

FORECASTING MODELS FOR COVID-19 CASES OF TURKEY USING ARTIFICIAL NEURAL NETWORKS AND DEEP LEARNING

Yunus EROĞLU^{1*}

¹İskenderun Technical University, Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Hatay, Turkey

ORCID No : <https://orcid.org/0000-0002-8354-6783>

Keywords	Abstract
<p>Covid-19, Pandemics, Forecasting, Artificial Neural Networks, Deep Learning, LSTM</p>	<p>Governments face a dilemma between public health and the economy while making strategic decisions on health during a pandemic outbreak. It is of great importance to forecast the number of cases in terms of strategic decisions to be taken by governments especially in outbreak periods and to manage the dilemma mentioned. One of the important issues today is the Covid-19 outbreak for almost all countries. Unfortunately, no effective vaccine or treatment has been found for Covid-19 yet. At the time of this study, however, it was reported that the total number of reported cases by the World Health Organization worldwide was more than thirteen million. Various quarantine measures have been necessary to deal with such a large epidemic. Quarantine measures taken by governments bring countries to face to face with the economic crisis. This creates economic uncertainties and puts governments under tremendous pressure to make accurate and least harmful strategic decisions. For these reasons, governments prefer to make strategic decisions for Covid-19 step by step observing the situation rather than making a sudden decision. If the number of pandemic cases could be predicted before a predetermined time, it would be used as an important guide for governments to manage public health and economic dilemma more accurately. Therefore, this study provides artificial neural network (ANN) and deep learning models (long-short term memory, LSTM networks) to forecast Covid-19 cases before 7-day. The proposed models were tested on real data for Turkey. The results showed that LSTM models performed better than ANN models in both cumulative cases and new cases on the training data set. Comparing the performance of the proposed models over the whole data set, it was observed that the ANN and LSTM algorithms gave competitive results. In addition, the cumulative case forecast performances of both ANN and LSTM models were observed to be better than the new case forecast.</p>

YAPAY SİNİR AĞLARI VE DERİN ÖĞRENME KULLANARAK TÜRKİYE'NİN COVID-19 VAKALARI İÇİN TAHMİN MODELLERİ

Anahtar Kelimeler	Öz
<p>Covid-19, Pandemiler, Tahminleme, Yapay Sinir Ağları, Derin Öğrenme, LSTM</p>	<p>Hükümetler, bir pandemi salgını sırasında stratejik kararlar alırken, halk sağlığı ve ekonomi arasında bir ikileme karşı karşıyadır. Özellikle salgın dönemlerinde hükümetler tarafından alınacak stratejik kararlar açısından vaka sayısını tahmin etmek ve belirtilen ikilemi yönetmek büyük önem taşımaktadır. Bugün neredeyse tüm ülkeler için önemli konulardan birisi de Covid-19 salgınıdır. Ne yazık ki, henüz Covid-19 için etkili bir aşı veya tedavi bulunamamıştır. Ayrıca, bu çalışmanın hazırlığı sırasında, Dünya Sağlık Örgütü tarafından dünya çapında toplam vaka sayısının on üç milyondan fazla olduğu bildirilmiştir. Böyle büyük bir salgınla başa çıkmak için çeşitli karantina önlemlerinin alınması gerekli olmuştur. Hükümetler tarafından alınan karantina önlemleri, ülkeleri ekonomik krizle karşı karşıya getirmiştir. Bu durum ekonomik belirsizlikler yaratmaktadır ve hükümetleri doğru ve en az zararlı stratejik kararlar almak için muazzam bir baskı altına sokmaktadır. Bu nedenlerle hükümetler, ani bir karar vermek yerine durumu adım adım gözlemleyerek Covid-19 için stratejik kararlar almayı tercih etmektedirler. Eğer pandemi vakalarının sayısı belirlenmiş bir zamandan önce tahmin edilebilirse, hükümetlerin halk sağlığı ve ekonomi ikilemini daha doğru bir şekilde yönetmeleri için önemli bir rehber olarak kullanılabilir. Bu nedenle, bu çalışmada 7 gün önceden Covid-19 vakalarını tahmin etmek için yapay sinir ağı (YSA) ve derin öğrenme (uzun-kısa süreli bellek, LSTM ağları) modelleri sunulmuştur. Önerilen modeller Türkiye'nin gerçek verileri üzerinde test edilmiştir. Sonuçlar LSTM modellerinin eğitim seti için hem kümülatif hem de yeni vaka tahminlerinde YSA modellerinden daha iyi performans gösterdiğini göstermiştir. Önerilen modellerin tüm veri seti üzerindeki performansları kıyaslandığında YSA ve LSTM algoritmalarının birbirleri ile rekabet edebilir sonuçlar verdiği gözlemlenmiştir. Ayrıca hem YSA hem de LSTM modellerinin kümülatif vaka tahmini performanslarının yeni vaka tahminlerinden daha iyi olduğu gözlemlenmiştir.</p>
Araştırma Makalesi	Research Article
Başvuru Tarihi : 19.07.2020	Submission Date : 19.07.2020
Kabul Tarihi : 24.10.2020	Accepted Date : 24.10.2020

*Sorumlu yazar; e-posta : yunus.eroglu@iste.edu.tr

1. Introduction

The World Health Organization (WHO) received notification from China for many cases of respiratory disease associated with some people visiting a seafood market in Wuhan in December 2019. It was a new type of Coronavirus which is a kind of large family of viruses appearing among humans, birds, livestock, mice, bats, and other wild animals (Wang, L. F., Shi, Z., Zhang, S., Field, H., Daszak, P., & Eaton, B. T., 2006). They are serious pathogens that infect respiratory, hepatic, gastrointestinal and neurological diseases for humans (Chen, Liu, and Guo, 2020). Unfortunately, this uncontrollable new type of coronavirus (after named as Covid-19) has begun to spread worldwide as an epidemic and declared a pandemic by the World Health Organization on March 11, 2020 (WHO, 2020). Ruan (2020) reports that the Covid-19 has not mortality rates as high as other types of coronaviruses such as severe acute respiratory syndrome (*SARS-Cov*) or Middle East respiratory syndrome *Mers-Cov* where the approximately mortality rates for 14–15% for *SARS-Cov* and 35% for *MERS-Cov*, on the other hand, 3% for *Covid-19*. Moreover, Covid-19 is a virus that can spread very quickly and can be transmitted easily. This situation causes a sudden increase in health-care system demand. Since the infection rate is quite high, it causes disruptions in the treatment process and consequently increases deaths. In fact, the best measure that can be taken for such epidemic diseases without effective treatment and vaccine is tight quarantine practices. But not knowing how long this tight quarantine process can last raises concerns about how big the economic crisis may be. For this reason, it is difficult for governments to make strict strategic decisions in terms of Covid-19 measures, as they may threaten their economies. By the way, the spread of coronavirus is very dangerous and requires strict policies and plans that have been implemented in many parts of the world by considering local economy. Therefore, it is essential to forecast the daily confirmed cases in the coming days to implement the necessary protection plans step by step.

In this study, the estimation of daily Covid-19 cases one week in advance is modeled and it is aimed to help the authorities to make more accurate and safe decisions about the measures to be taken, the measures to be applied, or on easing the restrictions applied. In the next section, current publications on

disease forecast and especially Covid-19 case forecast are summarized. The ANN and LSTM models proposed in this study differ from the current literature due to the following aspects, by keeping the literature research of the authors limited;

- There is a gap on modeling of daily Covid-19 case forecast for Turkey,
- Current Covid-19 case forecast studies are modeled on the forecast of cumulative cases, differently, this study presents daily case forecast approach,
- ANN and LSTM models have not previously been used to predict Covid-19 cases,
- The most suitable one was chosen by comparing the performance of ANN training algorithms,
- The most suitable optimizer and hidden layer number were determined with various models designed for LSTM.

2. Literature Review

Disease forecasting is essential for economic reasons of countries, public safety, and justified use of medicines. Forecasting is also pivotal to disease control. The forecasting fits the skills required by a pathologist, from the identification of the problem through to advising on control measures. Also, it requires an understanding of all of them for forecasting systems to work at all (Hardwick, 2006). There are many forecasting models for epidemic diseases in the current literature. Table 1 presents the summary of forecast studies on epidemics and applied methodology.

