Soft Computing (2023) 27:2509–2535
https://doi.org/10.1007/s00500-022-07800-7

MATHEMATICAL METHODS IN DATA SCIENCE

Forecast combination approach with meta-fuzzy functions for forecasting the number of immigrants within the maritime line security project in Turkey

Fatma Carman Cevik[1,2] · **Basak Gever**[3] · **Nihat Tak**[4] · **Tahir Khaniyev**[1,5]

Accepted: 26 December 2022 / Published online: 5 January 2023

Abstract
In this study, forecasting the number of immigrants on the Turkey's maritime line for use in a national security project carried out by Turkish Government within the scope of fight against uncontrolled immigration is discussed for the first time. Handling with the immigration problem is one of the biggest concerns of Turkey as unsupervised immigration can adversely affect the demographic and economic structure of the country. Precautions are needed as the short-, medium- and long-term impacts of undetected immigrants on the country's ecosystem are unpredictable, but due to the uncertainties inherent in immigration, the cost of using government resources such as patrol vehicles to capture undocumented immigrants can be extremely high. In order to both minimize the expenditure problem and keep immigration under control by providing a proper scan, forecasting the number of immigrants on the maritime line route is seen as an important problem and studied by probabilistic and non-probabilistic models. Since the data for 2020 and 2021 could not be attained yet due to COVID-19, in order to obtain forecasts and compare actual observations for 2019, which is the primarily focus of the research in this study, the dataset of interest on the number of daily immigrants between years 2016 and 2019 is obtained from Turkish Coast Guard Command within Ministry of Interior of Republic of Turkey. To obtain the most accurate forecasts, seven distinguished forecasting methods, from simple to complex, are implemented. Then, the forecast combination approach with meta-fuzzy functions which combines all methods is proposed. Consequently, the forecasting results are acquired and evaluated by using R. The evaluation of the results is made by using widely considered measurement accuracy metric root mean square error. According to the final assessments, the proposed approach gives more accurate forecasting results for the expected number of immigrants on the Turkey's maritime line and these results become an input to the national security project.

Keywords Immigration · Forecasting · Meta-fuzzy functions · Long short-term memory · Fuzzy inference systems and Artificial neural network

1 Introduction

Forecasting the number of immigrants is a branch of time series forecasting, and it is widespread research topic addressed to get an inkling of future population growth or decrease. However, countries struggling with immigration may need immigrant numbers prediction for other concerns and measures, besides just understanding population growth or decline, for instance, the fact that uncontrolled immigration can be predicted in advance, decision makers (DMs) such as governments take precautionary measures against immigration and facilitate the identification of immigrants. To achieve this in practice, regular patrolling on maritime line route can be planned, based on forecasts data in the fight

✉ Fatma Carman Cevik
fcarman@aselsan.com.tr

1 Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, Turkey

2 Aselsan A.Ş., Project Department, Ankara, Turkey

3 Department of Industrial Engineering, University of Turkish Aeronautical Association, Ankara, Turkey

4 Business Administration Department, Bursa Technical University, Bursa, Turkey

5 The Center of Digital Economics, Azerbaijan State University of Economics, Baku, Azerbaijan

against unsupervised immigration, and the use of government resources can be optimized.

In order to forecast the number of immigrants, time series forecasting methods are used. These methods can be divided into two categories: probabilistic and non-probabilistic methods. Each method, gathered under two basic groups, has strengths and weaknesses relative to each other, depending on the dataset they are applied to. For instance, while some methods take into consideration the linearity of the time series, some consider the scarcity of the data, and some consider its stationarity. In this study, autoregressive integrated moving average (ARIMA), simple exponential smoothing (SES) and Holt–Winters exponential smoothing (HES), which are probabilistic methods, and adaptive neuro fuzzy inference system (ANFIS), recurrent Type-1 fuzzy functions (R-T1FFs), artificial neural network (ANN) and long short-term memory (LSTM), which are non-probabilistic state-of-the-art methods, are briefly explained in Sect. 1.2 considering their advantages and disadvantages and these methods are discussed in the application section.

Unlike other time series, working on immigrant dataset and forecasting the number of immigrants require considering demographic values while developing an analytical approach by nature of immigrant dataset. In addition to the uncertainty of the concept of migration, which is affected by meteorological, political and economic developments, uncertainty increases in events involving human elements and numerical analysis becomes difficult. With all these aspects, it becomes difficult to analyze and make predictions on immigration dataset. As a way of considering demographic values, time series containing immigrant data should be carefully analyzed and forecasting methods should be used with this precision in time series. At the same time, the characteristic of the time series cannot be known in the first place, so it is not known which method will give the best forecasting results. Moreover, time series containing immigrant data can have more than one characteristic at the same time. Some part of the series may be stationary while others may be non-stationary.

In this study, the expected number of immigrants on Turkey's maritime line within the scope of a national security project is forecasted for the first time and compared with actual data to obtain the most consistent forecasting results. Firstly, the forecasting results are obtained by 7 existing methods, respectively, traditional methods such as ARIMA, SES, HES and alternative methods such as ANFIS, ANN, LSTM and R-T1FFs. ARIMA, SES and HES are used to deal with the probabilistic part of the time series. ARIMA is chosen since it is traditional and often gives better results than more complex approaches such as non-probabilistic methods. SES and HES are applied while handling with the probabilistic part of the time series, and it requires less data than ARIMA. As the complexity of immigration movements is very high due to various factors as mentioned in Sect. 2, the ability of ANN to perform nonlinear analysis is seen as an advantage in forecasting the number of immigrants. ANFIS is considered worthy of implementation because of its proven superiority over ANN when there is sufficient data to deal with the non-probabilistic part of the time series and to train it (Atmaca et al. 2001). LSTM and R-T1FFs are chosen because they are non-probabilistic, and their prediction performance is important. The LSTM method is applied because it preserves the characteristics of the given data for a longer period of time and uses them in its own training. R-T1FFs is considered valuable of implementation because it is not rule based; hence, it is easy to apply. Since the structure of the data cannot be known at the beginning, it is not known which method will give better results.

From this point of view, we consider that by combining the forecasting methods, the strengths and weaknesses of which are explained in detail in Sect. 1.2., it is remarked that more accurate and better performance results are obtained than the individual use of these existing methods. With this motivation, forecast combination approach with meta-fuzzy functions (FCA-MFFs) based on 7 existing forecasting methods is proposed and applied to obtain more robust forecasting results and to compare the performance of existing methods with the proposed method in this study. By choosing these 7 methods for FCA-MFFs, it is ensured that all situations such as stationarity, linearity or nonlinearity of the data, less or enough data, long-term dependencies or not, trending or not are handled with in the first place. In addition, more reliable and robust results with a lower Root Mean Square Error (RMSE) are aimed with FCA-MFFs. Thus, the ground is prepared with reliable forecasting results so that the Turkish Coast Guard Command can catch the maximum number of immigrants. Moreover, obtaining more accurate forecasting results will provide a more solid basis for decisions to be taken by the DMs such as governments. Also, the forecasting results obtained by FCA-MFFs are also compared with traditional and alternative methods for the first time in this dataset. Furthermore, this study is a pioneer on forecasting the number of immigrants on the maritime line in Turkey.

This paper is organized as follows. The introduction is given in Sect. 1. Literature review is given in Sect. 1.1, and forecasting methods are given in Sect. 1.2. The algorithm and pseudo-code of the proposed method are given in Sect. 2. In Sect. 3, the forecasting methods mentioned above are presented to evaluate the performance of the proposed method. Finally, the obtained results are discussed in Sect. 4.

1.1 Literature review

In the literature, there are various studies for forecasting the expected number of immigrants, the future population due to immigration and the factors affecting immigration. Bijak (2006) attempts to summarize selected theoretical bases of forecasting on international immigration. Alho et al. (2006) present the results of possible forecasts for the population by 2050 in 18 European countries. Bijak and Wisniowski (2010) consider the forecasting of the number of immigrants on the selected immigration countries in 2025 by the help of Bayesian Forecasting Method. According to Martineau (2010), the best way to stop the immigration crisis is to intervene before immigration begins. Using a dataset that contains information about 170 countries, a simple binary logistics model is created. Abel et al. (2013) investigates the forecasting of the uncertainty about immigration to England and immigration up to 2060. According to Cappelen et al. (2015), three different alternatives are determined for income variables and three different forecasting results are obtained for gross immigration until 2100 in Norway. Raymer and Wisniowski (2018) discuss the need for better predictive models of international immigration by testing a hierarchical model within the framework of the Bayes theorem to predict the number of immigrants based on age and gender in UK. Wicke et al. (2019) studies on scenarios that address how such socio-political events might occur, based on the Syrian civil war and immigration. According to Suriani et al. (2019), the fact that the number of immigrants can be forecasted in advance allows the host country to be prepared for immigrants. The model also considers various input factors affecting the immigrant population growth and predicts the number of immigrants between 2017 and 2022.

It is seen that these studies in the literature were conducted to investigate the factors affecting immigration and to determine future population density by traditional methods. Since these methods are traditional, they give results by considering the entire time series as a single type. However, a time series may have more than one structure such as some of it linear and some of it nonlinear. Recently, various methods have been developed to forecast time series datasets, but these methods have not yet been used for immigrants prediction. Moreover, these alternative methods ensure that the special nature of immigration data described in Sect. 1 can also be taken into consideration to forecast on. Therefore, there is a research gap in the literature on approaching immigration studies with alternative forecasting methods. The use of alternative forecasting methods should be considered to ensure that questions such as "Which method is more consistent for understanding immigration with its own uncertainties?" or "What is the structure of time series composed of immigrants data?" or "Can more consistent immigrants prediction be made using more than one method?".

1.2 Methods

Probabilistic or stochastic forecasting methods, also called traditional methods, make some assumptions about time series. Stationarity is an important assumption in probabilistic time series forecasting methods. This assumption states that time series have a constant mean, variance and covariance function (Tak et al. 2018). ARIMA is a probabilistic time series modeling method widely used in the literature, especially in social sciences (Ma 2020), and was developed by Box and Jenkins (1976). ARIMA assumes that the studied time series are generated from linear processes. The prediction does not need to assume any underlying models or related equations since ARIMA's forecast results are derived from the values of the input variables and error terms (Ma 2020). ARIMA often outperforms more complex structural models in short-term predictive ability. Advanced forecasting models use ARIMA because ARIMA models give results in terms of forecasting performance (Matamoros et al. 2020). However, ARIMA is "backward looking" and the model identification technique is subjective, which may cause the reliability of the chosen model to depend on the experience and choice of the predictor. Moreover, if a time series has a nonlinear structure, ARIMA models are not capable of dealing with it (Tak et al. 2018).

Another popular probabilistic forecasting model is exponential smoothing (ES) proposed in the late 1950s (SES by Holt in 1957, HES by Brown in 1959 and Winters in 1960). The mechanism of the ES method is mainly to smooth the time series. If a time series has a stochastic trend, it is practical to apply the ES method to forecast the problems of this time series. The biggest advantage of ES methods is that they are simple to implement and intuitive and easy to understand. In addition, data requirements are minimal, which makes ES suitable for real-time applications. However, the biggest disadvantage of the method stems from its basic premise about the model. To be applicable, the level of the time series must fluctuate at a constant level or change gradually over time. When the time series has an obvious trend, they do not perform well for forecasting (Li et al. (2008)).

However, datasets encountered in applications often do not meet the assumption of stationarity or linearity. Therefore, the literature subscribes to three main alternatives. First one is fitting a model on the historical data to simulate the future data. The second one is selecting the most appropriate forecasting model. Last but not least, the third one is averaging a set of candidate methods based on the properties of time-series as opposed to traditional time-series forecasting (Li et al. 2020). In recent years, many researchers have been working on non-probabilistic forecasting methods as alternative forecasting methods to deal with nonlinear time series data or non-stationarity. ANFIS, R-T1FF, ANN and

LSTM approaches can be counted as some examples to non-probabilistic forecasting methods.

