



Forecasting Future Monthly Patient Volume using Deep Learning and Statistical Models

Zeydin Pala¹ · Ramazan Atıcı² · Erkan Yıldız³

Accepted: 25 February 2023 / Published online: 10 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

The variety of diseases is increasing day by day, and the demand for hospitals, especially for emergency and radiology units, is also increasing. As in other units, it is necessary to prepare the radiology unit for the future, to take into account the needs and to plan for the future. Due to the radiation emitted by the devices in the radiology unit, minimizing the time spent by the patients for the radiological image is of vital importance both for the unit staff and the patient. In order to solve the aforementioned problem, in this study, it is desired to estimate the monthly number of images in the radiology unit by using deep learning models and statistical-based models, and thus to be prepared for the future in a more planned way. For prediction processes, both deep learning models such as LSTM, MLP, NNAR and ELM, as well as statistical based prediction models such as ARIMA, SES, TBATS, HOLT and THETAF were used. In order to evaluate the performance of the models, the symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE) metrics, which have been in demand recently, were preferred. The results showed that the LSTM model outperformed the deep learning group in estimating the monthly number of radiological case images, while the AUTO.ARIMA model performed better in the statistical-based group. It is believed that the findings obtained will speed up the procedures of the patients who come to the hospital and are referred to the radiology unit, and will facilitate the hospital managers in managing the patient flow more efficiently, increasing both the service quality and patient satisfaction, and making important contributions to the future planning of the hospital.

Keywords Radiology unit · Radiological data · Monthly radiology patient volume · Time series · Deep learning models · Statistics-based models

✉ Zeydin Pala
z.pala@alparslan.edu.tr

¹ Department of Software Engineering, Engineering Faculty, Mus Alparslan University, Mus, Turkey

² Department of Electricity and Automation, Technical Sciences Vocational School, Mus Alparslan University, Mus, Turkey

³ Halkbank IT Assistant Specialist, Istanbul, Turkey

1 Introduction

Today, many innovations that facilitate business and operation processes, which play an important role in all areas of life, are also used intensively in radiology departments. With the spread of diseases and their global impact, the need for imaging services such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, x-ray and mammography has increased significantly in recent years [1], [2]. However, it is not possible for many hospitals to meet such requirements in a short time [3]. It is very important that this requisition is planned accurately and inspired by past experiences and data. In the event that this demand cannot be made in place and on time, it will prevent adequate service to patients, especially during the ongoing global epidemics such as COVID-19.

High-tech products, which make life easier by increasing service quality and become a part of daily life, are widely used in all areas of health services, from organizational management to patient care and health promotion.

For example, artificial intelligence (AI) is used extensively in the field of radiology as in many other fields [4], [5]. In the future, health services will be able to be delivered with the help of more AI on a global basis [6–8]. Health Informatics (HI), by examining the effective use of probabilistic information in decision making, is of great importance for the improvement of human health and well-being, combining these two fields. As a matter of fact, the health discipline is gradually turning into a data science and health systems around the world are becoming a source of big data [9]. The COVID-19 pandemic once again showed the whole world how inadequate hospitals, patient management and hospital units are for emergencies, and that it is not possible to plan the future without benefiting from past experiences [10]. Since human health is more valuable than anything else, unit capacity values should be included in future planning as well as hospital-based institutional planning for health. With the COVID-19 pandemic, it has been realized once again that the changes in the supply and demand of radiology resources make the efficient allocation of unit resources more and more important. Forecasting is a vital activity that guides decision making in many areas of economic, industrial and scientific planning [11].

In general, there are three different patient flows to the radiology unit of hospitals: outpatients, inpatients and emergency patients [2], [12]. Patients admitted to the hospital, here expressed as three groups, potentially indicate a serious workload on the radiology unit. Therefore, accurate estimation of patient flow referred to the radiology unit is of great importance for the hospital's performance in response to both immediate demand and potential future demand [2]. A satisfactory level of estimation can only be possible with the data of real radiological cases from the relevant unit.

One of the considerations in the radiology unit is the calibration of the devices used. In particular, the devices in the radiology unit are periodically calibrated to ensure that the radiation dose given to both hospital staff and patients is in a range that will not harm health. [13].

Institutional units make significant efforts to see the future by looking at the past. To contribute to this, this study deals with the precision estimation process.

The aim of this study is to predict the future planning of the radiology unit, which is an important unit in hospitals, to make it prepared for the future and to contribute to the literature. To achieve this, the following operations were performed:

1. Actual radiology data between the years 2010–2020 obtained from the devices in the radiology unit were used.

2. The data used in this study were handled as a time series in accordance with the explanation, modeling and estimation process.
3. In order for the estimation process to be successful, the number of radiological images taken with different devices in the unit where the data were supplied was taken into account and a multiple estimation model was proposed for this.
4. Both statistical-based and deep learning-based models are used for future predictions.
5. While LSTM, NNAR, MLP and ELM models are preferred for deep learning predictions, AUTO.ARIMA, TBATS, HOLT-WINTERS, THETAF and SES models are preferred for statistical-based predictions.
6. The sMAPE and MASE metrics, which are more sensitive than the MAPE metric, are preferred to evaluate the predictions of deep learning and statistical-based models.

2 Related Works

In this section, the future predictions of patients who applied to both the emergency unit and the radiology unit using daily and monthly time series were examined. As given in Table 1, although there are sufficient studies in the literature on the estimation of the volume of patients who enter the emergency unit daily [14–23], the number of estimation studies on daily [18] and monthly [24] admission to the radiology unit is quite insufficient. This study has an important place among the studies carried out to make up for such a deficiency and to contribute to the hospital management in terms of predicting the future. In addition, unlike the studies in the literature, this study provides important contributions to the literature for making monthly radiological case estimations of patients referred to the radiology unit from all points of the hospital with multiple time series models and estimating the future patient potential.

3 Methodology

In this study, the data of radiological cases obtained from the radiology unit were converted into monthly time series format and the potential for future radiological cases was tried to be estimated numerically with the help of time series estimation models. In this section, firstly, the subjects of data acquisition, data preprocessing, modeling of data with the help of time series are given, and then the forecasting models used in the forecasting process are briefly introduced. In the last part, the performance metrics used to evaluate the forecast results were introduced and then forecast analyzes were performed.

3.1 Radiological Data Sources

This study was carried out based on the monthly radiological case numbers between 2010–2020 of the Muş State Hospital, which is located in the east of Turkey and started to serve as a 50-bed hospital in 1954. Located in the city center of Muş, the hospital provides tertiary care as a 445-bed hospital. The data used in this study consists of the number of images taken in the radiology unit between January 1, 2010 and December 31, 2020. Numerical data consists of computed tomography, X-ray, ultrasound, magnetic resonance and mammography images taken in the radiology unit. As given in Table 2, the statistical distribution of 11-year,

Table 1 Studies conducted in recent years to estimate the workload on the emergency and radiology unit

