IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 27, 2019, accepted October 14, 2019, date of publication October 16, 2019, date of current version November 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947712

Hybrid Artificial Intelligence Based Automatic Determination of Travel Preferences of Chinese Tourists

İBRAHIM TOPAL[1] AND MUHAMMED KÜRŞAD UÇAR[2]

[1]Department of Computer Technologies, Naval Petty Officer Vocational School, National Defence University, 77720 Yalova, Turkey
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering, Sakarya University, 54050 Sakarya, Turkey

Corresponding author: Muhammed Kürşad Uçar (mucar@sakarya.edu.tr)

ABSTRACT **Background and Objective:** Tourism and travel sector continues to grow by gaining an important place in the world economy and many countries want to increase their share in this sector. At the same time, it is known that today's consumer tourism and travel purchase decisions are influenced by social media. By examining the data of consumers on social media, it is possible for businesses to reach the right person and get more efficiency from high-cost promotion activities. The study aims to analyze the historical data of users on TripAdvisor with artificial intelligence methods to reveal a profile of consumers who might prefer Turkey. **Methods:** In this context, TripAdvisor, which is one of the best-known websites in the tourism sector, is an important source of data for countries to increase their share in the tourism market. Inferences can be made by using artificial intelligence methods and the data in TripAdvisor together. In this study, as a case study, the potentials of Chinese tourists to prefer Turkey are dealt because Turkey has increased its tourism targets ten folds for China and the year 2018 was declared as "Turkey Tourism Year" in China. In this context, this study aims to determine the potentials of Chinese tourists to prefer Turkey, by processing travel data histories obtained from TripAdvisor with artificial intelligence methods. It is expected that the study will contribute to the tourism sector as well as the academic literature. The study used the travel data history of Chinese tourists taken from TripAdvisor. Significant travel histories were selected by the F-score method. Depending on the selected and all travel histories of users, their travel preferences (Turkey/France) were classified by artificial intelligence algorithms. The developed model was tested with performance criteria. **Results:** At the end of the study, it was ensured that the Chinese, who would prefer Turkey, were determined with an accuracy rate of 75.25% and sensitivity rate of 0.76. **Conclusions:** It was observed that it is possible to find the tourists who will prefer Turkey by using the developed system. In other words, the study revealed that the countries can reach the individual instead of masses in their promotional activities.

INDEX TERMS Marketing, tourism, travel, user profile, turkey tourism year, Chinese tourism, tripadvisor, artificial intelligence algorithms, ensemble classification.

I. INTRODUCTION

The developments in mobile devices and web technologies that have started with the spread of the internet and the information and communication technologies that have reached its present form have also affected the tourism and travel sector as well as other sectors. Information and communication technologies manifest themselves in many different ways because of the structure that can influence the shift in the power balance in the competitive environment of tourism [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Hiram Ponce.

Tourism expenditures, which have increased in recent years due to the decrease in transportation costs [2], [3], make this sector attractive and require the countries to benefit from developments in information and communication technology in order to increase their market share. As an important tourism route, Turkey is seen as one of these countries with its targets put and investments made. One of Turkey's goals includes China in which the number of abroad travels is increasing and which is the most crowded country in the world. Turkey has revealed its targets to raise the number of Chinese tourists to 3 million by increasing it approximately 10-fold, on the occasion of the declaration of 2018 as "Turkey

Tourism Year'' in China [4]. It is likely to contribute to determining the Chinese tourists' Turkey preferences, with data mining and artificial intelligence that are the opportunities offered by information and communication technologies.

The internet is the first tool that comes to mind when developments in information and communication technologies are mentioned. The internet is experienced by 4.2 billion people, which corresponds to 54% of the world population, due to its rapid spread around the world [5]. For users, the internet has got meaning with the web. The internet structure, in which initially only a small number of information providers were available and a large number of people were only able to read these resources, has changed over time. The technology that is called Web 2.0 [6] and allows social media sites to be used has made it easier for users to share content. Today's users can share their ideas and opinions effortlessly through social media and mobile technologies at any time, from anywhere.

Social media, which has an indisputable effect on human life, can be used at every stage of purchasing. Consumers use social media to realize the need, evaluate alternatives, determine what when and where to buy, and make recommendations after purchasing [7]. For example, if a consumer shares his/her comments, location, and photographs during the break of the cinema film, this will excite the other consumers to watch the movie, besides creating a response to the questions like where and when they can watch the film. As revealed in the study of Nielsen (2013), the high trust of consumers in online consumer ideas with a ratio of 68% [8] increases the importance of these websites. Besides, the fact that social media is used by 2.8 billion people which constitute 73% of all internet users [9] increases the number of comments and readers. Many social media users benefit from this structure consisting of different types of websites that can appeal to their needs.

Social media is a set of web-based applications, based on Web 2.0 that allow content to be shared and exchanged [10], and it has a variety of tools. These tools can be classified as blogs, microblogs, social networks, media sharing, social news and labeling, voting and rating, forums and virtual worlds [11]. Although consumers can evaluate products or services on most of these tools, voting and rating sites are the sites mainly established for this purpose. While rating on voting and rating websites can be done with stars and points, open-ended texts can be used to convey the opinions and evaluations to other people. In this context, in addition to the electronic word-of-mouth being influential in the consumer purchasing decision-making process [12], the abstract, inextricable and non-standardizable structure of the service sector increases the value of communication for consumers in this field [13]. Voting and rating sites help consumers to make decisions in the service sector and, in particular, in the tourism and travel sector, in which comparison with technical data is difficult.

In 2018, Martin-Fuentes et al. conducted a comprehensive hotel classification study. This study aims to determine the parameters used in the classification of hotels. The hotel features such as the number of reviews, scoring, price, room rates, cleanliness and location, which were not used in the standard procedure but were useful in the evaluation, were revealed with support vector machines, one of the machine learning techniques [14]. From this point of view, we can think that new information can be revealed in the tourism sector with machine learning methods.

The expectations from tourism and travel, and the satisfaction arising after purchasing the service may vary according to consumers' culture, age, and experience. In this context, consumers can change their travel plans according to the comments made in social media by people thinking similar to them [15], and the travel intention and trust in the destination [16] can change. TripAdvisor is one of the most popular voting and rating sites in the tourism and travel sector. There are 630 million consumer reviews on TripAdvisor under the name of hotels, airline tickets, restaurants and things to do. The number of visitors to the site per month is 455 million [17]. Besides, according to New Oxford Economics Study, 10.3% of the world's travelers are influenced by TripAdvisor, revealing that TripAdvisor is a social media platform for both travelers and information exchange for those who will travel [18]. On the TripAdvisor site, there are also places the users have gone before in addition to the photographs, comments, and data, which are presented by users to all other users as open. Besides the benefits consumers provide to other consumers, the data generated by being ''visible'' [19], while providing data for many academical studies [20]–[24], can be used by the actors of the tourism sector in according with their marketing objectives. Countries wishing to increase their share in the growing world tourism market are likely to make inferences with various methods from the information provided by tourism-oriented voting and rating sites.

In 2016, the tourism and travel sector globally reached the volume of 7.6 trillion dollars corresponding to 10.2% of the total gross national product, and its employment capacity became 292 million people. Furthermore, the number of international travelers is approximately 1.3 billion people [25], [26]. Tourism is an attractive sector, in which the countries want to increase their share since it provides a considerable amount of employment and income. Turkey, besides the advantages of its geographical and cultural diversity, also shows itself in the areas such as health and congress tourism. In the year 2017, only Istanbul attracted 9.24 million visitors and became the 11th city among the most visited cities [27]. Turkey's income from tourism reached 26.3 billion dollars in 2017 with an increase of 27% compared to 2016 [28]. Despite this increase, Turkey has quite a small share in the world tourism. Turkey carries out various promotional activities, such as opening promotion offices abroad, participation in tourism fairs and giving advertisements, in order to increase its income from tourism. The proclamation of 2018 as ''Turkey Tourism Year'' in China caused these activities to increase in China. China holds many opportunities as a country with a significant population and tourism expenditures.

Although China has the world's largest population with 1.4 billion people, it sends a relatively small number of tourists to Turkey compared to other countries. The number of Chinese tourists that preferred Turkey in 2017 was approximately 248,000 people [29]. Increasing welfare in China in comparison with the past increases the desire of people to travel abroad. 98% of Chinese consumers want to go on a trip. Only in the first half of 2017, 62.03 million people traveled abroad. Moreover, China's international tourism expenditures were 13 billion dollars in 2000, 55 billion dollars in 2010 and 258 billion dollars in 2017. In this regard, in international tourism, China is the most spending country since 2012. In addition, Chinese tourists prefer France, Italy, the United Kingdom, Spain and Germany as the European country, and Paris, Amsterdam, Rome, Frankfurt and Barcelona as the city [26], [30], [31]. The arrival of 3 million tourists from the Chinese market is among the targets of Turkey. It may be useful to make estimations by using artificial intelligence over consumer data, in order to meet various tourism objectives.

Recently, studies in tourism are being carried out with artificial intelligence. In these studies, to determine the future tourism demand with the economy, population [32], airline [33] or tourism agencies [34] data, to predict the next destination of tourists with Call Detail Records data [35], and to contribute to hotel suggestions with TripAdvisor data [36], artificial intelligence was used. This study aims to find the tourists, who will prefer Turkey, by taking the consumers' past travel routes from TripAdvisor. With the empirical study done, it is expected that significant contributions will be made both to the literature and to the promotional activities in the tourism sector. This study is well developed compared to the articles in the literature. The essential features of the article can be listed as follows. (1) Evaluation of tourism data by machine learning methods, (2) Estimation of the new travel point by machine learning according to the travel history, (3) Hybrid machine learning method, (4) Machine learning-based customer identification, and (5) Determine the customer's travel preference.