ANFIS, which is proposed by Jang (1993), is a neuro fuzzy technique that combines the learning abilities of neural networks (NNs) and fuzzy inference systems (FIS) (Celikyilmaz and Turksen 2009). Combining NNs with FIS has two major benefits. First, the method is data driven; therefore, it does not make limiting assumptions on the form of the base model. Secondly, it overcomes some disadvantages of the individual use cases (Atmaca et al. 2001). ANFIS is used for modeling and defining many systems in various fields due to its nature by individual or hybridizing other methods. For instance, ARIMA and ANFIS hybrid models are implemented in the study of Barak and Sadegh (2016). Karaboga and Kaya (2020) use ANFIS to forecast the number of foreign visitors from Iran and Russia to Turkey, and ANFIS gives better results than the other methods discussed in the study.

Another non-probabilistic method is Type-1 fuzzy functions (T1FFs) that is introduced by Turksen (2008). Then, Tak (2018) proposes an advanced model that combines T1FFs and autoregressive and moving average (ARMA) model in the same algorithm. This modern method, which is called R-T1FFs, is an assumption free approach. Additionally, the prediction ability of R-T1FF is better than T1FFs due to the contribution of the moving average (MA) model. Moreover, T1FFs is not a rule-based system. Since the detection of rules is an important problem, R-T1FFs may be advantageous over FIS-based methods such as ANFIS. The other non-probabilistic forecasting methods is artificial intelligence and deep learning algorithms. One of the best-known artificial intelligence algorithms is ANN (McCulloch and Pitts 1943). ANN is often referred to as NNs and is computing systems vaguely inspired by the biological neural networks that basically make up animal brains (Chen et al. 2019). The prediction of ANN is based on the results obtained from the original data to make broad observations and then infer the potential part of the whole. Unlike the ARIMA model, it is very effective in solving nonlinear problems (Ma 2020). ANN, as an individual or hybridized with other methods, is often used in forecasts on social, economic, engineering, foreign exchange, stock problems, etc. (Tseng et al. 2002; Zhang 2003; BuHamra and Smaoui 2003; Jain and Kumar 2007; Egrioglu et al. 2009 and Uslu et al. 2010). ANN's ability to perform nonlinear analysis on complex data is high (Ahmad et al. 2014). Moreover, ANN may be advantageous over ANFIS when it comes to a short dataset that is not enough for good learning (Atmaca et al. 2001).

Another forecasting method is LSTM which is based on deep learning algorithm. LSTM (Hochreiter and Schmidhuber (1997) can learn long-term addictions. In this respect, it eliminates the disadvantage of NNs. Unlike other methods, its feedback connection makes it easier to find development trends through the backpropagation of historical data and current data (Ma 2020). It has been successfully applied to various sequence prediction and sequence tagging tasks (Sak et al. 2014). The LSTM algorithm has been recently successfully applied in forecasting problems by Namini et al. (2018), Wang et al. (2020), Chimmula and Zhang (2020), Shahid et al. (2020) and Elsheikh et al. (2021)

Probabilistic and non-probabilistic forecasting methods have been improved over the years to eliminate each other's disadvantages in order to give better performance results. In fact, it has recently become more effective to hybridize methods rather than using a single method that is considered to be the best. It results in obtaining consistent forecasts by getting benefit from the advantages of each method. For example, meta-fuzzy functions by Tak (2018), meta-fuzzy index functions by Tak (2020), meta-possibilistic index functions by Tak and Gok (2020) and meta-possibilistic fuzzy functions by Tak (2021) show that it was possible to boost the prediction accuracy by aggregating different methods. Thus, besides the performance of the forecasting results, it is observed that they are reliable and robust.

2 Proposed forecast combination approach with meta-fuzzy functions

Numerical analysis of immigration data is difficult due to the high uncertainty in events involving human elements such as cultural, individual choices or political reasons, etc. As well as uncertainty in immigration forecasting, how to include uncertainty in forecasts is also an important issue. The concept of uncertainty refers to the fuzziness or randomness of the process under consideration. According to Bijak (2010), these uncertainties can be summarized under three main headings. The first is inherent uncertainty about future events. Some errors are always inevitable in immigration forecasting because future inferences are made under uncertainty. In particular, there is uncertainty in all three components of demographic change, especially fertility, death and immigration. The second source of uncertainty is that available data are often inaccurate, inconsistent and incomplete. Sources of immigration data from different countries are often based on different definitions (Raymer et al. 2013). It is difficult to measure the exact size of international immigration flows. Data collection systems used to enroll immigrants often generate biased and incorrect forecasts that need to be corrected (Wisniowski 2013 and Disney 2014). The third source of uncertainty comes from the forecast models themselves. Applying different models to the same data can produce different forecasts, including different assessments of the forecasting uncertainty. There is no perfect model and choosing which model to apply is a matter of judgment. Even if forecasts from various competing models are combined using official metrics, the uncertainty about the additional model has been demonstrated by Bijak and Wisniowski (2010). As a result, there are many uncertainties in forecasting the number of immigrants that must first be fully understood and then considered in an empirical analysis. According to Lutz and Goldstein (2004), expert opinion plays a key role in the development of forecasts, but the success of the forecast depends on the approach chosen. For example, the role of the expert may be limited to selecting the forecast model and selecting the underlying data sources or providing expert opinion that is explicitly included as a parameter in the model.

In addition to the difficulties such as uncertainty and the availability of immigrant data described above, there are some situations due to the nature of the time series. Time series in practice can often include both trend and seasonal variations. It is almost impossible for a time series to be pure linear or pure nonlinear. For some time series, linear models can produce satisfactory results when the linear part of the time series is superior to the nonlinear part. Similarly, nonlinear models can give satisfactory results if the nonlinear part is superior with respect to the linear part. This causes the non-probabilistic part of the series to be ignored when the probabilistic method is applied, and the probabilistic part of the series to be ignored when the non-probabilistic method is applied.

To overcome those problems, forecast combination approaches Bates and Granger (1969) have been proposed in the literature. There are various forecast combination types of time series such as simple and weighted. The simple average combination method is a simple combination technique that assigns the same weight to each forecasts. Empirical results show that the simple average combination method can produce more accurate forecasting results in many cases and combining linear and nonlinear models improves forecasting performance (Buyuksahin and Ertekin 2019; Thomson et al. 2019; Shaub 2020). Various weighted average combination methods have been proposed to obtain different weights to the forecast methods over time. Manso et al. (2020) proposes an automated method to obtain weighted forecasting combinations using time series features.

However, there are two challenges even when applying the forecast combination: determining the optimum weights and determining forecasting methods. To overcome these challenges, the forecast combination method with meta-fuzzy functions (MFFs) was proposed by Tak (2018). MFFs gather methods that perform better in one cluster thanks to the underlying fuzzy clustering mechanism, while gather worse-performing methods into another cluster.

The purpose of MFFs is to unify methods for the same objective together with their weights into functions. The idea behind putting the methods together is the assumption that each method has too much or partial information for a given dataset, or no knowledge at all. Tak (2018), instead of bringing together the results of dissimilar studies for an objective, unifies dissimilar methods. It shows that unifying different methods has better forecasting precision. The aim for forecast combination with MFFs is to obtain reliable forecasting results by using the power of many individual methods.

There are three basic constituents in the MFFs algorithm that need to be defined. First one is the *"Function"*. The function consists of a linear combination of methods and their weights. *"Function"* represents the *"cluster"* in Fuzzy C-Means (FCM) clustering method . Second is the weights of the methods in functions. The weights of methods are simply derived from the membership degrees of a method in a cluster. The third one is the function that has the best suitability value among MFFs. Since we have as many *"Functions"* as the number of clusters, we look for a function with the best evaluation criteria. In this case, the function with the best evaluation criteria is called "MFF''_{best} and forecasting results are obtained with this.

Proposed method steps are given below and pseudo-code of FCA-MFFs is given in Algorithm.

- *Step 1* Determine available existing forecasting methods for the problem.
- *Step 2* The dataset X is divided into three groups as train (X_{train}), validation ($X_{\text{validation}}$) and test (X_{test}).

  $$X = [X_i] \ ; \ \ i = 1, 2, \ldots, n.$$

  where n is the number of total data.

  $$X_{\text{train}} = [X_i] \ ; \ \ i = 1, 2, \ldots, ntrain.$$

  where $ntrain$ is the number of training data. X_{train} is used in training existing methods for the FCA-MFFs training phase.

  $$X_{\text{validation}} = [X_i] \ ; \ \ i = 1, 2 \ldots, nvalidation.$$

  where $nvalidation$ is the number of validation data. $X_{\text{validation}}$ is used to choose the best MFF for FCA-MFFs.

  $$X_{\text{test}} = [X_i] \ ; \ \ i = 1, 2 \ldots, ntest.$$

  where $ntest$ is the number of test data. X_{test} is used to test the performance of the proposed FCA-MFFs method.
- *Step 3* Train the related existing methods by using X_{train} for a validation dataset $X_{\text{validation}}$.
- *Step 4* Obtain the forecasting results by using the trained methods for a validation dataset $X_{\text{validation}}$. Forecasting result matrix refers to the FCA-MFFs's input matrix which is called $Z_{\text{validation}}$. $Z_{\text{validation}}$ is training dataset of FCA-MFFs.

$$Z_{\text{validation}} = [Z_{\text{ij}}] \; ; \; i = 1, 2 \ldots p \; ; \; j = 1, 2, \ldots, nvalidation.$$

$$Z_{\text{validation}} = \begin{bmatrix} Z_{1,1} & Z_{1,2} & Z_{1,3} & \ldots & Z_{1,nvalidation} \\ Z_{2,1} & Z_{2,2} & Z_{2,3} & \ldots & Z_{2,nvalidation} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_{p,1} & Z_{p,2} & Z_{p,3} & \ldots & Z_{p,nvalidation} \end{bmatrix}$$

Here, Z_{ij} is the results of the ith method for the jth data point.

- *Step 5* Determine the weights of the methods in functions. The input matrix $Z_{\text{validation}}$ is clustered using FCM. Membership degrees in each cluster are used to calculate the weights of methods in functions. In this case, a cluster represents a function.
- *Step 5.1* Fuzziness index parameter (m) and number of clusters (c) are initialized. Cluster centers (v_i) are randomly initialized.
- *Step 5.2* Membership degree is calculated in Eq. (1).

$$\mu_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d(z_k, v_i)}{d(z_k, v_j)}\right)^{\frac{2}{m-1}} \right]^{-1} ; \quad i = 1, 2, \ldots, c \; ; \; k = 1, 2, \ldots, nvalidation \tag{1}$$

Under the constraint $\sum_{i=1}^{c} \mu_{ik} = 1$, if $\mu_{ik} < \alpha$-cut, then the value will be taken as zero. Z is the input matrix, $v_i, \; v_j$ is the cluster centers, $d(.)$ means the Euclidean distance function, c is the number of clusters and m is the fuzzy index parameter in Eqs. (1) and (2).

- *Step 5.3* New cluster centers are calculated in Eq. (2).

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{m} z_k}{\sum_{k=1}^{n} \mu_{ik}^{m}} \tag{2}$$

- *Step 5.4* Repeat Steps 3–5 until the number of repeats is reached or the difference of clusters between the two iterations falls below a defined threshold.
- *Step 6* Obtain the MFFs which are given in Eq. (3) by using the weights of the methods given in Eq. (4).

$$MFF_i(z) = \sum_{j=1}^{m} w_{ij} z_j \; ; \; i = 1, 2, \ldots, c \tag{3}$$

$$w_{ij} = \frac{\mu_{ij}}{\sum_{j=1}^{m} \mu_{ij}} \; ; \; i = 1, 2, \ldots, c \tag{4}$$

Here, MFF_i represents the ith meta-fuzzy function, while μ_{ij} represents the membership value of the jth method in the ith cluster, w_{ij} represent the weight of the jth method in the ith cluster and c is the number of clusters.