Literature	Forecasting model	Contributions/Conclusions/Limitations
Jones et al. [15]	SARIMA Time series regression (TSR) Time series regression with climatic variables (TSR-CV) Exponential smoothing (ES) Artificial neural network (ANN)	In the study, they made 1, 7, 14, 21 and 30 day estimations with the help of the data of the patients who entered the emergency department of three different hospitals between January 1, 2005 and March 31, 2007. SARIMA, time series regression (TSR), time series regression with climatic variables (TSR-CV), exponential smoothing (ES) and artificial neural network (ANN) models were used in estimation processes. They stated that regression-based models were more successful in daily forecasting. However, using a limited number of models and making only daily forecasts are considered as constraints
Jilani et al. [14]	ARIMA, neural networks Fuzzy time series models	In their study, they defined a heuristic-based fuzzy logic model to predict patients coming to the emergency outpatient clinic using daily time series data between January-2011-December-2015, which they obtained from four hospitals in the UK. While ARIMA, neural networks and fuzzy time series models are used by the model for estimation, mean absolute percentage errors (MAPE) and root mean square errors (RMSE) metrics are used to evaluate their performance. They set the forecast period as 4 weeks and 4 months. They calculated the MAPE value between 2.5% and 7% in daily forecasts and between 2.09% and 2.81% in monthly forecasts. The fact that the time series used in the study was univariate was accepted as a limitation
Harrou et al. [15]	VAE, RNN, LSTM, bidirectional LSTM, convolutional LSTM, RBM, GRU, CNN	In another study, data from daily and hourly visits in the emergency department of a hospital in France were used and predictions were made with the variational autoencoder (VAE), a deep learning-based algorithm. In their results, the authors stated that the VAE model outperformed other deep learning models used in the study
Golmohammadi et al. [5]	Logistic regression and neural networks	In their study, they aimed to provide operational hospitalization planning for patients coming to emergency services, which can be used as a decision support tool, using secondary data obtained from the emergency department of a local hospital, with the help of logistic regression and neural networks. Ultimately, they stated that their study was able to accurately predict the probability of patient admission so that the proposed model could be used in practice

Table 1 (continued)

Literature	Forecasting model	Contributions/Conclusions/Limitations
Xu et al. [45]	ARIMA-LR GLM, ARIMAX and ANN	The ARIMA-LR model was used as a hybrid using daily data from 2012–2013 obtained from the emergency outpatient clinic of a hospital in China, and it was emphasized that it was better than the independently used GLM, ARIMAX and ANN models
Sun et al. [40]	ARIMA	Using the data between July 2005 and Mar 2008, patients' admission to the emergency service due to seasonal acute illness were analyzed under three categories. In the estimation process using the ARIMA model in time series analysis, they estimated the overall rate of referral to the emergency service with a MAPE value of 4.8%
Luo et al. [21]	ARIMA, SES	The authors used a combination of ARIMA and SES models for the daily patient estimation of a hospital in Chengdu, suggesting that the combinatorial model outperforms the models used alone
Zhu et al. [48]	ARIMA, MSARIMA, weighted Markov Chain models	In the study, three different models such as seasonal regression and ARIMA, a multiplicative seasonal ARIMA, MSARIMA and weighted Markov Chain models were used to predict daily departures from the hospital. They emphasized that the results will make significant contributions to the planning point for hospital management
Jones et al. [14]	VAR models	In this study, the temporal relationships between the demands for basic resources in the emergency room and inpatients were examined. For this purpose, univariate and multivariate models were compared using daily data obtained from three different hospitals in 2006. It was emphasized that multivariate time series showed better results than univariate time series
Harrour et al. [7]	Deep Belief Network (DBN), RBM, LSTM, GRU, CNN-GRU, LSTM-CNN, GAN-RNN	Using the daily data of the diatric emergency department (PED) of Lille central hospital, France, between January 2011 and December 2012, daily admissions to the unit were tried to be estimated with hybrid and non-hybrid models. As a result, it has been revealed that the DBM model makes better predictions than other models with a MAPE value of 4.097%

Table 1 (continued)

Literature	Forecasting model	Contributions/Conclusions/Limitations
Zhang et al. [45]	ARIMA, lasso, linear-SVR, Radial-SVR, random forest, and adaboost	ARIMA, lasso, linear-SVR, Radial-SVR, random forest, and adaboost models were used to estimate patients entering the radiology unit of a hospital in China. They emphasized that the lasso model came to the fore with a MAPE value of 7.06% in the estimates made with the help of 4-year daily data
Shahid et al. [37]	ARIMA	In the study, the data of the patients coming to the radiology unit of a hospital in Lahore for ultrasound between the years 2001–2015 were provided as a monthly time series and a monthly forecast was made between June 2015–May 2017 using the ARIMA model. As a result, they stated that the ARIMA model performed quite well in the predictions

Table 2 Statistical information of the experimental datasets

RD Datasets	Time range	Length	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Total RD	Jan 2010–Dec 2020	132	289	1854	6480	17517	3637	47608
RD train 1	Jan 2010–Dec 2019	120	289	1595	6073	15784	32550	47608
RD test 1	Jan 2020–Dec 2020	12	14654	36075	37210	34854	38611	45406
RD train 2	Jan 2010–Dec 2018	108	289	1204	5020	12886	28100	43061
RD test 2	Jan 2019–Dec 2020	24	14654	37015	38693	38359	42700	47608
RD train 3	Jan 2010–Jun 2018	102	289	1190	4522	11419	27079	40833
RD test 3	Jul 2018–Dec 2020	30	14654	36385	38322	38253	42498	47608

in other words, 132-month data, is 289, 47,608, 6480 and 17,517 as min, max, median and mean, respectively.

Figure 1. shows the graph of the 132-month number of radiological cases, trend, seasonal, and random components between 2010 and 2020. As seen in Fig. 1, the number of radiological cases, which continued with a normal increasing trend between 2010 and 2015, entered a rapid increase trend after the half of 2015. Because of the limited resources available at that time, radiological examination services could not be provided to enough patients.

As seen in Fig. 1, the number of radiological cases, which continued with a normal upward trend between 2010 and 2015, entered a rapid upward trend after the half of 2015. Due to the limited resources available at that time, adequate radiological examination services could not be provided to the patients. The average number of monthly radiological cases in 2010, 2011, 2012, 2013, 2014 and 2015 were 526, 937, 1761, 3217, 5165 and 5976, respectively. The average number of radiological cases per day for the same years was 17, 31, 58, 107, 172 and 199, respectively. However, after 2015, a rapid increase was observed in the number of patients served with the development of the radiological infrastructure and the increase in services accordingly. The average monthly numbers of radiological services in 2016, 2017, 2018, 2019 and 2020 are 27478, 32995, 37919, 41865 and 34854, respectively. Especially due to the COVID-19 epidemic that broke out in the world at the end of 2019, the number of radiological images in 2020 was also affected [25]. In addition, the collection of data in the radiology unit of the hospital, preprocessing, dividing the data for training and testing at different rates, selection of estimation models and evaluation processes are given schematically in Fig. 2.

Especially with the curfew imposed in March, April and May of 2020, the radiology unit was adversely affected as in all units of the hospital. The radiological case declines in March, April and May, which were most affected by the curfew, were 14%, 63% and 52%, respectively, from the previous years. This decrease is 18%, 14% and 18% for the last three months of 2020, October, November and December compared to the same months of 2019, respectively. In general, the trend of the number of radiological cases, which continues with a certain increase every year, has tended to decrease due to the fact that a large part of the hospital is allocated to COVID-19 patients in 2020.

3.2 Modeling of Radiological Cases as Monthly Time Series and Model Selection

Because of the significant relationship between the past and the future, time series analysis is widely used in a variety of fields. Time series models can be used to model data that is obtained and indexed by time [26]. In many fields [27], time series are used to predict

