

## Covid-19'un Yayılım Tahminine Yönelik Makine Öğrenmesi ve Derin Öğrenme Tabanlı Karşılaştırmalı Bir Analiz: Türkiye İçin Örnek Bir Çalışma

Anıl UTKU<sup>1\*</sup>, Ümit CAN<sup>2</sup>

<sup>1,2</sup> Bilgisayar Mühendisliği, Mühendislik Fakültesi, Munzur Üniversitesi, Tunceli, Türkiye

\*<sup>1</sup> anilutku@munzur.edu.tr, <sup>2</sup> ucan@munzur.edu.tr

(Geliş/Received: 03/06/2022;

Kabul/Accepted: 08/08/2022)

**Öz:** Salgınlar tarih boyunca çeşitli zaman dilimlerinde ortaya çıkmış ve insan topluluklarına ciddi zararlar vermiştir. Günümüzde ise bu salgınların modern versiyonu Covid-19 milyonlarca insanın hayatını kaybetmesine ve bir o kadarının da sağlık sorunları yaşamasına yol açmıştır. Tüm dünya, altyapı, finans, veri kaynakları, koruyucu donanımlar, hayati risk tedavileri ve diğer birçok kaynak açısından bu ölümcül hastalığın yayılmasına karşı mücadele etmek için inanılmaz çaba sarf etmektedir. Araştırmacılar ülke çapında paylaşılan verileri kullanarak bu salgın durumunu analiz etmek için matematiksel modeller geliştirmektedirler. Ülkeler salgın hızına bağlı olarak bu salgınla mücadele etmeye çalışmaktadırlar. Bu çalışmada Türkiye özelinde vaka ve ölüm sayılarının tahmin edilmesi için LSTM tabanlı bir tahmin modeli oluşturulmuştur. Bu modelin tahmin başarısını ölçmek için RF, SVM, XGBoost, MLP, CNN ve RNN olmak üzere popüler derin öğrenme yöntemleri dâhil altı makine öğrenmesi yöntemi kullanılmıştır. LSTM modeli vaka sayılarının tahmininde, MSE: 16670823,040 RMSE: 4082,991 MAE: 2543,651 R<sup>2</sup>: 0,975 sonuçlarını ve ölüm sayılarının tahmininde ise MSE: 331,351 RMSE: 18,203 MAE: 14,891 R<sup>2</sup>: 0,740 sonuçlarını alarak en başarılı model olmuştur.

**Anahtar kelimeler:** Covid-19, makine öğrenmesi, derin öğrenme, vaka tahmini.

### A Comparative Analysis for the Spread Prediction of Covid-19 Based on Machine Learning and Deep Learning: A Case Study for Turkey

**Abstract:** Epidemics have occurred in various time periods throughout history and have caused serious damage to human communities. Today, the modern version of these epidemics, Covid-19, has caused millions of people to die and have health problems as well. The entire world is making incredible efforts to combat the spread of this deadly disease in terms of infrastructure, finance, data resources, protective equipment, life-threatening treatments, and many other resources. Researchers are developing mathematical models to analyze this epidemic situation using data shared across the country. Countries are trying to fight this epidemic depending on the speed of the epidemic. In this study, an LSTM-based prediction model was created to predict the number of cases and deaths in Turkey. Six machine learning methods, including popular deep learning methods, RF, SVM, XGBoost, MLP, CNN, and RNN, were used to measure the prediction success of this model. The LSTM model has been the most successful model in predicting the number of cases, with MSE: 16670823,040 RMSE: 4082,991 MAE: 2543,651 R<sup>2</sup>: 0.975, and in predicting the number of deaths, MSE: 331,351 RMSE: 18,203 MAE: 14,891 R<sup>2</sup>: 0.740.

**Key words:** Covid-19, machine learning, deep learning, case prediction.

#### 1. Giriş

Çin'in Wuhan şehrinde Aralık 2019 itibariyle ortaya çıkan ve genetik yapısı henüz belirlenemeyen Covid-19, enfekte insanların sağlıklı kişilerle temas etmeye başlaması ve farklı bölgelere seyahat etmesiyle birlikte dünyadaki tüm ülkelere yayılmıştır. Covid-19, 10 Ocak 2020'de bildirilen ilk ölüm ile tüm dünya için bir pandemi haline gelmiştir [1]. Covid-19, akciğerleri ve solunum sistemini ciddi bir şekilde etkilemektedir. Korona virüsün insan sağlığını önemli ölçüde etkilediği ve önceden var olan sağlık sorunlarını şiddetlendirerek ölümlere neden olduğu belirtilmektedir [2, 3].

Tüm ülkeler, seyahat kısıtlamaları, karantinalar, etkinlik ertelemeleri ve iptalleri, sosyal mesafe uygulamaları ile karantina gibi önlemleri uygulayarak Covid-19'un yayılımına azaltmaya çalışmıştır. Virüsün hastalık ve ölüm boyutunun yanı sıra ekonomik ve sosyal boyutu da ülkeleri zor duruma düşürmektedir [4-6]. 15 Aralık 2021 itibariyle, 500 milyona yakın doğrulanmış vaka ve 6 milyona yakın Covid-19 kaynaklı ölüm tespit edilmiştir. Covid-19 için geliştirilen aşılama çalışmaları devam ederken, virüs mutasyona uğradığı için aşının etkinliği ve

\* anilutku@munzur.edu.tr. Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-7240-8713, <sup>2</sup> 0000-0002-8832-6317

koruma süresi gibi konular ön plana çıkmaktadır. Virüsün yayılmasını yavaşlatmak, artan hasta yüküyle mücadele etmek, tıbbi müdahale sistemleri hazırlamak ve sağlık personellerinin yükünü hafifletmek pandemi sürecinde oldukça önemlidir [7, 8]. Dünya genelinde Covid-19 pandemisinden çok sayıda insan etkilendiğinden ve hastalığın tedavisi bulunmadığından, mevcut veriler kullanılarak oluşabilecek potansiyel vakaların sayısının tahmin edilmesi önem kazanmaktadır. Covid-19'un yayılımının tahminine yönelik literatürdeki çeşitli çalışmalar yapılmıştır.