The organization of the article consists of four main chapters: "Materials and Methods", "Results", "Discussion and Conclusion" and "Future Work". The Materials and Methods section provides a detailed overview of data collection, data selection, and classification processes. The Results section presents the results of the application obtained. Discussion and Conclusion section describes the meaning of the results compared to the literature. Finally, the Future Work section offers suggestions for future research on the subject of this article to improve the study.

II. MATERIALS AND METHODS

The operational steps of the study are given in Fig. 1. According to this, firstly, the targeted audience was determined through TripAdvisor. Later, the travel histories of individuals were gathered in four different groups. These are the individuals' travel histories to Europe (E), World (W) Countries and China (C) City/Province and all (EWC). Then, "One Zero Matrix (OZ)" and "Frequency Matrix (F)" were created for each group. Thus, the number of matrices belonging to four groups increased to eight. Before classification, the matrices were selected 0, 1 or 2 times by the F-score feature selection method, and the distributions were balanced. By this means, the number of cities and countries decreased, and the number of matrices increased to 24.

A. COLLECTION OF DATA

By Chinese tourists, the most preferred city in Europe is Paris, France [26], [30], [31] and the most preferred city in Turkey is Istanbul [37]. For this reason, during the process of data collection, the most commented structures in Paris and Istanbul were evaluated in the study. The data were taken from TripAdvisor based on the assumption that the commentators who wrote in Chinese were Chinese. The data belong to a total of 624 users who made everyone open comments under the titles of "Eiffel Tower in France," "Hagia Sophia in Istanbul," "Yerebatan Cistern" and "Topkapi Museum and Historical Istanbul". The acquisition of historical data took place between 27 April and 11 May 2018. The locations were taken as TripAdvisor defined locations. In their travel histories, there is location information of E (45), W other than Europe (97) and C (544) and EWC (687). Since the study aims to determine travel preferences to Turkey or France, the individuals who had a travel history both to Turkey and France were excluded from the study. Of 624 individuals, 254 individuals traveled only to Turkey, and 370 individuals traveled only to France. The reason for the selection of France is that its distance to China is similar to that of Turkey.

The matrices created for the records were called as OZ and F. The table 1 shows the examples of OZ and F. The columns of the table are named as follows. "ID" is the code that is given to those who comment. There are 45 European countries in the E column, 97 other world countries in the W column, and 544 Chinese cities or provinces in the C column. In the label column, the information on whether the user traveled to Turkey or France is available. Turkey and France labels are 1 and 2, respectively.

The OZ was created in the following way. A user was digitized as 1, if he/she had already been in the location stated in that column, otherwise as 0 (Table 1). The F contains numbers indicating how many times the users have been in those locations (Table 1).

B. F-SCORE FEATURE SELECTION ALGORITHM

The F-score classes are one of the feature selection algorithms that help to reveal distinctive features from each other [38]–[40]. To select a feature, an F-score (F_i) of each feature is calculated (Equation1). The F-score threshold value (F_E) is determined by taking the average of all F-score values. For i feature, if $F_i > F_E$, then i feature is selected. This step

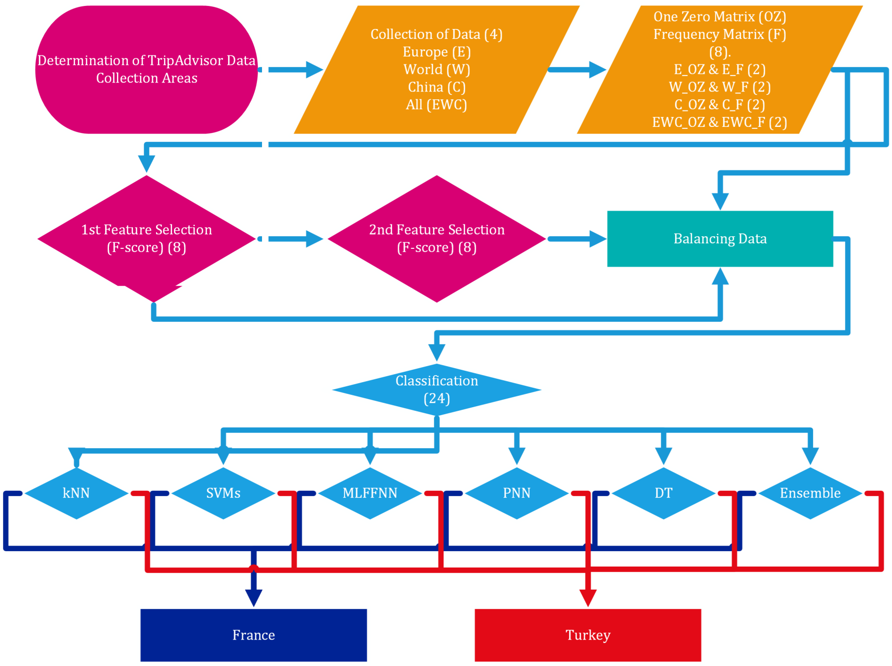

FIGURE 1. Process flow diagram.

TABLE 1. Example "One Zero Matrix" (OZ) and "Frequency Matrix" (F) table .

One Zero Matrix (OZ)

ID	Europe (E)				World (W)				China (C)				Label
	1	2	…	45	1	2	…	97	1	2	…	544	
T0001	1	0	…	1	0	0	…	1	0	1	…	0	1
T0002	1	0	…	1	1	0	…	0	0	0	…	1	1
T0003	1	1	…	1	1	1	…	1	1	1	…	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
T0254	0	1	1	0	0	0	…	1	0	0	…	3	1
F0001	1	0	0	1	1	0	…	0	0	1	…	0	2
F0002	0	1	0	1	1	1	…	0	1	0	…	0	2
F0003	1	1	1	1	0	1	…	1	1	0	…	1	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
F0370	1	1	…	1	1	0	…	1	1	0	…	1	2

Frequency Matrix (F)

ID	Europe (E)				World (W)				China (C)				Label
	1	2	…	45	1	2	…	97	1	2	…	544	
T0001	2	0	…	6	0	0	…	3	0	3	…	0	1
T0002	3	0	…	4	1	0	…	0	0	0	…	2	1
T0003	1	1	…	1	2	1	…	1	1	1	…	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
T0254	0	1	3	0	0	0	…	3	0	0	…	3	1
F0001	1	0	0	2	5	0	…	0	0	4	…	0	2
F0002	0	5	0	1	1	4	…	0	4	0	…	0	2
F0003	3	1	1	1	0	5	…	1	5	0	…	1	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
F0370	1	2	…	3	5	0	…	1	2	0	…	3	2

Labels, 1: Turkey, 2: France

is repeated for each feature.

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \tag{1}$$

The variables in Eq. 1 are as follows: (1) $x_{k,i}$ feature vector $k = 1, 2\ldots, m$, (2) $m = n_+ + n_-$ n_+ positive $(+)$ and n_- negative $(-)$ the total element number of classes, (3) i feature number. (4) $\bar{x}_i$, $\bar{x}_i^{(-)}$ and $\bar{x}_i^{(+)}$, respectively, are the mean value of the feature i., the mean value in the negative class, and the mean value in the positive class (5) $x_{k,i}^{(+)}$, represents the k. positive example of i. feature, (6) $x_{k,i}^{(-)}$, represents the k. negative example of i. feature.

In the study, each location owned by OZ and F, belonging to E, W, C, and EWC, is thought to be a feature. Locations were selected using the F-score (Table 2). The selection process ensures the reduction of the workload. In this study, feature selection was applied two times successively. In each selection to be eight matrices, a total of 24 matrix clusters were formed. For example, the World's OZ and F matrices initially have 97 locations. After the first feature selection, 34 locations from OZ and 31 locations from F were selected. In the second feature selection, 12 locations from OZ and 12 locations from F were selected.

C. ARTIFICIAL INTELLIGENCE CLASSIFICATION METHODS

The classifier can be defined as mathematical methods used to separate groups from each other. The classifiers used in this study are explained in detail in subheadings. As the inputs of the classifiers, tourists' travel histories and a total of 24 feature matrices selected by the F-score were used (Table 2).

TABLE 2. Selected features with F-Score .

Location	Features		1st Feature Selection		2nd Feature Selection	
	Number of Location		Number of Selected Location		Number of Selected Location	
	OZ	F	OZ	F	OZ	F
Europe (E)	45	45	17	18	9	9
World (W)	97	97	34	31	12	12
China (C)	544	544	214	214	60	60
EWC	686	686	150	175	67	68

EWC Europe World China, OZ One Zero Matrix, F Frequency Matrix

Each matrix was classified with five different classifiers, and the performance criteria of the classifiers were calculated in order to evaluate the classifier. The five different classifiers used are Decision trees (DT), k Nearest Neighbors Classification Algorithm (kNN), Multilayer Feedforward Artificial Neural Networks (MLFFNN), Probabilistic Neural Networks (PNN), and Support Vector Machines (SVMs) (Figure 1). Besides, with the collective decision of all these classifiers, the working community classifier was used.

There are two main reasons why these methods are preferred. The first is that their performance is quite good compared to other machine learning methods. The second reason is the rapid completion of training and testing processes. Speed and success are the most critical factors in the selection of machine methods. These choices were made according to our studies and the information given in the literature [41], [42].

