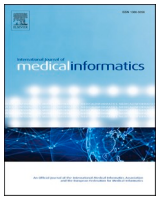
International Journal of Medical Informatics 177 (2023) 105163

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Predicting emergency health care demands due to respiratory diseases

J.C. Arias [a], M.I. Ramos [b,*], J.J. Cubillas [c]

[a] Group TIC-144 of the Andalusian Research Plan, University of Jaen, Spain
[b] Department of Cartographic, Geodetic and Photogrammetric Engineering, University of Jaen, Spain
[c] Department of Information and Communication Technologies applied to Education, International University of La Rioja, Spain

ARTICLE INFO

Keywords:
Machine Learning
Prediction
Health Emergency Service
Geospatial data

ABSTRACT

Background: Timely care in the health sector is essential for the recovery of patients, and even more so in the case of a health emergency. In these cases, appropriate management of human and technical resources is essential. These are limited and must be mobilised in an optimal and efficient manner.
Objective: This paper analyses the use of the health emergency service in a city, Jaén, in the south of Spain. The study is focused on the most recurrent case in this service, respiratory diseases.
Methods: Machine Learning algorithms are used in which the input variables are multisource data and the target attribute is the prediction of the number of health emergency demands that will occur for a selected date. Health, social, economic, environmental, and geospatial data related to each of the emergency demands were integrated and related. Linear and nonlinear regression algorithms were used: support vector machine (SVM) with linear kernel and generated linear model (GLM), and the nonlinear SVM with Gaussian kernel.
Results: Predictive models of emergency demand due to respiratory disseases were generated with am absolute error better than 35 %.
Conclusions: This model helps to make decisions on the efficient sizing of emergency health resources to manage and respond in the shortest possible time to patients with respiratory diseases requiring urgent care in the city of Jaén.

1. Introduction

One of the most important factors in the emergency healthcare sector is the time in which the patient receives care. Delays in arriving at the place where care is sought and in receiving treatment can lead to serious negative consequences and a poor prognosis for the patient [1]. The patient's life and possible sequelae depend directly on this time interval. Various medical studies stress the importance of this factor, and list the various complications that appear in patients as the waiting time for care progresses [2,3]. The reasons for the delay in patient care time are diverse, ranging from those inherent to the population seeking care, such as demographic characteristics, socioeconomic level, ethnicity, etc… [5,6], to the type of pathology. However, factors related to health systems also play an important role. These include the availability of services, the accessibility of healthcare facilities, the acceptability and adequacy of hospital resources [7]. The interaction of both individual and external factors could lead to further delays in arrival and treatment.

One of the most important problems that tends to occur in an emergency is the saturation of resources in the area where the event occurs, leading to a delay in care. Therefore, a crucial factor in solving this problem is to have correctly sized health resources to minimise the risk involved in this type of situation. Those responsible for managing the availability of these resources need to take into account multiple factors that need to be considered in order to provide an adequate emergency health service. These include the management experience of the person in charge of the service, who normally relies on data from previous years' time series, such as using the average number of health demands at the same time in previous years. In this sense, it should be noted that the reasons for healthcare demands are multiple, which is why there is fluctuation in the number of demands that occur. For this reason, in order to manage resources appropriately, it is also important to have a forecast of the type of health demands that can be produced. In order to obtain this predictive information, it is very useful to use the recent advances in Artificial Intelligence (AI) and Machine Learning (ML). This technology have led to substantial advances in the prediction and identification of health emergencies, disease populations, disease status and immune response, among others [4]. These are used to generate the predictive model. In the field of medicine there are

* Corresponding author.
E-mail address: miramos@ujaen.es (M.I. Ramos).

https://doi.org/10.1016/j.ijmedinf.2023.105163
Received 22 May 2023; Received in revised form 26 June 2023; Accepted 24 July 2023
Available online 24 July 2023

numerous researches where these algoritms have been used to predict key aspects of health, such as managing appointments, predict the number of patients attending health centers, etc. [5,6]. There are different algorithms applicable to this type of task, each of which is more effective depending on the type of data to be predicted. In fact, there are authors who have carried out work on the prediction of infectious diseases and the emergency health care that has been derived for this reason. In these studies, Bayesian networks have been used to obtain the prediction data [7]. Other studies have used classification algorithms in hospital emergency departments to make an initial classification of patients according to their medical situation. Even in the case of the supply of medicines in health facilities, it is important to have information in advance about which patients will require which medicines. In this field, there are also studies that have inferred data about the amount of demands for drugs related to diseases about respiratory insufficiency [8].

A study is presented on the analysis of the use of the health emergency service in a city, Jaén, in the south of Spain. In this work, firstly, an analysis of the data on emergency demands has been carried out. After a study of the different pathologies attended to in the Accident and Emergency Department, it has been observed that a significant number, 10 %, are caused by respiratory diseases and it is known that these are influenced by different factors: contamination, temperature, humidity, etc. [9–15]. The aim of this work is to study the factors that influence the demand for respiratory diseases and that generate a health emergency leading to the mobilisation of health resources for their care. The input data used are the history of the number of health demands made in each area, population data for each area and the meteorological and environmental factors occurring in the area. Based on this data, the number of demands that will occur in the immediate future is predicted. This generates a useful tool for those responsible for sizing these health services, providing data to help in decision-making in this task.

Actually there are several researches where ML has been used in healthcare to support a doctor's or analyst's ability to perform their functions, identify health trends and develop disease prediction models [16–20]. Others have used these techniques to manage appointment schedules in primary care health centres [5,6,21]. The work presented here addresses one of the predominant challenges in emergency departments, the optimal sizing of available resources. The increasing availability of clinical, resource mobilisation and other external multisource data means that the use of computational techniques such as ML enables the meaningful processing of large amounts of complex data. Thus contributing to the generation of expert systems capable of anticipating critical situations in emergencies.

2. Methods

2.1. Study area

The municipality of Jaén is located in the province of Jaén, in the autonomous community of Andalusia, Spain. It is located in the south of the country, Fig. 1. Jaén is situated on a hill, which gives it a mountainous relief. The city spreads over a hillside and is surrounded by a series of hills and mountains and is at the confluence of the Guadalbullón and Guadalquivir rivers.

The natural environment surrounding Jaén is characteristic of the Mediterranean area, with predominantly scrubland and olive groves. Olive groves are an integral part of the region's landscape, and olive oil production is an important economic activity in the area. It is precisely the flowering of this crop that causes numerous episodes of asthma among the population. The province of Jaén has more than 70 million olive trees, which, during the flowering season, gives rise to very high pollen concentrations, causing great damage to the health of the allergic population.