Punn ve diğerleri [9] makine öğrenimi ve derin öğrenme modellerini, Covid-19'un günlük üstel davranışını anlamak ve COVID-19'un uluslararası gelecekte erişilebileceği seviyeyi tahmin etmek amacıyla Johns Hopkins panosundan gerçek zamanlı bilgilerden faydalanarak kullanmayı önermişlerdir. Alazab ve diğerleri [10] Covid-19 hastalarını belirlemek için göğüs röntgeni görüntülerini makine öğrenmesi yöntemleri ile analiz etmişlerdir. Elde edilen sonuçlar X-ışınlarının hızlı ve düşük maliyetle kolayca elde edilebildiği için COVID-19 tanısında önemli olduğunu göstermiştir. Gerçek hastaların 1000 röntgen görüntüsünden elde edilen ampirik bulgular, önerilen sistemin COVID-19'u tespit etmede faydalı olduğunu doğrulamıştır. Bu çalışmada ayrıca Otoresif Entegre Hareketli Ortalama (Autoregressive Integrated Moving Average) (ARIMA) ve Uzun Kısa Süreli Bellek Sinir Ağı (Long Short-Term Memory Neural Network) (LSTM) gibi yöntemler kullanılarak bir haftalık periyot için başarılı vaka sayısı tahmini yapılmıştır. Malki ve diğerleri [11] birçok ülkedeki Covid-19 vakalarının yayılımını tahmin etmek için makine öğrenmesi yöntemi kullanmış ve önerdikleri yaklaşım Rastgele Orman (Random Forest) (RF) ve ARIMA gibi yöntemlerle karşılaştırılmış ve 0.99 R2 başarı oranını yakalamıştır. Macaristan özelinde yapılan bir çalışmada ise hibrit bir yöntem kullanılarak Covid-19 tahmini yapılmış ve yeni önerilen yöntem başarılı sonuçlar elde etmiştir [12]. Bir diğer çalışmada ise teyit edilen vaka sayısı ile belirli bölgelerdeki hava durumu değişkenleri arasındaki ilişki çıkartılarak, sıcaklık ve nem gibi hava değişkenlerinin COVID-19'un bulaşması üzerindeki etkisi tahmin edilmiştir. Deneysel sonuçlardan, nüfus, yaş ve kentleşme gibi diğer nüfus sayımı değişkenleriyle karşılaştırıldığında, hava durumu değişkenlerinin ölüm oranını tahmin etmede daha alakalı olduğu gösterilmiştir [13]. Shadid ve diğerleri [14] tarafından yapılan çalışmada ise COVID-19'dan etkilenen on büyük ülkede doğrulanmış vakaların, ölümlerin ve iyileşmelerin zaman serisi tahmini için ARIMA, Destek Vektör Makinesi (Support Vector Machines) (SVM), LSTM ve Bi-LSTM modelleri kullanılmış ve COVID-19 nedeniyle etkilenen on büyük ülkede ölümler ve iyileşmeler tahmin edilmiştir. Çoğunlukla Bi-LSTM modeli en başarılı sonuçları elde etmiştir. Pakash ve diğerleri [15] yaptıkları çalışmada hangi yaş gruplarının Covid-19 salgınından daha fazla etkilendiğini bulmak için makine öğrenmesi yöntemlerini kullanarak çalışma yapmışlardır. Satu ve diğerleri [16] ise yaptıkları başka bir çalışmada Bangladeş için bir tahmin modeli geliştirmişlerdir. Bu model çeşitli regresyon yöntemleri kullanarak bir haftalık bir periyot için Covid-19 hastalığı bulaşan insanları tahmin etmeye çalışmışlardır.

Bu çalışmada, geliştirilen LSTM tabanlı derin öğrenme modeli ile Türkiye için Covid-19 vaka ve ölüm sayılarının tahmin edilmesi amaçlanmıştır. LSTM, zaman serisi verilerinin analizinde gösterdiği başarı ile gelecekteki eğilimlerin doğru bir şekilde tahmin edilmesindeki başarısı sebebiyle seçilmiştir. Geliştirilen LSTM tabanlı model RF, SVM, Aşırı Gradyan Artırma (Extreme Gradient Boosting) (XGBoost), Çok Katmanlı Algılayıcılar (Multi-layer perceptrons) (MLP), Evrimsel Sinir Ağı (Convolutional Neural Network) (CNN) ve Yinelemeli Sinir Ağı (Recurrent Neural Network) (RNN) modelleri ile uygulamalı olarak karşılaştırılmıştır. MSE, RMSE, MAE ve R2 metrikleri kullanılarak yapılan deneysel çalışmalar LSTM'in karşılaştırılan diğer modellere göre vaka ve ölüm sayısı tahmininde daha başarılı olduğunu göstermiştir. Ayrıca literatürde genelde vaka sayılarının tahminine dönük çalışmalar çoğunlukla yapılmıştır. Bu çalışma vaka sayılarına ek olarak ölüm sayısını da başarılı bir şekilde tahmin ederek literatüre katkı sunmuştur. Böylece Türkiye özelinde başarı oranı yüksek derin öğrenme tabanlı bir model geliştirilmiştir.