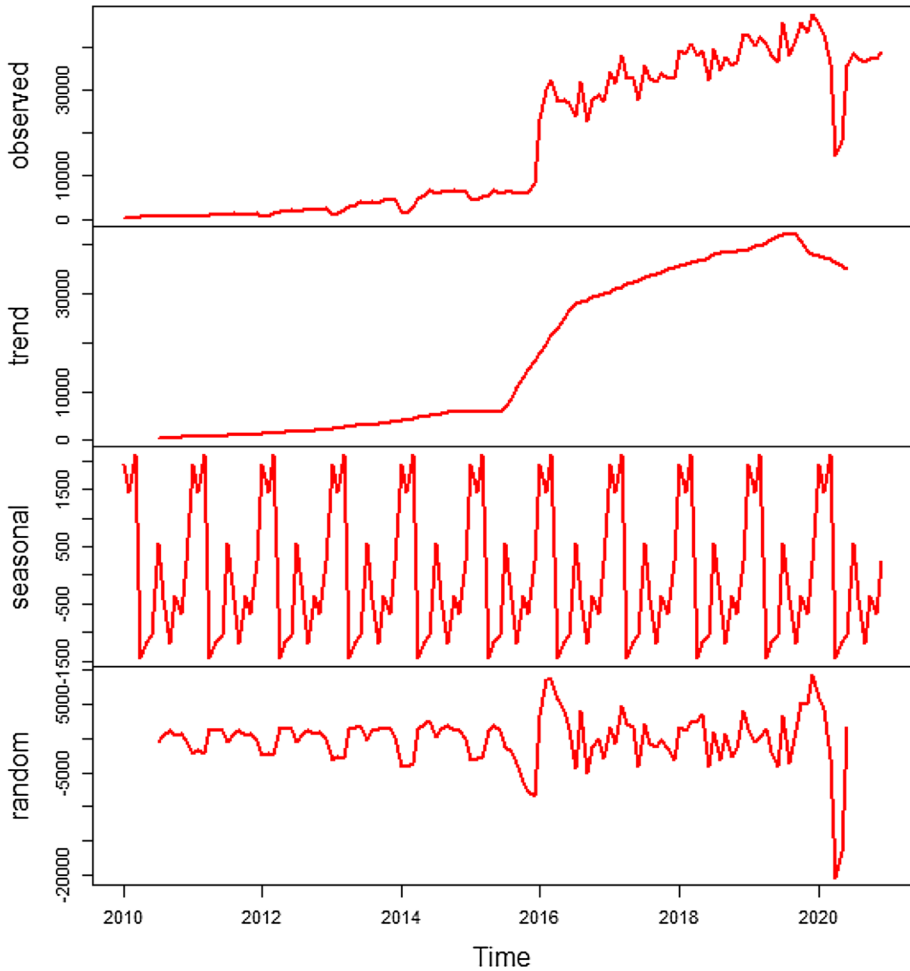


Fig. 1 132-month radiological case numbers, trend component, seasonal component, and random component

the future such as health [3] and [30], sensor data analysis [30], speech recognition [31], economics [32], radiation forecasting [33] and [33], sunspots [35] and [28], and natural gas forecasting [37]

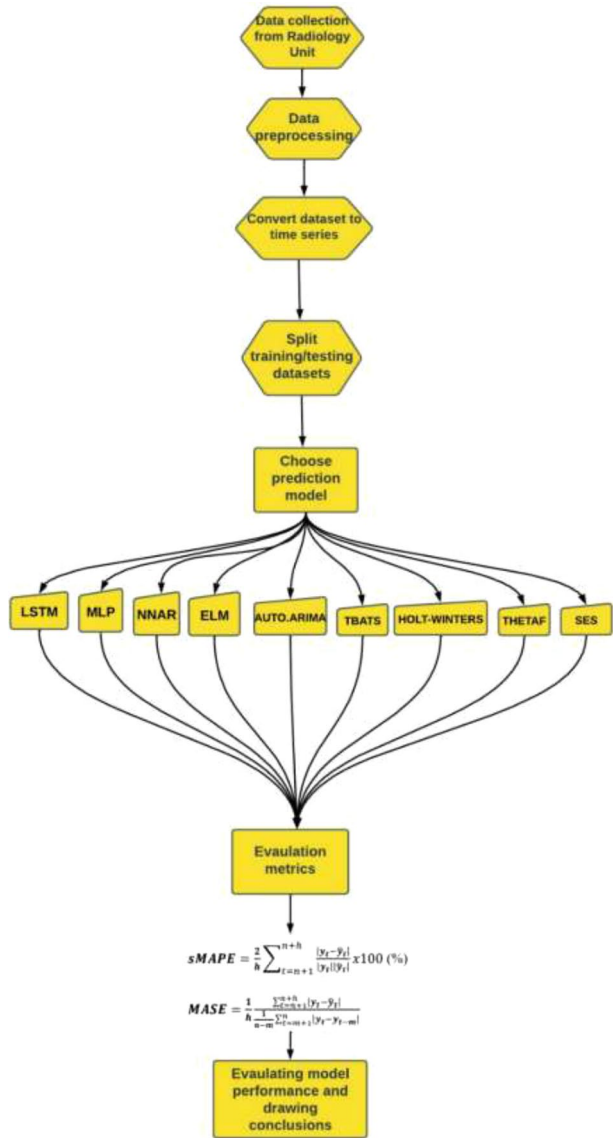
The following predictor variables can be used to model data supplied by the Radiology unit in time series format:

$$\text{RTS} = f(\text{radiologic_cases}, \text{time_of_month}, \text{error}) \quad (1)$$

As shown in Fig. 1, the "error" word on the right allows for random variation. As a result, the time series prediction equation will be as follows:

$$\text{RTS}_{n+1} = f(\text{RTS}_n, \text{RTS}_{n-1}, \text{RTS}_{n-2}, \text{RTS}_{n-3}, \dots, \epsilon) \quad (2)$$

Fig. 2 Schematic structure of the proposed forecasting model



where subscript n is the present month, subscript $n + 1$ is the next month, subscript $n - 1$ is the previous month, subscript $n - 2$ is the 2 months ago, subscript $n - 3$ is the 3 months ago, and so on. ϵ represents the random variation and the effects of variables not included in the model. Here, rather than using outside variables that might affect the system, future forecasting is based on a variable’s past values. The number of radiologic cases obtained each month at time t is displayed with y_t in the models that follow.

In this study, while using long short-term memory (LSTM), neural network autoregressive (NNAR), multilayer perceptrons (MLP) and extreme learning machines (ELM),

among the deep learning (DL) methods, AUTO.ARIMA, TBATS, HOLT-WINTERS, THETAF and SES models were used statistically [38–40].

Here, these statistical-based models and DL models, which are particularly simple but highly effective in estimation processes, are briefly explained.

The AUTO.ARIMA function in R combines unit root tests, which make use of a variant of the Hyndman–Khandakar algorithm [35] and offer a variety of modifications on the algorithm, minimization of the AICc, and maximum likelihood estimation (MLE) to produce ARIMA models. ARIMA models allow for the use of several parameters to make predictions. Non-negative integers make up the model's parameters (p , d , and q), which are employed in the ARIMA (p , d , q) algorithm. Where d denotes the severity of the trend difference, q denotes the moving average order, and p denotes the automatic regression order. The ARIMA model can be considered a “cascade” of two models. The first model is Eq. (3)'s non-stationary variant, as follows:

$$Y_t = (1 - L)^d X_t \quad (3)$$

The second is a stationary with a wide field of view:

$$\left(1 - \sum_{i=1}^p \alpha_i L^i\right) Y_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t \quad (4)$$

where L stands for the delay operator, θ stands for the model's moving average parameters, α stands for the autoregressive parameters, and ϵ stands for error terms. The degrees of the ARIMA model (p , q) are given by [38]:

$$x_t = c + \sum_{i=1}^p \theta_i x_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (5)$$

where $\theta_i \neq 0$, $\alpha_i \neq 0$, and $\sigma_\epsilon^2 > 0$. AR and MA models are named after the p and q parameters, respectively.

In the TBATS model, the letter T indicates trigonometric terms for seasonality, the letter B indicates Box-Cox transformations for heterogeneity, the letter A indicates ARMA errors for short-term dynamics, the letter T denotes damped trend, and the letter S indicates seasonality [35].

The lagged values of the time series can be used as inputs to a neural network, just as we use the lagged values in a linear autoregression model. This is also called autoregression. In this model, lagged inputs can be used to predict univariate time series. The same model can also work on seasonal time series. With this model, data conversion is performed using the Box-Cox method. The NNAR model is used in the R environment as NNAR (p , k) with the parameters k and p as a feed-forward and single hidden layer. Here, the p and k parameters represent the number of neurons in the delayed input and hidden layer, respectively. The neural network trained on radiological data (RD) can be expressed as

$$y_t = f(y_{t-1}) + \epsilon_t \quad (6)$$

where $y_{t-1} = (y_{t-1}, y_{t-2}, \dots)$ is a vector containing series lagged values.