In terms of weather conditions in Jaén, the area experiences a Mediterranean climate. Summers tend to be hot and dry, with temperatures often exceeding 30 degrees Celsius (86 degrees Fahrenheit). Winters are mild, with temperatures ranging from 8 to 15 degrees Celsius (46 to 59 degrees Fahrenheit). Precipitation is relatively low

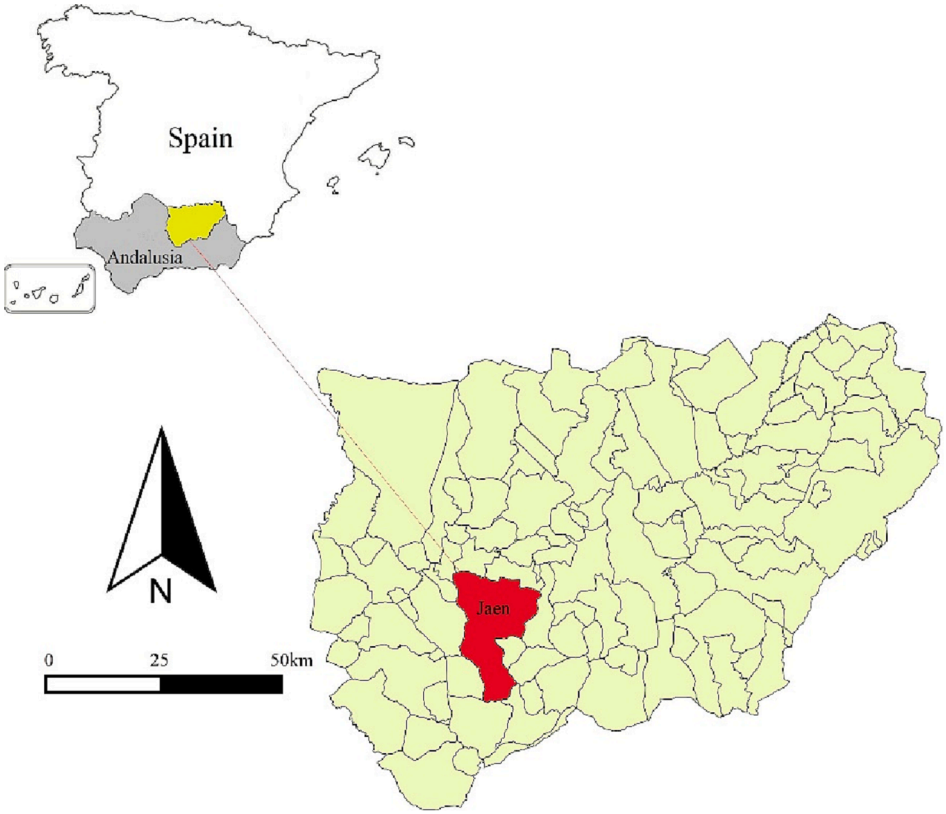


Fig. 1. Location of study area in Jaen, Andalusia (Spain).

throughout the year, with occasional rainfall occurring mainly during the autumn and spring seasons.

2.2. Data integration and management

The dataset required for this research includes all the information corresponding to the healthcare demands received by the emergency services in the city of Jaén, in southern Spain, from 2012 to 2019 inclusive. In this sense, it is important to bear in mind that the data is geolocated, i.e., its spatial component has been taken into account, i.e., where the demand for emergency health care has occurred. In addition, a series of secondary data necessary to contextualise the particular conditions in which each emergency episode occurs have been included. Thus, several multi-source variables have been added, such as environmental and meteorological data for the area in the years previously indicated, as well as socioeconomic data for each of the zones into which the city of Jaén is divided [22]. The following is a description of each of the variables that make up the dataset:

- Emergency healthcare data: All information related to the health care that the patient has received is stored, since the telephone call is received in the coordinating centre, until the medical team states the case as finished. These data include the user's requests for assistance in urgency and emergency situations including diagnosis, clinical trial, treatment, antecedents, resources mobilised and detailed action times as well as the geolocation and their resolution.
- Atmospheric data: Data collected by the environmental information network of Andalusia (REDIAM) and provided by the Regional Ministry of the environment and regional planning [23]. These data include values for temperature, precipitation, humidity, wind speed and solar radiation.
- Sociological data: Data related to the personal information of users were divided into:
- Economic level of the patients in each area: Analysing, on one hand, the cadastral value of real estate, extracted from the Directorate General for Cadastre of Spain website [24]. We also added an analysis of the current price of housing in each district of the city through real estate web portals.
- Level of unemployment in patients and their family units: This variable was obtained from data provided by the Spanish National Statistics Institute [25]. This public institution provided us with the type of population in each census tract. In Spain, a census tract composes a small region of the city, 1000 and 2500 residents.
- Education level, age of citizens, and members of the family unit: These data were obtained from the website of the Institute of Statistics and Cartography of Andalusia [26].

The methodology followed in this work is sequenced in several phases, from data understanding, data preparation and model generation to model validation. The output of the model, i.e. the target attribute, is the number of emergency health demands due to respiratory diseases that will be received at the control centre for a selected date. The aim is to predict an unknown data, but which can be deduced from other variables that are known and which are directly or inversely related to the target attribute. In this sense, the study is based on a supervised data mining analysis, where the value of an unknown variable is deduced from a few known variables.

The general flow of the methodology carried out consists of ML algorithms and techniques used as follows in each phase:

1. Data extraction and uploading. Meteorological data are downloaded from public web servers. The data relating to emergency health care provided are supplied by the Empresa Pública de Emergencias Sanitarias (Public Company for Health Emergencies) [27]. Both will be loaded into the database management system.
2. First analysis of the data. All the data are explored and analysed using distribution techniques (histograms), with the aim of reviewing the data and purging those whose dispersion or variability may cause inconsistencies in the study.
3. Anomaly detection. Anomaly detection is implemented as a class classification algorithm where the algorithm is able to predict, with a certain probability, whether a record in the data is typical of the distribution. The aim of this phase is to identify those cases that are not common within our data.
4. Data transformations. In this section, both yield data and meteorological meteorological data are transformed, i.e. formats, units, rescaled, etc., are adapted so that they can be optimally exploited by predictive models. optimally exploited by predictive models.
5. Aggregation of data. The downloaded information is aggregated on a monthly basis; therefore, the rest of the data to be added to the study should be aggregated in the same way.
6. Data integration. For our work, we have heterogeneous information from different sources. Thus, in order to carry out the data mining study, it is necessary to integrate all the information into a single source that serves as input for the predictive models
7. Detection of the level of influence of the input attributes on the target attribute. Before generating the model, the influence of each attribute on the target attribute (number of emergency health demands) is analysed in order to include or exclude attributes from the study based on their level of influence on the prediction.
8. Application of regression algorithms [28–30]. Different regression algorithms are tested to predict health care attendance. The objective of regression analysis is to determine the parameter values of a function that best fit an observational data set. There are different families of regression functions and different ways of measuring error. In this paper we have analysed linear and non-linear regression functions, as well as different parameters to assess the goodness of fit of the models.