## 2. Tahmin Modelleri

Bu çalışma çerçevesinde oluşturulan LSTM tabanlı model RF, SVM, XGBoost, MLP, CNN ve RNN ile uygulamalı olarak karşılaştırılmıştır. Bu modellerden bu bölümde kısa bahsedilmiştir

- **RF:** Topluluk denetimli bir makine öğrenimi tekniğidir. Torbalama (Bagging) ve rastgele öznitelik seçimine dayalı olarak, karar ağacı sayısı (temel sınıflandırıcılar) oluşturulur ve sınıflandırma için çoğunluk oylaması yapılır [17]. RF birçok alanda başarıyla uygulanmaktadır. Örneğin ağa izinsiz giriş tespiti, E-posta spam tespiti, gen sınıflandırması, kredi kartı sahtekârlığı tespiti, metin sınıflandırması ve olayların tahmin edilmesi gibi problemlerin çözümünde başarı ile kullanılmaktadır [18-20].
- **SVM:** Vapnik, sınıflandırma ve regresyon problemlerini çözmek için Vektör Makinelerini önermiştir [21]. Support Vector Regresyon (SVR) ise denetimli bir makine öğrenmesi yöntemidir ve SVM'nin regresyon problemlerinde kullanılan versiyonu olarak adlandırılabilir [22]. SVR, girdi tahmin değerlerini, gözlemlenen çıkış değerlerine eşleyen bir regresyon fonksiyonunu öğrenmek için bir optimizasyon

problemini formüle eder. SVR, model karmaşıklığı ile tahmin hatasını dengeler ve yüksek boyutlu verileri işlemek için iyi bir performansa sahiptir [23].

- *XGBoost*: Chen ve Guestrin [24] tarafından önerilmiştir ve gradyan arttırma algoritmasına dayanmaktadır. XGBoost, enerji, finans, sağlık vb. alanlarda uygulama alanı bulan popüler ve başarılı bir algoritmadır ve XGBoost hız ve performans açısından diğer algoritmalara göre çok avantajlı bir konumdadır [25]. XGBoost, hastalık teşhisi, görüntü tanıma, trafik akışı tahmini, spor analitiği, güç sistemleri, kişisel kredi, hava tahmini kamu güvenliği, arıza tespiti vb. gibi birçok alanda kullanılmıştır ve başarılı sonuçlar elde edilmiştir [26].
- *MLP*: Tipik bir parametrik olmayan sinir ağı sınıflandırıcısıdır ve diğer sınıflandırıcılara kıyasla üstün performans sağlayabilen güçlü sınıflandırıcılardır [27]. Bu yöntemde parametreler bir doğrulama seti veya çapraz doğrulama teknikleri yardımıyla ayarlanır [28]. Temel bir MLP modelinin yapısı üç farklı katmandan oluşur: girdi, gizli ve çıktı. Girdi katmanı, verilerin okunduğu katmandır. Her nöron bir özelliği temsil ettiğinden, özellik sayısı kadar nöron içerir. Çıktı katmanında sınıflandırma yapılır. Gizli katman, giriş ve çıkış katmanları arasındaki verilerin ara işlemeye maruz kaldığı yerdir. Gizli katman sayısı ve gizli katmandaki nöron sayısı kesin olmasa da eğitimin kalitesini etkileyen iki önemli faktördür [29, 30].
- *CNN*: MLP'nin bir versiyonudur. Derin öğrenme ağlarının en popüler ve en çok kullanılanlarından biridir. CNN'nin diğer derin öğrenme modellerine göre ana avantajı, herhangi bir insan denetimi olmadan önemli özellikleri otomatik olarak algılamasıdır [31,32]. İleriye dönük bir sinir ağı olan CNN algoritması, insanların görsel algısından esinlenen araştırmacıların, görsel alanda bilgisayar uygulamalarında ortaya çıkan problemlerin üstesinden gelmeye çalışmasıyla ortaya çıkmıştır [33].
- *RNN*: Sıralı ve zaman serisi verilerden özellikleri ve uzun vadeli bağımlılıkları öğrenme yeteneğine sahip bir yapay sinir ağı modelidir. RNN'ler, birimler arasında en az bir bağlantının yönlendirilmiş bir döngü oluşturduğu bir doğrusal olmayan birimler yığına sahiptir. İyi eğitilmiş bir RNN herhangi bir dinamik sistemi modelleyebilir. RNN'nin temel özelliği, ağın en az bir geri besleme bağlantısı içermesidir, böylece aktivasyonlar bir döngü içinde akabilir. Bu, ağların zamansal işleme yapmasını ve dizileri öğrenmesini sağlar [34, 35].

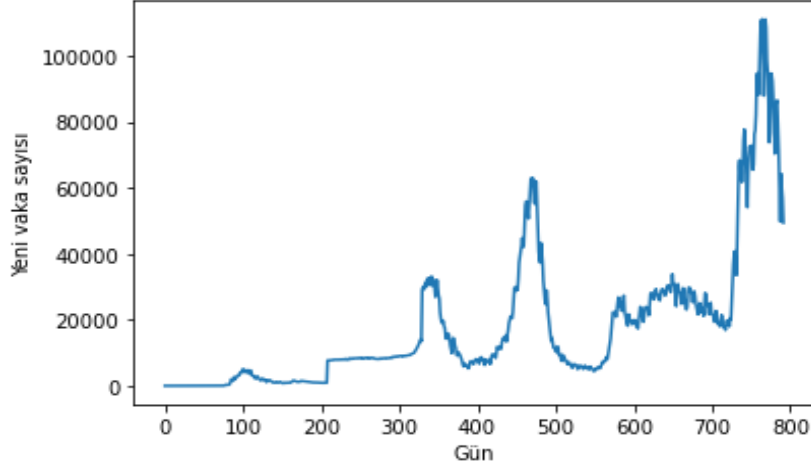
### 3. Geliştirilen Derin Öğrenme Tabanlı Tahmin Modeli

Teknolojik gelişmelerle birlikte makine öğrenmesi ve derin öğrenme teknolojileri birçok araştırma alanında uygulanmaktadır. Bu teknolojiler, Covid-19'un tahmin edilmesi araştırmalarında da büyük rol oynamaktadır. Günümüzde salgın sürecinin devam etmesi sebebiyle, hastalığın epidemiyolojik özellikleri konusunda ve Covid-19'un radyolojik görüntülerden tahmini konusunda çalışmalar yapılmaktadır. Etkili ve zamanında izleme, hastalığın tahmini ve ayrıca hastalık etkilerinin önlenmesi stratejilerinin temel bileşenleridir. Derin öğrenme, özellikle zamana bağımlı verilerde diğer modellere kıyasla daha başarılı olan bir analiz yöntemidir.