- *Step 7* Repeat Step 5 and Step 6 for different m and c.
- *Step 8* Select the best combination of the methods in a function. The function with the lowest RMSE is selected as the MFF_{best} with the help of $Z_{\text{validation}}$.
- *Step 9* Obtain the forecasting results with the FCA-MFFs algorithm.
- *Step 9.1* Since the MFF_{best} is obtained according to $Z_{\text{validation}}$, this function can be used for forecasting.
- *Step 9.2* Train the related existing methods by using X^*_{train} for a test dataset (X_{test}) and obtain the forecasting results matrix (Z_{test}).

$$X^*_{train} = [X_i] \; ; \; i = 1, 2, \ldots, ntrain + nvalidation.$$

$$Z_{\text{test}} = [Z_{\text{ij}}] \; ; \; i = 1, 2 \ldots p; \; j = 1, 2, \ldots, ntest.$$

$$Z_{\text{test}} = \begin{bmatrix} Z_{1,1} & Z_{1,2} & Z_{1,3} & \ldots & Z_{1,ntest} \\ Z_{2,1} & Z_{2,2} & Z_{2,3} & \ldots & Z_{2,ntest} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_{p,1} & Z_{p,2} & Z_{p,3} & \ldots & Z_{p,ntest} \end{bmatrix}$$

Here, Z_{ij} is the results of the ith method for the jth data point.

- *Step 9.3* Forecasting results of the FCA-MFFs method are calculated by using the MFF_{best} for the Z_{test} in Eq. (5).

$$F = \text{MFF}_{\text{best}}(Z_{\text{test}}) \tag{5}$$

where F is the outcomes (forecasting results) of the FCA-MFFs.

Algorithm : Pseudo code of the FCA-MFFs

ARIMA, SES, HES, ANN, ANFIS, R-T1FFs, LSTM are selected
The time series is divided into three group as train (X_{train}), validation ($X_{validation}$) and test (X_{test}) datasets
Selected methods are trained by using X_{train} and the forecasting results are obtained from the trained methods
The forecasting results of the methods are used as the input matrix ($Z_{validation}$) for FCA-MFFs
The number of clusters (c) and fuzzy index parameter (m) are specified
for (α from 1.5 to m) **do**
  for (i from 2 to c) **do**
    Weights of the methods in each function are obtained by using FCM
    MFFs are obtained by using Equation (3)-(4) for FCA-MFFs
    RMSE values of MFFs are calculated
  $c = c + 1$
$\alpha = \alpha + 0.5$
Returned the best MFF (MFF_{best}) which has the minimum RMSE value
Selected methods are trained by using X^*_{train} and the forecasting results matrix (Z_{test}) are obtained for X_{test}
The forecasting results (F) are calculated by using MFF_{best} for the Z_{test}

3 Forecasting the number of immigrants on the Turkey's maritime line

In this study, the expected number of immigrants is forecasted by 8 different forecasting methods in 5 different regions on Turkey's maritime line using the past immigrant data detected between April 2016 and July 2019. Before explaining the datasets and the data preparation process, it is important to describe the regions on the number of immigrants are forecasted. The Republic of Turkey has a continental shelf with legal jurisdiction at maritime. We refer to this entire continental shelf in this study as the Continental Shelf Region. We consider that predicting the expected number of immigrants for the Continental Shelf Region provides an overview for DMs such as government. Then, we divide the entire Continental Shelf Region into 4 zone depending on the 4 main military ports for which Turkish Coast Guard Command within Ministry of Interior of Republic of Turkey is responsible. Thus, we obtain Region 1, Region 2, Region 3 and Region 4 on maritime line that is represented in Fig. 1.

The raw form of the data source in this study comes from the daily number of immigrants caught between April 2016 and July 2019 on the maritime line route of Turkey. In the study, the daily number of immigrants data is used by converting it to a weekly basis. The number of all past immigrants caught weekly basis on the Continental Shelf Region between April 2016 and July 2019 is considered as the Continental Shelf Region dataset. Afterward, the captured immigrants are divided on the basis of regions and 4 separate datasets are obtained as the Region 1 dataset, Region 2 dataset, Region 3 dataset and Region 4 dataset. With this regioning, we prepare the infrastructure for the efficient patrol vehicle routing for the relevant patrol vehicles at the regions they are authorized, according to the forecasts obtained region based.

In practice, the following steps are followed in order to data pre-processing:

- *Step 1* A daily-based time series is created between April 2016 and July 2019.
- *Step 2* The number of immigrants determined for each day in the relevant period is recorded. If the number of immigrants detected at different times on the same day, they are merged and the daily number of immigrants is obtained. If no immigrants are detected during the relevant day, that day is recorded as zero. Thus, for the time series from April 2016 to July 2019, each day will have corresponding values in the time series.
- *Step 3* Then, the number of immigrants for every 7 days from the beginning is summed to convert the daily based data into weekly based data, and the number of immigrants on a weekly basis is obtained. Thus, the daily-based immigrants data for the range of 2016 April–2019 July are transformed on a weekly basis and 155 data are obtained.
- *Step 4* Here, however, the total number of immigrants in the relevant week may be zero, as in some weeks no immigrants may be caught. Since zero, that is, missing data, may cause problems to analyze the data; it is appropriate to assign meaningful values to these areas, rather than deleting the data with missing values in a list. If there are such weeks, imputation using with mean applies to missing data in these weeks. This works by calculating the average of the non-missing values in the weekly immigrants count column and then changing the missing

Fig. 1 Region information

values with this average value individually and independently of the others.

ARIMA, SES, HES, ANFIS, ANN, LSTM, R-T1FFs and the proposed FCA-MFFs are used in the study to forecast the expected number of immigrants. All calculations are done by using R, a statistical programming language. The performance of the forecasting results is evaluated using the RMSE given in Eq. (6).

RMSE aims to obtain the average model prediction error in observations of a time series and ranges from zero to infinity. The lower RMSE value means better forecasting precision (Tak 2018).

RMSE can be expressed mathematically as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \hat{X}_t\right)^2} \tag{6}$$

Here, X_t is the actual values of the time series and $\hat{X}_t$ is the forecasts.

The purpose of this study is to forecast the first 26 weeks of year 2019 with traditional methods such as ARIMA, SES, HES and alternative methods such as ANFIS, ANN, LSTM and R-T1FFs and then apply the proposed FCA-MFFs method. With FCA-MFFs, the aim is not to choose the best method, but to obtain more accurate and robust results and reduce the RMSE. In the study, the forecasting results given by FCA-MFFs are compared with traditional and alternative methods.

The first 103 data are taken as X_{train} and data between 104-129 are taken as $X_{\text{validation}}$, and the last 26 data of 155 are taken as X_{test}.

X_{train} is used for trained the methods and to obtain FCA-MFFs input matrix which is called $Z_{\text{validation}}$. $X_{\text{validation}}$ is used for choosing the best function with the lowest RMSE value. X_{test} is the actual data representing the number of weekly captured immigrants.

X_{test} is tried to be forecasted with the existing and the proposed method and is used to compare the performance of the forecasting results. In order to obtain $Z_{\text{validation}}$, 7 existing methods are trained using X_{train} and the forecasting results are calculated for $X_{\text{validation}}$. To train R-T1FFs, the number of clusters is set from 2 to 5, the AR lag is set from 1 to 5, the MA lag is set from 1 to 2, fuzzy index parameter is set from 1.3 to 3. When training ARIMA, Akaike Information Criteria (AIC) are used for determining the optimum lag parameters. Finally, when training SES, HES, LSTM and ANN, sum of square errors (SSEs) are searched iteratively in R to obtain their optimum parameters.

The results of the trained methods form the input matrix ($Z_{\text{validation}}$) of FCA-MFFs. $Z_{\text{validation}}$ is clustered with FCM and MFFs are obtained. There are two parameters to consider for constituting MFFs: fuzzy index parameter (m) and number of clusters (c). The fuzzy index parameter is varied from 1.5 to 2.5 with the increase rate of 0.5, and the number of clusters is varied from 2 to 4.

The best MFF (MFF_{best}) which give the most accurate results in terms of RMSE is obtained for use in predictions in FCA-MFFs algorithm. Lastly, the forecasting results are obtained by using MFF_{best} for the Z_{test}. The hyperparameters considered for each method are presented in Table 1.

3.1 Continental Shelf Region immigrants dataset

Continental Shelf Region immigrants dataset is used as the first calculation dataset. It consists of data on immigrants detected on the maritime line located within the continen-

Table 1 Hyper-parameters for methods

Methods	Hyper-parameters	Values' range
MFFs	Number of clusters (c)	From 2 to 4
	Fuzziness index parameter (m)	From 1.5 to 2.5
R-T1FFs	Number of clusters (c)	From 2 to 5
	Autoregressive order (p)	From 1 to 5
	Moving Average order (q)	From 1 to 2
	Fuzziness index parameter (m)	From 1.3 to 3
LSTM	Dropout	Between 0.2 and 0.5
	Decay rate	Between 0.97 and 1
	Learning rate	Between 0.01 and 0.1
	Momentum	Between 0.5 and 0.9
	Number of epochs	By a pre-set threshold
	Batch size	32
ANFIS	Dropout	Between 0.2 and 0.5
	Learning rate	Between 0.01 and 0.1
	Momentum	Between 0.5 and 0.9
	Number of epochs	By a pre-set threshold
	Batch size	32
ANN	Learning rate	Between 0.01 and 0.1
	Momentum	Between 0.5 and 0.9.
	Mini batch size	32
ARIMA	Autoregressive order (p)	From 0 to 9
	Degree of differencing (d)	From 0 to 1
	Moving Average order (q)	From 0 to 4
SES	Smoothing level (alpha)	Set for optimal values
	Smoothing slope (Beta)	Set for optimal values
HES	Smoothing level (alpha)	Set for optimal values
	Smoothing slope (Beta)	Set for optimal values
	Smoothing seasonal (Gamma)	Set for optimal values

tal shelf of Turkey. This region geographically refers to the maritime zone in the south, north and west of Turkey and within Republic of Turkey's intervention borders. The relevant dataset consists of the number of immigrants on a weekly basis detected between April 2016 and July 2019. There are 155 observation data belong to Continental Shelf Region. Line plot of these data is presented in Fig. 2.

In FCA-MFFs algorithm, traditional and alternative methods are trained by using X_{train} which is Continental Shelf Region training dataset, and the forecasting results are obtained for $X_{\text{validation}}$. These forecasting results form the $Z_{\text{validation}}$ which is the input matrix for FCA-MFFs. The $Z_{\text{validation}}$ that is given in Table 2 is used to generate MFFs and select the MFFs's best during the training phase of the FCA-MFFs algorithm.

To constitute all functions, FCM is performed using the $Z_{\text{validation}}$ when the number of clusters is set from 2 to 4 and the fuzzy index parameter is set from 1.5 to 2.5 with the increase rate of 0.5. The FCA-MFFs algorithm searches for the best function under these conditions for Continental Shelf Region dataset. The function which has the lowest RMSE value among all iterations is obtained when the number of clusters is 4 and the fuzzy index parameter is 2. This means that the methods are gathered under 4 clusters, that is, we obtain 4 functions. These functions are MFF_1 in Eq. (7), MFF_2 in Eq. (8), MFF_3 in Eq. (9) and MFF_4 in Eq. (10), one of which is the best function with the lowest RMSE. They are constituted according to the weights in Table 3. The weights of the methods which are obtained with Eq. (4), in cluster, that is, in functions, and the RMSE values of the functions for $Z_{\text{validation}}$ are given in Table 3.

