These predicted values of deaths may differ from the official data
which shall be obtained at the end of August 2020 because of the awareness of people and daily upgrading health
infrastructure towards novel corona virus.

In Example 6.2.1 and 6.2.2, errors are calculated month-wise. For checking the accuracy in forecasted COVID-
19 infected cases and deaths in India by the proposed ANHFTS model, the error percentage is calculated on daily
basis by taking a interval of 5 days for the month of July 2020. The forecasted values of COVID-19 infected cases
and deaths obtained by the proposed model, official data and calculated error percentage are shown in Table V. The
result of this table shows that the error between these two data for infected cases and deaths are minor. Hence, the
proposed model can be applied to predict the approximate infected cases and deaths of COVID-19 for upcoming
month August 2020 with fewer errors.

6.3. Estimation of approximated isolation bed's and ICU's

The recovery rate in India from COVID-19 is 64% [57] of the total infected cases on 30 July 2020. If it is
assumed that this recovery rate will remain constant till 31 August 2020. According to the proposed ANHFTS
model of COVID-19, the expected number of recovered peoples will be approximately 2341879 till 31 August
2020. The active cases may require hospitalization, quarantine and ICU’s in case of emergency. The number of
active cases will be calculated by equation (17).

$$A(t) = C(t) - R(t) - D(t). \quad (17)$$

From equation (17), the number of active cases will be approximately 1252896 at the end of August 2020.

According to the recent report of the Ministry of Health and Family Welfare (MoHFW), a total of 944170 iso-
lation beds, 31258 ICU’s and 114638 oxygen supported beds are available to fight against COVID-19 by including
930 COVID hospital, 2362 COVID health-centre, 10341 quarantine centres and 7195 COVID centres [58]. At
present, India has successfully prevented plenty of infected cases and deaths from COVID-19 due to their awareness and advance health infrastructure. However, COVID-19 infected cases continuously shows growth. Therefore, it is essential to increase the number of isolation beds, the number of ICUs or ventilator devices for the struggle against COVID-19 pandemic diseases. According to the present study, the expected number of active cases will be approximately 1252896 at the end of August 2020. It is assumed that only 2% to 5% [6, 56] of the active cases are critical who required Ventilators. So, the estimated number of required ventilators for the treatment of COVID-19 infection will be approximately 25058 to 62645 and the rest of the active cases approximately 1190252 to 1227838 should be hospitalized or quarantined. Therefore, India will require approximately 12.5 lakh beds for infected persons and 65 thousand ICU’s for critical infected persons at the end of August 2020. Indian government should impose a strict lockdown to make a breakdown in new cases. Also, the Indian people should follow the guidelines of the Health Ministry and should adopt protective measurements like hand wash, wearing a mask, use sanitizer, follow social distancing, etc.

6.4. Analysis of variance

The results obtained by the proposed ANHFTS model and official data of COVID-19 infected cases and deaths in India for the month of July 2020 were tested by one-way ANOVA and it has been carried out in MINITAB 19. The results of analysis of variance are shown in Table VI. The result of ANOVA between ANHFTS model and official data of COVID-19 infected cases shows that the P-value is 0.9855 which is greater than the F-value 0.0003 at 95% confidence level. Similarly, the P-value is also greater than the F-value for the number of COVID-19 deaths in India for July 2020. Hence, the mean value of the proposed ANHFTS model and official data do not differ significantly at 95% level of significant for infected cases as well as deaths in India, which conform the better accuracy of the proposed model.

7. Conclusion

Daily growing magnitude of COVID-19 cases put the whole world humanity at high risk. Thus, it becomes necessary to control the outbreak of COVID-19 disease and forecast the infected cases and deaths in the upcoming days to execute the necessary plans. Therefore, this study presents a hybrid predictive FTS model based on the MFCM clustering technique. The main purpose of this article is to develop an effective model for estimating the number of COVID-19 infected cases and deaths in India for next coming 31 days. The proposed ANHFTS model is tested against available COVID-19 data of India for the measurement of its performance based on MSE, RMSE and AFER. Tables II and IV, show that the proposed ANHFTS model is capable for the prediction of infection and deaths. According to Table III, the number of newly infected cases and deaths at the end of August 2020 in India will be approximately 3659185 and 64410 respectively. Also, the proposed model predict the requirement of isolation bed’s and ICU’s to deal with COVID-19 in coming days. The output of the proposed model suggest that there will be requirement of approximately 12.5 lakh beds for infected persons and 65 thousand ICU’s for critical infected persons at the end of August 2020 in India.