Bu çalışmada, Covid-19'un yayılım dağılımının tahminine yönelik LSTM tabanlı bir derin öğrenme modeli geliştirilmiştir. Geliştirilen model ile Covid-19'un ortaya çıktığı ilk tarihten 8 Nisan 2022'ye kadar olan günlük vaka ve ölüm sayılarının tahmin edilmesi amaçlanmıştır. Geliştirilen LSTM tabanlı model RF, SVM, XGBoost, MLP, CNN ve RNN ile uygulanmalı olarak karşılaştırılmıştır.

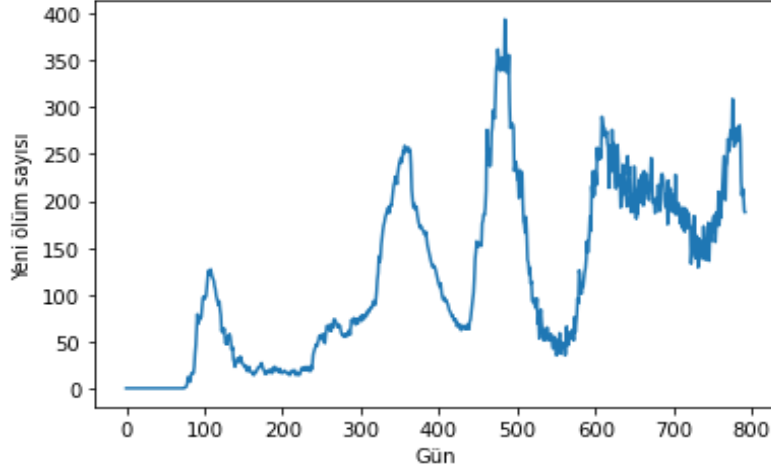
#### 3.1. Veriseti

Bu çalışmada Dünya Sağlık Örgütü (DSÖ) tarafından sunulan, Türkiye'nin günlük resmi Covid-19 vaka ve ölüm sayıları veriseti olarak kullanılmıştır. Kullanılan veriseti Date\_reported, Country\_code, Country, WHO\_region, New\_cases, Cumulative\_cases, New\_deaths ve Cumulative\_deaths olmak üzere 8 sütundan oluşmaktadır. Date\_reported, günlük verilerin DSÖ'ye bildirildiği zamanı ifade eder. Ülke\_kodu, ISO Alpha-2 ülke kodunu ifade eder. Country, ülke adını ifade etmektedir. WHO\_region, Afrika, Amerika, Güneydoğu Asya, Avrupa, Doğu Akdeniz ve Batı Pasifik olmak üzere DSÖ tarafından tahsis edilen bölge ofislerini ifade eder. New\_cases, doğrulanmış günlük yeni vaka sayısını ifade eder. Cumulative\_cases, belirli bir tarihe kadar rapor edilen doğrulanmış toplam vaka sayısını ifade eder. New\_deaths, doğrulanmış Covid-19 kaynaklı yeni ölümlerin sayısını ifade eder. Cumulative\_deaths, belirli bir tarihe kadar bildirilen Covid-19 kaynaklı toplam ölüm sayısını ifade eder. Şekil 1'de 8 Nisan 2022'ye kadar olan günlük vaka sayıları görülmektedir.



Şekil 1. Günlük vaka sayıları

Şekil 2’de ise 8 Nisan 2022’ye kadar olan Covid-19 kaynaklı günlük ölüm sayıları görülmektedir.



Şekil 2. Günlük ölüm sayıları

### 3.2. Veri Ön-İşleme

Bu çalışmada kullanılan veriseti zaman serisi verilerinden oluşmaktadır. Zaman serisi verileri, belirli bir zaman değişkenine bağımlı olarak değer alan verilerdir. Makine öğrenmesi ve derin öğrenme modellerinin genel itibariyle işlediği problemler ise denetimli öğrenme problemleridir. Denetimli öğrenme problemleri, bir algoritmanın girdi değişkenlerinden çıktı değişkenlerini bir fonksiyon yardımıyla nasıl tahmin edeceğini öğrenebildiği problemlerdir. Zaman serisi verileri, bir sonraki zaman adımındaki değeri tahmin etmek için önceki zaman adımındaki değerler kullanılarak kayan pencere yöntemi ile denetimli öğrenme problemi şeklinde yapılandırılabilir. Kayan pencere yöntemi, bir sonraki zaman adımını tahmin etmek için önceki zaman adımlarının kullanılmasına dayanan bir yöntemdir. Kayan pencerenin boyutu, önceki zaman adımlarının sayısına göre belirlenir. Bu çalışmada sürgülü pencere boyutu 3 olarak seçilmiştir. Bu sayede t-3, t-2 ve t-1 zaman adımlarındaki veriler girdi, t zaman adımındaki veri ise çıktı olacak şekilde zaman serisi verileri denetimli öğrenme problemine dönüştürülmüştür.

Daha sonra vaka ve ölüm sayılarından oluşan veriler normalize edilmiştir. Normalizasyon, genellikle bir veri ön-işleme süreci olarak uygulanan bir makine öğrenmesi tekniğidir. Normalizasyonun amacı, verisetindeki sayısal sütunların değerlerini, değer aralığındaki farklılıkları bozmadan ortak bir ölçeğe değiştirmektir. Veriler 0-1 aralığında normalize edilmiştir. Normalizasyon adımından sonra veriler %80 eğitim ve %20 test olarak ayrılmıştır. Eğitim verilerinin %10'u doğrulama için ayrılmıştır. Model parametrelerinin optimizasyonu için doğrulama

verileri kullanılmıştır. Karşılaştırılan modellerin en iyi sonuçları vermesi için GridSearchCV kütüphanesi kullanılarak parametre optimizasyonu yapılmıştır. GridSearch yönteminde, modelde test edilecek hiperparametreler ve değerleri için tüm kombinasyonlarla ayrı ayrı bir model kurulur ve belirlenen metriğe göre en başarılı hiperparametre seti belirlenir. Parametre optimizasyonu ile MSE değerleri en düşük olan parametreler belirlenmiş ve modeller oluşturulmuştur.