$$MFF_1 = 0.026 \times ARIMA + 0.112 \times SES + \cdots + 0.730 \times LSTM + 0.041 \times R-T1FFs \tag{7}$$

$$MFF_2 = 0.045 \times ARIMA + 0.078 \times SES + \cdots + 0.001 \times LSTM + 0.020 \times R-T1FFs \tag{8}$$

$$MFF_3 = 0.073 \times ARIMA + 0.164 \times SES + \cdots + 0.001$$

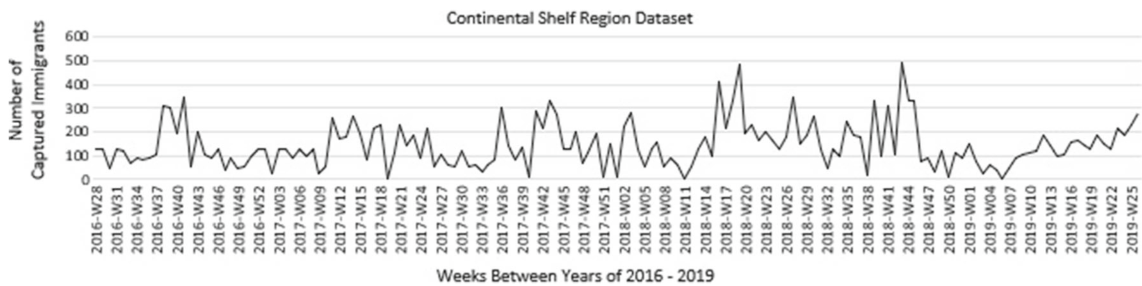

Fig. 2 Line plot of Continental Shelf Region dataset

Table 2 Continental Shelf Region training dataset ($Z_{validation}$) for MFFs

Observation no	t	$X_{validation}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
104	2018-week27	15	34	81	221	62	24	24	39
105	2018-week28	39	29	79	154	71	25	53	10
106	2018-week29	13	26	76	214	67	31	33	30
107	2018-week30	110	28	74	187	69	41	80	38
108	2018-week31	195	30	71	167	73	54	128	39
109	2018-week32	125	50	73	284	74	70	98	82
110	2018-week33	46	74	79	123	31	84	96	74
111	2018-week34	103	81	81	100	61	95	90	41
112	2018-week35	37	76	79	137	73	101	83	99
113	2018-week36	141	79	81	127	72	102	89	11
114	2018-week37	119	74	78	161	73	99	87	78
115	2018-week38	74	83	82	214	71	92	80	89
116	2018-week39	92	88	83	121	73	84	94	69
117	2018-week40	60	86	83	192	74	77	76	64
118	2018-week41	150	87	83	82	74	69	69	100
119	2018-week42	39	84	82	134	72	63	61	90
120	2018-week43	19	91	86	144	69	58	58	77
121	2018-week44	45	85	83	134	73	55	57	39
122	2018-week45	88	77	80	213	71	52	50	22
123	2018-week46	152	73	78	111	74	50	59	130
124	2018-week47	60	75	79	106	74	49	59	65
125	2018-week48	24	84	82	129	73	48	60	51
126	2018-week49	29	81	81	76	74	48	49	34
127	2018-week50	60	74	78	83	74	48	61	66
128	2018-week51	60	69	76	122	74	49	56	83
129	2018-week52	60	68	75	113	74	50	51	47

$$\times LSTM + 0.037 \times R - T1FFs \quad (9)$$

$$MFF_4 = 0.256 \times ARIMA + 0.175 \times SES + \cdots + 0.001 \times LSTM + 0.304 \times R - T1FFs \quad (10)$$

Table 3 indicates that the MFF_{best} is the first function in terms of RMSE. Therefore, forecasts are determined by using MFF_1 for Z_{test}. To obtain Z_{test} which is given in Table 4, existing methods are trained by using the first 129 out of 155 data and the last 26 observations which is test dataset (X_{test}) is forecasted.

Using the Z_{test} matrix and the function MFF_1, MFF_{best} is obtained as in Eq. (11) and the forecasting results of MFF_{best} for FCA-MFFs and the other existing methods are given in Table 5.

$$\text{MFF}_{\text{best}} = MFF_1(Z_{\text{test}}) = 0.026 \times ARIMA + \cdots + 0.730 \times LSTM + 0.041 \times R - T1FFs \quad (11)$$

Table 3 Weights of the methods in functions and RMSE values

Method	MFF_1	MFF_2	MFF_3	MFF_4
ARIMA	0.026	0.045	0.073	0.256
SES	0.112	0.078	0.164	0.175
HES	0.002	0.002	0.659	0.002
ANN	0.000	0.827	0.000	0.000
ANFIS	0.088	0.027	0.065	0.262
LSTM	0.730	0.001	0.001	0.001
R-T1FFs	0.041	0.020	0.037	0.304
RMSE	**200.05***	296.88	252.09	235.20

For the Continental Shelf Region dataset, the LSTM method makes the greatest contribution to the proposed FCA-MFFs method, with a weight of 0.730, as seen in Eq. (11) and Table 3. The LSTM method is followed by ARIMA, SES, HES, ANFIS, R-T1FFs methods with weights of 0.026, 0.112, 0.002, 0.088, 0.041, respectively. For this dataset, ANN does not contribute to the proposed method.

In Table 5, the evaluation of the performance of the proposed method comparing with the existing methods is presented. In terms of RMSE values, the best forecasting results are obtained from the proposed method. The forecasting results of the methods are shown in Fig. 3.

3.2 Region 1 immigrants dataset

Region 1 immigrants dataset is used as the second calculation dataset and refers to the past weekly observation values obtained from the sea area on the west and south coast of Turkey, which coincides with latitudes between 35 degrees and 37 degrees. The relevant dataset consists of the number of immigrants on a weekly basis detected on the maritime line in Region 1 between April 2016 and July 2019. There are 155 observation data belonging to Region 1. Line plot of these data is presented in Fig. 4

In FCA-MFFs algorithm, traditional and alternative methods are trained by using X_{train} which is Region 1 training dataset and the forecasting results are obtained for $X_{\text{validation}}$.

Table 4 Forecasts of Continental Shelf Region dataset (Z_{test}) and RMSE values of the methods

X_{test}	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
403	535	402	238	380	514	372	457
173	506	448	674	366	365	428	535
409	436	433	323	339	298	356	450
388	366	342	292	359	270	377	355
122	377	365	351	317	262	350	248
427	373	373	518	349	267	463	378
385	323	286	481	322	280	323	346
242	344	335	304	363	298	506	200
244	351	352	350	327	318	308	407
555	330	314	439	342	340	461	391
346	314	290	530	336	360	329	482
392	360	382	475	340	378	584	234
502	357	370	644	359	394	426	527
157	364	377	746	793	407	403	415
758	390	421	544	590	417	518	397
568	346	329	688	592	425	458	423
497	423	478	600	582	432	375	319
864	450	510	612	583	439	748	809
87	459	505	524	608	447	220	355
773	534	630	437	604	458	658	717
715	451	441	572	630	472	356	787
499	511	557	770	582	493	486	443
580	549	612	685	578	520	544	553
488	540	573	370	539	558	698	607
749	547	575	929	575	608	534	505
1034	536	545	692	576	675	962	585
RMSE	220.02	210.01	240.82	219.93	198.93	160.20	182.69

Table 5 Forecasting results and RMSE values of existing methods and FCA-MFFs

Actual	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs	FCA-MFFs
403	535	402	238	380	514	372	403	395
173	506	448	674	366	365	428	173	432
409	436	433	323	339	298	356	409	366
388	366	342	292	359	270	377	388	362
122	377	365	351	317	262	350	122	340
427	373	373	518	349	267	463	427	430
385	323	286	481	322	280	323	385	316
242	344	335	304	363	298	506	242	451
244	351	352	350	327	318	308	244	319
555	330	314	439	342	340	461	555	427
346	314	290	530	336	360	329	346	334
392	360	382	475	340	378	584	392	523
502	357	370	644	359	394	426	502	419
157	364	377	746	793	407	403	157	400
758	390	421	544	590	417	518	758	490
568	346	329	688	592	425	458	568	436
497	423	478	600	582	432	375	497	391
864	450	510	612	583	439	748	864	688
87	459	505	524	608	447	220	87	284
773	534	630	437	604	458	658	773	636
715	451	441	572	630	472	356	715	396
499	511	557	770	582	493	486	499	494
580	549	612	685	578	520	544	580	550
488	540	573	370	539	558	698	488	663
749	547	575	929	575	608	534	749	545
1034	536	545	692	576	675	962	1034	862
RMSE	220.02	210.01	240.82	219.93	198.93	160.20	182.69	**159.85***

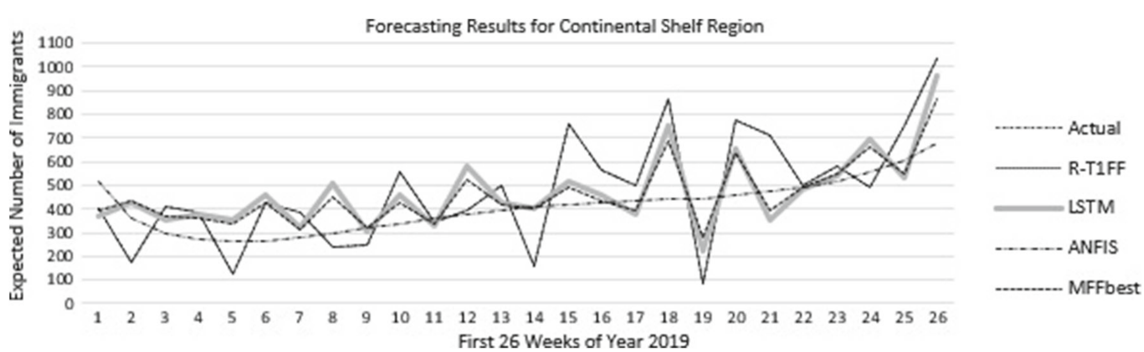

Fig. 3 Comparison forecasting results of methods

These forecasting results form the $Z_{\text{validation}}$ which is the input matrix for FCA-MFFs. The $Z_{\text{validation}}$ that is given in Table 6 is used to generate MFFs and select the MFFs's best during the training phase of the FCA-MFFs algorithm.

To constitute all functions, FCM is performed using the $Z_{\text{validation}}$ when the number of clusters is set from 2 to 4 and the fuzzy index parameter is set from 1.5 to 2.5 with the increase rate of 0.5. The FCA-MFFs algorithm searches for the best function under these conditions for Region 1 dataset. The function which has the lowest RMSE value among all iterations is obtained when the number of clusters is 3 and the fuzzy index parameter is 2. This means that the methods are gathered under 3 clusters, that is, we obtain 3 functions. These functions are MFF_1, MFF_2 and MFF_3, respectively, one of which is MFF_{best} with the lowest RMSE value. The weights of the methods which are obtained with Eq. (4), in cluster,

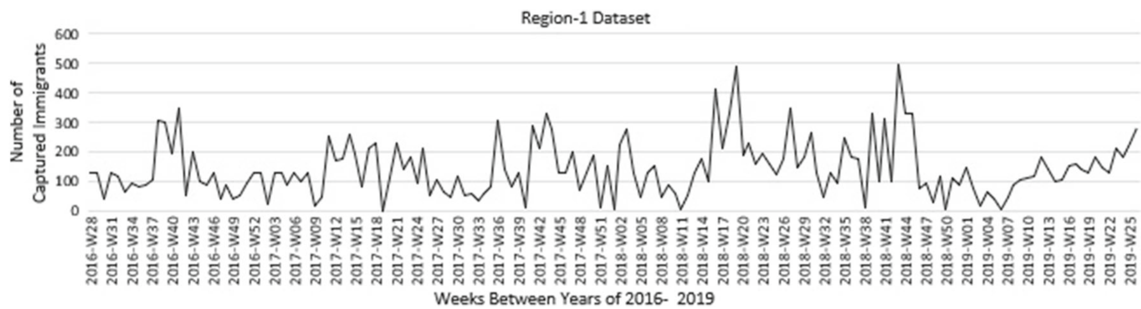

Fig. 4 Line plot of Region 1 dataset

Table 6 Region 1 training dataset ($Z_{\text{validation}}$) for MFFs

Observation no	$X_{\text{validation}}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
104	15	34	81	221	62	24	24	39
105	39	29	79	154	71	25	53	10
106	13	26	76	214	67	31	33	30
107	110	28	74	187	69	41	80	38
108	195	30	71	167	73	54	128	39
109	125	50	73	284	74	70	98	82
110	46	74	79	123	31	84	96	74
111	103	81	81	100	61	95	90	41
112	37	76	79	137	73	101	83	99
113	141	79	81	127	72	102	89	118
114	119	74	78	161	73	99	87	78
115	74	83	82	214	71	92	80	89
116	92	88	83	121	73	84	94	69
117	60	86	83	192	74	77	76	64
118	150	87	83	82	74	69	69	100
119	39	84	82	134	72	63	61	90
120	19	91	86	144	69	58	58	77
121	45	85	83	134	73	55	57	39
122	88	77	80	213	71	52	50	22
123	152	73	78	111	74	50	59	130
124	60	75	79	106	74	49	59	65
125	24	84	82	129	73	48	60	51
126	29	81	81	76	74	48	49	34
127	60	74	78	83	74	48	61	66
128	60	69	76	122	74	49	56	83
129	60	68	75	113	74	50	51	47

that is, in functions, and the RMSE values of the functions for $Z_{\text{validation}}$ are given in Table 7. Table 7 reveals that the MFF_{best} is the first function in terms of RMSE. Therefore, the forecasts are obtained by using MFF_1 for Z_{test}. To obtain Z_{test} which is given in Table 8, existing methods are trained by using the first 129 out of 155 data and the last 26 observations which is test dataset (X_{test}) is forecasted. Using the Z_{test} matrix and the function MFF_1, MFF_{best} is obtained as in Eq. (12) and the forecasting results of MFF_{best} for FCA-MFFs and the other existing methods are given in Table 9.

$$\text{MFF}_{\text{best}} = MFF_1(Z_{\text{test}}) = 0.085 \times ARIMA + \cdots + 0.334 \times LSTM + 0.293 \times R-T1FFs \quad (12)$$

For the Region 1 dataset, the LSTM method makes the greatest contribution to the proposed FCA-MFFs method, with a weight of 0.334, as seen in Eq. (12) and Table 7. The LSTM method is followed by ARIMA, SES, ANN, ANFIS, R-T1FFs methods with weights of 0.085, 0.005, 0.007, 0.276, 0.293, respectively. For this dataset, HES does not contribute to the proposed method.