Following the workflow of the methodology described above, the initial data were analysed. In this case, it is the data on the number of resources mobilised to attend to the demands of emergencies due to respiratory pathologies. Several analyses are carried out at different levels of grouping. Firstly, the data were grouped by month, as shown in Table 1, and then the evolution of these data over time was analysed, Fig. 2. The latter shows that in the first months of the year and in the last months, corresponding to autumn and winter, the number of resources mobilised is higher than in the months with good weather conditions, such as spring and especially in summer.

When analysing the series in Fig. 2 in detail, an anomalous behaviour is observed in the trend during 2014 in the months of February to May, as a very sharp increase in the activation of health resources is observed. This type of situation supports the need to use data mining to obtain models capable of predicting this type of situation. Since, if we follow a naive model, i.e. averaging, the decisions taken to size health resources in these months of 2014 were not adequate to provide a quality emergency service.

The initial study also analysed the distribution of care demands throughout the week, Table 2. It can be seen that the distribution of demands increases at weekends. Fig. 3. This is due to the fact that on weekdays patients have other health care alternatives available, such as going to primary health care centres, and do not have to resort to the emergency health care service. Although the difference is appreciable, it cannot be excessively significant, as we must not forget that the emergency telephone number is intended for patients with pathology and important symptoms, in which cases patients go directly to the emergency services.

On analysing the demands by day of the week, it can be seen that on Mondays there is a slight increase in demands with respect to the rest of the working days. This may be due to the fact that primary care health centres are not open on Saturdays and Sundays, and these patients have

Table 1
Emergency health resources activated to care for patients with respiratory diseases, grouped by month from 2012 to 2019.

Year	January	February	March	April	May	June	July	August	September	October	November	December
2012	222	165	206	150	153	127	88	109	89	160	129	132
2013	163	144	189	190	131	119	112	120	99	99	129	162
2014	188	292	281	200	246	98	116	101	121	144	169	139
2015	214	158	196	213	199	184	126	101	120	145	170	193
2016	357	192	188	202	171	132	88	100	158	152	168	237
2017	370	240	233	167	129	99	37	53	43	76	99	81
2018	154	154	176	133	142	127	123	85	115	97	143	124
2019	233	160	122	133	143	100	121	93	112	101	131	193

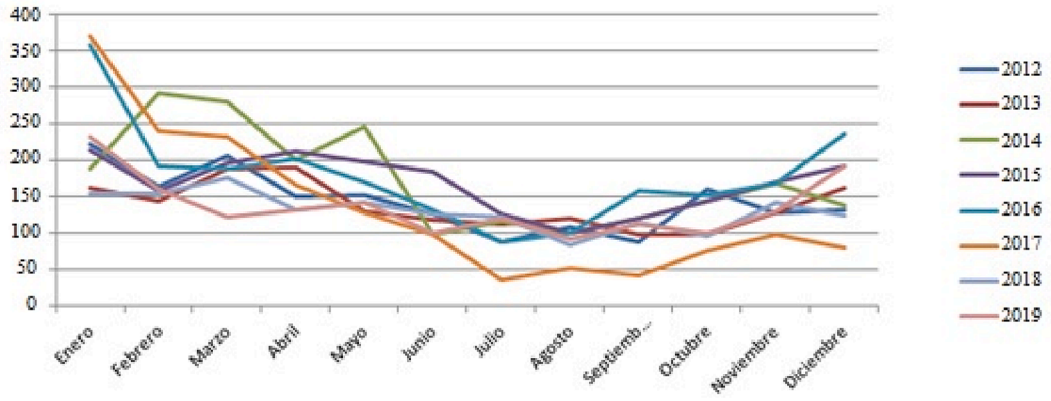


Fig. 2. Graph of the distribution of emergency health resources activated to attend patients with respiratory diseases, grouped by month from 2012 to 2019.

Table 2
Emergency health resources activated to care for patients with respiratory diseases, grouped by week days from 2012 to 2019.

Year	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2012	274	242	212	194	268	261	279
2013	260	220	199	229	225	257	267
2014	257	277	288	307	335	277	354
2015	279	270	271	270	279	350	300
2016	283	279	307	286	328	299	363
2017	266	213	210	193	210	245	290
2018	225	185	212	221	223	254	253
2019	253	222	191	237	229	256	254

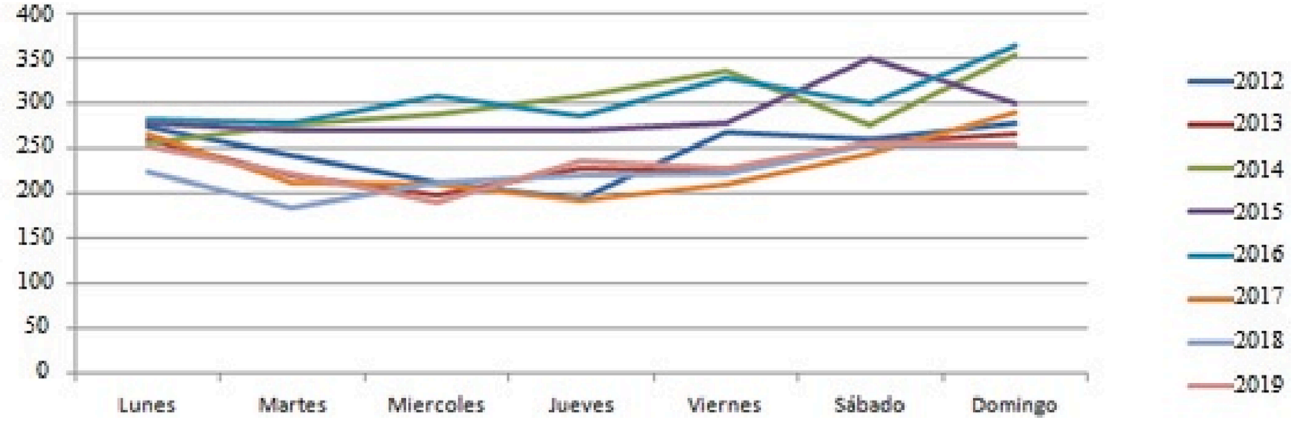


Fig. 3. Graph of the distribution of emergency health resources activated to attend patients with respiratory diseases, grouped by week days from 2012 to 2019.

not been able to be seen by their doctor over the weekend. Consequently, demand is lower on the other days, as the primary care health centres are open on weekdays and many patients can be seen by the doctors at these centres. Thus, it has also been observed that the evolution of these data on public holidays is similar to that of weekends.