Table 1
Summary of current literature related with disease forecast

Research Paper	Disease	Method(s)
Hufnagel et al. (2004)	SARS	Stochastic Simulation
Ture and Kurt (2006)	Hepatitis A	Time Series, ANN
Kaundal et al. (2007)	Rice Blast	SVM
McDonald and Boland (2007)	Sclerotinia spp.	Regression - Correlation
Shaman et al. (2013)	Influenza	Data assimilation
Nsoesie et al. (2013)	Influenza	Simulation Optimization
Sannakki et al. (2013)	Grape Diseases	k - NN* and FFNN*
Shaman et al. (2014)	West African Ebola	Dynamic modeling, Bayesian approach
DeFelice et al. (2017)	West Nile Virus	Epidemiological model
Ceylan, Z (2020)	Covid-19	Time Series
Elmousalami and Hassanien (2020)	Covid-19	Time Series
Petropoulos and Makridakis (2020).	Covid-19	Time Series
Fong et al. (2020)	Covid-19	Data mining, PNN + cf*
Gamero et al. (2020)	Covid-19	Logarithmic polynomial approach
Wu et al. (2020)	Covid-19	SEIR*
Roosa et al. (2020)	Covid-19	Logistic growth model
Santosh, K. C. (2020)	Covid-19	AI*
Al-qaness et al. (2020)	Covid-19	ANFIS*, FPA*, SSA*
Ghazaly et al. (2020)	Covid-19	NAR*
Fanelli and Piazza (2020)	Covid-19	SIR*
Perc et al. (2020)	Covid-19	Exponential model
Stübinger and Schneider (2020)	Covid-19	Dynamic Time Warping
Proposed approach	Covid-19	ANN*, LSTM*

* ANN, Artificial Neural Network; k-NN, k means Neural Network; SVM, Support Vector Machines; FFNN, Feed Forward Neural Network; PNN + cf, Polynomial Neural Network with corrective feedback; SEIR, Susceptible-Exposed-Infectious-Recovered; AI, Artificial Intelligence; ANFIS, Adaptive Neuro-Fuzzy Inference System; FPA, Flower Pollination Algorithm; SSA, Salp Swarm Algorithm; NAR, Non-Linear Regressive Network ; SIR, Susceptible, Infected and Recovered; LSTM, Long-Short Term Memory

A support vector machines for developing weather-based prediction models of plant diseases was contributed by Kaundal, Kapoor, and Raghava (2007). Forecasting diseases caused by Sclerotinia spp. in eastern Canada using air temperature, rate of cooling, surface wetness, and preexisting infection data was contributed in another study (McDonald and Boland, 2004). Hufnagel, Brockmann, and Geisel (2004) provided a probabilistic model that describes the worldwide spread of infectious diseases and demonstrated that a forecast of the geographical spread of epidemics

was indeed possible. They used a combination of stochastic local infection dynamics among individuals and stochastic transport in a worldwide network, taking into account national and international civil aviation traffic. In another study, a framework was contributed for near real-time forecast of influenza epidemics using a simulation optimization approach which combines an individual-based model and a simple root finding optimization method for parameter estimation and forecasting by Nsoesie, Mararthe, and Brownstein (2013). Moran, Fairchild, Generous, Hickmann,

Osthus, Priedhorsky, and Del Valle (2016) searched some of similarities and differences between these weather forecasting and disease forecasting fields and how the epidemic modeling community was rising to the challenges posed by forecasting to help anticipate and guide the mitigation of epidemics. They concluded that human behavior makes disease forecasting more challenging than weather forecasting. Sannakki, Rajpurohit, Sumira, and Venkatesh (2013) contributed a neural network approach using temperature, rainfall, and humidity data to forecast diseases in grapes. They believed that weather forecast could provide valuable and timely information for evaluation of various crop management techniques, to avoid potential losses. Moreover, their proposed approach intended to predict the weather using a modified k-Nearest Neighbor (NN) approach, and ANN utilizing parameters such as humidity and temperature to predict the disease outbreaks in grapes. A model was developed depicting West Nile Virus (WNV) transmission dynamics and optimized using a data assimilation method by DeFelice, Little, Campbell, and Shaman (2017). In that study, mosquito infection rates and reported human WNV cases were used. Also, the coupled model-inference framework was then used to generate retrospective ensemble forecasts of historical WNV outbreaks in Long Island, New York for 2001–2014. The contributed method provided the forecasts of mosquito infection rates generated before peak infection, and more than 65% of forecasts accurately predicted seasonal total human WNV cases up to 9 weeks before the past reported case. Ture and Kurt (2006) compared time series prediction capabilities of three ANN algorithms (multi-layer perceptron, radial basis function, and time delay neural networks), and auto-regressive integrated moving average model to Hepatitis A Virus (HAV) forecasting. They concluded that multi-layer perceptron ANN model was more accurate and performed better than other contributed models. A seasonal influenza prediction system using an advanced data assimilation technique and real-time estimates of influenza incidence was developed to optimize and initialize a population-based mathematical model of influenza transmission dynamics by Shaman, Karspeck, Yang, Tamerius, and Lipsitch (2013). Their proposed system was used to generate and evaluate retrospective forecasts of influenza peak timing in New York City and they presented weekly forecasts of seasonal influenza developed and run in real time for 108 cities in the USA during the recent 2012–

2013 season. Reliable ensemble forecasts of influenza outbreak peak timing with leads of up to 9 weeks were produced with 63% accuracy. The weekly forecast of West African Ebola (WAE) outbreak was studied with using case observations, dynamic modeling and Bayesian inference in Guinea, Liberia and Sierra Leone (Shaman, Yang, and Kandula, 2014). They have reported that the forecast projecting no future change in intervention efficacy has been more accurate for Guinea and Sierra Leone, but have overestimated incidence and mortality for Liberia.

Besides the all current studies related with disease forecasting mentioned above, there are many different methodologic studies focusing on forecasting for new coronavirus pandemics. A simple and an objective method was introduced to predicting the continuation of the Covid-19 using a simple time series forecasting approaches by Petropoulos and Makridakis (2020). Fong, Li, Dey, Crespo, and Herrera-Viedma (2020) proposed a data mining methodology that embraces augmenting the existing little data, using a panel selection to pick the best forecasting model from several models, and fine-tuning the parameters of an individual forecasting model for the highest possible accuracy techniques from a small dataset. They used the cases of Wuhan by applying their proposed methodology. Their results showed that an optimized forecasting model namely polynomial neural network with corrective feedback (PNN+cf) was able to make a forecast that had relatively the lowest prediction error. A comparison of day level forecasting models on Covid-19 affected cases using time series models and mathematical formulation was presented by Elmousalimi and Hassanien (2020). Based on their study, they recommended that all world countries must mandate substantially more invasive quarantines, restrictions on travel and public gatherings, and closing of schools, universities, and workplaces ("Social Distancing") in the near term when Intensive Care Unit (ICU) beds are unavailable and patient deaths begin to rise precipitously. Another time series approach was used to forecast Covid-19 prevalence in Italy, Spain, and France by Ceylan (2020). ARIMA models were tested on forecasting for cumulative Covid-19 prevalence of Italy, Spain, and France. Their results showed that ARIMA models have 4,75 to 6,71 mean absolute percentage errors. Gamero, Tamayo, and Martinez-Roman (2020) analyzed the temporal series of confirmed cases through a first order polynomial for the first difference of the series of