Thus, the proposed model could be significantly important for government and health care decision-makers for making protection plan during this pandemic. The Indian government substantially controls COVID-19 disease but they have to plan a strict strategy against the increment of COVID-19 cases in India and reduces the spread of virus significantly otherwise it will affect a large population of India. If there is no outbreak in the spreading of COVID-19 then, this figure 3659185 may be converted into another big figure or even up to crore in upcoming months. Through the developed ANHFTS model we may be able for calculating important parameter such as infection rates and deaths rates, which will help us to have a more accurate grasp of the transmission trained of COVID-19 type disease, if occurs in future.

In recent years, several membership functions have been developed. Each membership grade has some advantages as well as disadvantages. It is impossible to develop a general framework for a membership function because each model possesses its limitations and characteristic. In the context of each application, some membership functions have been seen more appropriate than other. However, the issue of choosing a general membership grade is still a subject of research. In future study, we will extend this present work and study the impact of different membership functions on the predictive results.
Compliance with ethical standards
Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References
[1] Acharya R, Porwal A. A vulnerability index for the management of and response to the COVID-19 epidemic in India: An ecological study. The Lancet Global Health 2020;8(9):1142-1151. doi:10.1016/S2214-109X(20)30300-4.
[2] Al-Qaess MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. Journal of Clinical Medicine 2020;9(3):1-15. doi:10.3390/jcm9030674.
[3] Alkhatami BST, Alzaid SS. A novel mathematics model of COVID-19 with fractional derivative, stability and numerical analysis. Chaos, Solitons & Fractals 2020;139:1-9. doi:10.1016/j.chaos.2020.110017.
[4] Ambikapathy B, KrishnaMurthy K. Mathematical modelling to assess the impact of lockdown on COVID-19 transmission in India: Model development and validation. JMRI Public Health and Surveillance 2020;6(2):1-8. doi:10.2196/19368.
[5] Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos, Solitons & Fractals 2020;139:1-9. doi:10.1016/j.chaos.2020.110017.
[6] Auld S, Caridi-Scheible M, Blum JM, Robichaux CJ, Kraft CS, Jacob JT, Jabaley CS, Carpenter D, Kaplow R, Hernandez AC. ICU and ventilator mortality among critically ill adults with coronavirus disease 2019. medRxiv 2020;48(9):799-804. doi:10.3390/cns000000004457.
[7] Bezdek JC. Objective function clustering. In Pattern recognition with fuzzy objective function algorithms. Springer, Boston, MA 1981;43-93. doi:10.1007/978-1-4757-0450-1_3.
[8] Bhattacharya R. A predictive model for the evolution of COVID-19. Transaction of the Indian National Academy of Engineering 2020;5(2):133-140. doi:10.7411/jti.2020.00130-w.
[9] Bhola J, Venkatateswaran VK, Koul M. Corona epidemic in Indian context: Predictive mathematical modelling. medRxiv 2020. doi:10.1200/jco.20.041837.
[10] Cakar S. Dynamic analysis of a mathematical model with health care capacity for COVID-19 pandemic. Chaos, Solitons & Fractals 2020;139:1-8. doi:10.1016/j.chaos.2020.110033.
[11] Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons & Fractals 2020;135:1-15. doi:10.1016/j.chaos.2020.109850.
[12] Chatterjee K, Chatterjee K, Kumar A, Shankar S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. Medical Journal Armed Forces India 2020;76(2):147-155. doi:10.1016/j.mjafi.2020.03.022.
[13] Chen SM. Forecasting enrollments based on fuzzy time series. Fuzzy Sets and Systems 1996;81(3):311-319. doi:10.1016/0165-0114(95)00220-0.
[14] Chen SM, Hwang JR. Temperature prediction using fuzzy time series. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 2000;30(2):263-275. doi:10.1109/41403-020-00130-w.
[15] Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals 2020;139:1-14. doi:10.1016/j.chaos.2020.110057.
[16] Data. Coronavirus outbreak in India. 2020. https://www.covid19india.org.
[17] Deh-S, Majumdar M. A time series method to analyze incidence pattern and estimate reproduction number of COVID-19.arXiv preprint 2020. doi:arXiv:2003.10655.
[18] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics 1973;3(3):32-57. doi:10.1080/01969727308546046.
[19] El Koutfi A, Adnani J, Bennar A, Yousfi N. Analysis of a stochastic SIR model with vaccination and nonlinear incidence rate. International Journal of Differential Equations 2019:2019.1:9. doi:10.1155/2019/9275051.
[20] Garg B, Beg MS, Ansari AQ, Imran BM. Fuzzy time series prediction model. In International Conference on Information Intelligence, Systems, Technology and Management. Springer, 2011:141-126-127. doi:10.1007/978-3-642-19423-8_14.
[21] Giri S, Yadav PK, Kumar H. A neural network based model with lockdown condition for checking the danger stage level of COVID-19 infection risk. International Journal of Recent Technology and Engineering 2020;9(2):1111-1118. doi:10.3910/jire.0407.097220.
[22] Huang K. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems 2001;123(3):387-394. doi:10.1016/S0165-0114(00)00057-6.
[23] Kaushik S, Kaushik S, Sharma Y, Kumar R, Yadav JP. The Indian perspective of COVID-19 outbreak. VirusDisease 2020;31(2):146-153. doi:10.1016/j.virusd.2020.03.004.
[24] Krishna MV, Prakash J. Mathematical modelling on phase based transmissibility of corona virus. Infectious Disease Modelling 2020;5:375-385. doi:10.1016/j.idm.2020.06.005.
[25] Kumar N, Kumar H, Sharma K. Extension of FCM by introducing new distance metric. SN Applied Sciences 2020;2(4):1-21. doi:10.1007/s42452-020-2417-9.
[26] Kuniya T. Prediction of the epidemic peak of coronavirus disease in Japan. 2020. Journal of Clinical Medicine 2020;9(3):1-7. doi:10.3390/jcm9030789.
[27] Labwani S, Sahni G, Mewara B, Kutiar R. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. Chaos, Solitons & Fractals 2020;138:1-8. doi:10.1016/j.chaos.2020.109939.
[28] Li ST, Cheng YC, Lin SY. A FCM-based deterministic forecasting model for fuzzy time series. Computers & Mathematics with Applications 2008;56(12):3052-3063. doi:10.1016/j.camwa.2008.07.033.
Zhang Y, Qu H, Wang W, Zhao J. A novel fuzzy time series forecasting model based on multiple linear regression and time series clustering. Mathematical Problems in Engineering 2020;2020:1-17. doi:10.1155/2020/9546792.
Table I
Selected sub-intervals with their mid-points for April 2020