Taking all this information into account and after observing that there is a great difference in the behaviour of emergency demands on working days and public holidays, this predictive study will treat these

data separately, generating different models. One prediction for weekdays, i.e. weekdays, and another predictive model for public holidays, i.e. weekends. Having said that, for the construction of these models it has been estimated that the best grouping of the data is by: Year, month, week, type of day (weekday or weekend/holiday). The Table 3 shows an example of the table that will be used to make one of the models, with the data that will serve as input to generate the predictive model. It shows the level of grouping by days, distinguishing between weekday or weekend/holiday.

A classical data mining life cycle is used in this study: Once cleaning algorithms have been applied to detect anomalies in the data and the data has been clustered, several regression models are tested. The results of each model are statistically analysed by cross-validation to identify the best fit.

2.3. Data mining techniques used

Data mining is a combination of statistics, computer science and ML. The techniques used in data mining involve a set of calculations to create a model from a set of selected data. In this sense, to create such a model, an algorithm first analyses the data provided, looking for specific types of patterns or trends. It uses the results of this analysis over many iterations to find the optimal parameters for generating the model. These parameters are then applied to the entire dataset to extract actionable patterns and detailed statistics. In this work, the complete data analysis has been carried out with an analysis module included in Oracle Data Mining software [31]. In this phase the challenge is to choose the algorithm that best performs the prediction of the target attribute.

Once the dataset is available, anomaly detection has to be carried out on the data before it becomes part of the input data. This is to identify data that appear to be homogeneous with the rest of the data but are not. Anomaly detection is important to detect fraud, outliers and other rare events that can have a major negative impact on the overall process and are a priori difficult to detect. A classification algorithm is used in this phase because anomaly detection can be considered as a type of classification. A classifier of a class develops a profile that generally describes a typical case in the training data. Deviation from the profile is identified as an anomaly. Specifically, in this phase, the algorithm used has been the Support Vector Machine algorithm (SVM) [32–34]. It is an algorithm that has recently been used in prediction work related to medical data [35–40]. SVM works on the basic idea of minimising the hypersphere of the single class of examples in the training data and considers all other samples outside this hypersphere as outliers or outside the training data distribution. This algorithm produces a prediction and a probability for each case in the score data. If the prediction is 1, the case is considered typical. If the prediction is 0, the case is considered anomalous. This behaviour reflects the fact that the model is trained on normal data.

Following the workflow already explained in previous sections, before creating the model, the level of influence that each of the input variables has on the target attribute was calculated using the Minimum Description Length (MDL) algorithm [32]. The execution of this algorithm returns a value between −1 and 1. The value −1 means that this variable has no relationship with the target attribute and that it can introduce noise in the model, so it should be discarded from the calculation of the model. The value 0 means that the variable is not related to the target and 1 means maximum relationship between the variable and the target attribute. The MDL algorithm treats each attribute as a predictive model of the target class. Each predictive model of these attributes is compared and ranked in the MDL metric. MDL penalises the complexity of the model to avoid overfitting. In this sense, prior to the input of the data into the algorithms, as indicated in section 2.2, an exploration and analysis of the data was performed using distribution techniques, histograms, as well as anomaly detection. This avoided both overfitting and underfitting of the models when applying the MDL. Only those attributes with weight greater than 0 are considered in this study, discarding all those with 0 or negative values. As a result, the most suitable variables for this research are obtained, which can be used for the generation of predictive models.

Finally, regression analysis algorithms have been used in order to generate the model, that is the prediction of the target, number of emergency health demands due to respiratory diseases will be requested on a certain date. A regression task starts with a data set in which the target values are known. A regression algorithm estimates the target value as a function of the predictors for each instance in the data set. The relationships between the predictors and the target are summarised in a model that can be applied to another data set where target values are unknown. Regression models are tested by calculating various statistics that measure the difference between predicted and actual values [41,42]. The historical data for a regression project is usually divided into two data sets: one to build the model and the other to test the model. It is necessary to specify that the date used for model testing is not included in the data set. is not included in the training data set. For this study, a pure linear model of the algorithm Generalized Linear Models (GLM) [43] is used and other models applying Support Vector Machines (SVM) with Gaussian and Linear Kernel respectively [44–46].

3. Results and Discussion

The pre-analysis of the dataset in section 2.2 showed that in this study it is necessary to separate the prediction for weekdays from that for public holidays or weekends. The reason for this is that the reasons that lead to demand for emergency services are different in both cases. Therefore, a separate study is required.

3.1. Weekday model generation

Based on the algorithms described above, a dataset from 2012 to 2019 is used as training data. In this regard, it is important to note that data corresponding to the period of the COVID-19 pandemic have not been included. The reason for this is that, during those years in Jaén, primary health care services were closed and only telephone care was provided. As a result, the data on emergency health demands for this period are abnormally saturated. Including this period in the study would distort the trend of the model since there is no homogeneity in the conditions under which emergency service demand occurs.

The predictive models are designed using the algorithms mentioned

Table 3
Example table of the Dataset used to generate the predictive models.

Year	Month	Week	Typeof day	Max Temp.	Min Temp.	Mean Temp.	Max Hum.	MinHum.	Mean Hum.	Wind speed	Rad.	Rainfall	Num Resources	Num Days	Mean Res.
2019	1	1	Weekday	14.6	1.6	7.7	95.8	49.5	80.8	0.5	6.1	0.0	23	3	7.7
2019	1	2	Weekday	15.3	−1.3	5.8	97.3	47.4	81.0	0.6	9.0	0.1	40	5	8.0
2019	1	3	Weekday	9.5	−1.3	3.8	91.0	40.1	70.4	0.9	7.3	0.7	33	5	6.6
2019	1	4	Weekday	13.5	−1.0	5.4	96.3	42.1	77.3	0.9	9.9	1.6	25	5	5.0
2019	2	1	Weekday	17.0	4.4	10.1	93.1	47.6	76.2	1.1	9.5	0.1	30	5	6.0
2019	2	2	Weekday	15.3	1.0	7.8	95.7	40.1	76.9	0.9	11.1	3.7	23	5	4.6
2019	2	3	Weekday	18.8	3.9	11.0	90.9	34.5	65.7	1.0	12.8	0.0	38	5	7.6
2019	2	4	Weekday	18.9	7.6	13.5	92.5	39.5	67.9	1.0	9.1	1.8	15	4	3.8

in the previous sections. We have made eight different models for these working day data, in each of them we have taken a different year to test the goodness of fit of the model designed.