daily accumulated confirmed Covid-19 cases expressed as a logarithm. They concluded that their results could serve to evaluate risks and control the evolution of this disease. Wu, Leung, and Leung (2020) provided an estimate of the size of the epidemic in Wuhan on the basis of the number of cases exported from Wuhan to cities outside mainland China and forecasted the extent of the domestic and global public health risks of epidemics, accounting for social and non-pharmaceutical prevention intervention. Roosa, Lee, Luo, Kirpich, Rothenberg, Hyman, and Chowell (2020) used phenomenological models that have been validated during previous outbreaks to generate and assess short-term forecasts of the cumulative number of confirmed reported cases in Hubei province, the epicenter of the epidemic, and for the overall trajectory in China, excluding the province of Hubei. They collected the daily reported cumulative confirmed cases for the Covid-19 outbreak for each Chinese province from the National Health Commission of China and provided five-day, ten-day, and fifteen-day forecasts for five consecutive days with an acceptable error rate. The importance of the AI-driven tools and their appropriate train and test models have been introduced and discussed by Santoch (2020). An improved adaptive neuro-fuzzy inference system (ANFIS) using an enhanced flower pollination algorithm (FPA) by using the salp swarm algorithm (SSA) model was contributed as a forecasting model to estimate and forecast the number of confirmed cases of Covid-19 in the upcoming ten days based on the previously confirmed cases recorded in China by Al-qaness, Ewees, Fan, H., and Abd El Aziz (2020). The proposed model was compared to several existing models, and it showed better performance in terms of Mean Absolute Percentage Error (MAPE), Root Mean Squared Relative Error (RMSRE), Root Mean Squared Relative Error (RMSRE), coefficient of determination (R^2), and computing time. An artificial intelligence and deep learning methodology were presented in another study to forecast of Covid-19 through time series using Non-Linear Regressive Network (NAR) (Ghazaly, Abdel-Fattah, and Abd El-Aziz, 2020). Another analysis and forecast of Covid-19 spreading in China, Italy and France was presented by Fanelli and Piazza (2020) using SIR model, recursive relations and non-linear fitting. Perc, Gorišek Miksić, Slavinec, and Stožer (2020) contributed a forecasting model with a simple iteration method that needs only the daily values of confirmed cases as input taking into account expected recoveries

and deaths, and determining maximally allowed daily growth rates that led away from exponential increase toward stable and declining numbers. The future spread of Covid-19 by exploiting the identified lead-lag effects between different countries was forecasted by Stübinger and Schneider (2020).

In this study, a deep learning method, long short-term memory (LSTM), and Artificial Neural Network (ANN) models with different learning algorithms using MATLAB software were compared to forecast daily Covid-19 cases for Turkey before one week. The daily confirmed Covid-19 cases are reported by the Governments to the World Health Organizations (WHO) day by day. We used daily new cases, cumulative cases, new deaths, and cumulative deaths data getting from the web page titled "WHO Coronavirus Disease (COVID-19) Dashboard" at <https://covid19.who.int/> (World Health Organization, 2020). By the way, there may be unreported Covid-19 positive cases due to lack of tests. Zhao, Musa, Lin, Ran, Yang, Wang, and Wang (2020) proposed the maximum likelihood estimation for the real number of Covid-19 cases in the first half of January 2020, which had not been reported. Moreover, this study does not cover the unreported cases. The rest of this study is organized as follows. The data gathering and descriptive statistics, the proposed LSTM and ANN models and their structures are given in the second section. The results were summarized in the third section. Finally, discussions and conclusions were made with recommendations.

3. Methodology

Research and publication ethics were followed in this study.

3.1 Covid-19 general situation in Turkey, data gathering, and descriptive statistics

In this study, an effective Covid-19 case forecasting model for Turkey is aimed. The first case was reported on 11/03/2020 by the Ministry of Health of Turkey. About a month later, on 12/04/2020, the number of reported cases per day reached its highest level with 5138. However, the first death was reported 7 days after the first case, on 18/03/2020, and the daily reported number of deaths was reported at 127 on 20/04/2020.

Due to the Covid-19 pandemic, the scientific committee, chaired by the Ministry of Health, convened day-to-day to assess the situation and as a result of these meetings, various economic and social measures were taken to protect public health and slow down the spread of the pandemic. Immediately after the first case, on 12/03/2020, the Government announced its first decisions to prevent the spread of the epidemic. In the circular issued by the Government, there were decisions affecting both social life and the economy, such as taking a break from schools, playing sports competitions without spectators, not allowing public officials to leave abroad if the subject is not necessary. Then, bars, nightclubs, theaters, cinemas, gyms and cafes were closed, mass worship was interrupted in mosques, and two weeks quarantine obligation was brought to everyone who returned from abroad on 16/03/2020. The first economic support package against Covid-19 was announced on 18/03/2020. After that, all private hospitals were declared as pandemic hospitals and cultural and scientific events were postponed on 20/03/2020. Hairdressers and beauty salons closed and flights to abroad were stopped on 21/03/2020.

The Government took several decisions to avoid spread of Covid-19, taking into account the recommendations made by the scientific committee from day to day. With the decline of daily reported cases, as of June 2020, decisions that adversely affect social life and the economy were loosened

and the normalization process was initiated. As can be seen, in order to control the spread of an epidemic that is extremely dangerous for public health, governments make decisions that may have social and economic adverse effects as a result of the observations made by scientists. Predicting how the number of cases may be at the end of a certain period will be guiding governments in their decisions. Therefore, predicting Covid-19 case numbers will make an important contribution in this area. In this study, prediction models are established for Turkey. Data provided by WHO was used in these models. The Covid-19 daily cases and death data were taken from the "WHO Coronavirus Disease (Covid-19) Dashboard" titled website at <https://covid19.who.int/> (World Health Organization, 2020). Descriptive statistics of the Covid-19 data between 11/03/2020–15/07/2020 for Turkey are given in Table 2. Daily reported new cases and cumulative cases are graphed in figure 1, daily reported new deaths and cumulative deaths are graphed in figure 2. On the other hand, according to data gathered from the WHO, the United States of America has the highest cumulative cases with 3405494 on the world. Brazil has the second highest with 1926824 cases, then Russian Federation comes with 746369 cases, and Mexico has 311486 cumulative cases as seen in table 3. By the way, the United Kingdom, Spain, and Italy have the higher number of reported Covid-19 cases with 291377, 256619, and 243344 respectively in the Europe zone at 16/07/2020.

Table 2
Descriptive statistics on used data types of Covid-19 in Turkey

Data Type	Minimum	Maximum	Mean	St. Dev	Variance	Skewness	Kurtosis
New Cases	0	5138	1692.85	1226.5494	1504423.41	1.10	0.60
Cumulative Cases	1	214993	118313	72459.56	5250388472	-0.49	-1.20
New Deaths	0	127	42.53	36.01	1296.91	0.94	-0.35
Cumulative Deaths	0	5402	3104.32	1955.76	3824999.68	-0.52	-1.34

Table 3
Covid-19 case statistics on the world reported to the WHO at 16/07/2020

Country	New cases	Cumulative cases	New deaths	Cumulative deaths
Australia	244	10495	3	111
Austria	201	19060	1	709
Brazil	41857	1926824	1300	74133
Canada	331	108486	8	8798
China	20	85697	2	4651
France	0	162390	0	29925
Germany	763	199726	7	9071
Iran (Islamic Republic of)	2521	262173	179	13211
Italy	114	243344	17	34984
Japan	382	22890	1	985
Mexico	7051	311486	836	36327
Netherlands	53	51091	0	6128
Pakistan	2165	255769	66	5386
Qatar	517	104533	1	150
Russian Federation	6422	746369	156	11770
Saudi Arabia	2692	237803	40	2283
Spain	666	256619	3	28409
The United Kingdom	1240	291377	138	44968
Turkey	992	214993	20	5402
United States of America	60711	3405494	754	135807