| Variables | Sub-intervals                      | Corresponding elements | Mid-points     |
|-----------|------------------------------------|------------------------|----------------|
| b₁        | [1531.3185, 2165.7626]             | 2059                   | 1848.5405      |
| b₂        | [2165.7626, 2800.2066]             | 2545                   | 2482.9846      |
| b₃        | [2800.2066, 3434.6507]             | 3105                   | 3117.4286      |
| b₄        | [3434.6507, 4069.0947]             | 3684                   | 3751.8727      |
| b₅        | [4069.0947, 4660.3387]             | 4293                   | 4364.7167      |
| b₆        | [4660.3387, 5251.5828]             | 4777                   | 4955.9608      |
| b₇        | [5251.5828, 5842.8268]             | 5350                   | 5547.2048      |
| b₈        | [5842.8268, 6434.0709]             | 5915                   | 6138.4488      |
| b₉        | [6434.0709, 7025.3149]             | 6728                   | 6729.6929      |
| b₁₀       | [7025.3149, 7986.8055]             | 7599                   | 7506.0602      |
| b₁₁       | [7986.8055, 8948.2961]             | 8453                   | 8467.5508      |
| b₁₂       | [8948.2961, 9909.7866]             | 9211                   | 9429.0413      |
| b₁₃       | [9909.7866, 10871.2772]            | 10454                  | 10390.5319     |
| b₁₄       | [10871.2772, 11844.1885]           | 11485                  | 11387.7329     |
| b₁₅       | [11844.1885, 12817.0999]           | 12971                  | 12330.6442     |
| b₁₆       | [12817.0999, 13790.0112]           | 13432                  | 13303.5556     |
| b₁₇       | [13790.0112, 14762.9226]           | 14354                  | 14276.4669     |
| b₁₈       | [14762.9226, 15735.8339]           | 15725                  | 15249.3782     |
| b₁₉       | [15735.8339, 16708.7452]           | 17305                  | 17877.2092     |
| b₂₀       | [16708.7452, 17681.6565]           | 18544                  | 18544          |
| b₂₁       | [17681.6565, 18654.5678]           | 20081                  | 20732.3762     |
| b₂₂       | [18654.5678, 19627.4790]           | 21373                  | 20732.3762     |
| b₂₃       | [19627.4790, 20591.3903]           | 21373                  | 20732.3762     |
| b₂₄       | [20591.3903, 21544.3016]           | 22804                  | 22804.9933     |
| b₂₅       | [21544.3016, 22506.3128]           | 24448                  | 23950.6441     |
| b₂₆       | [22506.3128, 23469.3240]           | 26283                  | 25620.2960     |
| b₂₇       | [23469.3240, 24432.3353]           | 27990                  | 27289.9458     |
| b₂₈       | [24432.3353, 25395.3466]           | 29458                  | 29043.3201     |
| b₂₉       | [25395.3466, 26358.3579]           | 31360                  | 30880.4179     |
| b₃₀       | [26358.3579, 27321.3692]           | 33065                  | 32717.5157     |
| b₃₁       | [27321.3692, 28284.3804]           | 34866                  | 34554.6135     |
Table II
Forecasted infected cases of COVID-19 for the month of April, May, June and July 2020 in India