Thetaf method, returns the prediction and the forecast intervals for the theta method forecast. The Theta-model takes a distinct method to decomposition, such as breaking down seasonally adjusted series into short and long term components. The model is built on the concept of altering the time series' local curvatures [1].

Holt’s approach was extended to capture seasonality by Holt (1957) and Winters (1960). The Holt-Winters model is used to calculate the average, trend, and seasonality of a time series. This model is an extension of the simple exponential smoothing model [18]. The equations of this model for average, trend and seasonality are given Equations 7, 8 and 9, respectively.

$$y_{t+h|t} = l_t + hb_t \tag{7}$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{8}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \tag{9}$$

Where l_t shows level estimation of the series at time t , and b_t shows the trend of the series at time t . α and β^* are located in the range of $0 \leq \alpha \leq 1$ and $0 \leq \beta^* \leq 1$, respectively, as smoothing parameters.

Multilayer perceptrons (MLPs) are fully connected feed forward networks that are commonly utilized in network architectures. Typically, training is accomplished using error post-propagation or a similar process. The MLP function in the R environment trains it by constructing a multi-layer perceptron.

The R environment’s extreme learning machines (ELM) model is comparable to the MLP model, however the number of nodes in the number of hidden layers is greater.

Long short-term memory (LSTM) is a form of RNN model that can learn long-term dependencies such as time series [8]. Adding LSTM to the network is like to adding a memory unit that can recall context from the start. As a result, LSTMs are meant to prevent the issue of long-term addiction and to make knowledge easy to recall over lengthy time stages.

In the LSTM model, information is held only by cells and memory manipulations are performed by gates [46]. An LSTM cell whose block diagram is given in Fig. 3 contains three gates: forget (f_t), input (i_t) and output (o_t). The forget gate is where information that is no longer needed in the cell is pulled. This gate is powered by two inputs $x(t)$ and $h(t-1)$.

When the f_t output, which represents the forget gate, sees a new data, it forgets the old data as in Eq. (10).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{10}$$

The process of saving a new information in the case of a cell consists of two parts. In the first step, the sigmoid layer named as the “input gate layer” decides which values it will update.

$$i_t = \sigma W_f \cdot h_{t-1, x_t} + b_i \tag{11}$$

In the second stage, the tanh layer creates the \tilde{C}_t vector for new candidate values to be added to the situation.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{12}$$

Equation 11 and 12 are combined to update the previous cell state, C_{t-1} , as a new C_t case.

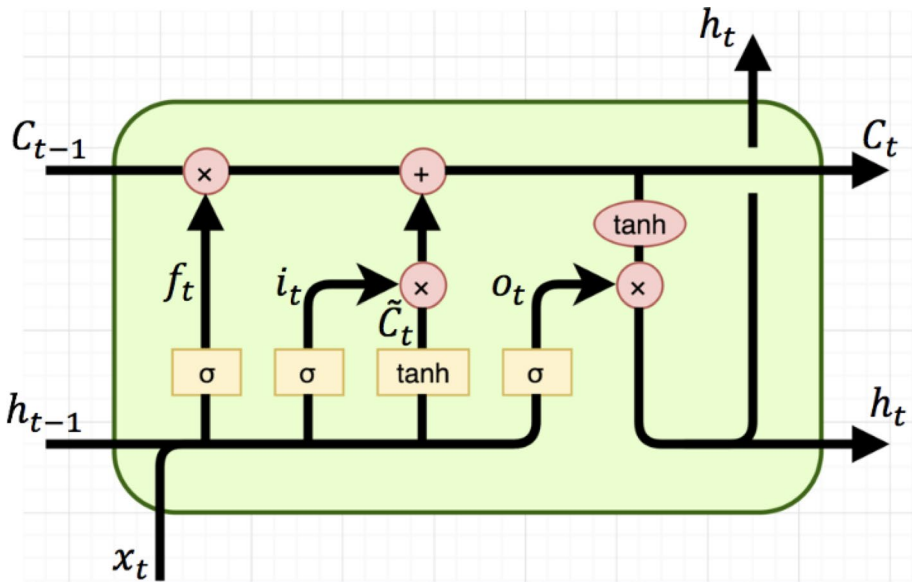


Fig. 3 LSTM architecture

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{13}$$

After this stage, it is decided what will be transmitted to the output depending on the cell status. For this process, first of all, a sigmoid layer is run and the output of the cell state is determined. Then, after the cell state is passed through the tanh function and multiplied by the output of the sigmoid gate, only the decided parts are subtracted.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{14}$$

$$h_t = o_t * \tanh(C_t) \tag{15}$$

3.3 Performance Evaluation

The accuracy of our calculations, missed values, corrupted data, and data entry errors will all be significantly impacted by how consistently our data is collected. Data accuracy is crucial in time series forecasting because the sample size is constrained.

In order to prevent overfitting and be able to assess the model’s predictive performance on fresh data, we first split the data into training and test periods for time series forecasting, just as we do for cross-sectional results. There is, however, one significant distinction between cross-sectional and time series data partitioning. The partitioning typically takes place randomly in cross-sectional data, with a random group of measurements designated as training data and the others as test data. A random partition, however, causes two problems in time series: first, it does not mimic the temporal ambiguity where we use the past and present to predict the future.

Furthermore, with “holes” it yields two time series, while many traditional forecasting strategies do not accommodate time series with missing values. Therefore, splitting the time series into training and test sets is handled differently: the series is split into two cycles of which the earlier period ($t=1, 2, \dots, n$) is designated as the training period and the latter period ($t=n+1, n+2, \dots$) is designated as the test period. Forecast accuracy measurement is important for a number of reasons, including investigating existing or potential problems in the supply chain and keeping the forecast system in check.

In the literature, many different metrics can be used to evaluate the performance of models [11], [11].

In this study, the symmetric mean absolute percentage error (sMAPE) [22] and mean absolute scaled error (MASE) [11] metrics were preferred. Symmetric MAPE provides an alternative way of overcoming the MAPE division by zero problem. For example, instead of using a base that is actual sales or estimated sales in the calculation of percentage error, it uses the average of actual sales and estimated sales.

MASE has recently been proposed as an alternative to MAPE variants. The scaled error is calculated as the ratio of the error from ‘a simple estimation technique’ to the MAE. The MASE is then computed as the mean of absolute scaled errors. The lower the MASE value, in other words, the lower the relative absolute prediction error, the better the method. In Equations 15 and 16 respectively, equations for sMAPE, and MASE metrics are given [22].

$$sMAPE = \frac{2}{h} \sum_{t=n+1}^{n+h} \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \times 100(\%) \tag{16}$$

$$MASE = \frac{1}{h} \frac{\sum_{t=n+1}^{n+h} |y_t - \hat{y}_t|}{\frac{1}{n-m} \sum_{t=m+1}^n |y_t - y_{t-m}|} \tag{17}$$

The y_t , is the value of the time series at point t , \hat{y}_t the estimated forecast, h the forecasting horizon, n the number data points available in-sample, and m the time interval. The forecast error for time period n , denoted e_n , is defined as the difference between the actual value, y_n and the forecast value, \hat{y}_n at time n .

$$e_n = y_n - \hat{y}_n \tag{18}$$

Here, e_n does not mean a misunderstanding, it means an unforeseeable aspect of the observation. It can be written down as

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h} | T \tag{19}$$

where the training data is given by $\{y_1, y_2, y_3, \dots, y_T\}$ and the test data is given by $\{y_{T+1}, y_{T+2}, y_{T+3}, \dots\}$. Prediction errors differ from residuals in two ways. First, the residues are measured on the training set, followed by the predicted errors on the test set. Second, residues are based on single-step predictions, whereas forecast errors will necessitate multi-stage predictions.