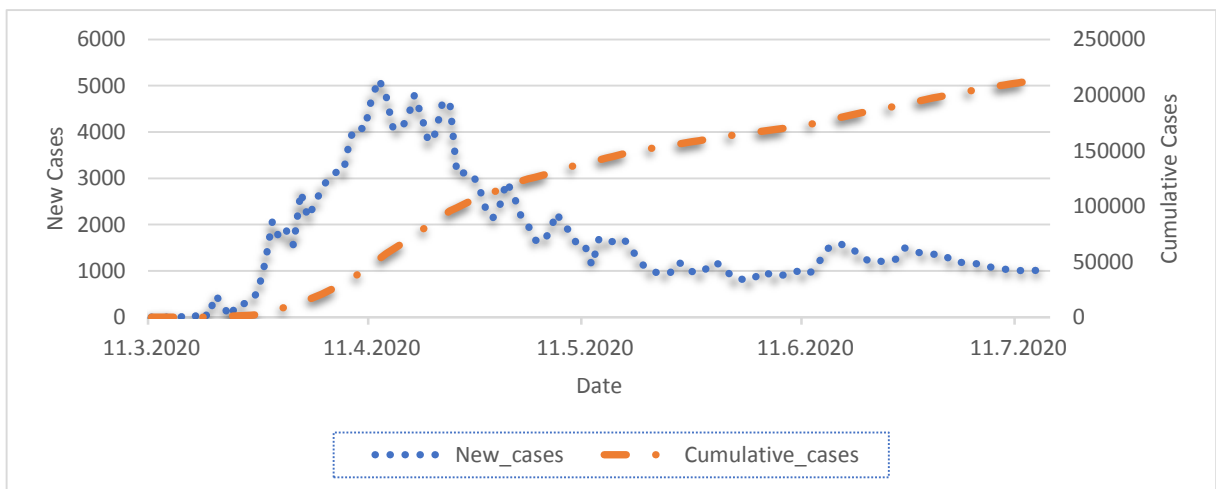


Figure 1. Daily Reported Covid-19 Cases for Turkey

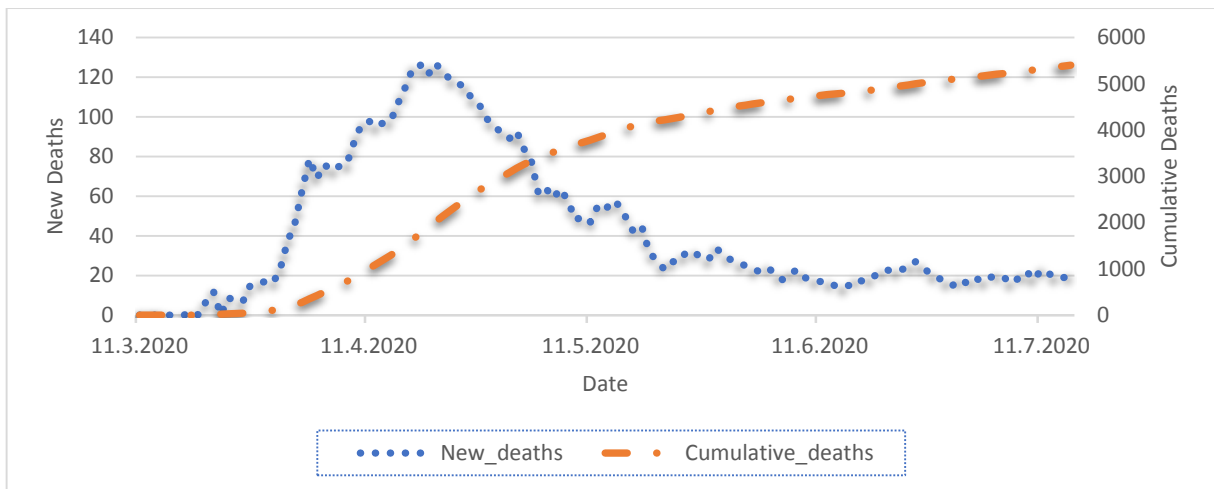


Figure 2. Daily Reported Covid-19 Deaths for Turkey

3.2 Model creation to forecast Covid-19 cases

Since the first reported case of Covidien-19 by the World Health Organization did not exceed one year. In addition, worldwide cases did not occur simultaneously. The first case was reported on March 11, 2020 in Turkey. It is not expected to save a lot of data to make forward-looking analysis in such a short time. Moreover, the data set discussed in this study covers the period between 18 March 2020, where the provincial death incident was reported, and 15 July 2020, where the last current situation was prepared, and only 120 days (four-month) of data were processed. In the current literature, time series approaches, which are generally not applicable with small historical dataset, are generally used to forecast Covid-19 cases. On the other hand, ANN and Deep Neural Network approaches are powerful tools in modelling less data occurrence. Those kinds of methodologies are smarter approaches due to their nature than time series methods and linear regression models. The main advantage of neural network-based methodologies is being able to model even nonlinear and or non-related functions.

3.2.1 Model type 1: ANN approaches

ANN is a kind of artificial intelligence that simulates the functioning of a human brain. Processing units form ANNs, which are inputs and outputs, respectively. Inputs are what ANN learned to produce the desired output. There is need a rule-based guidance for learning procedure of ANN, the Backpropagation. Hence, ANN method has basically three layers as input, hidden, and output with an input vector, bias vector, and weight matrices that tie the layers. The main procedure of ANN is reaching to the target output by converting inputs using a kind of training approach based on backpropagation which tries to finds bias vector and weight matrix. Figure 3 represents the general architecture of the ANN models. Equation 1 gives the general mathematical model of the ANN;

$$Y=ANN(x)+\epsilon \quad (1)$$

where Y is the prediction, $ANN(x)$ is the network function, and ϵ is the forecast error.

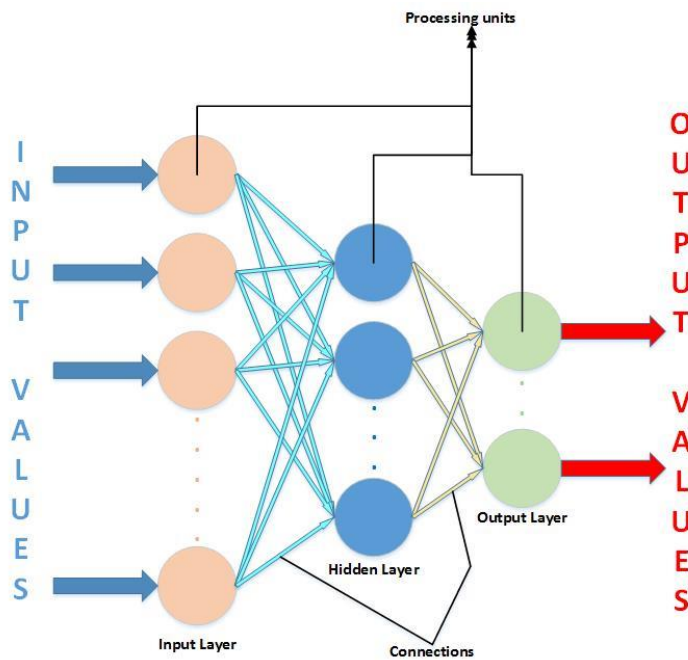


Figure 3. General architecture of ANN

In this study, six different ANN models with different learning algorithms were modelled to forecast next week’s daily Covid-19 new cases and daily Covid-19 cumulative cases for Turkey, separately. The architecture of the ANN models was set to 4 – 5 – 1 (four input, 1 hidden layer with five hidden neurons, and 1 output) as seen in the Figure 4. The inputs of the model are daily new cases, cumulative cases, new deaths, and cumulative deaths for day t , and the outputs are daily new cases at $t+7$ day and daily cumulative cases at $t+7$ day as modelled in Equation 2 – 3,

$$Y_1 = ANN(x) + \epsilon \tag{2}$$

$$Y_2 = ANN(x) + \epsilon \tag{3}$$

where Y_1 is daily new cases on day $t+7$, Y_2 is daily cumulative cases on day $t+7$, and x is a vector consisting of NC_t, CC_t, ND_t, CD_t . Here, NC_t is daily new cases on day t , CC_t refers to the daily cumulative cases on day t , ND_t refers to the daily new deaths on day t , CD_t is daily cumulative deaths on day t , and. The proposed ANN models were performed on data between 18 March and 15 July 2020 for Turkey when the first death was reported. Therefore, we examined 120 days (four-month) for Turkey. The dataset was randomly partitioned in three subsets such as training, validation, and test datasets with a

portion of 70%, 15%, and 15% respectively. The test data set was only used to check the performance of the proposed models and never used during training phase of the ANN procedure.

4-5-1

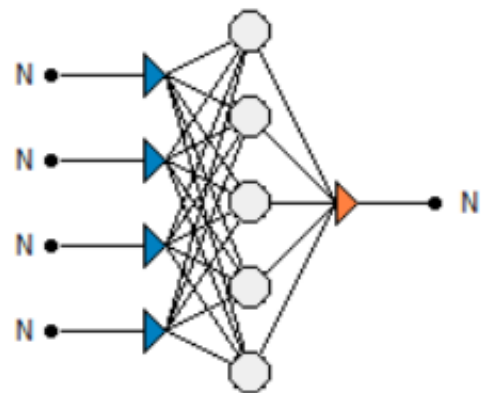


Figure 4. Architecture of Proposed ANN Models

The learning algorithms compared in this study are as follows; *Quick Propagation, Conjugate Gradient Descent, Quasi – Newton, Levenberg – Marquardt,*

Limited Memory Quasi-Newton, and Online Back Propagation.