| Date | April 2020 | May 2020 | June 2020 | July 2020 |
|------|------------|----------|-----------|-----------|
|      | Linguistic variable | Forecasted infected cases of COVID-19 | Linguistic variable | Forecasted infected cases of COVID-19 | Linguistic variable | Forecasted infected cases of COVID-19 | Linguistic variable | Forecasted infected cases of COVID-19 |
| 1    | I₁       | 2021     | I₁       | 37471     | I₁       | 198485     | I₁       | 609538     |
| 2    | I₁       | 2399     | I₁       | 39560     | I₁       | 205184     | I₁       | 626775     |
| 3    | I₁       | 3052     | I₁       | 42829     | I₁       | 215456     | I₁       | 652376     |
| 4    | I₁       | 3694     | I₁       | 46155     | I₁       | 229330     | I₁       | 675651     |
| 5    | I₁       | 4317     | I₁       | 49544     | I₁       | 236013     | I₁       | 696495     |
| 6    | I₁       | 4920     | I₁       | 53001     | I₁       | 243070     | I₁       | 716497     |
| 7    | I₁       | 5516     | I₁       | 56603     | I₁       | 256004     | I₁       | 737468     |
| 8    | I₁       | 6110     | I₁       | 60366     | I₁       | 266959     | I₁       | 761458     |
| 9    | I₁       | 6742     | I₁       | 64105     | I₁       | 277496     | I₁       | 788599     |
| 10   | I₁       | 7503     | I₁       | 67583     | I₁       | 282226     | I₁       | 817035     |
| 11   | I₁       | 8413     | I₁       | 70780     | I₁       | 290023     | I₁       | 846076     |
| 12   | I₁       | 9380     | I₁       | 73879     | I₁       | 300807     | I₁       | 875750     |
| 13   | I₁       | 10347    | I₁       | 77148     | I₁       | 312065     | I₁       | 905893     |
| 14   | I₁       | 11318    | I₁       | 80959     | I₁       | 325111     | I₁       | 936820     |
| 15   | I₁       | 12292    | I₁       | 85188     | I₁       | 344057     | I₁       | 968566     |
| 16   | I₁₁      | 13268    | I₁₁      | 90050     | I₁₁      | 355730     | I₁₁      | 1000985    |
| 17   | I₁₁      | 14243    | I₁₁      | 95092     | I₁₁      | 367800     | I₁₁      | 1034623    |
| 18   | I₁₁      | 15556    | I₁₁      | 100500    | I₁₁      | 380285     | I₁₁      | 1069536    |
| 19   | I₁₁      | 17724    | I₁₁      | 106146    | I₁₁      | 392915     | I₁₁      | 1105449    |
| 20   | I₁₁      | 17724    | I₁₁      | 112150    | I₁₁      | 406281     | I₁₁      | 1143113    |
| 21   | I₁₁      | 20275    | I₁₁      | 118833    | I₁₁      | 421951     | I₁₁      | 1182607    |
| 22   | I₁₁      | 20275    | I₁₁      | 126509    | I₁₁      | 440045     | I₁₁      | 1225814    |
| 23   | I₁₁      | 22253    | I₁₁      | 131770    | I₁₁      | 458627     | I₁₁      | 1278650    |
| 24   | I₁₁      | 23892    | I₁₁      | 138581    | I₁₁      | 476037     | I₁₁      | 1339135    |
| 25   | I₁₁      | 25566    | I₁₁      | 145884    | I₁₁      | 492216     | I₁₁      | 1395061    |
| 26   | I₁₁      | 27257    | I₁₁      | 152533    | I₁₁      | 507968     | I₁₁      | 1442334    |
| 27   | I₁₁      | 29009    | I₁₁      | 159739    | I₁₁      | 524454     | I₁₁      | 1486467    |
| 28   | I₁₁      | 30826    | I₁₁      | 167056    | I₁₁      | 543236     | I₁₁      | 1530135    |
| 29   | I₁₁      | 32666    | I₁₁      | 174555    | I₁₁      | 564426     | I₁₁      | 1573321    |
| 30   | I₁₁      | 33920    | I₁₁      | 182246    | I₁₁      | 579347     | I₁₁      | 1602320    |
| 31   | —        | —        | —        | —         | —        | —          | —        | —          |
|MSE  | 191360.0033 | 837093.6694 | 7095004.0344 | 47860278.3793 |
|RMSE | 437.4471   | 914.9282   | 2663.6449   | 6918.1123  |
|AFER (%)| 2.0093 | 0.6061   | 0.4802   | 0.5561   |
Table III
Predicted COVID-19 infected cases and deaths for upcoming month August 2020 in India