Table 7 Weights of the methods in functions and RMSE values

Method	MFF_1	MFF_2	MFF_3
ARIMA	0.085	0.000	0.243
SES	0.005	0.000	0.318
HES	0.000	0.998	0.000
ANN	0.007	0.000	0.316
ANFIS	0.276	0.000	0.011
LSTM	0.334	0.001	0.064
R-T1FFs	0.293	0.001	0.048
RMSE	**42.16***	100.05	47.26

In order to evaluate the performance of the proposed method with the existing methods, Table 9 is given. In terms of RMSE values, it is obvious that the best forecasting results are get from the proposed method. The forecasting results of the methods are shown in Fig. 5.

3.3 Region 2 immigrants dataset

Region 2 immigrants dataset is used as the third calculation dataset and refers to the past weekly observation values obtained from the sea area on the west coast of Turkey, which coincides with latitudes between 37 degrees and 38 degrees. The relevant dataset consists of the number of immigrants on a weekly basis detected on the maritime line in Region 2 between April 2016 and July 2019. There are 155 observation data belonging to Region 2. Line plot of these data is presented in Fig. 6.

In FCA-MFFs algorithm, traditional and alternative methods are trained by using X_{train} which is Region 2 training dataset and the forecasting results are obtained for $X_{\text{validation}}$. These forecasting results form the $Z_{\text{validation}}$ which is the input matrix for FCA-MFFs. The $Z_{\text{validation}}$ that is given in Table 10 is used to generate MFFs and select the MFFs's best during the training phase of the FCA-MFFs algorithm.

To constitute all functions, FCM is performed using the $Z_{\text{validation}}$ when the number of clusters is set from 2 to 4 and

Table 8 Forecasts of Region 1 dataset (Z_{test}) and RMSE values of the methods

X_{test}	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
17	60	46	73	64	38	50	32
60	47	46	76	66	43	58	63
60	44	45	80	79	48	46	27
60	45	46	81	80	52	52	61
60	52	46	76	68	54	33	32
60	50	47	303	52	55	49	58
60	54	48	81	46	54	38	56
46	53	48	53	38	52	47	57
60	55	49	95	50	49	42	56
20	53	49	64	53	45	47	49
18	53	49	41	50	42	46	55
60	51	48	198	50	38	43	33
68	49	46	36	54	35	48	39
11	49	47	189	53	32	33	29
60	49	48	46	147	30	38	39
21	50	46	81	64	29	40	35
60	46	47	48	61	28	48	58
68	50	46	94	98	27	29	22
6	45	46	189	90	27	52	48
7	50	47	33	61	28	25	9
100	46	46	69	74	29	55	94
53	48	44	59	80	31	41	8
51	44	46	58	185	33	40	38
18	51	47	67	63	36	32	24
60	44	47	103	60	41	60	78
60	50	46	128	74	46	43	35
RMSE	25.01	23.75	80.42	43.65	24.56	22.72	23.29

Table 9 Forecasting results and RMSE values of $Z_{\text{validation}}$ and FCA-MFF

Actual	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FF	FCA-MFFs
17	60	46	73	64	38	50	32	42
60	47	46	76	66	43	58	63	53
60	44	45	80	79	48	46	27	41
60	45	46	81	80	52	52	61	54
60	52	46	76	68	54	33	32	42
60	50	47	303	52	55	49	58	54
60	54	48	81	46	54	38	56	50
46	53	48	53	38	52	47	57	52
60	55	49	95	50	49	42	56	50
20	53	49	64	53	45	47	49	48
18	53	49	41	50	42	46	55	48
60	51	48	198	50	38	43	33	39
68	49	46	36	54	35	48	39	41
11	49	47	189	53	32	33	29	33
60	49	48	46	147	30	38	39	38
21	50	46	81	64	29	40	35	36
60	46	47	48	61	28	48	58	44
68	50	46	94	98	27	29	22	29
6	45	46	189	90	27	52	48	42
7	50	47	33	61	28	25	9	24
100	46	46	69	74	29	55	94	57
53	48	44	59	80	31	41	8	29
51	44	46	58	185	33	40	38	39
18	51	47	67	63	36	32	24	33
60	44	47	103	60	41	60	78	58
60	50	46	128	74	46	43	35	42
RMSE	25.01	23.75	80.42	43.65	24.56	22.72	23.29	**21.65***

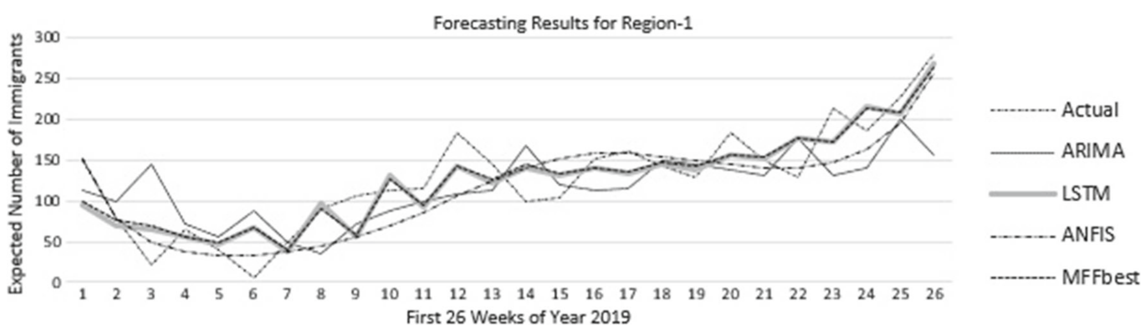

Fig. 5 Comparison forecasting results of methods

the fuzzy index parameter is set from 1.5 to 2.5 with the increase rate of 0.5. The FCA-MFFs algorithm searches for the best function under these conditions for Region 2 dataset. The function which has the lowest RMSE value among all iterations is obtained when the number of clusters is 4 and the fuzzy index parameter is 2. This means that the methods are gathered under 4 clusters, that is, we obtain 4 functions. These functions are MFF_1, MFF_2, MFF_3 and MFF_4, respectively, one of which is MFF_{best} with the lowest RMSE value. The weights of the methods which are obtained with Eq. (4), in cluster, that is, in functions, and the RMSE values of the functions for $Z_{\text{validation}}$ are given in Table 11.

Table 11 reveals that the MFF_{best} is the second function in terms of RMSE. Therefore, the forecasts are obtained by using MFF_2 for Z_{test}. To obtain Z_{test} which is given in Table 12, existing methods are trained by using the first 129

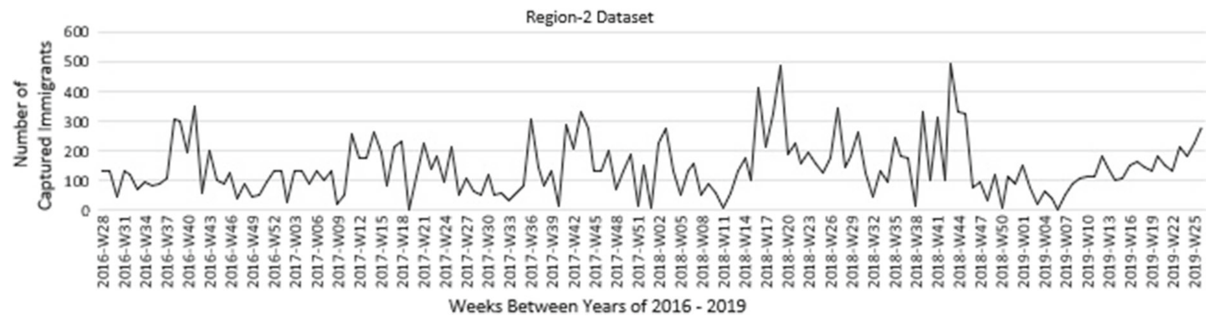

Fig. 6 Line plot of Region 2 dataset

Table 10 Region 2 training dataset ($Z_{\text{validation}}$) for MFFs

Observation no	$X_{\text{validation}}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
104	27	146	68	123	74	66	118	113
105	93	144	78	90	102	81	133	46
106	90	130	70	112	76	90	122	57
107	116	136	73	61	76	93	94	57
108	143	137	76	59	85	91	115	91
109	124	141	82	117	76	88	64	88
110	41	145	92	93	76	85	59	87
111	52	143	97	85	76	82	72	93
112	99	132	88	111	76	80	55	86
113	136	131	82	106	76	80	106	91
114	58	137	85	93	76	81	89	104
115	113	144	93	92	106	83	157	86
116	92	135	87	84	76	87	106	70
117	51	139	91	102	106	92	139	79
118	214	138	92	145	76	97	128	95
119	144	132	85	170	94	103	101	108
120	154	152	105	95	76	109	116	101
121	263	148	111	132	76	115	107	132
122	16	147	118	96	106	120	91	128
123	54	162	141	103	101	124	68	127
124	99	133	121	126	92	127	128	145
125	127	130	111	147	76	129	93	100
126	152	137	109	65	76	129	190	138
127	219	142	112	59	76	128	165	115
128	445	147	118	82	85	127	251	90
129	242	156	134	112	103	124	172	167

out of 155 data and the last 26 observations which is test dataset (X_{test}) is forecasted.

Using the Z_{test} matrix and the function MFF_2, MFF_{best} is obtained as in Eq. (13) and the forecasting results of MFF_{best} for FCA-MFFs and the other existing methods are given in Table 13.

For the Region 2 dataset, the LSTM method makes the greatest contribution to the proposed FCA-MFFs method, with a weight of 0.971, as seen in Eq. (13) and Table 11. The LSTM method is followed by ANN, ANFIS, R-T1FFs methods with weights of 0.010, 0.009, 0.010, respectively. For this dataset, SES and HES do not contribute to the proposed method.

$$\text{MFF}_{\text{best}} = MFF_2(Z_{\text{test}}) = 0.01 \times ANN + 0.009 \times ANFIS + 0.971 \times LSTM \quad (13)$$

In order to evaluate the performance of the proposed method with the existing methods, Table 13 is given. In terms of RMSE values, it is obvious that the best forecasting results are

Table 11 Weights of the methods in functions and RMSE values

Method	MFF_1	MFF_2	MFF_3	MFF_4
ARIMA	0.000	0.000	0.974	0.000
SES	0.002	0.000	0.000	0.264
HES	0.854	0.000	0.000	0.000
ANN	0.091	0.010	0.011	0.230
ANFIS	0.010	0.009	0.003	0.260
LSTM	0.000	0.971	0.000	0.000
R-T1FFs	0.043	0.010	0.011	0.246
RMSE	96.73	**68.26***	89.28	92.14

obtained from the proposed method. The forecasting results of the methods are shown in Fig. 7.