1) DECISION TREES

DT are one of the most common machine learning algorithms. This algorithm consists of two different methods. The first one is classification, and the second one is the regression (CRT) method [43]. In the classification process, the data are prepared in the tree format. The model consists of roots, branches, and leaves. The distribution of the data is from the root to the leaves. The leaves represent the labels of the data, in other words, the names of the data. Branching regions are called nodes. Nodes are prepared according to the training algorithm [43].

The purpose of the CRT method is to group the data set in the tree model according to the variables. The variables with the best discrimination ability are located in the nodes. As in the classification method, leaves represent class labels. From the root to the leaves, the nodes are repeated with the same or different variables. When this process is performed, the group variances are tried to be minimized. A rule-based classification tree arises when all the operations are completed. If the class label of the data is a categorical variable, the classification tree model is used, and if it is a continuous numeric variable, the CRT model is used. In this study, the classification decision trees model was used because class labels were categorical variables.

2) K NEAREST NEIGHBORS CLASSIFICATION ALGORITHM

The kNN is one of the methods of classification of machine learning that has an advisory learning infrastructure [44]. Following the structure of the training data set, classification

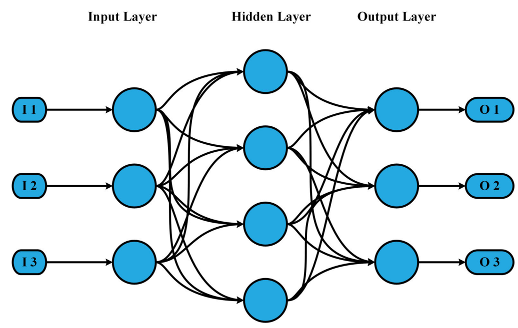

FIGURE 2. Network structure.

is made according to the nearest k data to the data to be newly classified. The classifier's performance k value depends on the number of nearest neighbors and distribution of the data in the distance formula training set. Initially, the k value can typically be selected as 3, 5, or 7 [45]. The selection of the big k value causes similar clusters to be aggregated [45]. In this study $k = 5$ was selected, and ten distance calculation formulae were used. These are Spearman, Seuclidean, Minkowski, Mahalanobis, Jaccard, Hamming, Euclidean, Cosine, Correlation, and Cityblock.

3) MULTILAYER FEEDFORWARD ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are bringing artificial neural cells together to process data by inspiring from biological nerves [46]. MLFFNN consist of single and forward-oriented channels for processing of the data (Figure 2). The feedforward network consists of three layers. These are input, hidden and output layer. The data start from the input layer and continue to the output layer in turn.

Although the MLFFNN operates with different training algorithms, the Scaled Conjugate Gradient (trainscg) training algorithm was used in this study. This algorithm allows the training process to be done quickly in large data sets. Except for the training algorithm, the data are connected to each other through neurons due to the structure. In this study, different neurons were used between 1 and 60 neurons.

4) PROBABILISTIC NEURAL NETWORKS

The PNN is a Kernel and Bayesian-based statistical methods based classification algorithm [47]. The method was developed on feedforward networks [47]. The classifier does processing by taking all class elements into consideration [48]. The distance between class elements is calculated by the radial basis kernel function.

The PNN network structure is similar to the feedforward network structures (Figure 2). In this network structure, the number of neurons at the input layer is equal to the number of features given to the network. The number of features that can be used as network input in the study is summarized in the Table 2. There are two hidden layers in the network structure. The number of neurons in the first layer is equal to the number of samples given to the network, and the second one has

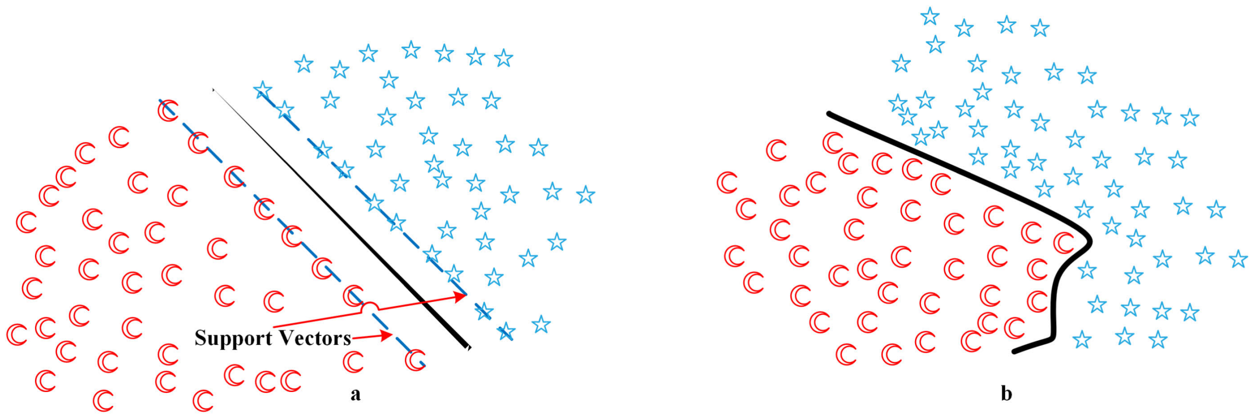

FIGURE 3. Separation of data with (a) linear and (b) nonlinear lines.

two neurons. Finally, the output layer has two neurons, since two different output values (Turkey/France) are owned in this study.

For the PNN classifier, only the spread initialization parameter can be interfered. As the spread parameter approaches to zero, the network begins to behave like the nearest neighbor classifier [49]. As this value moves away from zero, the classifier classifies by considering several vectors which separate the data from each other [49]. In the study, the networks are designed with the spread parameter between 0.01 and 5, with a step length of 0.01 and a total of 500 different values. At the end of the study, the best performing network parameters and performance criteria were calculated.

5) SUPPORT VECTOR MACHINES

SVMs are among the best machine learning algorithms [50]. In addition to its classification ability, it can also be used in regression analysis [50]. SVMs try to separate data sets from each other with linear and non-linear lines (Figure 3).

The purpose of the SVMs algorithm is to be able to separate the data from each other with the minimum error [51]. For this purpose, it uses the closest data as the support vector machine. The curve is adjusted where the distance between the support vector machines is maximum [51]. The curve is the solution set that separates the data set from each other [51]. The goal is to separate the data sets from each other optimally on the hyperplane and to classify new data with a minimum error rate [51]. The learning data nearest to the hyperplane are called the support vectors. The position where the distance between the support vectors is maximum is determined, and a curve is adjusted between them. This curve is accepted as the generalized solution that divides the data set into two.

Because Radial Basis Function (RBF) kernel is faster and more successful than the Gaussian, rbf was used in the study. The BoxConstraint box limit was set by changing between 1-100 so that the best performance could be achieved.

6) ENSEMBLE CLASSIFIER

The community classifier is a system created by combining different classifiers to produce safer or more stable estimates [42]. The system is built with N classifiers (Figure 4). N can be single or double. While classifying according to the feature vector, for the first feature vector, each classifier generates an output value. The produced output values are counted. Then, the decision of the community classifier is determined by the majority of the votes. If the number of classifiers is even, the decision of the community classifier is determined, by taking the average of the classifiers' decision values and rounding it off. This process is applied to the entire feature vector. The community classifier was prepared in the MATLAB environment using four different classifiers as kNN, MLFFNN, PNN, SVMs [52].

The distribution of the data in the study is unbalanced because the number of Turkey's data is 254 and France is 370. In machine learning, the data must be balanced before classification. Therefore, France data were reduced to 255 by selecting the data according to the systematic sampling theorem [53]. In each classifier, the database is divided into two sections as training and test data set. In the training dataset, there are 203 Turkey and 205 France data, while in the test set, there are 51 Turkey and 50 France data.

D. THE USED PERFORMANCE CRITERIA

Different performance assessment criteria were used to test the accuracy rates of the proposed systems. These are accuracy rates, sensitivity, specificity, kappa value, Receiver Operating Characteristic (ROC), area under a ROC (AUC), and k(10)-fold cross-validation accuracy rate.

1) k-FOLD CROSS VALIDATION

Cross-validation tests the performance of machine learning systems. In this method, all data are used in training and testing of machine learning. For cross-validation, all data is subdivided into k. The $k-1$ subset is used for training, while another cluster is used for testing. This process is repeated

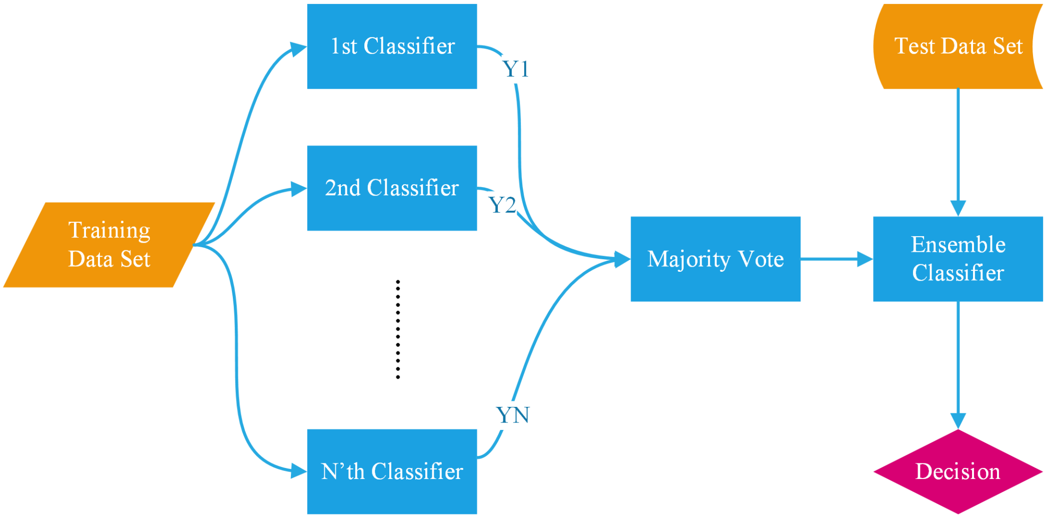

FIGURE 4. Workflow for the ensemble classifier.