Firstly, as a starting point, the years 2013 to 2019 are taken as training data for our model, leaving 2012 as a comparison year to compare the results obtained in our model with the actual resource demands received in 2012. Then, to assess the reliability of the model, the k-fold cross validation technique was used to evaluate results in statistical analyses [47]. This evaluation methodology consists of assessing the quality of each year's prediction by separating the data of the year to be predicted from the training data. In particular, for this research, it was used to test the reliability of the model in predicting the number of emergency health demands to be received at the health monitoring centre. As indicated above, production data from 2012 to 2019 was used for this study. Specifically, a predictive model was generated using data from seven years and then data from the eighth year was used to assess the reliability of the model.

The theoretical quality of the models generated from each of the regression algorithms was assessed on the basis of the differences between the predicted values and the actual values, Table 4.

Theoretically, the results obtained by the two variants of the SVM algorithm are better than those obtained by the GLM algorithm. Table 5 shows the parameters used for the application of the SVM algorithm. Although the values obtained in the two variants of the SVM algorithm are very similar, it will be shown later that if the data averaged over all years is used, the SVM algorithm with Gaussian kernel will give a better predictive result. In the SVM model configuration, the complexity factor parameter has been specified. This parameter allows balancing the model error (measured with respect to the training data) and the model complexity in order to avoid over-fitting or under-fitting of the data. Larger values provide a larger penalty to errors, leading to a higher risk of over-fitting the data; smaller values provide a smaller penalty to errors and may lead to under-fitting. After multiple tests this default value is 1 and in our study it is lowered to 0.589 precisely to avoid overfitting, Table 5.

The second part of this analysis is to check whether the theoretical results obtained in the first part correspond to reality. As a measure of the goodness of the models obtained, we have taken each week of 2012 and calculated the absolute value of the difference between the actual number of resources demanded and the value predicted by our three models. In order to confirm that the SVM algorithm with Gaussian kernel is the best performer for working days, the average value of the absolute errors of the three models is calculated for all years, Table 6. It is confirmed that the SVM algorithm performs better than the GLM. It is also confirmed that the SVM with Gaussian kernel performs better than the Linear one, although without great difference.

Having verified that the SVM algorithm with Gaussian kernel is the best fit, its behaviour is analysed in more detail over all the weeks of the selected year, 2012, Fig. 4.

The results shown in Fig. 4 confirm that the prediction of the Gaussian SVM model is consistent in values and trend over all weeks of 2012. In the first part of the year the number of emergency resources mobilised due to respiratory illnesses was higher than the rest of the year, due to the low temperatures and the increase in humidity in the environment. In the middle of the year there was a slight decrease in the number of resources required compared to the beginning of the year. It should be borne in mind that, as we are dealing with emergencies and emergencies, a very pronounced decrease is not to be expected. Finally, as expected, in the last period of the year the number of resources increased slightly again as weather conditions worsened again, with the arrival of winter.

It is interesting how our model's prediction fits the general trend throughout the year, as well as the small peaks that occur during the weeks.

Table 4
Error metrics of the predictive models obtained for weekdays using 2012 as control year.

Algorithm	Mean Absolute Error	Root Mean Square Error	Mean Actual Value	Mean predictive Value
GLM	1.823	2.207	6.241	6.649
SVM with lineal kernel	1.612	2.025	6.241	6.326
SVM with Gaussian kernel	1.629	2.069	6.241	6.326

Table 5
Parameters of the SVM algorithm with Gaussian kernel used to generate the predictive model. Data for the 2012 prediction.

SVM with Gaussian kernel	Parameters
Standard Deviation	2.028.306
Complexity Factor	0.589921
Kernel Function	Gausiano
Algorithm Name	Support Vector Machine
Active Optimisation	Enable
Automatic Preparation	Enable
SVMS_EPSILON	0.040708
Core Cache Size	50,000.000
Tolerance	0.001

Table 6
Mean results absolute error for weekdays models.

2012–2019	Mean absolute error
GLM	30.4 %
SVM with lineal kernel	27.8 %
SVM with Gaussian kernel	27.3 %

3.2. Generation of models on holidays and weekends

The methodology followed is identical to that for weekdays, with seven of the eight years of available training data being used to generate the models and one being left to check the quality of the prediction. In order to make a consistent comparison with the case of weekdays, the year 2012 was again selected as the control year and the period from 2013 to 2019 as the training data for the models.

The theoretical results on the quality of the models obtained are shown in the Table 7. In this case, according to the metrics of the models, they all obtain very similar results. However, theoretically, the one that best matches the real value is the SVM algorithm with linear kernel with an average absolute error of 2.326.

In order to corroborate whether the SVM algorithm with linear kernel is the one that obtains the best results for weekends and holidays, the average value of the absolute errors of the three models is calculated for all the years used as training data set. Thus, it can be confirmed that on this occasion the models behave in a very similar way, although as can be seen in Table 8, the average value corresponding to the SVM model with linear kernel is slightly better than that of the other two.

The Fig. 5 allows a detailed graphical analysis of the predictive behaviour of the model generated with the SVM algorithm with linear kernel, which, as has been shown, gives the best results.