3.2.2 Model type 2: Deep learning approaches

Deep Learning algorithms have gained importance among the common machine learning tools of recent times. Deep Neural Networks were also used in this study as a Deep Learning Tool, an improved version of Artificial Neural Networks. Deep learning uses multiple layers to progressively extract higher level features from the raw input compared with traditional ANN models. Long Short-Term Memory (LSTM) approach-based models were compared with ANN models considering the performance of

forecasting daily Covid-19 new cases and cumulative cases of Turkey before 7-day.

LSTM is a kind of artificial recurrent neural network architecture (Hochreiter and Schmidhuber, 1997). Unlike standard feedforward neural networks, LSTM has feedback connections. It has the advantage of processing the entire sequence of data rather than single data points. Figure 5 gives the general structure of LSTM algorithm which consists of a cell which remembers values over arbitrary time intervals, and three gates regulate the flow of information into and out of the cell which are input, output and, forget gates.

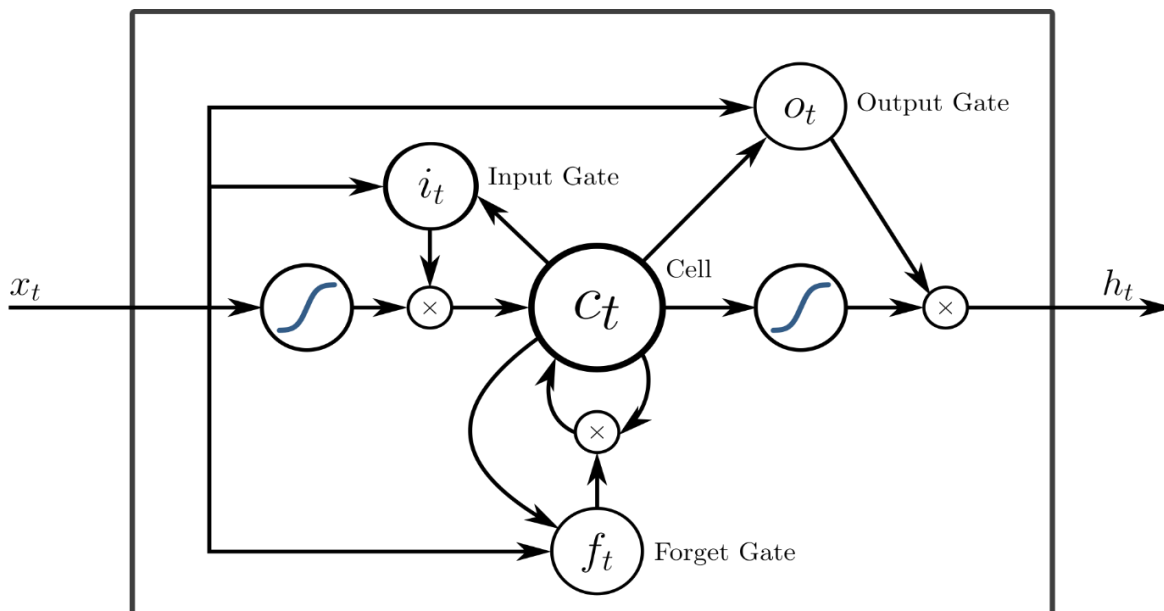


Figure 5. General structure of LSTM algorithm

The main challenge in designing deep neural networks is making a decision on the hidden layer numbers. Therefore, in this study, we have tried different numbers of hidden layers to find a better result. Also, three different optimizer, adaptive moment estimation - Adam, stochastic gradient descent with momentum - SGDM, and root mean square propagation - RMSProp, were compared to get more accurate forecasts.

4. Results

Forty-eight different models were contributed to check the ability of forecasting for daily Covid-19 new cases and cumulative cases before 7 day. Twelve of the proposed models were based on ANN models with six different learning algorithms in the training phase for daily new cases and cumulative cases, separately. Thirty-six of the proposed models were based on LST models with three different learning algorithms and six different hidden layer numbers in the training phase for daily new cases and cumulative cases, separately.

The performance comparison of those models was done by using mean absolute percentage error (MAPE) values of proposed models. MAPE, also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting methods in statistics used as a loss

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left(\frac{(|Actual Data_t - Forecasted Data_t|)}{Actual Data_t} \right) * 100 \quad (4)$$

where n is the total number of observations, t is the period index.

4.1 Results of ANN models

ANN structure is set to four input, five hidden neurons, and one output with 5000 epoch numbers for each model. The number of hidden neurons and epochs were set via trial and error approach.

The ANN models are initially fit on a training dataset to find the best values of weights of connections between neurons. The model is trained on the training dataset using an optimization method such as gradient descent then produces a result for target value. The result is compared with original target value and the learning adjust the parameters. Then, the model is used to predict the responses on validation dataset which is used for regularization by early stopping of training process (James, Hastie, and Tibshirani, 2013; Prechelt, 1998). Hence, MAPE values of validation data sets are better than training data sets. In addition, test datasets were never used during the training stages. The proposed models may have different performance ratios in test data sets, for this reason.

function for regression problems in machine learning. The basic calculation of MAPE is the mean of the ratio of actual data to the absolute deviation of forecasted data and actual data multiplied by 100 as given in Equation 4,

4.1.1 Results for daily cumulative case forecast using ANN

Table 4 presents the MAPE values for proposed learning algorithms for 7-day ahead daily Covid-19 cumulative case forecasts for subsets of used data as training, validation, test, and as a whole dataset respectively. *Conjugate Gradient Descent* algorithm performs better than others for cumulative case forecasting on training data set with only 4.79% mean error. *Online Back Propagation* gives the worst MAPE value on the training data set with 14.52% mean error.

Table 4

Mean absolute percentage error (MAPE) values of proposed ANN learning algorithms for daily cumulative case forecasts

Training Algorithm	MAPE			
	Training	Validation	Test	All
Quick Propagation	8.64	0.71	0.54	6.09
Conjugate Gradient Descent	4.79	0.75	1.26	3.58
Quasi - Newton	8.43	1.15	0.62	6.02
Limited Memory Quasi-Newton	9.96	1.75	1.88	7.37
Levenberg - Marquardt	6.99	0.90	1.20	5.10
Online Back Propagation	14.52	3.60	2.08	10.8

The nature of ANN algorithms is designed to find best regression between inputs and outputs, so that, training and validation data sets are used by algorithm for this purpose. Moreover, the sensitivity of the training algorithm to errors may not tell much about the performance of the learning algorithm, alone. For this reason, the test data set is used in the trained network and the actual performance of the model can be interpreted according to the errors of the test set. Although the *Online Back Propagation* algorithm has the highest error rate in the training set, it is quite surprising that it only gives 2.08% error in the test set. Meanwhile, *Quick Propagation* is the best algorithm on test data set with only 0.54%.

Figure 6 gives the graphical comparison of actual daily cumulative cases and results of *Conjugate Gradient Descent* and *Quick Propagation* algorithms. Both proposed models have a near-perfect fit with actual target data. In addition, all ANN models have MAPE values on training data set with a range between 4.79% and 14.52% and on test data set with a range between 0.54% and 2.08%. According to these results, it can be said that ANN models can

compete with other methods with the performance of Covid-19 cumulative case numbers in predicting one week in advance.

4.1.2 Results for daily new case forecast using ANN

Table 5 gives the MAPE values for proposed learning algorithms for 7-day ahead daily Covid-19 new case forecasts for subsets of used data as training, validation, test, and as a whole dataset respectively. *Conjugate Gradient Descent* algorithm performs better than others for new case forecasting on training data set with 11.66% mean error. *Online Back Propagation* gives the highest mean error on the training data set with 14.52%. Moreover, *Limited Memory Quasi-Newton* has the least mean error on test data set with 11.05%.