TABLE 3. Comparison matrix for accuracy, specificity and sensitivity.

		Estimated Status	
		P	N
Actual Status	P	TP	FN
	N	FP	TN

TABLE 4. Kappa coefficients limit ranges.

Kappa coefficients	Description
0.81 – 1.00	Very well fit
0.61 – 0.80	Good fit
0.41 – 0.60	Medium fit
0.21 – 0.40	Low fit
0.00 – 0.20	Weak fit
<0.00	Very weak fit

k times. In this way, each subset is used in the test process. In this study, cross-validation was performed for $k = 10$ value.

2) CONFUSION MATRIX, KAPPA VALUE, F-MEASURE AND RECEIVER OPERATING CHARACTERIC

This section describes the performance evaluation criteria.

The sensitivity indicates the ability of the test to separate that class (Class 1) within a given class (Class 1). It ranges from 0 to 1. System quality increases, as sensitivity approach 1. A sensitivity value of 1 indicates that the test can correctly identify all class members. Specificity is the ability of the test to separate those class members (Class 2) within a given class (Class 2). It ranges from 0 to 1 and is desirably close to 1. It is used in cases where the detection needs to be verified. If the specificity of a test is 1, it indicates that the test is able to detect all Class 2 correctly. If a test has a specificity of 1, the system has correctly detected all Class 2 members. Accuracy rate, sensitivity and specificity parameters are calculated according to equations 2, 3, and 4, respectively. TP, TN, FP and FN in equations 2, 3, and 4 are True Positives, True Negatives, False Positives, and False Negatives, respectively. In addition, Table 1 shows the Confusion Matrix for accuracy, specificity, and sensitivity.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \times 100 \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{FP + TN} \tag{4}$$

F-Measurement is used to determine the effectiveness of the model. The value obtained is the weighted average of sensitivity and specificity values. The F-measurement is calculated as in equation 5. The F-measurement takes a value between 0 and 1. 1 indicates that the model is perfect, and 0 indicates that it is too bad.

$$F = 2 \times \frac{Specificity \times Sensitivity}{Specificity + Sensitivity} \tag{5}$$

The AUC value is used to evaluate the performance of diagnostic tests used to diagnose a disease [54]. The AUC value represents the area under the ROC curve.

The Kappa coefficient is a coefficient that provides information about reliability by correcting the "chance matches" that are purely dependent on chance. Different limit values for the Kappa coefficient have been defined in the literature regarding the degree of agreement [55].

III. RESULTS

This study aims to determine on the basis of machine learning the Chinese tourists, who may come to Turkey. Therefore, DT, kNN, MLFFNN, PNN, SVMs machine learning algorithms and Ensemble Classifier were used. As system inputs, Chinese tourists' travel histories were used as a feature (Table 2). The features were reduced by passing through the F-score feature selection algorithm in two stages, and the feature matrices were constructed again (Table 2). The resulting

TABLE 5. Classification results for "One Zero Matrix (OZ)" for Europe (E).

Decision Tree (DT)									
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.63	0.50	56.44	0.22	0.92	56.44	0.18	0.92	54.46
France	0.50	0.63		0.92	0.22		0.92	0.18	
AUC		0.56			0.57			0.55	
Kappa		0.13			0.13			0.10	
F-measure		0.56			0.35			0.30	
k(10)-fold (%)		58.35			57.76			55.21	
k-Nearest Neighbors Classification Algorithm (kNN)									
Network Parameters	k = 6, distance function = 'correlation'			k = 5, distance function = 'correlation'			k = 1, distance function = 'cityblock'		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.92	0.28	62.99	0.94	0.08	51.49	1.00	0.00	50.50
France	0.28	0.92		0.08	0.94		0.00	1.00	
AUC		0.60			0.51			0.50	
Kappa		0.20			0.02			0.00	
F-measure		0.43			0.15			0.00	
k(10)-fold (%)		57.37			51.87			49.12	
Multilayer Feedforward Artificial Neural Networks (MLFFNN)									
Network Parameters	Number of neurons = 45			Number of neurons = 4			Number of neurons = 7		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.69	0.54	61.39	0.39	0.96	67.33	0.37	0.94	65.35
France	0.54	0.69		0.96	0.39		0.94	0.37	
AUC		0.61			0.68			0.66	
Kappa		0.23			0.35			0.31	
F-measure		0.60			0.56			0.53	
k(10)-fold (%)		-			-			-	
Probabilistic Neural Networks (PNN)									
Network Parameters	Spread = 0.0910			Spread = 0.001			Spread = 0.001		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.88	0.22	55.45	0.33	0.92	62.38	0.20	0.92	55.45
France	0.22	0.88		0.92	0.33		0.92	0.20	
AUC		0.55			0.63			0.56	
Kappa		0.10			0.25			0.12	
F-measure		0.35			0.49			0.32	
k(10)-fold (%)		-			-			-	
Support Vector Machines (SMVs)									
Network Parameters	BoxConstraint = 10			BoxConstraint = 67			BoxConstraint = 67		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.82	0.42	62.38	0.45	0.94	69.31	0.33	0.90	61.39
France	0.42	0.82		0.94	0.45		0.90	0.33	
AUC		0.62			0.70			0.62	
Kappa		0.24			0.39			0.23	
F-measure		0.56			0.61			0.49	
k(10)-fold (%)		60.51			58.35			55.60	
Ensemble Classifier									
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 45			Number of Features = 17			Number of Features = 9		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.84	0.42	63.37	0.45	0.94	69.31	0.33	0.90	61.39
France	0.42	0.84		0.94	0.45		0.90	0.33	
AUC		0.63			0.70			0.62	
Kappa		0.26			0.39			0.23	
F-measure		0.56			0.61			0.49	
k(10)-fold (%)		-			-			-	

TABLE 6. Classification results for "Frequency Matrix (F)" for Europe (E).

Decision Tree (DT)									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.71	0.42	56.44	0.57	0.56	56.44	0.59	0.30	44.55
France	0.42	0.71		0.56	0.57		0.30	0.59	
AUC	0.56			0.56			0.44		
Kappa	0.13			0.13			-0.11		
F-measure	0.53			0.56			0.40		
k(10)-fold (%)	61.10			60.31			54.81		
k-Nearest Neighbors Classification Algorithm (kNN)									
Network Parameters	**k = 7, distance function = 'chebychev'**			**k = 10, distance function = 'chebychev'**			**k = 10, distance function = 'hamming'**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.90	0.38	64.36	0.92	0.34	0.63	0.94	0.12	53.47
France	0.38	0.90		0.34	0.92		0.12	0.94	
AUC	0.64			0.63			0.53		
Kappa	0.28			0.26			0.06		
F-measure	0.53			0.50			0.21		
k(10)-fold (%)	57.96			54.03			48.33		
Multilayer Feedforward Artificial Neural Networks (MLFFNN)									
Network Parameters	**Number of neurons = 12**			**Number of neurons = 8**			**Number of neurons = 4**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.39	0.90	64.36	0.43	0.88	65.35	0.94	0.94	64.36
France	0.90	0.39		0.88	0.43		0.35	0.35	
AUC	0.65			0.66			0.65		
Kappa	0.29			0.31			0.29		
F-measure	0.55			0.58			0.51		
k(10)-fold (%)	-			-			-		
Probabilistic Neural Networks (PNN)									
Network Parameters	**Spread = 0.521**			**Spread = 0.011**			**Spread = 0.121**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.35	0.80	57.43	0.86	0.26	56.44	0.37	0.90	63.37
France	0.80	0.35		0.26	0.86		0.90	0.37	
AUC	0.58			0.56			0.64		
Kappa	0.15			0.12			0.27		
F-measure	0.49			0.40			0.53		
k(10)-fold (%)	-			-			-		
Support Vector Machines (SMVs)									
Network Parameters	**BoxConstraint = 21**			**BoxConstraint = 18**			**BoxConstraint = 3**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.88	0.34	61.39	0.84	0.36	60.40	0.82	0.26	54.46
France	0.34	0.88		0.36	0.84		0.26	0.82	
AUC	0.61			0.60			0.54		
Kappa	0.22			0.20			0.08		
F-measure	0.49			0.50			0.40		
k(10)-fold (%)	60.12			57.76			52.85		
Ensemble Classifier									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 45			**Number of Features = 18**			**Number of Features = 9**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.82	0.46	64.36	0.82	0.44	63.37	0.80	0.30	55.45
France	0.46	0.82		0.44	0.82		0.30	0.80	
AUC	0.64			0.63			0.55		
Kappa	0.28			0.26			0.10		
F-measure	0.59			0.57			0.44		
k(10)-fold (%)	-			-			-		

TABLE 7. Classification results for "One Zero Matrix (OZ)" for World (W) .