| Date       | Predicted COVID-19 infected cases | Predicted COVID-19 deaths |
|------------|-----------------------------------|---------------------------|
| 31 July    | 1644409                           | 36259                     |
| 1 August   | 1687561                           | 36939                     |
| 2 August   | 1731761                           | 37631                     |
| 3 August   | 1777008                           | 38334                     |
| 4 August   | 1823443                           | 39052                     |
| 5 August   | 1871093                           | 39782                     |
| 6 August   | 1919990                           | 40526                     |
| 7 August   | 1970164                           | 41284                     |
| 8 August   | 2021649                           | 42056                     |
| 9 August   | 2074479                           | 42843                     |
| 10 August  | 2128691                           | 43645                     |
| 11 August  | 2184319                           | 44461                     |
| 12 August  | 2241400                           | 45293                     |
| 13 August  | 2299973                           | 46140                     |
| 14 August  | 2360077                           | 47003                     |
| 15 August  | 2421752                           | 47862                     |
| 16 August  | 2485038                           | 48778                     |
| 17 August  | 2549978                           | 49690                     |
| 18 August  | 2616616                           | 50620                     |
| 19 August  | 2684994                           | 51567                     |
| 20 August  | 2755160                           | 52531                     |
| 21 August  | 2827159                           | 53514                     |
| 22 August  | 28901039                          | 54515                     |
| 23 August  | 2976850                           | 55535                     |
| 24 August  | 3054643                           | 56574                     |
| 25 August  | 3134468                           | 57632                     |
| 26 August  | 3216379                           | 58710                     |
| 27 August  | 3300431                           | 59808                     |
| 28 August  | 3386680                           | 60927                     |
| 29 August  | 3475182                           | 62067                     |
| 30 August  | 3565997                           | 63228                     |
| 31 August  | 3659189                           | 64410                     |
Table IV
Forecasted COVID-19 deaths for the month of April, May, June and July 2020 in India