4 Results and Discussion

Time series analyzes in this study were carried out with two different applications. While python (v3.8.5) programming was used for prediction processes with the LSTM model, RStudio (v4.2.0 2022-04-22 ucrt) program was used for all the remaining deep learning and statistical-based models. On the other hand, the computer on which the analyzes were run had an Intel i5 PC 3.2 GHz CPU, 8 GB RAM, SATA disk and Windows 10 Pro operating system configuration.

Before the analyzes were made, the dataset consisting of 132 months for the training and testing processes was divided into three different ways as given in Table 3. While the last 12 months of the dataset were used for testing in the first analyzes for both the LSTM model and the prediction models used in the RStudio environment, the last 24 months of the dataset were reserved for testing in the second analysis.

The LSTM model used in this study consists of two LSTM layers and a dense layer. The first LSTM layer was created from 64 memory units and since it is connected to the next layer, the *return_sequence* parameter is set to *True*. The second LSTM layer consisted of 32 units and then a *dropout* layer was used again.

Since overfitting is a very common problem in dense networks like LSTM, a dropout layer was added immediately after the first layer to prevent the LSTM model from overfitting the data. What is meant by overfitting here is that good performance is achieved in the training data, but rather poor performance in the test data. The use of more data for deep learning models, in a sense, means better performance in dealing with overfitting. The fact that the data used in this study consists of only 132 months affects the model performance negatively. At this point, dropout appears as a solution [48].

When the LSTM model is trained, the *callback* method is defined to complete the training process before the epoch upper limit is reached and the desired accuracy threshold value is obtained. While compiling the LSTM model, *adam* was used as the optimizer and *mean_squared_error* was used as the loss. The results were obtained by running the model 100 epoch with the batch size set to 1. The effect of the callback method was investigated by making the epoch number 200 for the same LSTM model. As soon

Table 3 sMAPE and MASE metric performance values of models for 12 month and 24 month forecasts under two different approaches

Model name	Prediction period 12 months: 91% for training, 9% for testing		Prediction period 24 months: 82% for training, 18% for testing		Prediction period 30 months: 77% for training, 23% for testing	
	sMAPE test	MASE-test	sMAPE test	MASE test	sMAPE test	MASE test
LSTM	0.196	0.90	0.134	0.89	0.144	1.08
NNETAR	0.244	1.755	0.138	0.996	0.129	0.941
MLP	0.323	2.543	0.310	2.389	0.332	3.267
ELM	0.395	3.351	0.254	2.182	0.170	1.280
AUTO.ARIMA	0.318	2.497	0.145	1.067	0.179	1.321
SES	0.322	2.541	0.192	1.524	0.188	1.381
HOLT-WINTER	0.368	3.042	0.214	1.756	0.157	1.192
THETAF	0.393	3.338	0.257	2.175	0.155	1.149
TBATS	0.404	3.449	0.363	3.512	0.659	9.194

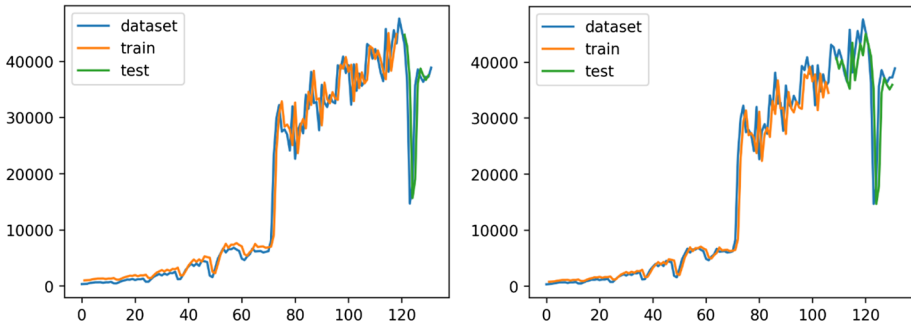


Fig. 4 Train/test graphs of the LSTM model when the dataset is reserved for 91% training and 9% testing in the first analysis (left), and 82% for training and 18% for testing in the second analysis (right)

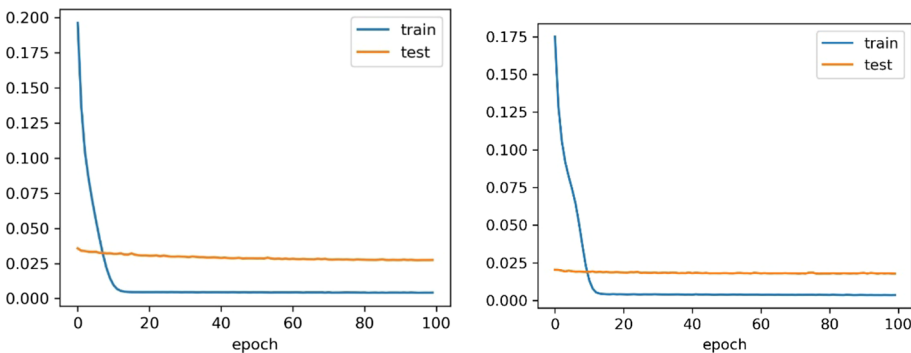


Fig. 5 Loss graphs for analyzes where the dataset is divided for 91% training, 9% testing (left), and 82% training, 18% testing (right)

as the training process reaches the 163rd epoch, the training process is stopped because the threshold value defined in the callback method has been reached.

Among the analysis results of the model, 120/12 and 108/24 month training/test graphics depending on two different approaches are given in Fig. 4., while the “mean_squared_error” metric graphics set as the loss parameter of the model are given in Fig. 5. The values of the sMAPE and MASE metrics, which express the actual performance of the model, are given in Table 3.

As seen in Table 3, against the 12 month and 24 month test runs of the LSTM model, the sMAPE metrics were obtained as 19.6% and 13.4%, respectively, while the MASE values were obtained as 0.90 and 0.89, respectively. In this case, the model accuracy rate for the first approach is 80.4%, while the accuracy rate for the second approach is 86.6%. Increasing the data allocated for the test from 12 to 24 months reduced the error of the LSTM model by 6.2%. The smaller MASE values given in Table 3 mean that the model makes better predictions. However, when the test data is increased to 23%, it is seen that the data used for model training is insufficient and many models are negatively affected.

As can be seen in Fig. 4, the training/test losses of the model decreased gradually over 100 epochs. While the “mean_squared_error” error values obtained on the training data are sharper, the “mean_squared_error” error values obtained on the test data are softer.

12 month and 24 month prediction graphs of deep learning models used in RStudio environment such as NNETAR, ELM and MLP along with statistical based models such as AUTO.ARIMA, TBATS, THETAF, HOLT and SES are given in Figs. 6 and 7, respectively. The estimation performance values of the models are given in Table 3 for three different approaches.

As seen in Figs. 8 and 9, the dark shaded area contains 80% of the expected intervals. In other words, each probable value is 80% likely to be in the dark shaded range. The 95% prediction intervals are displayed in the light shaded area. These forecasting intervals are a good approach to show variability in the prediction. In this situation, the forecast is presumed to be reliable, hence the forecast intervals are relatively narrow. The blue line represents the average of anticipated future values, which we refer to as point forecasts. The forecast interval is often stated as

$$\hat{y}_{T+h|T} \pm c\hat{\sigma}_h \tag{20}$$

where $\hat{\sigma}_h$ indicates an approximation of the standard deviation of the h-step forecast distribution and multiplier c is dependent on the likelihood of coverage. The standard deviation is expressed as

$$\hat{\sigma}_h = \sqrt{\frac{1}{T-k-1} \sum_{t=1}^T e_t^2} \tag{21}$$

where k is the model’s number of predictors, e_t represents the difference between the actual value and the predicted value. The value of the constant c is 1.28 for the 80% prediction interval and 1.96 for the 95% prediction interval.