This algorithm adjusts to the trend of resources requested throughout 2012, identifying even small peaks throughout this year. As was the case on weekdays, the trend in resources needed at weekends decreases slightly as the year progresses and weather conditions improve. It is worth mentioning the last week of the year, in which the difference between the real values and the forecast is slightly different. If this gap is analysed in detail, it is due to the fact that in this last week there was an extraordinarily small number of requests for emergency assistance compared to the same period of any other year. Another positive aspect

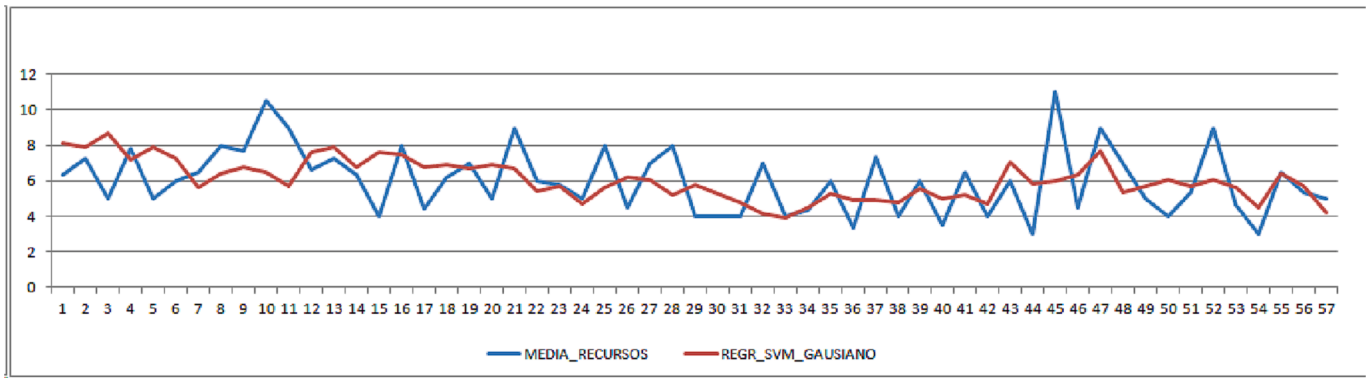


Fig. 4. Comparison between the real value of demands for emergency health resources and the value predicted by SVM with Gaussian kernel. Values per week considering only weekdays in 2012.

Table 7
Error metrics of the predictive models obtained for weekends and holydays using 2012 as control year.

Algorithm	Mean Absolute Error	Root Mean Square Error	Mean RealValue	Mean predictive Value
GLM	2.413	3.235	7.172	6.687
SVM with lineal kernel	2.326	3.213	7.172	6.185
SVM with Gaussian kernel	2.343	3.268	7.172	6..140

Table 8
Mean results absolute error for weekends and holidays models.

2012–2019	Mean absolute error
GLM	34.3 %
SVM with lineal kernel	32.8 %
SVM with Gaussian kernel	33.0 %

is that the prediction of our model detects the small change in the time trend of the actual cases in the middle of the year.

Another evaluation is made by analysing the coefficients resulted for the linear regression of the SVM predictive model with linear kernel, Table 9. These coefficients indicate the level of influence of each variable on the target attribute. It should be noted that this model takes as learning data the data corresponding to weekends and public holidays, a much smaller volume than the weekday model, but necessary due to the different functioning of the demand for resources.

If the variable Month is analysed, it can be seen that the first three or four months of the year have a greater weight than the following months, this is due to the fact that these are months with adverse weather conditions, which increases the presence of respiratory diseases, giving rise to a greater number of data available for the model. The greater availability of information is also influenced by the greater number of public holidays in the town of Jaen in the first four months of the year. It is also observed that weeks with 4 public holidays have a higher weight on the model than weeks with fewer days, as we have more training information available coinciding with holiday periods.

Finally, it can be seen that the variables average humidity, average temperature and wind speed have a very important positive weight. This is because they have a direct influence on the symptoms related to respiratory diseases, such as colds, bronchitis, asthma caused by pollen allergies, etc… Although these data correspond to the model in which the control year is 2012, a similar behaviour has been verified in the models corresponding to the other seven remaining years.

4. Conclusion

In this study, the challenge of generating an effective predictive model to estimate the health demands of emergencies to be received at the control centre of the municipality of Jaén, Spain, has been achieved. It has been calculated for the most recurrent specific case in this municipality, which are emergencies related to respiratory diseases. In addition, two prediction models have been generated, one for weekdays and the other for weekends and public holidays, since the behaviour of both models is different. The results indicate that the best algorithm for predicting the health resources needed to attend respiratory emergency episodes in the city of Jaén on weekdays is the SVM algorithm with Gaussian kernel. However, for the case of weekends and public holidays, the SVM algorithm with linear kernel was more efficient.

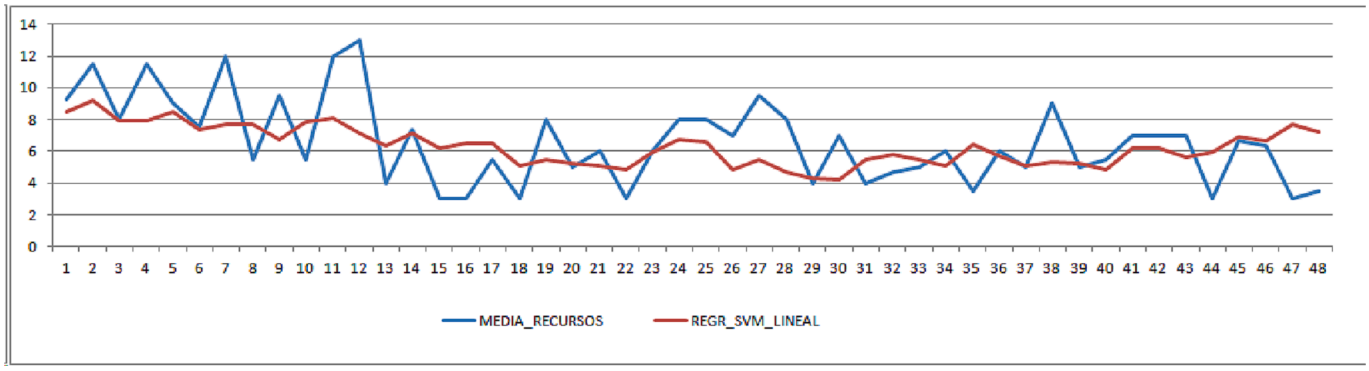


Fig. 5. Comparison between the real value of demands for emergency health resources and the value predicted by SVM with Gaussian kernel. Values per week considering weekends and holidays in 2012.

Table 9
Coefficients used for the SVM linear kernel algorithm.

Variable	Value	Coefficients
maximum_humidity		−0.028
mean_humidity		0.093
minimum_moisture		0
month	10	−0.086
month	1	0.079
month	11	−0.059
month	5	−0.049
month	9	0.038
month	3	0.034
month	4	0.029
month	8	0.029
month	12	−0.024
month	2	0.019
month	7	−0.01
n_days	1	−0.041
n_days	4	0.029
n_days	2	0.021
n_days	3	−0.009
precipitation		0.019
radiation		0.09
week_month	2	0.019
week_month	4	−0.017
week_month	5	0.008
week_month	3	−0.007
week_month	1	−0.003
maximum_temperature		−0.22
mean_temperature		0.039
minimum_temperature		−0.006
wind_speed		0.072

The regression coefficients of the models show a direct relationship between weather conditions and the increase of patients with respiratory failure, and, consequently, with the need for increased emergency resources to attend to this demand.