Decision Tree (DT)

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.49	0.80	64.36	0.51	0.72	61.39	0.53	0.76	64.36
France	0.80	0.49		0.72	0.51		0.76	0.53	
AUC	0.65			0.61			0.64		
Kappa	0.29			0.23			0.29		
F-measure	0.61			0.60			0.62		
k(10)-fold (%)	59.92			61.10			58.55		

k-Nearest Neighbors Classification Algorithm (kNN)

Network Parameters	k = 4, distance function = 'correlation'			k = 10, distance function = 'cityblock'			k = 7, distance function = 'chebychev'		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.90	0.32	61.39	0.88	0.22	55.45	1.00	0.04	52.48
France	0.32	0.90		0.22	0.88		0.04	1.00	
AUC	0.61			0.55			0.52		
Kappa	0.22			0.10			0.04		
F-measure	0.47			0.35			0.08		
k(10)-fold (%)	52.85			51.87			47.35		

Multilayer Feedforward Artificial Neural Networks (MLFFNN)

Network Parameters	Number of neurons = 47			Number of neurons = 10			Number of neurons = 11		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.53	0.80	66.34	0.49	0.90	69.31	0.51	0.84	67.33
France	0.80	0.53		0.90	0.49		0.84	0.51	
AUC	0.66			0.70			0.67		
Kappa	0.33			0.39			0.35		
F-measure	0.64			0.63			0.63		
k(10)-fold (%)	-			-			-		

Probabilistic Neural Networks (PNN)

Network Parameters	Spread = 0.191			Spread = 0.061			Spread = 0.001		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.78	0.52	65.35	0.67	0.68	67.33	0.49	0.80	64.36
France	0.52	0.78		0.68	0.67		0.80	0.49	
AUC	0.65			0.67			0.65		
Kappa	0.31			0.35			0.29		
F-measure	0.63			0.67			0.61		
k(10)-fold (%)	-			-			-		

Support Vector Machines (SMVs)

Network Parameters	BoxConstraint = 12			BoxConstraint = 1			BoxConstraint = 16		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.69	0.68	68.32	0.63	0.76	69.31	0.43	0.88	65.35
France	0.68	0.69		0.76	0.63		0.88	0.43	
AUC	0.68			0.69			0.66		
Kappa	0.37			0.39			0.31		
F-measure	0.68			0.69			0.58		
k(10)-fold (%)	60.90			62.48			58.94		

Ensemble Classifier

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 34			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.69	0.74	71.29	0.61	0.72	66.34	0.51	0.78	64.36
France	0.74	0.69		0.72	0.61		0.78	0.51	
AUC	0.71			0.66			0.64		
Kappa	0.43			0.33			0.29		
F-measure	0.71			0.66			0.62		
k(10)-fold (%)	-			-			-		

TABLE 8. Classification results for "Frequency Matrix (F)" for for World (W) .

Decision Tree (DT)

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.47	0.78	62.38	0.45	0.82	63.37	0.49	0.80	64.36
France	0.78	0.47		0.82	0.45		0.80	0.49	
AUC	0.63			0.64			0.65		
Kappa	0.25			0.27			0.29		
F-measure	0.59			0.58			0.61		
k(10)-fold (%)	60.51			59.92			62.08		

k-Nearest Neighbors Classification Algorithm (kNN)

Network Parameters	k = 3, distance function = 'chebychev'			k = 3, distance function = 'cosine'			k = 3, distance function = 'chebychev'		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.82	0.48	65.35	0.78	0.52	65.35	0.82	0.40	61.39
France	0.48	0.82		0.52	0.78		0.40	0.82	
AUC	0.65			0.65			0.61		
Kappa	0.30			0.31			0.22		
F-measure	0.61			0.63			0.54		
k(10)-fold (%)	52.06			55.40			54.22		

Multilayer Feedforward Artificial Neural Networks (MLFFNN)

Network Parameters	Number of neurons = 40			Number of neurons = 40			Number of neurons = 22		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.57	0.80	68.32	0.43	0.94	68.32	0.45	0.86	65.35
France	0.80	0.57		0.94	0.43		0.86	0.45	
AUC	0.68			0.69			0.66		
Kappa	0.37			0.37			0.31		
F-measure	0.66			0.59			0.59		
k(10)-fold (%)	-			-			-		

Probabilistic Neural Networks (PNN)

Network Parameters	Spread = 0.781			Spread = 0.051			Spread = 0.011		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.41	0.88	64.36	0.67	0.60	63.37	0.65	0.64	64.36
France	0.88	0.41		0.60	0.67		0.64	0.65	
AUC	0.65			0.63			0.64		
Kappa	0.29			0.27			0.29		
F-measure	0.56			0.63			0.64		
k(10)-fold (%)	-			-			-		

Support Vector Machines (SMVs)

Network Parameters	BoxConstraint = 5			BoxConstraint = 2			BoxConstraint = 90		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.73	0.66	69.31	0.67	0.64	65.35	0.63	0.66	64.36
France	0.66	0.73		0.64	0.67		0.66	0.63	
AUC	0.69			0.65			0.64		
Kappa	0.39			0.31			0.29		
F-measure	0.69			0.65			0.64		
k(10)-fold (%)	0.00			0.00			0.00		

Ensemble Classifier

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 97			Number of Features = 31			Number of Features = 12		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.61	0.78	69.31	0.59	0.76	67.33	0.61	0.76	68.32
France	0.78	0.61		0.76	0.59		0.76	0.61	
AUC	0.69			0.67			0.68		
Kappa	0.39			0.35			0.37		
F-measure	0.68			0.66			0.68		
k(10)-fold (%)	-			-			-		

TABLE 9. Classification results for "One Zero Matrix (OZ)" for China (C) .

Decision Tree (DT)									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.88	0.20	54.46	0.22	0.90	55.45	0.22	0.90	55.45
France	0.20	0.88		0.90	0.22		0.90	0.22	
AUC	0.54			0.56			0.56		
Kappa	0.08			0.11			0.11		
F-measure	0.33			0.35			0.35		
k(10)-fold (%)	51.47			53.83			53.44		
k-Nearest Neighbors Classification Algorithm (kNN)									
Network Parameters	**k = 10, distance function = 'spearman'**			**k = 4, distance function = 'cityblock'**			**k = 9, distance function = 'cityblock'**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.98	0.20	59.41	0.96	0.22	59.41	0.92	0.18	55.45
France	0.20	0.98		0.22	0.96		0.18	0.92	
AUC	0.59			0.59			0.55		
Kappa	0.18			0.18			0.10		
F-measure	0.33			0.36			0.30		
k(10)-fold (%)	50.69			54.81			52.26		
Multilayer Feedforward Artificial Neural Networks (MLFFNN)									
Network Parameters	**Number of neurons = 39**			**Number of neurons = 41**			**Number of neurons = 39**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.41	0.84	62.38	0.39	0.92	65.35	0.29	0.98	63.37
France	0.84	0.41		0.92	0.39		0.98	0.29	
AUC	0.63			0.66			0.64		
Kappa	0.25			0.31			0.27		
F-measure	0.55			0.55			0.45		
k(10)-fold (%)	-			-			-		
Probabilistic Neural Networks (PNN)									
Network Parameters	**Spread = 0.661**			**Spread = 1.111**			**Spread = 0.281**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.90	0.28	59.41	0.22	0.94	57.43	0.22	0.94	57.43
France	0.28	0.90		0.94	0.22		0.94	0.22	
AUC	0.59			0.58			0.58		
Kappa	0.18			0.15			0.15		
F-measure	0.43			0.35			0.35		
k(10)-fold (%)	-			-			-		
Support Vector Machines (SMVs)									
Network Parameters	**BoxConstraint = 55**			**BoxConstraint = 98**			**BoxConstraint = 60**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.67	0.38	52.48	0.33	0.78	55.45	0.29	0.76	52.48
France	0.38	0.67		0.78	0.33		0.76	0.29	
AUC	0.52			0.56			0.53		
Kappa	0.05			0.11			0.05		
F-measure	0.48			0.47			0.42		
k(10)-fold (%)	49.90			46.76			50.29		
Ensemble Classifier									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.98	0.30	64.36	0.31	0.90	60.40	0.29	0.94	61.39
France	0.30	0.98		0.90	0.31		0.94	0.29	
AUC	0.64			0.61			0.62		
Kappa	0.28			0.21			0.23		
F-measure	0.46			0.47			0.45		
k(10)-fold (%)	-			-			-		

TABLE 10. Classification results for "Frequency Matrix (F)" for China (C) .

Decision Tree (DT)									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.76	0.30	53.47	0.86	0.22	54.46	0.25	0.92	58.42
France	0.30	0.76		0.22	0.86		0.92	0.25	
AUC	0.53			0.54			0.59		
Kappa	0.06			0.08			0.17		
F-measure	0.43			0.35			0.40		
k(10)-fold (%)	51.08			50.88			51.08		

k-Nearest Neighbors Classification Algorithm (kNN)									
Network Parameters	**k = 10, distance function = 'spearman'**			**k = 4, distance function = 'cityblock'**			**k = 9, distance function = 'seuclidean'**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.98	0.20	59.41	0.96	0.22	59.41	0.92	0.18	55.45
France	0.20	0.98		0.22	0.96		0.18	0.92	
AUC	0.59			0.59			0.55		
Kappa	0.18			0.18			0.10		
F-measure	0.33			0.36			0.30		
k(10)-fold (%)	49.31			54.22			51.47		

Multilayer Feedforward Artificial Neural Networks (MLFFNN)									
Network Parameters	**Number of neurons = 56**			**Number of neurons = 27**			**Number of neurons = 13**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.76	0.44	60.40	0.29	0.94	61.39	0.31	0.96	63.37
France	0.44	0.76		0.94	0.29		0.96	0.31	
AUC	0.60			0.62			0.64		
Kappa	0.21			0.23			0.27		
F-measure	0.56			0.45			0.47		
k(10)-fold (%)	-			-			-		