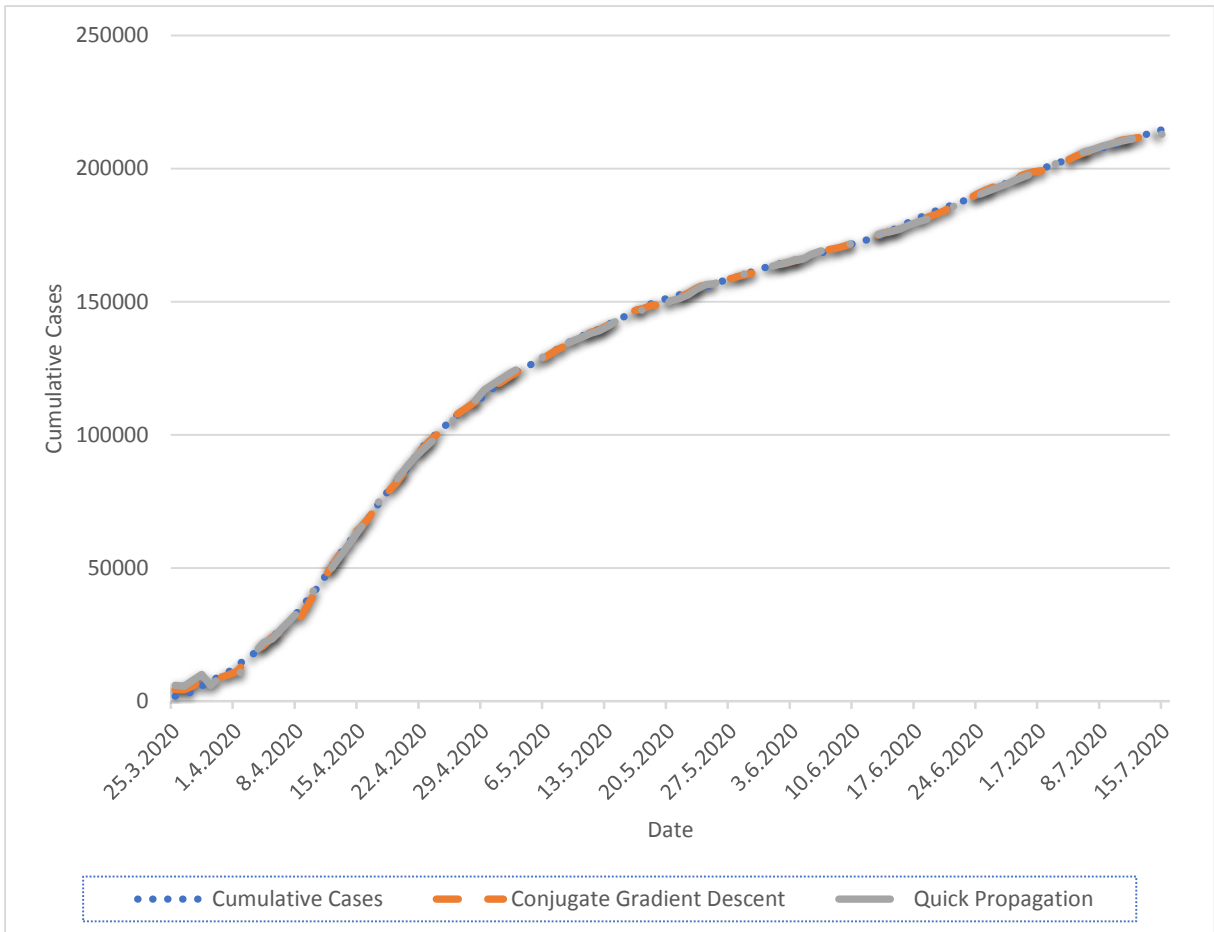


Figure 6. Daily Cumulative Cases vs Conjugate Gradient Descent and Quick Propagation Forecast Models

Table 5

Mean absolute percentage error (MAPE) values of proposed ANN learning algorithms for daily new case forecasts

Training Algorithm	MAPE			
	Training	Validation	Test	All
Quick Propagation	18.04	13.80	13.85	16.70
Conjugate Gradient Descent	11.66	12.28	15.38	12.30
Quasi - Newton	18.15	14.10	12.34	16.58
Limited Memory Quasi-Newton	13.32	12.74	11.05	12.86
Levenberg - Marquardt	15.64	14.95	14.27	15.31
Online Back Propagation	20.03	15.82	12.41	18.14

Figure 7 presents the graphical comparison of actual daily new cases and results of *Conjugate Gradient Descent* and *Limited Memory Quasi-Newton* algorithms. Both proposed models have a good performance on tracking actual new cases. In addition, MAPE values of proposed ANN models

range between 11.66% and 20.03% on training data set and range between 11.05% and 15.38% on test data. Considering that cumulative forecasts are generally more consistent rather than single forecasts in forecasting models, these results can be said to be as expected.

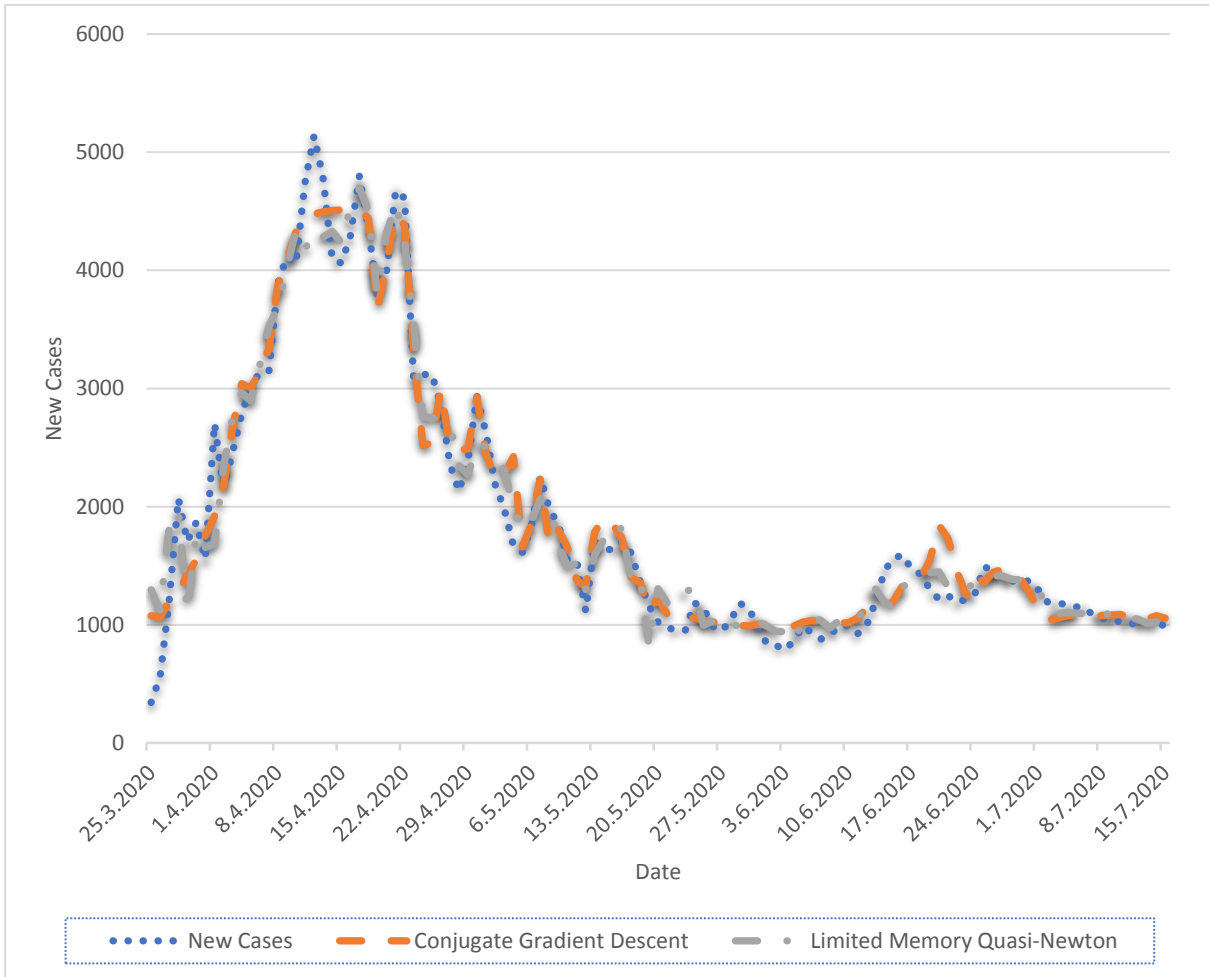


Figure 7. Daily New Cases vs Conjugate Gradient Descent and Limited Memory Quasi-Newton Forecast Models

4.2 Results of LSTM models

LSTM structure is set to four input, and one output with 250 epoch numbers for each model. The number of daily new cases, cumulative cases, new deaths, and cumulative deaths were used as input, and the number of daily new cases and cumulative cases after 7-days were used as outputs for the Covid-19 situation in Turkey, separately. Since the LSTM algorithm is a multi-layered neural network algorithm, how many layered models (1, 5, 10, 50, 100, or 500) should be used was also tested in this

study. In addition, three different optimization approaches (Adam, SGDM, and RMSProp) were studied in the learning phase.