For a better understanding of this issue, the 12-month forecast values of the AUTO. ARIMA model are given in Table 4 at 80% and 95% intervals. According to Eq. (21),

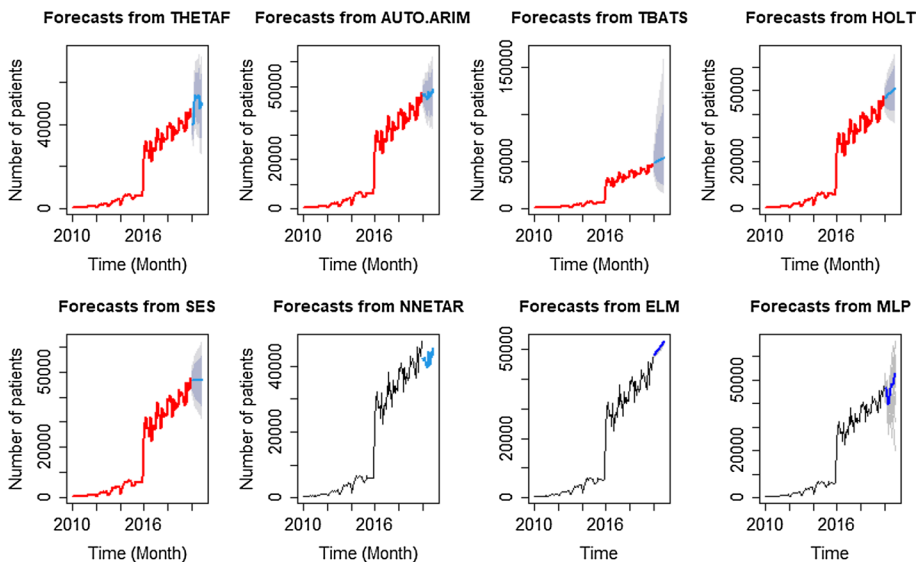


Fig. 6 Combining 12 month estimates with 80% and 95% prediction intervals for radiological cases, three of which are deep learning and five of which are statistical models

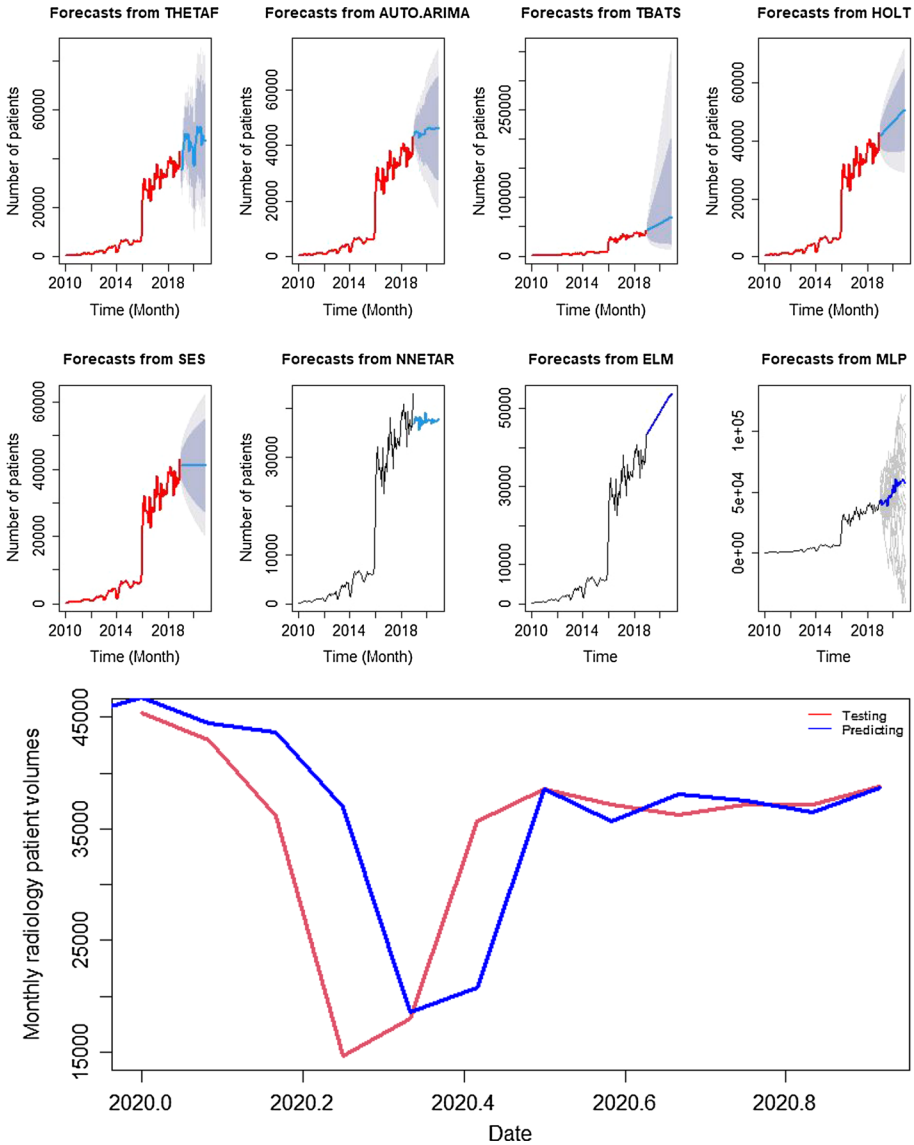
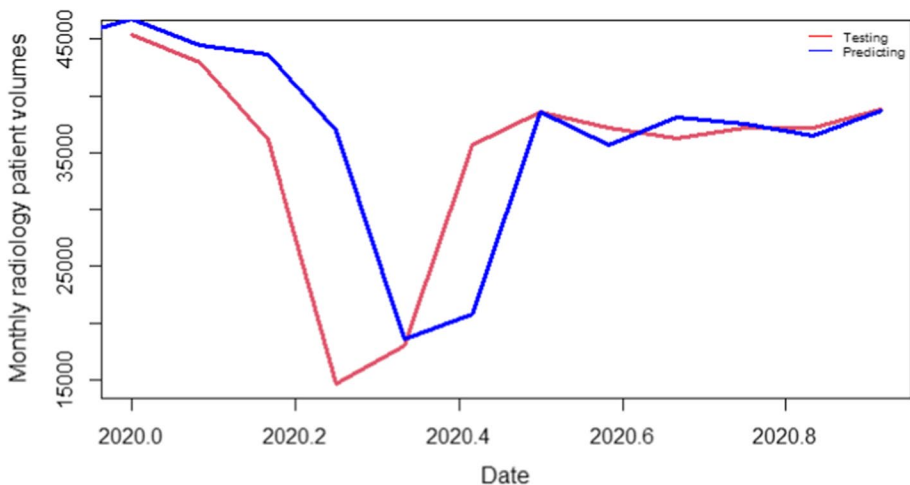


Fig. 7 Combining 24 month estimates with 80% and 95% prediction intervals for radiological cases, three of which are deep learning and five of which are statistical models

the standard deviation value used in calculating the intervals in Table 4 was previously calculated as 2856.008. In this case, for example, the monthly forecast range of 80% for January 2020 is $[46815.52 \mp 2856.008 * 1.28] = [43159.83, 50471.22]$. Similarly, the monthly forecast interval of 95% for January 2020 is $[46815.52 \mp 2856.008 * 1.96] = [41224.62, 52406.43]$. The remaining range values were calculated accordingly.