### 3.3. Değerlendirme Metrikler

Regresyon problemlerinde yapılan tahminler ile gerçek veriler arasındaki hata oranını ölçmek için yaygın olarak MSE, RMSE, MAE ve  $R^2$  metrikleri kullanılmaktadır. MSE, verisetindeki gerçek değerler ile tahmin edilen değerler arasındaki farkların karelerinin ortalamasını ifade eder. MSE,  $y$  gerçek değerleri,  $\hat{y}$  tahmin edilen değerleri ve  $n$  ise örnek sayısını ifade etmek üzere Denklem 1 kullanılarak hesaplanır.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2 \quad (1)$$

Hataların standart sapmasını ölçen RMSE ise MSE'nin karekökü alınarak Denklem 2'de görüldüğü gibi hesaplanır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (|y - \hat{y}|)^2} \quad (2)$$

MAE, verisetindeki gerçek değerler ile tahmin edilen değerler arasındaki farkların mutlak değerlerinin ortalamasını ifade eder. Verisetindeki hataların ortalamasını hesaplar. MAE, Denklem 3 kullanılarak hesaplanmaktadır.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (3)$$

$R^2$ , modellerin verisine gösterdiği uyumun bir ölçüsüdür.  $R^2$ , veri noktalarının regresyon çizgisi etrafındaki dağılımını değerlendirir. Aynı veriseti için daha yüksek  $R^2$  değerleri, gerçek değerler ile tahmin edilen değerler arasındaki daha düşük hataları ifade eder.  $R^2$ ,  $\hat{y}$  tahmin edilen  $y$  değerlerini,  $\bar{y}$  ise  $y$  değerlerinin ortalamasını ifade etmek üzere Denklem 4 kullanılarak hesaplanmaktadır.

$$R^2 = \frac{\sum (y - \hat{y})^2}{(y - \bar{y})^2} \quad (4)$$

### 3.4. Geliştirilen LSTM Tabanlı Tahmin Modeli

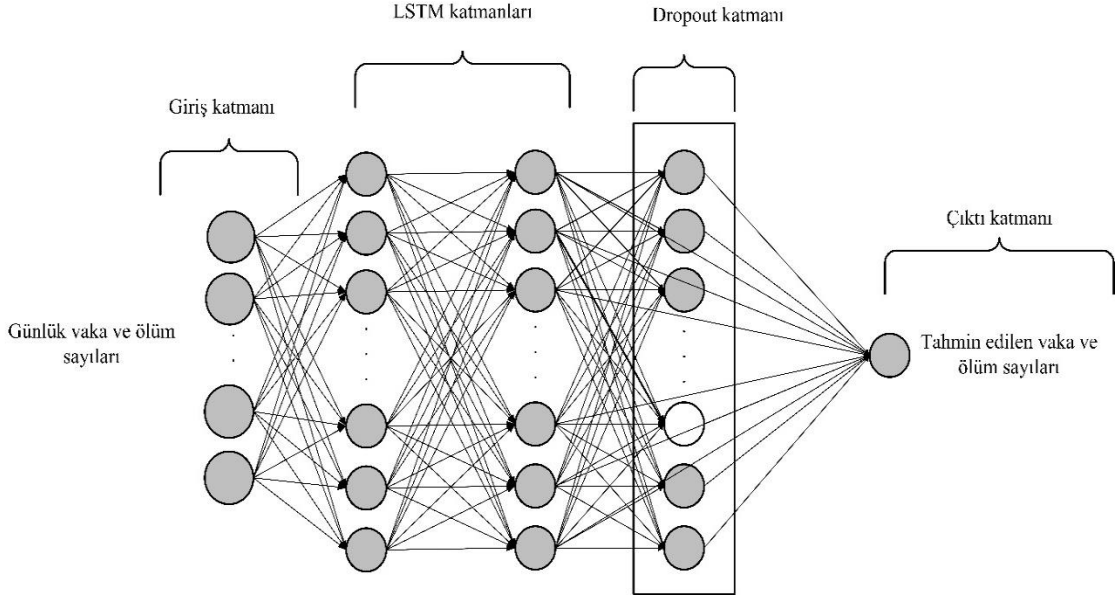
Derin öğrenme modelleri, girdi verisinden çıktı verisine doğrusal olmayan dönüşümler yaparak karmaşık fonksiyonları yüksek boyutlarda öğrenmeyi amaçlayan karmaşık mimarilere sahip yapay sinir ağı modelleridir. Derin öğrenme modelleri, görüntü işleme, ses ve konuşma tanıma ve doğal dil işleme gibi birçok uygulama alanında başarıyla kullanılmaktadır. Derin öğrenme modelleri, karmaşık işlevleri çözmek için insan beyninin faaliyetlerini taklit eden sinir ağı mimarileridir. Yapay sinir ağları temel olarak giriş katmanı, gizli katmanlar ve çıkış katmanı olmak üzere üç temel katmandan oluşur. Giriş katmanı, sinir ağında işlenecek giriş verilerini okur. Girdi katmanı ile çıktı katmanı arasındaki tüm katmanlara gizli katmanlar denir. Gizli katmanlardaki nöronlar, ağırlıklı girdileri ve aktivasyon fonksiyonlarını kullanarak bir çıktı üretir. Çıktı katmanı, sistemin çıktısını veren son sinir ağı katmanıdır.

CNN, görüntüleri girdi olarak alan ve görüntülerin çeşitli bölümlerine ağırlıklar ve önyargılar atayan bir derin öğrenme modelidir. CNN, görüntü tanıma, görüntü sınıflandırma, zaman serisi analizi ve regresyon problemlerine başarıyla uygulanabilir. CNN'ler, görüntülerde mevcut olan uzamsal bilgileri işleyebilen evrişim işlemlerini kullanırken RNN'ler, zaman serisi verilerinde mevcut olan zamansal bilgileri depolayabilen belleğe sahiptir. MLP, genellikle klasik bir sinir ağı olarak kullanılan temel bir modeldir. CNN'nin evrişim katmanları, metin verileri ve zaman serisi verileri gibi 1 boyutlu veriler üzerinde de başarılı bir şekilde çalışır.