The main achievement of this research is to have constructed a model capable of predicting the number of emergency health resources needed to respond to patients with respiratory diseases in a given area and period. Therefore, this model helps to make decisions about the efficient sizing of available health resources and their location in the most efficient way possible. The use of this model avoids shortages and over-dimensioning of human and material health resources. On the other hand, it allows resources to be located where they will really be needed in order to minimise patient waiting time, a vital aspect in emergencies, directly and positively influencing their health.

On the other hand, from a less vital but also important point of view, the use of this predictive model has a positive impact on the economic costs derived from the management of emergencies, as it is possible to adjust human and material health resources to those that are strictly necessary.

Consequently, the predictive model generated is a vital tool for managing and responding in the shortest possible time to patients with respiratory diseases requiring emergency care in the city of Jaén.

Declaration of Competing Interest

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

Acknowledgements

This work would not have been possible without the support of EPES (Public Company for Health Emergencies), a company belonging to the Andalusian Health System. Also, this work has been partially supported by Graphics and Geomatics Group of Jaén (TIC-144).

References

[1] A. Guttmann, M.J. Schull, M.J. Vermeulen, T.A. Stukel, Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada, BMJ (Clinical Research Ed.) 342 (2011) d2983, https://doi.org/10.1136/bmj.d2983.
[2] E.L.D.S. Cabral, W.R.S. Castro, D.R.M. Florentino, D.A. Viana, Costa Junior, R. P. Souza, A.C.M. Rêgo, I. Araújo-Filho, A.C. Medeiros, Response time in the emergency services. Systematic review, Acta cirurgica brasileira 33 (12) (2018) 1110–1121, https://doi.org/10.1590/s0102-865020180120000009.
[3] I. Beltrán Guzmán, J. Gil Cuesta, M. Trelles, O. Jaweed, S. Cherestal, van Loenhout, D. Guha-Sapir, Delays in arrival and treatment in emergency departments: Women, children and non-trauma consultations the most at risk in humanitarian settings, PloS one 14 (3) (2019), e0213362, https://doi.org/10.1371/journal.pone.0213362.
[4] H. Habehh, S. Gohel, Machine Learning in Healthcare, Current genomics 22 (4) (2021) 291–300, https://doi.org/10.2174/1389202922666210705124359.
[5] M.I. Ramos, J.J. Cubillas, J.M. Jurado, W. Lopez, F.R. Feito, M. Quero, J. M. Gonzalez, Prediction of the increase in health services demand based on the analysis of reasons of calls received by a customer relationship management, The International journal of health planning and management 34 (2) (2019), e1215–e1222, https://doi.org/10.1002/hpm.2763.
[6] J.J. Cubillas, M.I. Ramos, F.R. Feito, Use of Data Mining to Predict the Influx of Patients to Primary Healthcare Centres and Construction of an Expert System, Applied Sciences 12 (22) (2022) 11453, https://doi.org/10.3390/app122211453.
[7] Shan Gao, Wang Hanyi, Scenario prediction of public health emergencies using infectious disease dynamics model and dynamic Bayes, Future Generation Computer Systems 127 (2022) 334–346.
[8] J.J. Cubillas, M.I. Ramos, M.C. Gutierrez, M.C. Rrez, F.R. Feito, A. Parra, J.C. Arias, Use of meteorological, environmental and spatial variables to predict drug Use, in: Digital Healthcare Empowering Europeans, Studies in Health Technology and Informatics, IOS Press, 2015, p. 938.
[9] H. Qiu, K. Tan, F. Long, L. Wang, H. Yu, R. Deng, H. Long, Y. Zhang, J. Pan, The Burden of COPD Morbidity Attributable to the Interaction between Ambient Air Pollution and Temperature in Chengdu, China. Int J Environ Res Public Health. 2018 Mar 11;15(3):492. doi: 10.3390/ijerph15030492. PMID: 29534476; PMCID: PMC5877037.
[10] P. Almagro, C. Hernandez, P. Martinez-Cambor, R. Tresserras, J. Escarrabill, Seasonality, ambient temperatures and hospitalizations for acute exacerbation of COPD: a population-based study in a metropolitan area, International journal of chronic obstructive pulmonary disease 10 (2015) 899–908, https://doi.org/10.2147/COPD.S75710.
[11] R.R. Duan, K. Hao, T. Yang, Air pollution and chronic obstructive pulmonary disease, Chronic diseases and translational medicine 6 (4) (2020) 260–269, https://doi.org/10.1016/j.cdtm.2020.05.004.
[12] J. Dawson, C. Weir, F. Wright, C. Bryden, S. Aslanyan, K. Lees, W. Bird, M. Walters, Associations between meteorological variables and acute stroke hospital admissions in the west of Scotland, Acta neurologica Scandinavica 117 (2) (2008) 85–89, https://doi.org/10.1111/j.1600-0404.2007.00916.x.
[13] T.H. Oiamo, I.N. Luginaah, D.O. Atari, et al., Air pollution and general practitioner access and utilization: a population based study in Sarnia, 'Chemical Valley,' Ontario, Environ Health 10 (2011) 71, https://doi.org/10.1186/1476-069X-10-71.
[14] G.C. Donaldson, J.J. Goldring, J.A. Wedzicha, Influence of season on exacerbation characteristics in patients with COPD, Chest 141 (1) (2012) 94–100, https://doi.org/10.1378/chest.11-0281.
[15] U. Ferrari, T. Exner, E.R. Wanka, C. Bergemann, J. Meyer-Arnek, B. Hildenbrand, A. Tufman, C. Heumann, R.M. Huber, M. Bittner, R. Fischer, Influence of air pressure, humidity, solar radiation, temperature, and wind speed on ambulatory visits due to chronic obstructive pulmonary disease in Bavaria, Germany, International journal of biometeorology 56 (1) (2012) 137–143, https://doi.org/10.1007/s00484-011-0405-x.
[16] S.R. Rao, C.M. Desroches, K. Donelan, E.G. Campbell, P.D. Miralles, A.K. Jha. Electronic health records in small physician practices: availability, use, and perceived benefits. J Am Med Inform Assoc. 2011 May 1;18(3):271-5. doi: 10.1136/amiajnl-2010-000010. PMID: 21486885; PMCID: PMC3078653.
[17] M.K. Siddiqui, R. Morales-Menendez, X. Huang, et al., A review of epileptic seizure detection using machine learning classifiers, Brain Inf 7 (2020) 5, https://doi.org/10.1186/s40708-020-00105-1.
[18] A.Z. Woldaregay, E. Årsand, T. Botsis, D. Albers, L. Mamykina, G. Hartvigsen, Data-Driven Blood Glucose Pattern Classification and Anomalies Detection: Machine-Learning Applications in Type 1 Diabetes, Journal of medical Internet research 21 (5) (2019), e11030, https://doi.org/10.2196/11030.
[19] L. Tian, D. Zhang, S. Bao, P. Nie, D. Hao, Y. Liu, J. Zhang, H. Wang, Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas, Clinical radiology 76 (2) (2021) 158.e19–158.e25, https://doi.org/10.1016/j.crad.2020.08.038.
[20] J.H. Kaouk, J. Garisto, M. Eltemamy, R. Bertolo, Robot-assisted surgery for benign distal ureteral strictures: step-by-step technique using the SP® surgical system, BJU international 123 (4) (2019) 733–739, https://doi.org/10.1111/bju.14635.
[21] J.J. Cubillas, M.I. Ramos, F.R. Feito, T. Ureña, An improvement in the appointment scheduling in primary health care centers using data mining, Journal of medical systems 38 (8) (2014) 89, https://doi.org/10.1007/s10916-014-0089-y.
[22] J.C. Arias, J.J. Cubillas, M.I. Ramos, Optimising Health Emergency Resource Management from Multi-Model Databases, Electronics 11 (21) (2022) 3602, https://doi.org/10.3390/electronics11213602.