Table 4 The representation of the predictive values of the AUTO.ARIMA model for 12-month estimates between 80 and 95% intervals

Months	Years	Point forecast	Low 80%	High 80%	Low 95%	High 95%
Jan	2020	46815.52	43159.83	50471.22	41224.62	52406.43
Feb	2020	46270.08	41791.83	50748.32	39421.19	53118.97
Mar	2020	46797.96	41463.13	52132.79	38639.04	54956.88
Apr	2020	46270.87	40242.25	52299.50	37050.89	55490.86
May	2020	45426.58	38764.81	52088.34	35238.29	55614.87
Jun	2020	44875.06	37638.38	52111.75	33807.51	55942.62
Jul	2020	48021.38	40251.37	55791.39	36138.17	59904.59
Aug	2020	45441.85	37173.07	53710.63	32795.84	58087.85
Sep	2020	46731.64	37992.44	55470.83	33366.19	60097.08
Oct	2020	47948.18	38762.66	57133.71	33900.13	61996.24
Nov	2020	47146.39	37535.22	56757.55	32447.38	61845.40
Dec	2020	48648.65	38629.92	58667.39	33326.32	63970.98

**Fig. 8** Forecast plot of the AUTO.ARIMA model, where data is allocated as 12 months for testing in the first approach of the analysis

Since numerical estimation data in the range of 80% and 95% of the AUTO.ARIMA model are given in Table 4, the graphs of the 12-month and 24-month test period, which visually show these estimations more clearly and enable the process to be understood faster, are given in Figs. 8 and 9, respectively.

When the model performance metric values given in Table 3 are examined, it is seen that the results obtained as a result of dividing the dataset applied in the analyzes as training and testing, that is, using it as 120-months/12-months and 108-months/24-months, are also affected.

Considering the split ratio of the dataset, it is clearly seen that the LSTM model makes better predictions in the analyzes made using Python. In the analyzes made in the

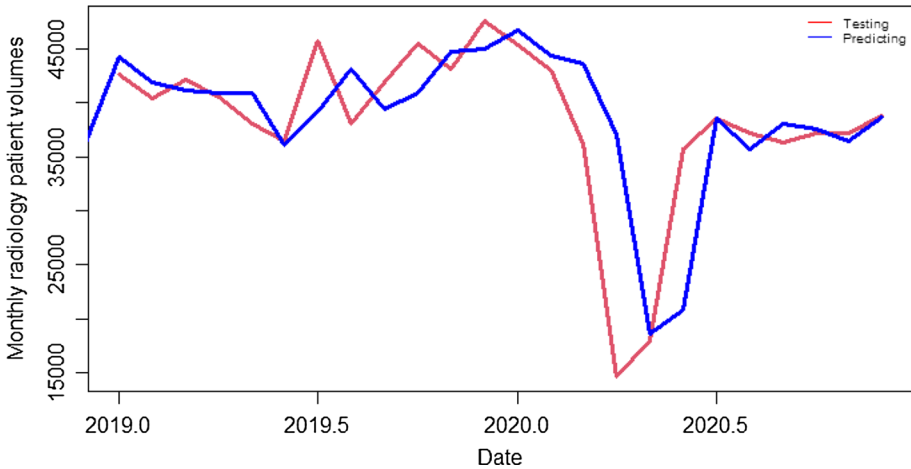


Fig. 9 Forecast graph of the AUTO.ARIMA model, where in the second approach of the analysis the data is reserved for 24 month testing

RStudio environment, it is seen that the NNETAR model performs better than both statistics-based and other deep learning models such as MLP and ELM. In statistics-based models, the AUTO.ARIMA model has come to the fore in both approaches.

In all models used in the analysis process, doubling the portion of the dataset used for testing and separating it into 24 months further increased the accuracy of the estimation. Among the main reasons for this, the curfew imposed due to COVID-19 since March 2020 can be shown as a reason for the decrease in radiological cases.

It is believed that the data, which continues with the upward trend, tends to decrease in the relevant months and the last months are left as a test process, which has a negative effect on the performances of the models. In other words, the use of limited data for testing does not give the desired performance.

The performances of the prediction models used in the R environment varied according to the size of the data allocated for the test process.

For example, in the case of 12-month test data, the sMAPE values of deep learning models such as NNETAR, MLP, and ELM were 24.4%, 32.3%, and 39.5%, respectively. In the same case, the sMAPE prediction values for the statistical-based AUTO.ARIMA, SES, HOLT-WINTER, THETAF and TBATS models were obtained as 31.8%, 32.2%, 36.8%, 39.3%, and 40.4%, respectively. On the other hand, in the case where 24-month data is allocated for the test process, the sMAPE values of the NNETAR, MLP and ELM models were 13.8%, 31.0% and 25.4%, respectively.

Similarly, the sMAPE prediction values for AUTO.ARIMA, SES, HOLT-WINTER, THETAF and TBATS were 14.5%, 19.2%, 21.4%, 25.7%, and 36.3%, respectively. The values given here are the percentages of the values in the table, in other words, the errors made by the models.

Some conclusions drawn from this study are listed below:

1. In the analyzes made here, analyzes with statistical-based models were completed in a short time, while more time was spent for deep learning-based models such as MLP,

- ELM and NNETAR. Among the models, the most computation time was used by the MLP model.
2. The 132-month dataset, which was obtained from the radiology unit and used as a time series, was a disadvantage for prediction models due to the limited number of data.
 3. When the results obtained from the analyzes made by dividing the dataset in three different ways are evaluated, the LSTM model performed better.
 4. As given in Table 3, when the dataset test rate was increased from 9 to 18%, the prediction performances of all models used were positively affected. However, when this rate was increased to 23%, 5 of the 9 models used in the estimation were positively affected, while 4 were negatively affected. This shows that the data allocated for training is insufficient for models that need more data such as LSTM, MLP, TBATS and AUTO.ARIMA.
 5. Here, the sMAPE values for the estimations made with limited data ranged from 19.6%-13.4% for the LSTM model to 31.8%-14.5% for the AUTO.ARIMA model.
 6. Among the statistical models, the AUTO.ARIMA model showed the best performance here as well as in [19, 20, 27] studies.
 7. Here the performance of the LSTM model was better than the AUTO.ARIMA model. This result was also supported by study [41] comparing the LSTM and ARIMA model.

5 Conclusion

In this article, future predictions were made on 132 months radiological data obtained from Muş State Hospital and used for this study. While deep learning models such as LSTM, MLP, NNAR and ELM were used in the 12 month, 24 month, and 30 month forecasting processes performed here, statistical-based models of the R development environment such as AUTO.ARIMA, SES, TBATS, HOLT and THETAF were also used.

In addition, recently preferred metrics by researchers such as MASE and sMAPE were used to evaluate model performances. In the radiological patient volume 12 month future predictions, the LSTM and AUTO.ARIMA models performed with 80.4% and 68.2% accuracy, respectively, while the model performances at the 24 month predictions were 86.6% and 85.5%, respectively. In all the predictions made, it is seen that the LSTM model outperforms the models used in the study.

Since some of the radiological data coincided with the COVID-19 pandemic, which was effective in Turkey as well as all over the world, the number of images in the radiology unit of the Muş State Hospital also decreased and was negatively affected by the curfew at that time.

A limited number of publications were found in the literature review related to the subject studied, and it was seen that a limited number of models were used in these studies. This increases the chance of success of this study, which uses both deep learning and statistical-based learning models.

It is thought that the estimation results obtained in this study will be useful for hospital management to predict future radiological case patient volume, to estimate the need for radiology unit personnel, to plan radiology devices according to potential needs, and for the well-being of the health care system.

Funding No funding was required for this study.

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Ethical Approval Ethical approval is not required for this study. Since the analyzes performed are only related to the number of radiological images, there was no direct interaction with any patient, and the personal identity information of the patients was not used.