LSTM, RNN'nin gelişmiş bir versiyonudur. RNN ve LSTM arasındaki temel fark, bilginin saklanma süresidir. LSTM, bilgileri RNN'den daha uzun süre hafızada işleyebildiği için RNN'den daha avantajlıdır. LSTM

hücreleri, okunan ve kendilerine yazılan hücre durumlarını korur. Giriş ve hücre durumu değerlerine bağlı olarak, 4 kapı okuma, yazma ve hücre durumu çıkışı düzenler. Birinci kapı, gizli durumun neyi unuttuğunu belirler. İkinci kapı, hücre durumunun hangi kısmına yazıldığını belirlemekten sorumludur. Üçüncü kapı, yazılan içeriğe karar verir. Son kapı, çıktı üretmek için hücre durumundan okur. LSTM, uzun vadeli bağımlılıkları hatırlamadaki başarısı nedeniyle diğer derin öğrenme modellerine göre öne çıkmaktadır.

Bu çalışmada, geliştirilen LSTM tabanlı derin öğrenme modeli, günlük Covid-19 vaka ve ölüm sayılarını girdi olarak almakta ve çıktı olarak ölüm ve vaka tahminleri üretmektedir. Geliştirilen LSTM tabanlı modelin mimarisi Şekil 3'te gösterilmektedir.



Şekil 3. Geliştirilen LSTM tabanlı modelin mimarisi

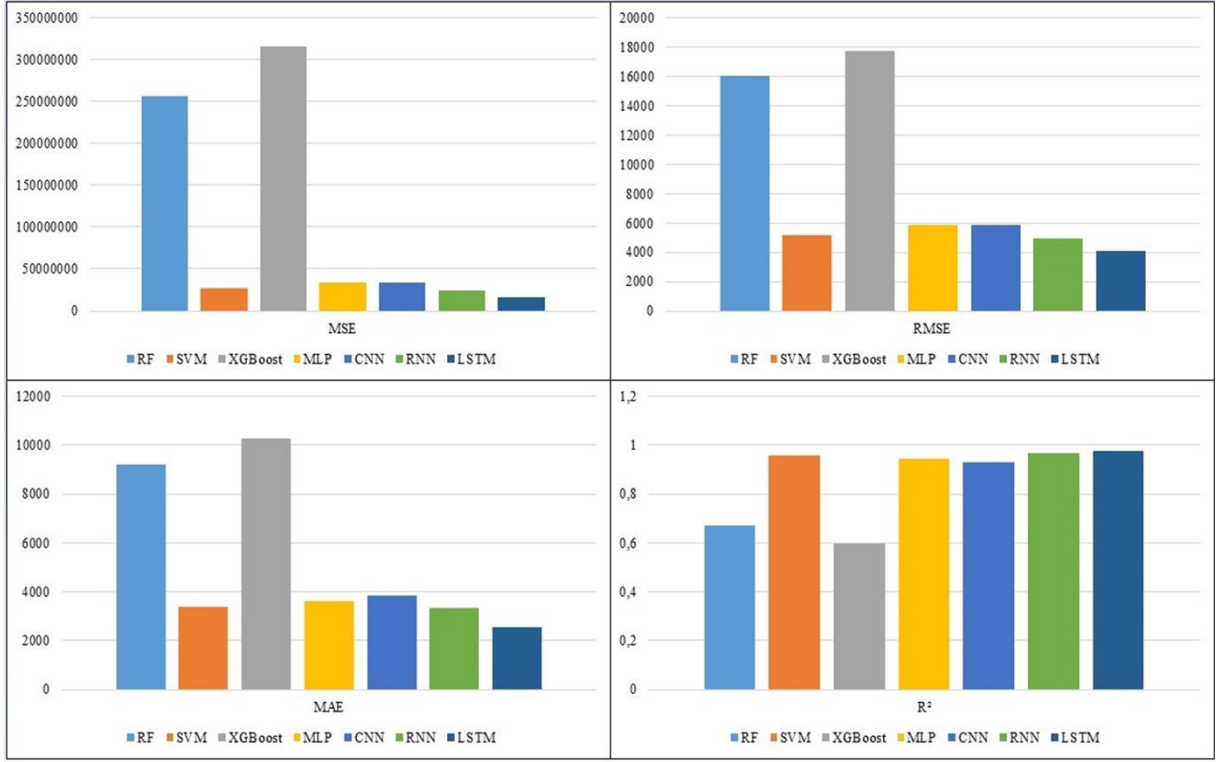
#### 4. Deneysel Sonuçlar

Bu çalışmada, Covid-19'un Türkiye'deki yayılım dağılımının tahminine yönelik RF, SVM, XGBoost, MLP, CNN, RNN ve LSTM modellerinin karşılaştırmalı bir analizi sunulmuştur. Uygulanan her bir algoritma ve model için MSE, RMSE, MAE ve R<sup>2</sup> metriklerine göre elde edilen sonuçlar karşılaştırmalı olarak analiz edilmiştir. Tablo 1 ve Şekil 4'te vaka sayısı tahmine yönelik karşılaştırmalı deneysel sonuçlar görülmektedir.

Tablo 1. Vaka sayısını tahminine yönelik deneysel sonuçlar

Modeller	Metrikler			
	MSE	RMSE	MAE	R <sup>2</sup>
RF	256806715,004	16025,190	9202,189	0,672
SVM	27080635,618	5203,906	3371,091	0,956
XGBoost	315752073,903	17769,414	10293,180	0,597
MLP	34066400,819	5836,642	3607,475	0,946
CNN	34253568,270	5852,655	3820,487	0,931
RNN	24346133,217	4934,180	3317,174	0,968
LSTM	16670823,040	4082,991	2543,651	0,975

Vaka sayısına sayı tahminine yönelik deneysel sonuçlar LSTM'in karşılaştırılan modellerden daha başarılı olduğunu göstermiştir. LSTM'in ardından RNN, SVM, MLP, CNN, RF ve XGBoost modelleri sırasıyla başarılı olmuştur.