[23] REDIAM Red de Información Ambiental de Andalucía - Portal Ambiental de Andalucía Available online: https://www.juntadeandalucia.es/medioambiente/portal/acceso-rediam (accessed on 16 April 2023).
[24] Sede Electrónica Del Catastro - Inicio Available online: http://www.sedecatastro.gob.es/ (accessed on 18 January 2020).
[25] INE INE. Instituto Nacional de Estadística Available online: https://www.ine.es/ (accessed on 16 April 2023).
[26] Instituto de Estadística y Cartografía de Andalucía Available online: https://www.juntadeandalucia.es/institutodeestadisticaycartografia (accessed on 13 January 2020).
[27] Empresa Pública de Emergencias Sanitarias EPES—061 | Gestión de las Emergencias y Urgencias Sanitarias en Andalucía; Málaga, Spain, 2023.
[28] Sonnberger, H. (1989). Regression diagnostics: Identifying influential data and sources of collinearity, by D. A. Belsley, K. Kuh and R. E. Welsch. (John Wiley & Sons, New York, 1980, pp. xv + 292, ISBN 0-471-05856-4, cloth $39.95. *Journal of Applied Econometrics*, *4*(1), 97-99. https://doi.org/10.1002/jae.3950040108.
[29] D.M. Allen, C.B. Foster, Analyzing Experimental Data by Regression, Belmont, Calif, ISBN 978-0-534-97963-8, 1982.
[30] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data; Econometric society monographs, Cambridge University Press, Cambridge, UK; New York, NY, USA, 1998. ISBN 978-0-521-63201-0.
[31] Oracle Data Miner Available online: https://www.oracle.com/big-data/technologies/dataminer/ (accessed on 19 May 2023).
[32] A.J. Dobson, An Introduction to Generalized Linear Models; Chapman & Hall/CRC texts in statistical science series, 2nd ed.;, Chapman & Hall/CRC:, Boca Raton, 2002. ISBN 978-1-58488-165-0.
[33] R. Chalapathy, A.K. Menon, S. Chawla Anomaly, Detection Using One-Class Neural Networks, arXiv preprint arXiv:1802.06360 **2018**.
[34] P. Oza, V.M. Patel, One-class convolutional neural network, IEEE Signal Process Lett. 26 (2019) 277–281, https://doi.org/10.1109/LSP.2018.2889273.
[35] P. Golpour, M. Ghayour-Mobarhan, A. Saki, H. Esmaily, A. Taghipour, M. Tajfard, H. Ghazizadeh, M. Moohebati, G.A. Ferns, Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography, Int. J. Environ. Res. Public Health 17 (2020) 6449, https://doi.org/10.3390/ijerph17186449.
[36] S. Singh, K.S. Parmar, S.J.S. Makkhan, J. Kaur, S. Peshoria, J. Kumar, Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries, Chaos Solitons Fractals 139 (2020), 110086, https://doi.org/10.1016/j.chaos.2020.110086.
[37] A.K. Gupta, V. Singh, P. Mathur, C.M. Travieso-Gonzalez, Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario, Journal of Interdisciplinary Mathematics 24 (1) (2021) 89–108, https://doi.org/10.1080/09720502.2020.1833458.
[38] H. Byeon, Predicting the Severity of Parkinson's Disease Dementia by Assessing the Neuropsychiatric Symptoms with an SVM Regression Model. Int J Environ Res Public Health. 2021 Mar 4;18(5):2551. doi: 10.3390/ijerph18052551. PMID: 33806474; PMCID: PMC7967659.
[39] K. Harimoorthy, M. Thangavelu, Retraction Note to: Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system, Journal of Ambient Intelligence and Humanized Computing 14 (1) (2023) 117, https://doi.org/10.1007/s12652-022-03971-1.
[40] P. Kumar, R.P. Chauhan, T. Stephan, A. Shankar, S. Thakur, A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* 2021, 568 – 574.
[41] P.D. Grünwald, J.I. Myung, M.A. Pitt, *Advances in Minimum Description Length: Theory and Applications*, A Bradford Book, Cambridge, M.A, USA. 2005 .
[42] B.M. Bolker, M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H. Stevens, J. S. White, Generalized linear mixed models: a practical guide for ecology and evolution, Trends in ecology & evolution 24 (3) (2009) 127–135, https://doi.org/10.1016/j.tree.2008.10.008.
[43] T.J. Hastie (Ed.), Statistical Models in S, 1st ed., Routledge, 1992 https://doi.org/10.1201/9780203738535.
[44] C. Ortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297, https://doi.org/10.1007/BF00994018.
[45] N. Cristianini, J. Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000, https://doi.org/10.1017/CBO9780511801389.
[46] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, New, 2007, https://doi.org/10.1142/S0218196799000199.
[47] J.D. Rodríguez, A. Pérez, J.A. Lozano, Sensitivity analysis of kappa-fold cross validation in prediction error estimation, IEEE transactions on pattern analysis and machine intelligence 32 (3) (2010) 569–575, https://doi.org/10.1109/TPAMI.2009.187.