References

1. Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16(4), 521–530.
2. Atici, R., & Pala, Z. (2021). "Prediction of the ionospheric foF2 parameter using R language forecast-hybrid model library convenient time. *Wireless Personal Communications*, 122(4), 1–20. <https://doi.org/10.1007/s11277-021-09050-6>
3. J. A. Doornik., J. L. Castle., & Hendry D. F. (2020). Short-term forecasting of the coronavirus pandemic. *International Journal of Forecasting*.
4. Gode, A. P., Tiwaskar, S., Lakhar, B. N., & Dhande, R. (2022). Artificial intelligence in the field of radiology. *A Review Article*, 13(8), 97–104.
5. Golmohammadi, D. (2016). Predicting hospital admissions to reduce emergency department boarding. *International Journal of Production Economics*, 182, 535–544.
6. Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2020). Forecasting emergency department overcrowding: A deep learning framework. *Chaos, Solitons and Fractals*, 139, 110247.
7. Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2022). Effective forecasting of key features in hospital emergency department: Hybrid deep learning-driven methods". *Machine Learning with Applications*, 7, 100200.
8. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
9. Holzinger, A. (2016). *Machine learning for health informatics state-of-the-art and future challenges*. Springer International Publishing.
10. Hyndman, R. J., Athanasopoulos G. (2018). *Forecasting : Principles and Practice*. In (2nd ed.). Monash University
11. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
12. Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.
13. Jilani, T., Housley, G., Figueredo, G., Tang, P. S., Hatton, J., & Shaw, D. (2019). Short and Long term predictions of Hospital emergency department attendances. *International Journal of Medical Informatics*, 129, 167–174.
14. Jones, S. S., et al. (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics*, 42(1), 123–139.
15. Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2019). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159–170.
16. Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159–170.
17. Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679.
18. Koehler, A. B., Snyder, R. D., & Ord, J. K. (2001). Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *International Journal of Forecasting*, 17(2), 269–286.
19. Kulkarni, S., Seneviratne, N., Baig, M. S., & Khan, A. H. A. (2020). Artificial intelligence in medicine: Where are we now? *Academic Radiology*, 27(1), 62–70.

20. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
21. Luo, L., Luo, L., Zhang, X., & He, X. (2017). Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research*, 17(1), 1–13.
22. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.
23. Maleki, M., Mahmoudi, M. R., Heydari, M. H., & Pho, K. H. (2020). Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models. *Chaos, Solitons and Fractals*, 140, 110151.
24. Moghadas, S. M., Shoukat, A. S., Fitzpatrick, M. C., Wells, C. R., Sah, P. S., Pandey, A., Sanchs, J. D., Galvani, A. P., & Zheng Wang, L. A. (2020). Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences*, 117(16), 9122–9126.
25. Nuti, S., & Vainieri, M. (2012). Managing waiting times in diagnostic medical imaging. *British Medical Journal Open*, 2, 1255.
26. Pala, Z. (2021). Examining EMF time series using prediction algorithms with R. *IEEE Canadian Journal of Electrical and Computer Engineering*, 44(2), 223–227.
27. Pala, Z. (2023). “Comparative study on monthly natural gas vehicle fuel consumption and industrial consumption using multi-hybrid forecast models. *Energy*, 263, 1–21.
28. Pala, Z., & Atici, R. (2019). Forecasting sunspot time series using deep learning methods. *Solar Physics*, 294(5), 50.
29. Pala, Z., & Özkan, O. (2020). Artificial Intelligence Helps Protect Smart Homes against Thieves. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(3), 945–952.
30. Pala, Z., & Pala, A. F. (2020). Perform time-series predictions in the r development environment by combining statistical-based models with a decomposition-based approach. *J. Muş Alparslan University Fac. Eng. Archit.*, 1(1), 1–13.
31. Pala, Z., & Pala, A. F. (2021). Comparison of ongoing COVID-19 pandemic confirmed cases/deaths weekly forecasts on continental basis using R statistical models. *Dicle University Journal Engineering*, 4, 635–644.
32. Pala, Z., & Şana, M. (2020). Attackdet: Combining web data parsing and real-time analysis with machine learning. *Journal of Advances in Technology and Engineering*, 6(1), 37–45.
33. Pala, Z., Ünlük, İH., & Yaldız, E. (2019). Forecasting of electromagnetic radiation time series: An empirical comparative approach. *Applied Computational Electromagnetics Society Journal*, 34(8), 1238–1241.
34. Poyiadji, N., Klochko, C., LaForce, J., Brown, M. L., & Griffith, B. (2021). COVID-19 and radiology resident imaging volumes-differential impact by resident training year and imaging modality. *Academic Radiology*, 28(1), 106–111.
35. Rob, J. (2008). Hyndman and yeasmin khandakar, “automatic time series forecasting: The forecast package for R.” *Journal of Statistical Software*, 27(3), 22.
36. Savin, L. V., & Wang, S. (2006). Managing Patient Service in a Diagnostic Medical Facility. *Diagnostic Medical Facility Operations Research*, 54(1), 11–25.
37. Shahid, K., Manzoor, T., Ibrahim, M., Ahmed, T., & Fiaz, M. (2016). Forecasting of monthly patient volume at radiology department coming for ultrasound: A time series approach. *Journal of University Medical Dental College JUMDC*, 7(3), 22–27.
38. S. Siami-Namini, N. Tavakoli, & A. Siami Namin (2019). A Comparison of ARIMA and LSTM in Forecasting Time Series. In: *Proceedings-17th IEEE International Conference Machine Learning Applications. ICMLA 2018* (pp. 1394–1401).
39. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
40. Sun, Y., Heng, B. H., Seow, Y. T., & Seow, E. (2009). “Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine*, 9, 1–9.
41. Tavares Thomé, A. M., Cyrino Oliveira, F. L., & Carvalho Ferrer, A. L. (2018). Time series analysis with explanatory variables: A systematic literature review”. *Environmental Modelling and Software*, 107, 199–209.
42. van Assen, M., Muscogiuri, G., Caruso, D., Lee, S. J., Laghi, A., & De Cecco, C. N. (2020). Artificial intelligence in cardiac radiology. *La Radiologia Medica*, 1, 3.

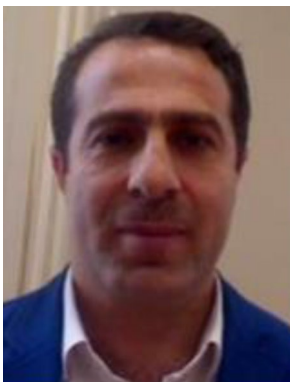
43. Van Lent, W. A. M., Deetman, J. W., Teertstra, H. J., Muller, S. H., Hans, E. W., & Van Harten, W. H. (2012). Reducing the throughput time of the diagnostic track involving CT scanning with computer simulation. *European Journal of Radiology*, *81*(11), 3131–3140.
44. Wanluk, N., Pintavirooj, C., & Treebupachatsakul, T. (2019). Image Processing for X-ray calibration phantom. *BMEiCON 2018–11th Biomedical Engineering International Conference*, *2*, 14–17.
45. Xu, Q., Tsui, K. L., Jiang, W., & Guo, H. (2016). A Hybrid Approach for Forecasting Patient Visits in Emergency Department. *Quality and Reliability Engineering International*, *32*(8), 2751–2759.
46. Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., & Wang, X. (2018). A CFCC-LSTM Model for Sea Surface Temperature Prediction. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 207–211.
47. Zhang, Y., et al. (2020). Emergency patient flow forecasting in the radiology department. *Health Informatics Journal*, *26*(4), 2362–2374.
48. Zhu, T., Luo, L., Zhang, X., Shi, Y., & Shen, W. (2017). Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients. *IEEE Journal Biomedical Health Informatics*, *21*(2), 515–526.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Zeydin Pala Zeydin Pala Was born in Turkey's van Province. He completed his undergraduate education in 1993, his master's degree's in 2007 and his doctorate education in 2012. The author, who has published many books and articles related to his field, is still working as an associate professor at Muş Alparslan University.



Ramazan Atıcı was born in the Province of Erzincan in eastern Turkey in 1980. He completed his undergraduate degree in 2004, master's degree in 2007 and doctorate in 2013. He investigated the changes in ionospheric plasma in his studies. He is currently working as an professor at Muş Alparslan University.



Erkan Yıldız was born in Türkiye's Muş province. He completed his undergraduate education in 2012, his master degree in 2020. Erkan Yıldız worked in Muş Municipality and Muş State Hospital, respectively. He now works as a data analysis specialist at Halkbank's head office in Istanbul.