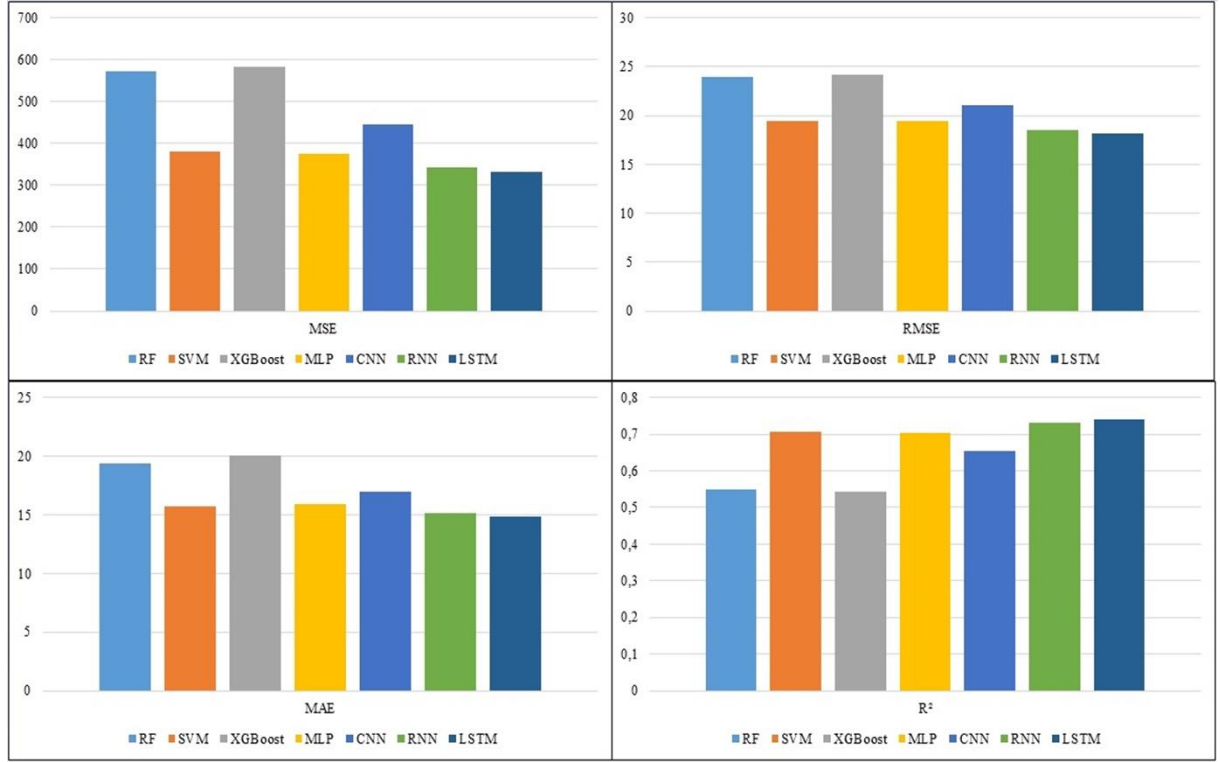
Şekil 4. Vaka sayısını tahminine yönelik deneysel sonuçlar

Tablo 2 ve Şekil 5'te ölüm sayısı tahmine yönelik karşılaştırmalı deneysel sonuçlar görülmektedir.

Tablo 2. Ölüm sayısını tahminine yönelik deneysel sonuçlar

Modeller	Metrikler			
	MSE	RMSE	MAE	R <sup>2</sup>
<b>RF</b>	572,090	23,918	19,371	0,550
<b>SVM</b>	379,332	19,476	15,690	0,705
<b>XGBoost</b>	582,543	24,135	20,032	0,542
<b>MLP</b>	375,899	19,388	15,883	0,704
<b>CNN</b>	444,557	21,084	16,938	0,653
<b>RNN</b>	342,931	18,518	15,141	0,731
<b>LSTM</b>	331,351	18,203	14,891	0,740

Ölüm sayısına sayı tahminine yönelik deneysel sonuçlar LSTM'in karşılaştırılan modellerden daha başarılı olduğunu göstermiştir. LSTM'in ardından RNN, SVM, MLP, CNN, RF ve XGBoost modelleri sırasıyla başarılı olmuştur.



Şekil 5. Ölüm sayısını tahminine yönelik deneysel sonuçlar

## 5. Sonuçlar

Covid-19 pandemisi, dünyanın her ülkesindeki insanların yaşamlarını ciddi şekilde etkilemiştir. Henüz kesin bir tedavisi olmayan bu pandeminin ciddiyeti gün geçtikçe artmaktadır. Virüsün dünya çapında yayılmasını kontrol altına almak için karantina uygulamaları gibi önlemler alınmaktadır. Ancak, enfeksiyonun hızlı ve öngörülemez yayılımı, sağlık kurumları ve çalışanları üzerinde ağır bir yük oluşturmuştur. Salgınla mücadele, kişisel hijyen ve sağlık kurumlarının belirlediği çeşitli kurallara uymayı, sağlık çalışanlarının fedakarlığını, toplumsal örgütlenmeyi ve teknolojik çözümleri gerektirir. Pandemi sürecinde hastanelerde yapılacak olan kapasite planlama çalışmaları, test miktarı ve kişisel koruyucu ekipman sayısı gibi pek çok problem ön plana çıkmaktadır. Bu sebeple pandemi sürecinde bu tür planlamaların yapılması ve ilgili tahmin modellerin oluşturulması büyük önem taşımaktadır.