Probabilistic Neural Networks (PNN)									
Network Parameters	**Spread = 0.661**			**Spread = 1.111**			**Spread = 0.281**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.90	0.28	59.41	0.22	0.94	57.43	0.22	0.94	57.43
France	0.28	0.90		0.94	0.22		0.94	0.22	
AUC	0.59			0.58			0.58		
Kappa	0.18			0.15			0.15		
F-measure	0.43			0.35			0.35		
k(10)-fold (%)	-			-			-		

Support Vector Machines (SMVs)									
Network Parameters	**BoxConstraint = 93**			**BoxConstraint = 98**			**BoxConstraint = 74**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.67	0.38	52.48	0.33	0.78	55.45	0.27	0.82	54.46
France	0.38	0.67		0.78	0.33		0.82	0.27	
AUC	0.52			0.56			0.55		
Kappa	0.05			0.11			0.09		
F-measure	0.48			0.47			0.41		
k(10)-fold (%)	51.08			49.51			51.08		

Ensemble Classifier									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 544			**Number of Features = 214**			**Number of Features = 60**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.86	0.32	59.41	0.31	0.90	60.40	0.29	0.92	60.40
France	0.32	0.86		0.90	0.31		0.92	0.29	
AUC	0.59			0.61			0.61		
Kappa	0.18			0.21			0.21		
F-measure	0.47			0.47			0.45		
k(10)-fold (%)	-			-			-		

TABLE 11. Classification results for "One Zero Matrix (OZ)" for Europe (E), World (W) and China (C) - (ADC) .

Decision Tree (DT)

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.55	0.60	57.43	0.39	0.68	53.47	0.55	0.76	65.35
France	0.60	0.55		0.68	0.39		0.76	0.55	
AUC	0.57			0.54			0.65		
Kappa	0.15			0.07			0.31		
F-measure	0.57			0.50			0.64		
k(10)-fold (%)	62.48			60.51			58.55		

k-Nearest Neighbors Classification Algorithm (kNN)

Network Parameters	k = 4, distance function = 'correlation'			k = 6, distance function = 'correlation'			k = 10, distance function = 'hamming'		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.80	0.50	65.35	0.88	0.40	64.36	0.98	0.22	60.40
France	0.50	0.80		0.40	0.88		0.22	0.98	
AUC	0.65			0.64			0.60		
Kappa	0.30			0.28			0.20		
F-measure	0.62			0.55			0.36		
k(10)-fold (%)	64.24			59.53			55.60		

Multilayer Feedforward Artificial Neural Networks (MLFFNN)

Network Parameters	Number of neurons = 17			Number of neurons = 38			Number of neurons = 32		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.86	0.48	67.33	0.61	0.84	72.28	0.47	0.92	69.31
France	0.48	0.86		0.84	0.61		0.92	0.47	
AUC	0.67			0.72			0.70		
Kappa	0.34			0.45			0.39		
F-measure	0.62			0.71			0.62		
k(10)-fold (%)	-			-			-		

Probabilistic Neural Networks (PNN)

Network Parameters	Spread = 0.731			Spread = 0.221			Spread = 0.131		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.59	0.64	61.39	0.80	0.60	70.30	0.75	0.68	71.29
France	0.64	0.59		0.60	0.80		0.68	0.75	
AUC	0.61			0.70			0.71		
Kappa	0.23			0.40			0.43		
F-measure	0.61			0.69			0.71		
k(10)-fold (%)	-			-			-		

Support Vector Machines (SMVs)

Network Parameters	BoxConstraint = 2			BoxConstraint = 1			BoxConstraint = 73		
Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.80	0.48	64.36	0.69	0.74	71.29	0.76	0.68	72.28
France	0.48	0.80		0.74	0.69		0.68	0.76	
AUC	0.64			0.71			0.72		
Kappa	0.28			0.43			0.45		
F-measure	0.60			0.71			0.72		
k(10)-fold (%)	60.31			60.31			63.26		

Ensemble Classifier

Class	All Features			1st Feature Selection			2nd Feature Selection		
	Number of Features = 686			Number of Features = 150			Number of Features = 67		
	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)	Sensivity	Specificity	Accuracy (%)
Turkey	0.78	0.54	66.34	0.75	0.70	72.28	0.76	0.74	75.25
France	0.54	0.78		0.70	0.75		0.74	0.76	
AUC	0.66			0.72			0.75		
Kappa	0.33			0.45			0.50		
F-measure	0.64			0.72			0.75		
k(10)-fold (%)	-			-			-		

TABLE 12. Classification results for "Frequency Matrix (F)" for Europe (E), World (W) and China (C) - (ADC) .

Decision Tree (DT)									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.71	0.58	64.36	0.57	0.48	52.48	0.57	0.44	50.50
France	0.58	0.71		0.48	0.57		0.44	0.57	
AUC	0.64			0.52			0.50		
Kappa	0.29			0.05			0.01		
F-measure	0.64			0.52			0.50		
k(10)-fold (%)	58.35			59.72			62.67		
k-Nearest Neighbors Classification Algorithm (kNN)									
Network Parameters	**k = 6, distance function = 'euclidean'**			**k = 7, distance function = 'cosine'**			**k = 6, distance function = 'cosine'**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.73	0.62	67.33	0.75	0.74	74.26	0.86	0.56	71.29
France	0.62	0.73		0.74	0.75		0.56	0.86	
AUC	0.67			0.74			0.71		
Kappa	0.35			0.49			0.42		
F-measure	0.67			0.74			0.68		
k(10)-fold (%)	61.69			62.48			61.30		
Multilayer Feedforward Artificial Neural Networks (MLFFNN)									
Network Parameters	**Number of neurons = 12**			**Number of neurons = 4**			**Number of neurons = 45**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.59	0.74	66.34	0.57	0.86	71.29	0.61	0.86	73.27
France	0.74	0.59		0.86	0.57		0.86	0.61	
AUC	0.66			0.71			0.73		
Kappa	0.33			0.43			0.47		
F-measure	0.66			0.68			0.71		
k(10)-fold (%)	-			-			-		
Probabilistic Neural Networks (PNN)									
Network Parameters	**Spread = 1.751**			**Spread = 0.701**			**Spread = 0.1510**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.47	0.74	60.40	0.31	0.88	59.41	0.63	0.72	67.33
France	0.74	0.47		0.88	0.31		0.72	0.63	
AUC	0.61			0.60			0.67		
Kappa	0.21			0.19			0.35		
F-measure	0.58			0.46			0.67		
k(10)-fold (%)	-			-			-		
Support Vector Machines (SMVs)									
Network Parameters	**BoxConstraint = 19**			**BoxConstraint = 6**			**BoxConstraint = 38**		
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.80	0.50	65.35	0.73	0.66	69.31	0.86	0.54	70.30
France	0.50	0.80		0.66	0.73		0.54	0.86	
AUC	0.65			0.69			0.70		
Kappa	0.30			0.39			0.40		
F-measure	0.62			0.69			0.66		
k(10)-fold (%)	55.01			57.76			61.10		
Ensemble Classifier									
Class	**All Features**			**1st Feature Selection**			**2nd Feature Selection**		
	Number of Features = 686			**Number of Features = 175**			**Number of Features = 68**		
	Sensivity	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**	**Sensivity**	**Specificity**	**Accuracy (%)**
Turkey	0.78	0.66	72.28	0.61	0.82	71.29	0.88	0.54	71.29
France	0.66	0.78		0.82	0.61		0.54	0.88	
AUC	0.72			0.71			0.71		
Kappa	0.44			0.43			0.42		
F-measure	0.72			0.70			0.67		
k(10)-fold (%)	-			-			-		

TABLE 13. Best classification results.

Ensemble Classifier						
Data Set	One Zero Matrix (OZ) for Europe (E), World (W), China (C) - (EWC)					
Class	1st Feature Selection			2nd Feature Selection		
	Number of Features = 150			Number of Features = 67		
	Sen	Spe	Acc (%)	Sen	Spe	Acc (%)
Turkey	0.75	0.70	72.28	0.76	0.74	75.25
France	0.70	0.75		0.74	0.76	
Kappa	0.45			0.50		
AUC	0.72			0.75		
F-measure	0.72			0.75		
k(10)-fold (%)	-			-		

Spe Specificity, Sen Sensivity, Acc Accuracy

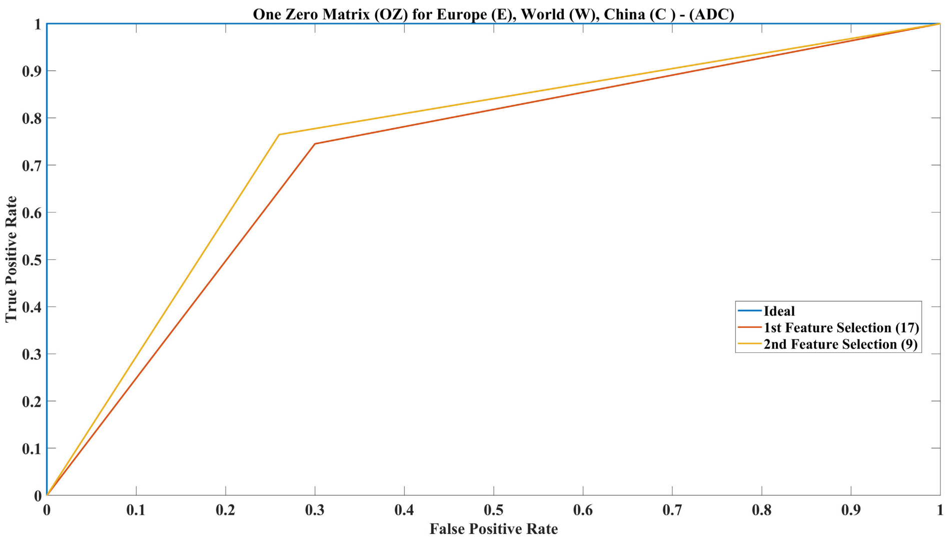

FIGURE 5. ROC for best classification results.