3.4 Region 3 immigrants dataset

Region 3 immigrants dataset is used as the fourth calculation dataset and refers to the past weekly observation values obtained from the sea area on the west coast of Turkey, which coincides with latitudes between 38 degrees and 39 degrees. The relevant dataset consists of the number of immigrants on a weekly basis detected on the maritime line in Region 3 between April 2016 and July 2019. There are 155 observation data belonging to Region 3. Line plot of these data is presented in Fig. 8.

In FCA-MFFs algorithm, traditional and alternative methods are trained by using X_{train} which is Region 3 training dataset and the forecasting results are obtained for $X_{\text{validation}}$. These forecasting results form the $Z_{\text{validation}}$ which is the input matrix for FCA-MFFs. The $Z_{\text{validation}}$ that is given in Table 14 is used to generate MFFs and select the MFFs's best during the training phase of the FCA-MFFs algorithm.

To constitute all functions, FCM is performed using the $Z_{\text{validation}}$ when the number of clusters is set from 2 to 4 and the fuzzy index parameter is set from 1.5 to 2.5 with the increase rate of 0.5. The FCA-MFFs algorithm searches for the best function under these conditions for Region 3 dataset. The function which has the lowest RMSE value

Table 12 Forecasts of Region 2 dataset (Z_{test}) and RMSE values of methods

X_{test}	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
56	168	128	102	33	155	96	224
94	165	151	48	34	124	96	220
256	173	132	49	39	111	144	167
77	172	124	64	27	106	138	121
43	164	151	49	41	105	153	85
274	172	136	46	42	105	147	121
168	174	117	60	44	103	120	135
99	164	149	106	32	101	110	87
71	168	153	36	54	98	91	140
55	171	142	34	43	95	73	114
82	173	127	85	34	93	78	77
70	173	113	129	54	93	62	60
163	172	106	70	54	94	108	52
47	173	99	74	39	98	92	65
293	169	112	112	32	105	187	73
353	174	99	245	22	116	212	99
54	163	138	30	41	132	170	122
304	160	182	135	41	151	465	196
52	173	156	31	43	174	78	216
502	162	186	138	22	198	427	114
354	174	159	35	3	220	149	205
102	153	229	125	10	232	95	163
252	160	254	178	26	229	313	234
213	171	223	65	51	206	172	254
101	165	229	117	43	164	110	152
257	166	226	212	23	114	277	149
RMSE	122.22	122.07	133.58	184.7	118.92	83	131.59

Table 13 Forecasting results and RMSE values of existing methods and FCA-MFFs

Actual	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs	FCA-MFFs
56	168	128	102	33	155	96	224	97
94	165	151	48	34	124	96	220	97
256	173	132	49	39	111	144	167	143
77	172	124	64	27	106	138	121	136
43	164	151	49	41	105	153	85	151
274	172	136	46	42	105	147	121	146
168	174	117	60	44	103	120	135	120
99	164	149	106	32	101	110	87	109
71	168	153	36	54	98	91	140	92
55	171	142	34	43	95	73	114	74
82	173	127	85	34	93	78	77	78
70	173	113	129	54	93	62	60	63
163	172	106	70	54	94	108	52	107
47	173	99	74	39	98	92	65	91
293	169	112	112	32	105	187	73	184
353	174	99	245	22	116	212	99	208
54	163	138	30	41	132	170	122	168
304	160	182	135	41	151	465	196	456
52	173	156	31	43	174	78	216	80
502	162	186	138	22	198	427	114	418
354	174	159	35	3	220	149	205	149
102	153	229	125	10	232	95	163	96
252	160	254	178	26	229	313	234	309
213	171	223	65	51	206	172	254	172
101	165	229	117	43	164	110	152	110
257	166	226	212	23	114	277	149	271
RMSE	122.22	122.07	133.58	184.7	118.92	83	131.59	**82.84***

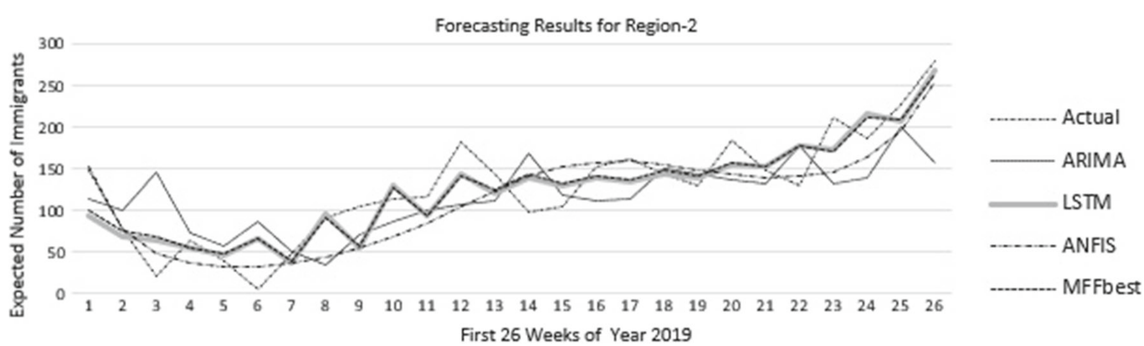

Fig. 7 Comparison forecasting results of methods

among all iterations is obtained when the number of clusters is 4 and the fuzzy index parameter is 2. This means that the methods are gathered under 4 clusters, that is, we obtain 4 functions. These functions are MFF_1, MFF_2, MFF_3 and MFF_4, respectively, one of which is $\mathrm{MFF}_{\mathrm{best}}$ with the lowest RMSE value. The weights of the methods which are obtained with Eq. (4), in cluster, that is, in functions and the RMSE values of the functions for $Z_{\mathrm{validation}}$ are given in Table 15. Table 15 reveals that the $\mathrm{MFF}_{\mathrm{best}}$ is the second function in terms of RMSE. Therefore, the forecasts are obtained by using MFF_2 for Z_{test}. To obtain Z_{test} which is given in Table 16, existing methods are trained by using the first 129 out of 155 data and the last 26 observations which is test dataset (X_{test}) is forecasted. Using the Z_{test} matrix and the function MFF_2, $\mathrm{MFF}_{\mathrm{best}}$ is obtained as in Eq. (14) and the forecasting results of $\mathrm{MFF}_{\mathrm{best}}$ for

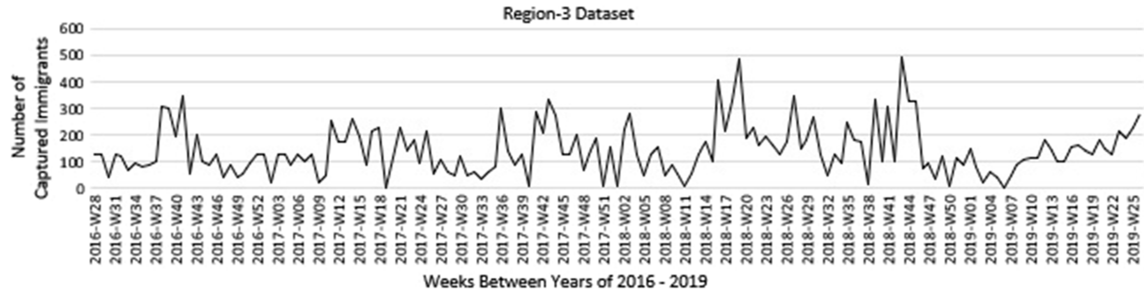

Fig. 8 Line plot of Region 3 dataset

Table 14 Region 3 training dataset ($Z_{\text{validation}}$) for MFFs

Observation no	$X_{\text{validation}}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
104	346	152	185	74	53	234	160	141
105	148	176	184	166	112	226	148	214
106	184	174	182	213	125	200	169	196
107	267	194	181	146	84	170	176	107
108	129	96	180	76	114	144	144	304
109	46	241	178	131	118	126	183	156
110	130	121	177	134	129	116	130	71
111	94	94	175	86	122	112	158	118
112	247	169	174	93	115	113	174	163
113	185	127	173	134	137	119	145	76
114	176	166	171	104	19	128	189	154
115	15	200	170	102	119	139	21	106
116	333	107	168	155	68	150	306	116
117	99	144	167	170	127	160	93	135
118	311	164	166	181	146	167	312	224
119	101	212	164	316	131	170	108	76
120	494	92	163	260	150	167	470	270
121	330	250	161	304	156	160	131	200
122	328	141	160	333	25	148	330	414
123	78	336	159	179	116	135	121	155
124	93	81	157	225	15	120	88	245
125	33	201	156	148	120	105	111	131
126	121	69	155	185	70	92	112	168
127	8	151	153	170	63	80	109	86
128	114	119	152	123	126	71	121	71
129	88	125	150	172	17	63	82	35

FCA-MFFs and the other existing methods are given in Table 17.

$$\text{MFF}_{\text{best}} = MFF_2(Z_{\text{test}}) = 0.062 \times ARIMA + \cdots + 0.879 \times LSTM + 0.019 \times R - T1FF \tag{14}$$

For the Region 3 dataset, the LSTM method makes the greatest contribution to the proposed FCA-MFFs method, with a weight of 0.879, as seen in Eq. (14) and Table 15. The LSTM method is followed by ARIMA, SES, HES, ANFIS, R-T1FFs methods with weights of 0.062, 0.001, 0.005, 0.052, 0.019, respectively. For this dataset, ANN does not contribute to the proposed method.

In order to evaluate the performance of the proposed method with the existing methods, Table 17 is given. In terms of RMSE values, it is obvious that the best forecasting results are obtained from the proposed method. The forecasting results of the methods are shown in Fig. 9.

Table 15 Weights of the methods in functions and RMSE values

Method	MFF_1	MFF_2	MFF_3	MFF_4
ARIMA	0.123	0.062	0.227	0.116
SES	0.041	0.001	0.353	0.046
HES	0.004	0.005	0.006	0.592
ANN	0.635	0.000	0.003	0.002
ANFIS	0.102	0.052	0.284	0.066
LSTM	0.003	0.879	0.003	0.004
R-T1FFs	0.093	0.019	0.124	0.175
RMSE	130.07	**72.01***	115.12	111.80

3.5 Region 4 immigrants dataset

Region 4 immigrants dataset is used as the last calculation dataset and refers to the past weekly observation values obtained from the sea area on the west and north coast of Turkey, which coincides with latitudes between 39 degrees and 41 degrees. The relevant dataset consists of the number of immigrants on a weekly basis detected on the maritime line in Region 4 between April 2016 and July 2019. There are 155 observation data belonging to Region 4. Line plot of these data is presented in Fig. 10.

In FCA-MFFs algorithm, traditional and alternative methods are trained by using X_{train} which is Region 4 training dataset and the forecasting results are obtained for $X_{\text{validation}}$. These forecasting results form the $Z_{\text{validation}}$ which is the input matrix for FCA-MFFs. The $Z_{\text{validation}}$ that is given in Table 18 is used to generate MFFs and select the MFFs's best during the training phase of the FCA-MFFs algorithm.