4.2.1 Results for daily cumulative case forecast using LSTM

The MAPE values of daily Covid-19 cumulative case forecast models using LSTM approach for training, test, and as a whole dataset for different optimizers

and different hidden layer numbers are presented in Table 6.

Table 6
Mean absolute percentage error (MAPE) values of proposed LSTM models for daily cumulative case forecasts

Optimizer	Hidden Layer	MAPE		
		Training	Test	All
Adam	500	23.61	10.54	21.53
SGDM		15.34	3.79	13.50
RMSProp		10,87	13.56	11.29
Adam	100	5.66	12.29	6.72
SGDM		9.03	2.74	8.03
RMSProp		8.74	18.63	10.32
Adam	50	9.15	7.34	8.86
SGDM		8.71	2.82	7.77
RMSProp		7.19	22.20	9.58
Adam	10	4.70	11.32	5.75
SGDM		7.80	2.92	7.02
RMSProp		9.14	18.40	9.77
Adam	5	20.47	11.91	19.11
SGDM		8.10	2.47	7.20
RMSProp		26.12	23.01	25.62
Adam	1	182.30	38.71	159.42
SGDM		10.38	2.90	9.19
RMSProp		175.22	36.39	153.11

The results showed that the models with *Adam* optimizer gave the best result for ten hidden layers with 4.70% MAPE value on training data set and 5.75% on all data. Also, the worst configuration has *Adam* optimizer with one hidden layer. No optimizer clearly outperformed the others, different optimizers for different hidden layer numbers performed better. *Adam* optimizer gave better results than others for 100 and 10 hidden layers, *SGDM* was better for five and one hidden layers, and

RMSProp gave better results for 500 and 50 hidden layers on training data set. Furthermore, *SGDM* has the best MAPE values for all hidden layers on test data set. The best value is reached with for 100 hidden layers with 2.74% on test data.

The graphical comparison of actual daily cumulative cases and forecast results of 10 hidden layered LSTM model with *Adam* optimizer, and 100 hidden layered LSTM model with *SGDM* optimizer are given in figure 8.



Figure 8. Daily Cumulative Cases vs Adam Optimizer - 10 Hidden Layers and SGDM Optimizer 100 Hidden Layers Forecast Models

4.2.2 Results for daily new case forecast using LSTM

The results of daily Covid-19 new case forecast models using LSTM approach for training, test, and as a whole dataset for different optimizers and different hidden layer numbers are given in Table 7. The new case forecast results have higher MAPE values than cumulative case forecasts. This situation showed that the prediction of individual situations is generally worse than the plural as for the ANN

models. Moreover, all the LSTM models have higher MAPE values on the test data set. The best value is reached using SGDM optimizer and five hidden layers. RMSProp with 100 hidden layers has been the best model with an obvious difference compared to others on the training data set and on all data set. The worst LSTM model on the training data set has 36.96% MAPE value using one hidden layered Adam optimizer.

Table 7

Mean absolute percentage error (MAPE) values of proposed LSTM models for daily new case forecasts

Optimizer	Hidden Layer	MAPE		
		Training	Test	All
Adam	500	29.56	50.90	32.96
SGDM		26.33	59.04	31.54
RMSProp		19.76	51.50	24.81
Adam	100	15.40	21.96	16.44
SGDM		24.10	51.30	28.44
RMSProp		7.62	21.62	9.85
Adam	50	9.19	31.91	12.81
SGDM		21.83	33.93	23.76
RMSProp		17.19	21.21	17.83
Adam	10	16.52	22.56	17.49
SGDM		18.96	19.02	18.97
RMSProp		17.61	29.71	19.53
Adam	5	15.29	20.09	16.05
SGDM		16.03	18.39	16.41
RMSProp		15.95	28.76	17.99
Adam	1	36.96	28.80	35.66
SGDM		15.10	18.54	15.65
RMSProp		36.05	29.46	35.00

Figure 9 presents the graphical comparison of actual daily new cases and forecast results of 100 hidden layered LSTM model with *RMSProp* optimizer, and five hidden layered LSTM model with *SGDM* optimizer.

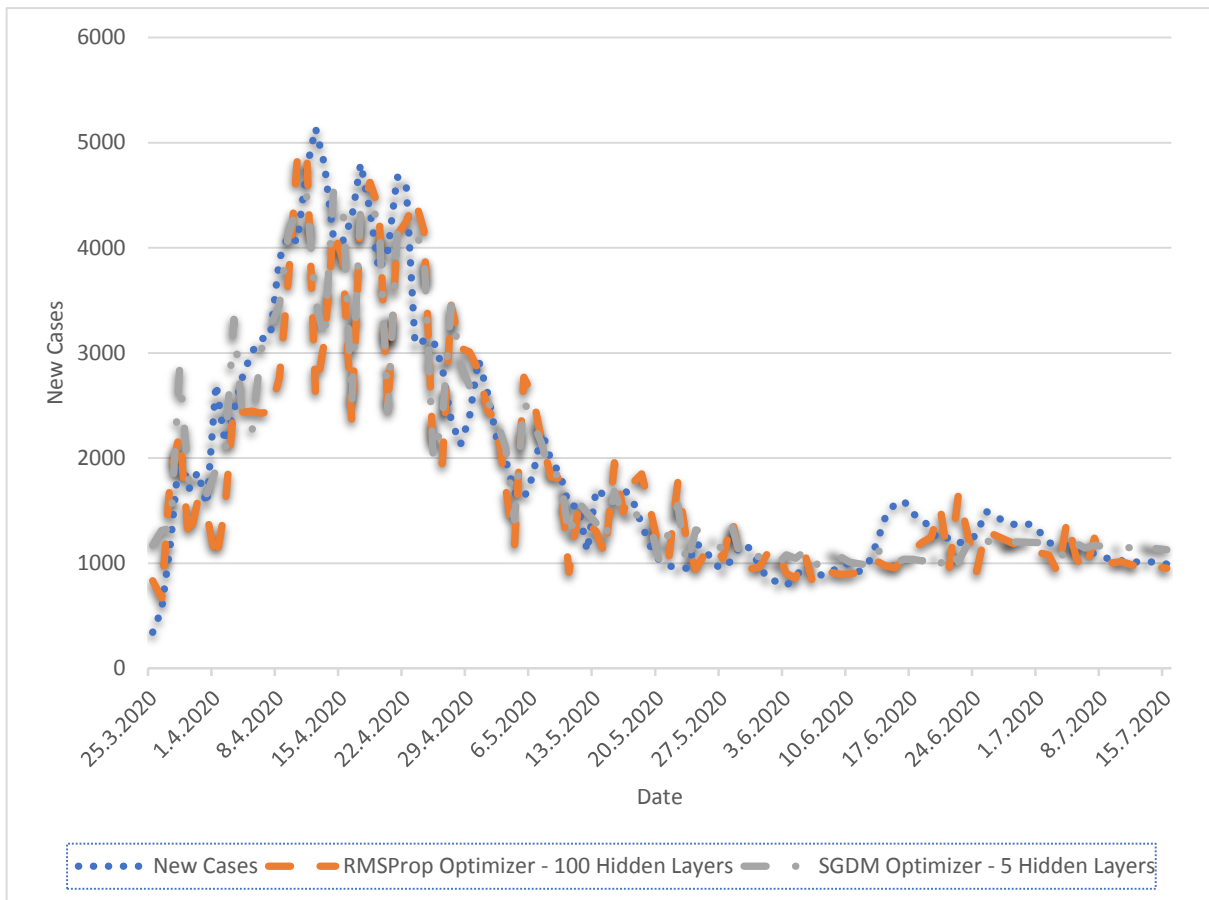


Figure 9. Daily New Cases vs RMSProp Optimizer - 100 Hidden Layers and SGDM Optimizer 5 Hidden Layers Forecast Models