24 matrices were classified with all classifiers, and performance evaluation criteria were calculated for each classifier. However, since giving the results of 24 matrices increases the number of tables, only the results of 6 matrices belonging to ADC were given (Table 5, 6, 7, 8, 9, 10, 11, 12). In this matrix, there are all the features and the selected features of the first and second levels. The best two results were summarized again (Table 13), to assess the results easier. The performance evaluation criteria used are as follows: accuracy rates of classes, sensitivity, specificity, AUC, Kappa coefficient, F - criterion values. Furthermore, the ROC curve analysis was also performed (Fig. 5). However, when the ROC analysis is performed for each dataset, 24 graphs are generated. In order to optimize the number of graphs, ROC analysis graphs of 2 classifiers, which give the best result, are given (Fig. 5).

In each table, the classifiers, network parameters of the classifiers and performance evaluation criteria are given (Table 5, 6, 7, 8, 9, 10, 11, 12, 13). There was no significant performance difference between them when the "OZ" and "F" data sets of A, D, C, and EWC were compared (OZ - Table 5, 7, 9, 11, F - Table 6, 8, 10, 12).

The number of operand features was decreased at each step with F-score feature selection (Tablo 5, 6, 7, 8, 9, 10, 11, 12). Decreasing the number of features will reduce the processing burden.

The best sensitivity for Turkey is calculated as 1 with kNN, and for France as 0.94 with MLFFNN (Table 5). It was observed that as the number of features increased in the "OZ" data sets belonging to EWC, the performance increased (Table 11). However, the same situation was not observed in the "F" data set (Table 12). While the Ensemble classifier displayed a performance at the average level of all classifiers in some cases (Table 12, %71.29), it performed better than all classifiers in some other cases (Table 11, %75.25).

According to the results obtained in the study, the travel preferences of Chinese tourists can be correctly determined by 75.25%. While the sensitivity rate is 0.76 in choosing Turkey, it is 0.74 in choosing France (Table 13).

IV. DISCUSSION AND CONCLUSIONS

In the tourism sector, many machine learning-based studies have been carried out, such as forecasting of tourism demands [56], classification of user comments [57], classification of hotel opinions [58], and prediction of skiing days [59]. The common result of these studies is that in this sector the performance of machine learning is acceptable.

It is difficult to model human behaviors and to make inferences accordingly because there are many variables. It is a different perspective to make inferences with only tourists' past travel experiences. The overall accuracy rate obtained in studies in the literature is around 50-60%, which is consistent with this study [59], [60].

The primary objective of the study is to determine Chinese tourists who may travel to Turkey. For this, the travel histories of Chinese tourists were used in different variations and with different classifiers. The aim is to reveal the different powerful features of each set of travel history with different classifiers. For example, in the ADC - OZ dataset, 67 features were classified with 60.40% accuracy with kNN, while the same features were classified with 72.28% accuracy with SVMs. This difference may be because the SVMs are better adapted to the data set than the KNN. The study has a wide range compared with the literature regarding using different classifiers and different data clusters [57], [58], [60], [61].

In this study, a total of 686 cities were used in classifiers as features. Thus, this study had a wide range of travel history information. In the literature, the number of features is considerably limited [57], [58], [60]. This study aims to determine Chinese tourists who may come directly to Turkey. Nevertheless, there are generally fuzzy logic and regression-based studies in the literature to estimate only the number of tourists [60]–[63]. While today's technologies make personalized production, this study is a pioneering step for customizing the advertisements and for displaying them specific to individuals.

Time series analyses are quite common in the literature [46], [56], [61], [62]. However, since the features used in this study are based solely on travel histories, it is thought to be useful in predictions independent of time. In the study, it was attempted to determine the powerful features of each dataset by using the F-score feature selection method. Thus, the workload in the system was reduced, and the system was simplified. However, all the features that were obtained from the studies on tourism in the literature were used directly [56], [62], [64]. The F-score added power to the study. As a matter of fact, the best results were obtained after the first and second feature selection (Table 13).

According to the results of the study, the travel preferences of Chinese tourists can be determined with 75.25% accuracy. It is desirable that a diagnostic system developed in healthcare has a minimum accuracy of 80% [55]. However, it can be said that 75.25% accuracy is satisfactory when the tourism sector is thought to be more tolerant than the healthcare sector. Upon examining the literature, the accuracy rate in the time series is 80-92% [60]. However, when the regression models are predicted by artificial intelligence methods, it is expected that the accuracy rate is already high because artificial intelligence methods have been developed for solving problems that cannot be solved by regression. Given this fact and the difficulty of the problem, the 75.25% accuracy ratio is thought to be satisfactory.

The study has contributed to a limited number of studies with the use of artificial intelligence for marketing in parallel to Turkey's tourism goals [59], [60]. Within the framework of the obtained results, it is expected to contribute to the effective and efficient use of the budget allocated to promotional activities by Turkey.

As a result, it was observed that a new travel plan of any Chinese tourist could be estimated with the proposed method by contributing to Turkey's promotional activities. It was found out that it is possible to reach the right person with the method tested in the study. In other words, it was demonstrated that it is possible to contribute to the success of Turkey by preventing the high cost of mass advertising done with traditional media in the tourism sector, or by preventing the price undercut done in order to attract a large number of people. Nevertheless, data obtained from other voting and rating sites besides TripAdvisor will allow carrying out healthier promotional activities to a larger number of people. The same method can be used in other target markets other than China. The study contributes to a limited number of academic studies discussing the use of artificial intelligence in the tourism sector and makes explicit a tool that Turkey can use in tourism targets.

V. FUTURE WORK

Within the scope of this study, the future travel preferences of the individual were estimated based on their previous travel information. Many parameters can affect the individual's choice of travel. Therefore, it would be beneficial to include the study history of the individual as well as different information that may affect the individual's choice of travel. We can list the new information that can be included in future studies, as follows.

- Demographic information
- Financial situation
- Education information
- Hobbies
- Habits etc.

Work can also be improved in different ways. For this, the most critical method is the development of predictive methods. A more advanced system design can be made with different machine learning algorithms.

REFERENCES

[1] C. Berne, M. Garcia-Gonzalez, and J. Mugica, "How ICT shifts the power balance of tourism distribution channels," *Tourism Manage.*, vol. 33, no. 1, pp. 205–214, Feb. 2012.

[2] D. Hummels, "Transportation costs and international trade in the second era of globalization," *J. Econ. Perspect.*, vol. 21, no. 3, pp. 131–154, Summer 2007.

[3] MFA. (2017). *Joint Press Conference of Foreign Minister HE Mevlüt Çavuşoğlu with Wang Yi, Minister of Foreign Affairs of the People's Republic of China, Ministry of Foreign Affairs*. Accessed: Jul. 4, 2018. [Online]. Available: http://www.mfa.gov.tr/disisleri-bakani-sayin-mevlut-cavusoglu_nun-cin-halk-cumhuriyeti-disisleri-bakani-wang-yi-ile-ortak-basin-toplantisi_-3-agustos.tr.mfa