To constitute all functions, FCM is performed using the $Z_{\text{validation}}$ when the number of clusters is set from 2 to 4 and the fuzzy index parameter is set from 1.5 to 2.5 with the increase rate of 0.5. The FCA-MFFs algorithm searches for the best function under these conditions for Region 4 dataset. The function which has the lowest RMSE value among all iterations is obtained when the number of clusters is 4 and the fuzzy index parameter is 2.5. This means that the methods are gathered under 4 clusters, that is, we

Table 16 Forecasts of Region 3 dataset (Z_{test}) and RMSE values of each method

X_{test}	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs
149	114	76	127	120	153	94	88
79	100	79	242	69	78	70	146
21	146	82	179	41	49	64	80
64	73	86	173	98	37	55	20
40	57	89	132	98	33	46	65
5	87	92	10	95	33	68	44
49	50	96	48	95	37	38	9
91	36	99	174	98	44	96	54
106	71	102	72	98	55	56	95
114	88	106	97	98	69	131	108
116	100	109	190	121	86	93	109
183	108	112	216	99	106	143	116
144	112	115	224	95	124	122	181
99	169	119	313	94	141	140	142
105	120	122	246	98	152	131	99
153	112	125	157	150	158	140	104
162	115	129	168	116	159	134	151
143	150	132	187	98	155	144	160
130	145	135	73	90	149	139	141
184	138	139	165	81	144	156	128
149	132	142	178	41	140	153	182
130	177	145	191	97	141	177	147
213	132	149	232	41	147	173	129
186	140	152	142	97	164	217	211
228	200	155	255	95	196	207	185
281	157	159	106	94	256	269	230
RMSE	38.06	40.63	88.38	74.68	33.61	30.47	42.18

Table 17 Forecasting results and RMSE values of existing methods and FCA-MFFs

Actual	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs	FCA-MFFs
149	114	76	127	120	153	94	88	100
79	100	79	242	69	78	70	146	76
21	146	82	179	41	49	64	80	70
64	73	86	173	98	37	55	20	56
40	57	89	132	98	33	46	65	48
5	87	92	10	95	33	68	44	68
49	50	96	48	95	37	38	9	39
91	36	99	174	98	44	96	54	91
106	71	102	72	98	55	56	95	58
114	88	106	97	98	69	131	108	127
116	100	109	190	121	86	93	109	95
183	108	112	216	99	106	143	116	142
144	112	115	224	95	124	122	181	126
99	169	119	313	94	141	140	142	145
105	120	122	246	98	152	131	99	133
153	112	125	157	150	158	140	104	141
162	115	129	168	116	159	134	151	137
143	150	132	187	98	155	144	160	148
130	145	135	73	90	149	139	141	142
184	138	139	165	81	144	156	128	157
149	132	142	178	41	140	153	182	154
130	177	145	191	97	141	177	147	178
213	132	149	232	41	147	173	129	172
186	140	152	142	97	164	217	211	213
228	200	155	255	95	196	207	185	210
281	157	159	106	94	256	269	230	265
RMSE	38.06	40.63	88.38	74.68	33.61	30.47	42.18	**30.26***

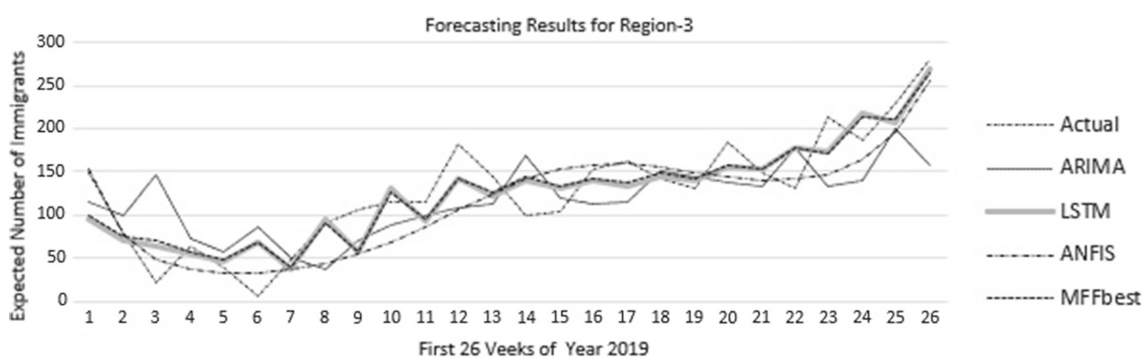

Fig. 9 Comparison forecasting results of methods

obtain 4 functions. These functions are MFF_1, MFF_2 and MFF_3 and MFF_4, respectively, one of which is $\mathrm{MFF}_{\mathrm{best}}$ with the lowest RMSE value. The weights of the methods which are obtained with Eq. (4), in cluster, that is, in functions and the RMSE values of the functions for $Z_{\mathrm{validation}}$ are given in Table 19.

Table 19 reveals that the $\mathrm{MFF}_{\mathrm{best}}$ is the first function in terms of RMSE. Therefore, the forecasts are obtained by using MFF_1 for Z_{test}. To obtain Z_{test} which is given in Table 20, existing methods are trained by using the first 129 out of 155 data and the last 26 observations which is test dataset (X_{test}) is forecasted. Using the Z_{test} matrix and the function MFF_1, $\mathrm{MFF}_{\mathrm{best}}$ is obtained as in Eq. (15) and the forecasting results of $\mathrm{MFF}_{\mathrm{best}}$ for FCA-MFFs and the other existing methods are given in Table 21.

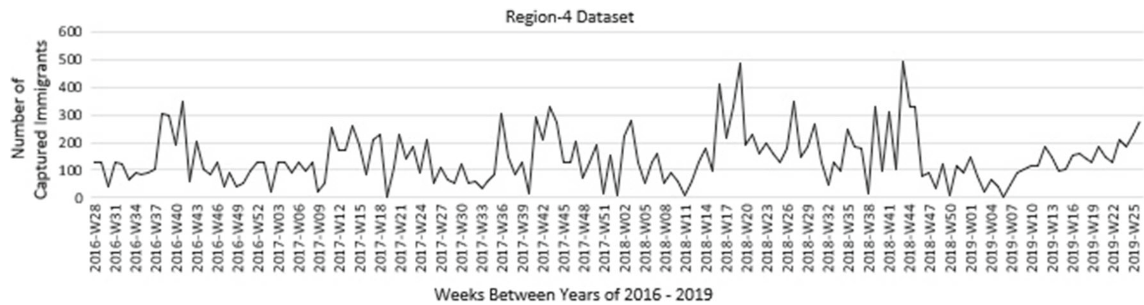

Fig. 10 Line plot of Region 4 dataset

Table 18 Region 4 training dataset ($Z_{\text{validation}}$) for MFFs

Observation no	$X_{\text{validation}}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FF
104	105	183	108	114	177	135	202	194
105	148	178	110	200	200	121	184	194
106	174	174	109	215	216	123	162	194
107	271	176	117	202	218	136	273	194
108	205	179	129	172	250	156	145	194
109	92	188	158	354	116	181	215	194
110	174	188	168	179	177	209	231	194
111	198	178	152	217	198	235	205	194
112	227	180	157	272	315	256	208	194
113	228	183	165	231	141	268	176	194
114	405	187	178	207	230	270	191	194
115	313	189	188	141	131	264	282	194
116	437	204	233	202	218	251	202	194
117	293	205	249	182	250	235	227	194
118	227	215	288	201	86	217	164	194
119	174	209	289	131	204	199	195	194
120	86	200	276	116	217	182	104	194
121	275	191	255	145	262	168	165	194
122	90	180	220	182	144	157	81	194
123	68	189	232	256	223	147	158	194
124	91	179	203	296	160	141	72	194
125	169	172	175	116	192	136	196	194
126	288	170	158	138	215	133	254	194
127	57	175	160	121	225	131	121	194
128	204	188	186	139	167	131	203	194
129	206	176	160	123	231	132	264	194

$$\text{MFF}_{\text{best}} = MFF_1(Z_{\text{test}}) = 0.001 \times ARIMA + +0.999 \times LSTM \tag{15}$$

For the Region 4 dataset, the LSTM method makes the greatest contribution to the proposed FCA-MFFs method, with a weight of 0.999, as seen in Eq. (15) and Table 19. The LSTM method is followed by ARIMA with a weight of 0.001. For this dataset, SES, HES, ANN, ANFIS and R-T1FF do not contribute to the proposed method.

In terms of RMSE values, it is obvious that the best forecasting results are obtained from the proposed method. The forecasting results of the methods are shown in Fig. 11.

As a summary, the RMSE performances obtained from the applied methods for each dataset are presented in Table 22. As seen in Table 22, the proposed FCA-MFFs method has

Table 19 Weights of the methods in functions and RMSE values

Method	MFF_1	MFF_2	MFF_3	MFF_4
ARIMA	0.001	0.301	0.018	0.004
SES	0.000	0.000	0.727	0.000
HES	0.000	0.000	0.000	0.857
ANN	0.000	0.249	0.048	0.037
ANFIS	0.000	0.140	0.203	0.099
LSTM	0.999	0.000	0.000	0.000
R-T1FF	0.000	0.309	0.004	0.003
RMSE	**82.55***	90.17	97.79	111.41

the lowest RMSE value and gives more accurate forecasting results in all datasets. Likewise, in all datasets handled in practice, LSTM method gives the best forecasting results after FCA-MFFs. Accordingly, LSTM gives the greatest contribution in the FCA-MFFs method in all datasets.

Furthermore, computational complexity with running time for FCA-MFFs is presented in Table 23.

4 Conclusions

With this study, as a part of a national security project in Turkey in order to make more effective immigration detection of the Turkish Coast Guard Command, the number of immigrants in the maritime line is forecasted for the first time. Thus, by forecasting immigration movement on a weekly basis before it happens, it is aimed to ensure that the intervention to immigration is operative for national security.

In the study, 7 existing forecasting methods which are traditional and alternative with different insights are applied on 5 different datasets consisting of real-world historical observation data, and since the data for 2020 and 2021 could not be obtained yet due to COVID-19, weekly forecasts of the number of immigrants for the first half of 2019 are obtained. The results show that each forecasting method has different performance. This makes it possible to obtain more accurate and reliable results by combining forecasting methods for each dataset. Therefore, in order to obtain more accurate results and increase the success of the study, the proposed

Table 20 Forecasts of Region 4 dataset (Z_{test}) and RMSE values of each method

$X_{\text{validation}}$	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FF
181	171	117	173	135	207	145	254
174	156	137	125	129	190	100	50
132	144	146	42	215	166	124	176
247	160	152	43	202	143	244	168
39	177	148	85	163	125	146	176
148	147	170	341	164	114	168	163
168	184	141	132	205	110	140	125
105	224	143	43	192	111	163	176
67	164	148	154	154	117	81	126
366	188	139	175	180	128	117	141
130	226	123	34	192	141	230	187
139	120	176	32	188	155	169	140
127	102	166	149	175	168	146	183
174	202	160	85	227	175	177	237
360	188	153	37	195	175	339	126
41	175	158	180	151	168	104	152
281	86	202	201	204	154	265	169
349	145	167	91	215	136	216	210
29	178	192	37	189	118	40	252
80	46	226	32	260	103	70	132
112	160	183	152	235	91	122	54
214	260	161	182	180	86	289	304
64	209	150	169	231	89	75	79
71	161	164	41	160	105	102	80
420	185	142	195	229	147	364	87
496	242	127	130	180	254	332	225
RMSE	123.65	134.68	147.79	124.16	117.35	77.39	133.98

Table 21 Forecasting results and RMSE values of existing methods and FCA-MFF

Actual	ARIMA	SES	HES	ANN	ANFIS	LSTM	R-T1FFs	FCA-MFFs
181	171	117	173	135	207	145	254	145
174	156	137	125	129	190	100	50	101
132	144	146	42	215	166	124	176	125
247	160	152	43	202	143	244	168	244
39	177	148	85	163	125	146	176	147
148	147	170	341	164	114	168	163	169
168	184	141	132	205	110	140	125	141
105	224	143	43	192	111	163	176	164
67	164	148	154	154	117	81	126	81
366	188	139	175	180	128	117	141	118
130	226	123	34	192	141	230	187	230
139	120	176	32	188	155	169	140	170
127	102	166	149	175	168	146	183	146
174	202	160	85	227	175	177	237	178
360	188	153	37	195	175	339	126	339
41	175	158	180	151	168	104	152	105
281	86	202	201	204	154	265	169	265
349	145	167	91	215	136	216	210	217
29	178	192	37	189	118	40	252	41
80	46	226	32	260	103	70	132	71
112	160	183	152	235	91	122	54	123
214	260	161	182	180	86	289	304	290
64	209	150	169	231	89	75	79	76
71	161	164	41	160	105	102	80	103
420	185	142	195	229	147	364	87	365
496	242	127	130	180	254	332	225	333
RMSE	123.65	134.68	147.79	124.16	117.35	77.39	133.98	**77.34***

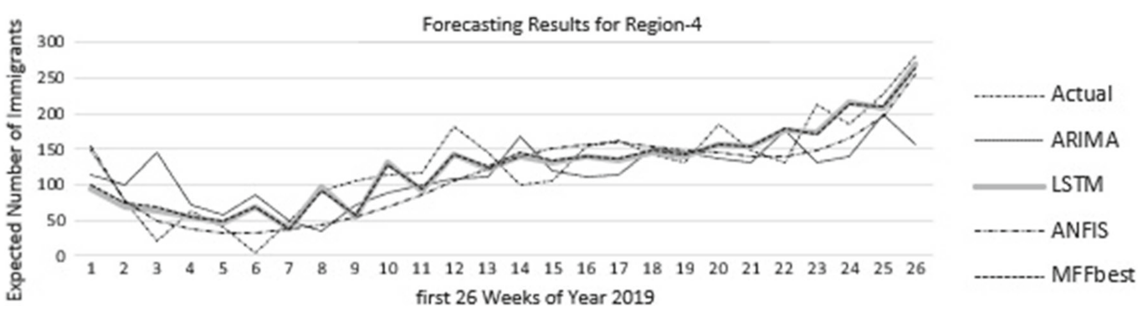

Fig. 11 Comparison forecasting results of methods

method FCA-MFFs, which is an FCM-based approach that combines existing forecasting methods, is applied.