5. Discussions and Conclusions

In this study, ANN models with different learning algorithms and LSTM models with different hidden-layer numbers and optimization algorithms were applied on real Covid-19 case datasets for Turkey to forecast daily new cases and cumulative cases 7-days in advance. The main purpose of this study was to contribute an effective short-term forecast methodology for the Covid-19 pandemic. Today, governments are making step-by-step decisions for the Covid-19 pandemic. These may include possible decisions that may lead to a new wave of epidemics, such as opening hairdressers, opening shopping malls, opening schools, so as not to worsen the economy. If the number of Covid-19 cases could be forecasted 7-days in advance, this might provide information shedding light on the threat of public health to the decisions taken to prevent the economic crisis. Unless an effective treatment

method or vaccine for Covid-19 can be developed, measures taken by governments will have to be relaxed, although they are indispensable for public health. In this case, estimating the impact of the decisions taken on the cases will also be very beneficial for their governments to see their way.

The results of this study show that daily cumulative cases can be forecasted 7-days in advance with approximately 4% to 10% margin of error. Because the nature of the forecasting science, models for individual forecasts have generally higher errors than for cumulative forecasts as in this study. The contributed successive models have approximately 10% to 18% margin of error. It was also observed that both ANN and LSTM algorithms gave competitive results in the short-term forecast of Covid-19 cases. To sum up, the proposed models may be effectively used for modeling and predicting to what extent Covid-19 cases will be the next week,

and this might be an important guideline for the strategic decisions that governments should take regarding Covid-19 measures.

Data Availability

The data that support the findings of this study are available in ref. [29] World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard. Retrieved from <https://covid19.who.int/> (accessed on 16 July 2020).

Conflict of Interest

The authors declared that there is no conflict of interest.

References

- Al-Qaness, M. A., Ewees, A. A., Fan, H., & Abd El Aziz, M. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. *Journal of Clinical Medicine*, 9(3), 674. doi: <https://doi.org/10.3390/jcm9030674>
- Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 138817. Doi: <https://doi.org/10.1016/j.scitotenv.2020.138817>
- Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of medical virology*, 92(4), 418-423. Doi: <https://doi.org/10.1002/jmv.25681>
- DeFelice, N. B., Little, E., Campbell, S. R., & Shaman, J. (2017). Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nature Communications*, 8(1), 1-6. Retrieved from: <https://www.nature.com/articles/ncomms14592/>
- Elmousalami, H. H., & Hassanien, A. E. (2020). Day level forecasting for Coronavirus Disease (COVID-19) spread: analysis, modeling and recommendations. *arXiv preprint arXiv:2003.07778*. Retrieved from: <https://arxiv.org/ftp/arxiv/papers/2003/2003.07778.pdf>
- Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, 134, 109761. doi: <https://doi.org/10.1016/j.chaos.2020.109761>
- Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. *arXiv preprint arXiv:2003.10776*. doi: <https://doi.org/10.9781/ijimai.2020.02.002>
- Gamero, J., Tamayo, J. A., & Martinez-Roman, J. A. (2020). Forecast of the evolution of the contagious disease caused by novel coronavirus (2019-nCoV) in China. *arXiv preprint arXiv:2002.04739*. Retrieved from: <https://arxiv.org/ftp/arxiv/papers/2002/2002.04739.pdf>
- Ghazaly, N. M., Abdel-Fattah, M. A., & Abd El-Aziz, A. A. (2020). Novel Coronavirus Forecasting Model using Nonlinear Autoregressive Artificial Neural Network. *Journal of advanced science*. Retrieved from: <http://serisc.org/journals/index.php/IJAST/article/view/8578/4797>
- Hardwick, N. V. (2006). Disease forecasting. In *The epidemiology of plant diseases* (pp. 239-267). Springer, Dordrecht. Retrieved from: <https://www.springer.com/gp/book/9781402045790>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. Retrieved from: <https://www.bioinf.jku.at/publications/older/2604.pdf>
- Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, 101(42), 15124-15129. Doi: <https://doi.org/10.1073/pnas.0308344101>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer. Retrieved from: <https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf>
- Kaundal, R., Kapoor, A. S., & Raghava, G. P. (2006). Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC bioinformatics*, 7(1), 485. Doi: <https://doi.org/10.1186/1471-2105-7-485>
- McDonald, M. R., & Boland, G. J. (2004). Forecasting diseases caused by *Sclerotinia* spp. in eastern Canada: fact or fiction?. *Canadian Journal of Plant Pathology*, 26(4), 480-488. Doi: <https://doi.org/10.1080/07060660409507168>

- Moran, K. R., Fairchild, G., Generous, N., Hickmann, K., Osthus, D., Priedhorsky, R., ... & Del Valle, S. Y. (2016). Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *The Journal of infectious diseases*, 214(suppl_4), S404-S408. Doi: <https://doi.org/10.1093/infdis/jiw375>
- Nsoesie, E., Mararthe, M., & Brownstein, J. (2013). Forecasting peaks of seasonal influenza epidemics. *PLoS currents*, 5. Doi: <https://doi.org/10.1371/currents.outbreaks.bb1e879a23137022ea79a8c508b030bc>
- Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting Covid-19. *Frontiers in Physics*, 8, 127. Doi: <https://doi.org/10.3389/fphy.2020.00127>
- Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS one*, 15(3), e0231236. Doi: <https://doi.org/10.1371/journal.pone.0231236>
- Prechelt, L. (1998). Early stopping-but when?. In *Neural Networks: Tricks of the trade* (pp. 55-69). Springer, Berlin, Heidelberg. Doi: https://doi.org/10.1007/978-3-642-35289-8_5
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., ... & Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, 256-263. doi: <https://doi.org/10.1016/j.idm.2020.02.002>
- Ruan, S. (2020). Likelihood of survival of coronavirus disease 2019. *The Lancet Infectious Diseases*, 20(6), 630-631. doi: [10.1016/S1473-3099\(20\)30257-7](https://doi.org/10.1016/S1473-3099(20)30257-7)
- Sannakki, S., Rajpurohit, V. S., Sumira, F., & Venkatesh, H. (2013, July). A neural network approach for disease forecasting in grapes using weather parameters. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE. doi: <https://doi.org/10.1109/ICCCNT.2013.6726613>
- Santosh, K. C. (2020). AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. *Journal of medical systems*, 44(5), 1-5. doi: <https://doi.org/10.1007/s10916-020-01562-1>
- Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012-2013 season. *Nature communications*, 4(1), 1-10. doi: <https://doi.org/10.1038/ncomms3837>
- Shaman, J., Yang, W., & Kandula, S. (2014). Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia. *PLoS currents*, 6. doi: <https://doi.org/10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6>
- Stübinger, J., & Schneider, L. (2020, June). Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. In *Healthcare* (Vol. 8, No. 2, p. 99). Multidisciplinary Digital Publishing Institute. doi: <https://doi.org/10.3390/healthcare8020099>
- Ture, M., & Kurt, I. (2006). Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Systems with Applications*, 31(1), 41-46. doi: <https://doi.org/10.1016/j.eswa.2005.09.002>
- Wang, L. F., Shi, Z., Zhang, S., Field, H., Daszak, P., & Eaton, B. T. (2006). Review of bats and SARS. *Emerging infectious diseases*, 12(12), 1834. doi: <https://doi.org/10.3201/eid1212.060401>
- WHO, Novel Coronavirus (2019-nCoV) 2020, 2020. Retrieved from: <https://www.who.int/>
- World Health Organization, (2020). WHO Coronavirus Disease (COVID-19) Dashboard. Retrieved from <https://covid19.who.int/>
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225), 689-697. doi: [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
- Zhao, S., Musa, S. S., Lin, Q., Ran, J., Yang, G., Wang, W., ... & Wang, M. H. (2020). Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. *Journal of clinical medicine*, 9(2), 388. doi: <https://doi.org/10.3390/jcm9020388>