[4] *Dışışleri Bakanı Sayın Mevlüt Çavuşoğlu'nun Çin Halk Cumhuriyeti Dışışleri Bakan. Wang Yi ile Ortak Basın Toplantısı*, TFM, Pekin, IL, USA, 2017.
[5] (2018) *Internet World Stats Usage and Population Statistics*. Accessed: Jun. 6, 2018. [Online]. Available: http://www.internetworldstats.com/stats.htm
[6] B. Getting. (2007). *Basic Definitions: Web 1.0, Web 2.0, Web 3.0*. [Online]. Available: http://www.practicalecommerce.com/Basic-Definitions-Web-1-0-Web-2-0-Web-3-0
[7] S. Manjit Yadav, K. D. Valck, T. Hennig-Thurau, L. Donna Hoffman, and M. Spann, "Social commerce: A contingency framework for assessing marketing potential," *J. Interact. Marketing*, vol. 27, no. 4, pp. 311–323, Nov. 2013.
[8] *Under The Influence: Consumer Trust in Advertising*, Nielsen, New York, NY, USA: 2013.
[9] S. Kemp. (2017). *Digital in 2017: Global Overview*. Accessed: Jul. 4, 2017. [Online]. Available: https://wearesocial.com/uk/special-reports/digital-in-2017-global-overview
[10] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus. Horizons*, vol. 53, no. 1, pp. 59–68, 2010.
[11] D. Zarrella, *The Social Media Marketing Book*. Sebastopol, ON, Canada: O'Reilly Media Inc., 2009.
[12] L. Heyne, "Electronic word of mouth—A new marketing tool," Economy Manage., Univ. Appl. Sci., Düsseldorf, Germany, 2009.
[13] U. Prashanth Nyer and M. Gopinath, "Effects of complaining versus negative word of mouth on subsequent changes in satisfaction: The role of public commitment," *Psychol. Marketing*, vol. 22, no. 12, pp. 937–953, Dec. 2005.
[14] E. Martin-Fuentes, C. Fernandez, C. Mateu, and E. Marine-Roig, "Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb," *Int. J. Hospitality Manage.*, vol. 69, pp. 75–83, Jan. 2018.
[15] Z. Xiang, V. P. Magnini, and D. R. Fesenmaier, "Information technology and consumer behavior in travel and tourism: Insights from travel planning using the Internet," *J. Retailing Consum. Services* vol. 22, no. 2014, pp. 244–249, Jan. 2015.
[16] A. M. Abubakar and M. Ilkan, "Impact of online WOM on destination trust and intention to travel: A medical tourism perspective," *J. Destination Marketing Manage.*, vol. 5, no. 3, pp. 192–201, Sep. 2016.
[17] *Media Center*, TripAdvisor, Needham, MA, USA, 2018.
[18] *Media Center*, TripAdvisor, Needham, MA, USA, 2017.
[19] T. Hennig-Thurau, C. Edward Malthouse, C. Friege, S. Gensler, L. Lobschat, A. Rangaswamy, and B. Skiera, "The impact of new media on customer relationships," *J. Service Res.*, vol. 13, no. 3, pp. 311–330, Aug. 2010.
[20] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, "Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews," *J. Hospitality Marketing Manage.*, vol. 25, no. 1, pp. 1–24, 2016.
[21] V. Cozza, M. Petrocchi, and A. Spognardi, "Mining implicit data association from Tripadvisor hotel reviews," in *Proc. CEUR Workshop*, 2018, pp. 56–61.
[22] B. Lu, M. Ott, C. Cardie, and B. K. Tsou, "Multi-aspect sentiment analysis with topic models," *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2011, pp. 81–88.
[23] S. Moro, P. Rita, and J. Coelho, "Stripping customers' feedback on hotels through data mining: The case of las vegas strip," *Tourism Manage. Perspect.*, vol. 23, pp. 41–52, Jul. 2017.
[24] Z. Xiang, Q. Du, Y. Ma, and W. Fan, "Assessing reliability of social media data: Lessons from mining TripAdvisor hotel reviews," *Inf. Technol. Tourism*, vol. 18, nos. 1–4, pp. 43–59, Apr. 2018.
[25] *Tourism Economic Impact 2017: World*, World Travel & Tourism Council, London, U.K., 2017, p. 1.
[26] *2017 Annual Report*, UNWTO, Madrid, Spain, 2017.
[27] A. Millington. (2017). The most visited cities around the world in 2017—Business insider, Business Insider. Accessed: Jun. 6, 2018. [Online]. Available: http://uk.businessinsider.com/the-most-visited-cities-around-the-world-in-2017-2017-9#1-bangkok-thailand-202-million-international-visitors-30
[28] *Tourism Statistics, IV. Quarter: September to December and Annual 2017*, Turkey Stat. Inst., Ankara, Turkey, 2018.
[29] Foreigners and Citizens. (2017). *Turkish Statistical Institute*. [Online]. Available: http://www.tuik.gov.tr/PreTablo.do?alt_id=1072
[30] (2017). *2017 China Tourism Facts & Figures*. Accessed: May 18, 2018. [Online]. Available: https://www.travelchinaguide.com/tourism/2017statistics/
[31] *2017 China Tourism Facts & Figures*, China Tourism Academy, Beijing, China, 2017.
[32] C. Goh and R. Law, "Incorporating the rough sets theory into travel demand analysis," *Tourism Manage.*, vol. 24, no. 5, pp. 511–517, Oct. 2003.
[33] P.-F. Pai, K.-C. Hung, and K.-P. Lin, "Tourism demand forecasting using novel hybrid system," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3691–3702, 2014.
[34] C.-F. Chen, M.-C. Lai, and C.-C. Yeh, "Forecasting tourism demand based on empirical mode decomposition and neural network," *Knowl.-Based Syst.*, vol. 26, pp. 281–287, Feb. 2012.
[35] N. C. Chen, W. Xie, R. E. Welsch, K. Larson, and J. Xie, "Comprehensive predictions of tourists' next visit location based on call detail records using machine learning and deep learning methods," in *Proc. IEEE Int. Congr. Big Data*, Jun. 2017, pp. 1–6.
[36] M. Nilashi, O. B. Ibrahim, N. Ithnin, and N. H. Sarmin, "A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS," *Electron. Commerce Res. Appl.*, vol. 14, no. 6, pp. 542–562, Oct./Nov. 2015.
[37] *Declared as Turkey Tourism Year*, TCA, Beijing China, 2018.
[38] K. Polat and S. Güneş, "A new feature selection method on classification of medical datasets: Kernel F-score feature selection," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10367–10373, Sep. 2009.
[39] J. Gao, Z. Wang, Y. Yang, W. Zhang, C. Tao, J. Guan, and N. Rao, "A novel approach for lie detection based on F-score and extreme learning machine," *PLoS ONE*, vol. 8, no. 6, Jun. 2013, Art. no. e64704.
[40] P. V. C. de Souza, "Pruning fuzzy neural networks based on unineuron for problems of classification of patterns," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 2597–2605, Aug. 2018.
[41] M. K. Uçar, M. R. Bozkurt, C. Bilgin, and K. Polat, "Automatic detection of respiratory arrests in OSA patients using PPG and machine learning techniques," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2931–2945, Oct. 2017.
[42] L. Rokach, A. Schclar, and E. Itach, "Ensemble methods for multi-label classification," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7507–7523, Nov. 2014.
[43] M. Topaloğlu and H. Sur, "Sarilik semptomlarinda yanliş teshisi azaltmak için karar ağaci uygulamasi," *Nobel Medicus*, vol. 11, no. 3, pp. 64–73, 2015.
[44] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and K-nn algorithm for breast cancer diagnosis," *Comput. Biol. Med.*, vol. 37, no. 3, pp. 415–423, Mar. 2007.
[45] M. Khan, Q. Ding, and W. Perrizo, "k-nearest Neighbor Classification on Spatial Data Streams Using P-trees," in *Advances in Knowledge Discovery and Data Mining*. (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2002, pp. 517–528.
[46] J. Farajzadeh, A. F. Fard, and S. Lotfi, "Modeling of monthly rainfall and runoff of Urmia lake basin using 'feed-forward neural network' and 'time series analysis' model," *Water Resour. Ind.*, vols. 7–8, pp. 38–48, Sep. 2016.
[47] A. Khamis, H. Shareef, A. Mohamed, and E. Bizkevelci, "Islanding detection in a distributed generation integrated power system using phase space technique and probabilistic neural network," *Neurocomputing*, vol. 148, pp. 587–599, Jan. 2015.
[48] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[49] D. Philip Wasserman, *Advanced Methods Neural Computing*. New York, NY, USA: Wiley, 1993.
[50] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[51] V. N. Mandhala, V. Sujatha, and B. R. Devi, "Scene classification using support vector machines," in *Proc. IEEE Int. Conf. Adv. Commun., Control Comput. Technol.*, May 2014, pp. 1807–1810.
[52] A. S. D. P. Wallisch, M. E. Lusignan, M. D. Benayoun, T. I. Baker, and N. G. Hatsopoulos, *MATLAB for Neuroscientists: An Introduction to Scientific Computing in MATLAB*, 2nd ed. San Diego, CA, USA: Academic, 2014.
[53] T. Yamane, *Elementary Sampling Theory*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1967.

[54] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets," *Expert Syst. Appl.*, vol. 64, pp. 239–246, Dec. 2016.
[55] R. Alpar, *Applied Statistic and Validation—Reliability*. Ankara, Turkey: Detay, 2010.
[56] S. Cankurt and A. Subasi, "Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components," *Balkan J. Electr. Comput. Eng.*, vol. 3, no. 1, pp. 45–49, 2015.
[57] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. K. Roy, "Predicting the 'helpfulness' of Online consumer reviews," *J. Bus. Res.*, vol. 70, pp. 346–355, Jan. 2017.
[58] Y.-H. Hu and K. Chen, "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings," *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 929–944, Dec. 2016.
[59] M. A. King, A. S. Abrahams, and C. T. Ragsdale, "Ensemble methods for advanced skier days prediction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1176–1188, Mar. 2014.
[60] G. S. Atsalakis, I. G. Atsalaki, and C. Zopounidis, "Forecasting the success of a new tourism service by a neuro-fuzzy technique," *Eur. J. Oper. Res.*, vol. 268, no. 2, pp. 716–727, Jul. 2018.
[61] G. Yu and Z. Schwartz, "Forecasting short time-series tourism demand with artificial intelligence models," *J. Travel Res.*, vol. 45, no. 2, pp. 194–203, Nov. 2006.
[62] R. Law and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," *Tourism Manage.*, vol. 20, no. 1, pp. 89–97, Feb. 1999.
[63] R. Law, "Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting," *Tourism Manage.*, vol. 21, no. 4, pp. 331–340, Aug. 2000.
[64] O. Claveria, E. Monte, and S. Torra, "Tourism demand forecasting with neural network models: Different ways of treating information," *Tourism*, vol. 17, no. 5, pp. 492–500, Sep./Oct. 2015.

İBRAHIM TOPAL was born in Çanakkale, Turkey. He received the Business degree from Anadolu University, Turkey, the master's degree in production management and marketing from Kocaeli University, and the Ph.D. degree from Sakarya University, in 2017. His research interests include marketing, social media, and electronic word-of-mouth.

MUHAMMED KÜRŞAD UÇAR was born in Gümüşhane, Turkey. He received the Electrical and Electronics Engineering degree from Mustafa Kemal University, Turkey, and the master's degree in electrical and electronic engineering and the Ph.D. degree from Sakarya University, in 2017.

He is currently a Research Assistant Ph.D. with the Department of Electrical and Electronics Engineering, Sakarya University. His research interests include biomedical signal processing, statistical signal processing, digital signal processing, artificial intelligence, and machine learning.