The results show that FCA-MFFs perform better than the other existing methods. The best performances for forecasting in terms of RMSE values for all datasets are obtained from FCA-MFFs. Moreover, for these datasets, it is seen that LSTM gives more performance results than other existing methods and provides the most contribution to the proposed FCA-MFFs method in this study. However, as with LSTM, a common consensus cannot be reached on their performance in all datasets for other state-of-the-art methods such as R-T1FFs, ANFIS and ANN methods in the study. R-T1FFs provide the largest contribution after LSTM for the Continental Shelf Region dataset and Region 1 dataset, while providing ANFIS for the Region 2, Region 3 and Region 4 dataset. It is remarkable that ANN is one of the methods that provide the least contribution in some datasets as it does not contribute at all in some datasets such as Continental Shelf

Table 22 RMSE values of methods for each region's dataset

Forecasting methods	Continental Shelf Region	Region 1	Region 2	Region 3	Region 4
ARIMA	220.02	25.01	122.22	38.06	123.65
SES	210.01	23.75	122.07	40.63	134.68
HES	240.82	80.42	133.58	88.38	147.79
ANN	219.93	43.65	184.70	74.68	124.16
ANFIS	198.93	24.56	118.92	33.61	117.35
LSTM	160.20	22.72	83.00	30.47	77.39
R-T1FFs	182.69	23.29	131.59	42.18	133.98
FCA-MFFs	**159.85***	**21.65***	**82.84***	**30.26***	**77.34***

Table 23 Computational complexity with running time (sec) for FCA-MFFs

Datasets	Running time (sec)
Continental Shelf Region	5.93
Region 1	5.11
Region 2	4.76
Region 3	5.16
Region 4	4.98

Region dataset, Region 3 dataset and Region 4 dataset. As for the traditional methods, ARIMA, SES and HES, it seems that they do not contribute much to the FCA-MFFs method for these datasets. However, considering different datasets, all these methods may perform differently and contribute to the proposed FCA-MFFs approach in a completely different weights. The FCA-MFFs method provides the most robust forecasting results for the dataset of interest by allowing the use of different forecasting methods in proportion to their contribution weights, rather than using a single method that is considered to give the best forecasting results.

Another point that draws attention in the study is that some methods show different performance in the regions where they are applied. The reason for this is that the structure of the immigrant dataset belonging to the regions is different. The same method may perform differently on different datasets. Therefore, in practice, it is difficult to choose the most suitable forecasting method since the structure of the datasets will not be known initially. To meet this challenge, it is necessary to combine various forecasting methods that can deal with any structure of data. We provide this with the FCA-MFFs algorithm in the study.

Although the purpose of the study is forecasting the number of immigrants on a weekly basis for a national security project, the study is unique in applying 7 different forecasting methods that are ARIMA, SES, HES, ANFIS, ANN, LSTM and R-T1FFs on the real immigrants data for the first time. Moreover, the proposed FCA-MFFs method, which is a combination of these 7 methods, is also applied for the first time and more reliable results are obtained with this method.

As a future study, it is aimed to conduct an efficient Mixed Patrol Vehicle Routing by using the forecasting results obtained weekly basis with this study. An efficient patrol vehicle route, which is arranged depending on the density of the expected number of immigrants, will ensure that more immigrants are caught and therefore more effective patrols.

Acknowledgements I would like to thank to Turkish Coast Guard Command within Ministry of Interior of Republic of Turkey and The Scientific and Technological Research Council of Turkey (TUBITAK) for their support.

Author Contributions FCC involved in conceptualization, investigation, methodology, formal analysis, resources, supervision, validation, software programmer, writing—original draft. BG took part in supervision, data curation, formal analysis, methodology, formal analysis, writing—review & editing. NT involved in supervision, software programmer, methodology, formal analysis, writing—review & editing. TK involved in conclusion, supervision, methodology & review.

Funding This study is funded by Turkish Coast Guard Command within Ministry of Interior of Republic of Turkey, The Scientific and Technological Research Council of Turkey (TUBITAK) and ASELSAN.

Availability of data and material Data used in this study are confidential by Turkish Coast Guard Command within Ministry of Interior of Republic of Turkey.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies involving animals performed by any of the authors.

Informal consent Informed consent was obtained from all individual participants included in the study.

References

Abel G, Bijak J, Findlay A, McCollum D, Wisniowski A (2013) Forecasting environmental migration to the United Kingdom: an exploration using Bayesian models. Popul Environ 35(2):183–203

Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, Saidur R (2014) A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 33:102–109

Alho J, Alders M, Cruijsen H, Keilman N, Nikander T, Pham DQ (2006) New forecast: population decline postponed in Europe. Stat J U N Econ Comm Eur 23:1–10

Atmaca H, Cetisli B, Yavuz HS (2001) The comparison of fuzzy inference systems and neural network approaches with ANFIS method for fuel consumption data. In: Second international conference on electrical and electronics engineering Papers ELECO'2001. Bursa Turkey

Barak S, Sadegh SS (2016) Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm. Int J Electr Power Energy Syst 82:92–104

Bates JM, Granger CWJ (1969) The combination of forecasts. Oper Res Q 20:451–468

Bijak J (2006) Forecasting international migration: selected theories, models, and methods. Central European Forum for Migration Research 4, Poland

Bijak J (2010) Forecasting international migration in Europe: a Bayesian view. Springer Series on Demographic Methods and Population Analysis 24

Bijak J, Wisniowski A (2010) Bayesian forecasting of immigration to selected European countries by using expert knowledge. J Roy Stat Soc 4:775–796

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control holden-day. San Francisco

Brown RG (1959) Statistical forecasting for inventory control. McGraw/Hill, USA

BuHamra N, Smaoui MG (2003) The Box-Jenkins analysis and neural networks: prediction and time series modeling. Appl Math Model 27:805–815

Buyuksahin UC, Ertekin S (2019) Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. Neurocomputing 361:151–163

Cappelen A, Skjerpen T, Tonnessen M (2015) Forecasting immigration in official population projections using an econometric model. Int Migr Rev 49(4):945–980

Celikyilmaz A, Turksen B (2009) Modeling uncertainty with fuzzy logic. Springer Book Series

Chen YY, Lin YH, Kung CC, Chung MH, Yen IH (2019) Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes. Sensors. 19(9):2047

Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons Fractals 135:109864

Disney G (2014) Model-based estimates of UK immigration. University of Southampton, UK

Egrioglu E, Aladag CH, Yolcu U, Basaran M, Uslu VR (2009) A new hybrid approach based on SARIMA and partial high order bivariate fuzzy time series forecasting model. Expert Syst Appl 36:7424–7434

Elsheikh AH, Katekar VP, Muskens OL, Deshmukh SS, Elaziz MA, Dabour SM (2021) Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. Process Saf Environ Prot 148:273–282

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Holt CE (1957) Forecasting seasonals and trends by exponentially weighted averages. Carnegie Institute of Technology, USA

Jain A, Kumar AM (2007) Hybrid neural network models for hydrological time series forecasting. Appl Soft Comput 7:585–592

Jang JSR (1993) ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

Karaboga D, Kaya E (2020) Estimation of number of foreign visitors with ANFIS by using ABC algorithm. Soft Comput 24:7579–7591

Li ZP, Yu H, Liu YC, Liu FQ (2008) An improved adaptive exponential smoothing model for short-term travel time forecasting of Urban Arterial Street. Acta Autom Sinica 34(11):1404–1409

Li X, Kang Y, Li F (2020) Forecasting with time series imaging. Exp Syst Appl 160:113680

Lutz W, Goldstein JR (2004) Introduction: how to deal with uncertainty in population forecasting. Int Stat Rev 72(1):1–4

Ma Q (2020) Comparison of ARIMA, ANN and LSTM for stock price prediction. E3S Web of Conferences 218, 01026

Manso PM, Athanasopoulos G, Hyndman RJ (2020) FFORMA: Feature-based forecast model averaging. Int J Forecast 36:86–92

Martineau JS (2010) Red flags: a model for the early warning of refugee outflows. J Immigr Refug Stud 8(2):135–157

Matamoros AH, Fujita H, Hayashi T, Meana HP (2020) Forecasting of COVID19 per regions using ARIMA models and polynomial functions. Appl Soft Comput 96:106610

McCulloch WS, Pitts W (1943) A logical calculus of ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133

Namini SS, Tavakoli N, Namin AS (2018) A comparison of ARIMA and LSTM in forecasting time series. In: IEEE international conference on machine learning and applications

Raymer J, Wisniowski A (2018) Applying and testing a forecasting model for age and sex patterns of immigration and emigration. J Demogr 72(3):339–355

Raymer J, Wisniowski A, Forster JJ, Smith PWF, Bijak J (2013) Integrated modelling of European migration. J Am Stat Assoc 108:801–819

Sak H, Senior A, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Proc, İnterspeech, USA

Shahid F, Zameer A, Muneeb M (2020) Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons Fractals 140:110212

Shaub D (2020) Fast and accurate yearly time series forecasting with forecast combinations. Int J Forecast 36:116–120

Suriani S, Ibn AU, Shaikot HM (2019) A predictive model for the population growth of refugees in Asia: a multiple linear regression approach. J Comput Theor Nanosci 16(3):1196–1202

Tak N (2018) Meta fuzzy functions: application of recurrent type-1 fuzzy functions. Appl Soft Comput 73:1–13

Tak N (2020) Meta fuzzy index functions. Comm Faculty Sci Univ Ankara Series A1 Math Statis 69(1):654–667

Tak N (2021) Forecast combination with meta possibilistic fuzzy functions. Inf Sci 160:168–182

Tak N, Gok A (2020) Dating currency crises and designing early warning systems: meta-possibilistic fuzzy index functions. Int J Financ Econ pp 1-18

Tak N, Tez M, Evren A, Egrioglu E (2018) Recurrent type-1 fuzzy functions approach for time series forecasting. Appl Intell 48(1):68–77

Thomson ME, Pollock AC, Onkal D, Gonul MS (2019) Combining forecasts: performance and coherence. Int J Forecast 35:474–484

Tseng M, Yu HC, Tzeng GH (2002) Combining neural network model with seasonal time series ARIMA model. Technol Forecast Soc Chang 69:71–87

Turksen B (2008) Fuzzy functions with LSE. Appl Soft Comput 8(3):1178–1188

Uslu VR, Aladag CH, Yolcu U, Egrioglu E (2010) A new hybrid approach for forecasting a seasonal fuzzy time series. In: Interna-

tional symposium computing science and engineering proceeding book pp 1152-1158
Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M (2020) A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Convers Manage 212:112766
Wicke L, Dhami MK, Onkal D, Belton IK (2019) Using scenarios to forecast outcomes of a refugee crisis. International Journal of Forecasting in Press, Corrected Proof, Available online
Winters PR (1960) Forecasting sales by exponentially weighted moving averages. Manage Sci 6:324–342
Wisniowski A (2013) Bayesian modelling of international migration with labour force survey data. Collegium of Economic Analyses, Warsaw, USA
Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175

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