Bu çalışmada, Covid-19 vaka ve ölüm sayılarının tahmini için makine öğrenmesi ve derin öğrenme tabanlı karşılaştırmalı bir analiz sunulmuştur. RF, SVM, XGBoost, MLP, CNN, RNN gibi popüler makine öğrenmesi ve derin öğrenme modelleri ile geliştirilen LSTM tabanlı derin öğrenme modeli karşılaştırılmıştır. Veri seti olarak 8 Nisan 2022 tarihine kadar olan ve DSÖ tarafından onaylanmış Türkiye'deki vaka ve ölüm sayıları kullanılmıştır. MSE, RMSE, MAE ve R<sup>2</sup> metrikleri kullanılarak deneysel çalışmalar yapılmıştır. Deneysel sonuçlar, uygulanan tüm hata metriklerine göre vaka ve ölüm sayıları için geliştirilen LSTM tabanlı modelin, karşılaştırılan diğer modellere göre daha iyi bir tahmin performansına sahip olduğunu göstermiştir.

## Kaynaklar

- [1] John CC, Ponnusamy V, Chandrasekaran SK, Nandakumar R. A survey on mathematical, machine learning and deep learning models for COVID-19 transmission and diagnosis. IEEE Rev. Biomed. Eng 2021; 15: 325-340.
- [2] Kreuz R ve diğerler. Hypertension, the renin-angiotensin system, and the risk of lower respiratory tract infections and lung injury: implications for COVID-19: European Society of Hypertension COVID-19 Task Force Review of Evidence. Cardiovasc. Res. 2020; 116(10): 1688-99.
- [3] Aguiar D, Lobrinus JA, Schibler M, Fracasso T, Lardi C. Inside the lungs of COVID-19 disease. Int. J. Legal Med 2020; 134(4): 1271-4.
- [4] Ebrahim SH, Ahmed QA, Gozzer E, Schlagenhauf P, Memish ZA. Covid-19 and community mitigation strategies in a pandemic. Bmj. 2020; 368.



- [5] Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Aff* 2020; 39(7): 1237-46.
- [6] Aleta A, Martin-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini Jr IM, Merler S, Pentland A. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav* 2020; 4(9): 964-71.
- [7] Heath C, Sommerfield A, von Ungern-Sternberg BS. Resilience strategies to manage psychological distress among healthcare workers during the COVID-19 pandemic: a narrative review. *Anaesthesia*. 2020; 75(10): 1364-71.
- [8] Melnick ER, Ioannidis JP. Should governments continue lockdown to slow the spread of covid-19?. *BMJ* 2020;369.
- [9] Punn NS, Sonbhadra SK, Agarwal S. COVID-19 epidemic analysis using machine learning and deep learning algorithms. *MedRxiv*. 2020 Jan 1.
- [10] Alazab M, Awajan A, Mesleh A, Abraham A, Jatana V, Alhyari S. COVID-19 prediction and detection using deep learning. *Int. J. Comput. Inf. Syst. Ind. Manage. Appl* 2020; 12:168-81.
- [11] Malki Z, Atlam ES, Ewis A, Dagnev G, Ghoneim OA, Mohamed AA, Abdel-Daim MM, Gad I. The COVID-19 pandemic: prediction study based on machine learning models. *Environ. Sci. Pollut. Res* 2021; 28(30): 40496-506.
- [12] Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Math* 2020; 8(6): 890.
- [13] Malki Z, Atlam ES, Hassanien AE, Dagnev G, Elhosseini MA, Gad I. Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons Fractals*. 2020; 138: 110137.
- [14] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons Fractals*. 2020; 140: 110212.
- [15] Prakash KB, Imambi SS, Ismail M, Kumar TP, Pawan YN. Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms. *Int. J. Emerging Trends Eng. Res*. 2020; 8(5), 2199-2204.
- [16] Satu M, Howlader KC, Mahmud M, Kaiser MS, Shariful Islam SM, Quinn JM., Moni, MA. Short-term prediction of COVID-19 cases using machine learning models. *Appl. Sci*, 2021; 11(9), 4266.
- [17] Kulkarni VY, Sinha PK. Pruning of random forest classifiers: A survey and future directions. In: 2012 International Conference on Data Science & Engineering (ICDSE); 18-20 July 2012; Cochin, India: IEEE. pp. 64-68.
- [18] Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Res* 2007; 35(suppl\_2): W339-44.
- [19] Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S. Random Forest based hourly building energy prediction. *Energy Build* 2018; 171: 11-25.
- [20] Zakariah M. Classification of large datasets using Random Forest Algorithm in various applications: Survey. *Int. J. Eng. Innovative Technol* 2014;4(3).
- [21] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst* 1997; 9: 155-161.
- [22] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat. Comput* 2004; 14(3): 199-222.
- [23] Zhang F, O'Donnell LJ. Support vector regression. *Machine Learning*, Academic Press, 2020.
- [24] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 13-17 August 2016; San Francisco California USA, pp. 785-794.
- [25] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev* 2021; 54(3): 1937-67.
- [26] Zhang Y, Chen L. A study on forecasting the default risk of bond based on XGBoost algorithm and over-sampling method. *Theor. Econ. Lett* 202; 11(02): 258.
- [27] Zhang C, Pan X, Li H, Gardiner A, Sargent I, Hare J, Atkinson PM. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens* 2018; 140: 133-44.
- [28] Windeatt T. Ensemble MLP classifier design. *Computational Intelligence Paradigms*, 2008: 133-147, Springer, Berlin, Heidelberg.
- [29] Car Z, Baressi Šegota S, Anđelić N, Lorencin I, Mrzljak V. Modeling the spread of COVID-19 infection using a multilayer perceptron. *Comput. Math. Methods Med* 2020.
- [30] Delashmit WH, Manry MT. Recent developments in multilayer perceptron neural networks. In Proceedings of the seventh Annual Memphis Area Engineering and Science Conference, MAESC 2005 May.
- [31] Yao G, Lei T, Zhong J. A review of convolutional-neural-network-based action recognition. *Pattern Recognit. Lett* 2019; 118: 14-22.
- [32] Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog Artif. Intell* 2020; 9(2): 85-112.
- [33] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst* 2012; 25.
- [34] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans* 1994; 5(2): 157-66.
- [35] Sutskever I, Martens J, Hinton GE. Generating text with recurrent neural networks. *InICML* 2011 Jan 1.