| Date | April 2020 | May 2020 | June 2020 | July 2020 |
|------|------------|----------|-----------|-----------|
|      | Linguistic variable | Forecasted COVID-19 deaths | Linguistic variable | Forecasted COVID-19 deaths | Linguistic variable | Forecasted COVID-19 deaths | Linguistic variable | Forecasted COVID-19 deaths |
| 1    | I_1        | 83       | I_1       | 1257      | I_1     | 17874     |
| 2    | I_1        | 53       | I_1       | 1347      | I_1     | 18210     |
| 3    | I_1        | 82       | I_1       | 1477      | I_1     | 18720     |
| 4    | I_1        | 101      | I_1       | 1588      | I_1     | 19232     |
| 5    | I_1        | 120      | I_1       | 1693      | I_1     | 19744     |
| 6    | I_1        | 141      | I_1       | 1802      | I_1     | 20260     |
| 7    | I_1        | 162      | I_1       | 1914      | I_1     | 20781     |
| 8    | I_1        | 189      | I_1       | 2026      | I_1     | 21309     |
| 9    | I_1        | 231      | I_1       | 2131      | I_1     | 21825     |
| 10   | I_1        | 231      | I_1       | 2229      | I_1     | 22298     |
| 11   | I_1        | 275      | I_1       | 2325      | I_1     | 22725     |
| 12   | I_1        | 314      | I_1       | 2421      | I_1     | 23137     |
| 13   | I_1        | 357      | I_1       | 2521      | I_1     | 23605     |
| 14   | I_1        | 395      | I_1       | 2626      | I_1     | 24244     |
| 15   | I_1        | 424      | I_1       | 2731      | I_1     | 25029     |
| 16   | I_1        | 450      | I_1       | 2842      | I_1     | 25776     |
| 17   | I_1        | 477      | I_1       | 2958      | I_1     | 26417     |
| 18   | I_1        | 510      | I_1       | 3111      | I_1     | 27034     |
| 19   | I_1        | 548      | I_1       | 3260      | I_1     | 27685     |
| 20   | I_1        | 589      | I_1       | 3412      | I_1     | 28371     |
| 21   | I_1        | 630      | I_1       | 3567      | I_1     | 29078     |
| 22   | I_1        | 672      | I_1       | 3724      | I_1     | 29611     |
| 23   | I_1        | 717      | I_1       | 3883      | I_1     | 30255     |
| 24   | I_1        | 751      | I_1       | 4051      | I_1     | 31335     |
| 25   | I_1        | 782      | I_1       | 4220      | I_1     | 32091     |
| 26   | I_1        | 891      | I_1       | 4402      | I_1     | 32838     |
| 27   | I_1        | 943      | I_1       | 4593      | I_1     | 33583     |
| 28   | I_1        | 998      | I_1       | 4790      | I_1     | 34332     |
| 29   | I_1        | 1064     | I_1       | 4994      | I_1     | 35085     |
| 30   | I_1        | 1117     | I_1       | 5205      | I_1     | 35593     |
| 31   | —          | —        | —         | —         | —       | —         |
| MSE  | 126.2271   | 1203.2048| 27223.6109| 17023.0030|
| RMSE | 11.2351    | 34.6872  | 164.9958  | 130.4722  |
| AFER (%) | 2.2989   | 1.0395   | 1.0560   | 0.4196   |
Table V
Comparison of official data and forecasted data of infected cases as well as deaths for July 2020 in India

| Date       | Official data | forecasted value | Error percentage |
|------------|---------------|------------------|------------------|
| 05-07-2020 | 697846        | 696495           | 0.1936           |
| 10-07-2020 | 822604        | 817035           | 0.6770           |
| 15-07-2020 | 970169        | 968566           | 0.1652           |
| 20-07-2020 | 1154913       | 1143113          | 0.6770           |
| 25-07-2020 | 1387087       | 1395061          | -0.5749          |
| 30-07-2020 | 1612354       | 1602320          | 0.6223           |

Deaths

| Date       | Official data | forecasted value | Error percentage |
|------------|---------------|------------------|------------------|
| 05-07-2020 | 19701         | 19744            | -0.2183          |
| 10-07-2020 | 22144         | 22298            | -0.6954          |
| 15-07-2020 | 24929         | 25029            | -0.4011          |
| 20-07-2020 | 28099         | 28371            | -0.9680          |
| 25-07-2020 | 32121         | 32091            | 0.0934           |
| 30-07-2020 | 35769         | 35593            | 0.4920           |

Table VI
ANOVA analysis of COVID-19 infected cases and deaths in July 2020

| Source                              | DF | Adj. SS     | Adj. MS    | F-value | P-value |
|-------------------------------------|----|-------------|------------|---------|---------|
| AHFTS model versus Official data for infected cases in India |    | 4.17E+07    | 1.20E+12   | 0.0003  | 0.9855  |
| Between-group                       | 1  | -1.0E+04    | 3.65E+08   |         |         |
| Within-group                        | 10 | 1.10E+04    | 3.65E+08   |         |         |
| Total                               | 11 | 1.10E+04    | 3.65E+08   |         |         |
| AHFTS model versus Official data for deaths in India |    | 4.17E+07    | 1.20E+11   | 0.0003  | 0.9865  |
| Between-group                       | 1  | 1.10E+04    | 3.65E+08   |         |         |
| Within-group                        | 10 | -1.0E+04    | 3.65E+08   |         |         |
| Total                               | 11 | 1.10E+04    | 3.65E+08   |         |         |
Fig. 1: Flow chart of proposed ANHFTS model
Fig. 2: Graphical representations of forecasted and actual infected cases of COVID-19 from April to July 2020 in India
Fig. 3: Graphical representations of predicted COVID-19 infected cases in August 2020 in India
Fig. 4: Graphical representations of forecasted and actual deaths due to COVID-19 from April to July 2020 in India
Fig. 5: Graphical representations of predicted COVID-19 deaths in August 2020 in India
Declaration of interests

✔ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: