

Makine Öğrenmesi Yöntemleri ile Covid-19 Verilerinin İncelenmesi: Türkiye Örneği

An Analysis of Covid-19 Data With Machine Learning Methods: The Case of Turkey

Zeynep ÖZEL*¹, Mert DEMİRSÖZ²

Öz

Covid-19 virüsü günümüzde sağlığımızı ve yaşamımızı etkileyen en önemli sorunların başındadır. Bu virüsün normal bir hastada etkisinin yaklaşık olarak bir ay sürdüğü düşünülmektedir. Buradan yola çıkarak Türkiye Sağlık Bakanlığının açıkladığı günlük vakalar, vefatlar, iyileşenler, testler ve ağır hasta sayıları verilerinden yola çıkarak aylık bir değerlendirme yaparak virüsün aylık olarak seyrinin değerlendirilmesi amaçlandı. Araştırmada Mart 2020 ile Mart 2021 arasında açıklanan 12 aylık veri seti kullanıldı. Bu veri seti Makine Öğrenmesi sınıflandırma yöntemlerinden Random Forest algoritması ile analiz edildi. Analiz sonucunda yöntem kesinlik, duyarlılık, F ölçüsü ve AUC-ROC performans ölçütleri ile sınılandı. Ayrıca kullanılan değişkenlerin model için önemi değerlendirildi. Yapılan analizler sonucunda modelin doğruluğu (OOB) ise %83 olarak bulundu. Performans ölçütleri de kesinlik oranı %90, duyarlılık oranı %89, F ölçüsü %89 ve eğrinin altında kalan alan (AUC-ROC) %99 olarak bulundu. Modelinin doğruluğu için en önemli değişken günlük iyileşen sayısı iken sınıf belirlemedeki en önemli değişken günlük iyileşen sayısı olarak belirlendi.

Anahtar Kelimeler: Covid-19, Türkiye, Makine öğrenmesi, Rasgele orman, Sınıflandırma yöntemleri

Abstract

Covid-19 virus is one of the most important problems affecting our health and life today. It is considered that the effect of this virus for a normal patient continues approximately one month. Turkey's Health Ministry declared that the daily cases of death, recovering patient, tests and the number of seriously ill and these cases aimed to evaluate on a monthly basis the progress of virus. In our research were used to announced between March 2020 and March 2021 the data set twelf montly. This data set was analysed with the Random Forest algorithmn, which is one of the machine learning classification methods. As a result of the analysis, the method was tested with precision, recall, score F and AUC-ROC performance criteria. In addition, the importance of the variables used for the model was evaluated. As a result of the analysis, the accuracy (OOB) of our model was found to be 83%. Performance criteria were found to have an precision rate of 90%, a recall rate of 89%, an F score of 89%, and an area under the curve (AUC-ROC) of 99%. While the most important variable for the accuracy of the model was the daily healing number, the most important variable in determining the grade was the Daily Healing Number.

Keywords: Covid-19, Turkey, Machine learning, Random forest, Classification methods

Received / Geliş	20.03.2021
Accepted / Kabul	20.07.2021
Publication Date	18.08.2021

*Sorumlu Yazar
Corresponding Author

Zeynep ÖZEL
¹Selçuk Üniversitesi
Veterinerlik Fakültesi
Biyostatistik Anabilim Dalı,
Konya, Türkiye

✉ zozel4225@gmail.com
<https://orcid.org/0000-0002-1077-1250>

Mert DEMİRSÖZ
²KTO Karatay Üniversitesi,
Teknoloji Transfer Ofisi
<https://orcid.org/0000-0002-4800-2529>

Giriş

Geçmişten günümüze kadar tüm toplumları etkisi altına alan salgın hastalıkların yaşandığı bilinmektedir. 21. yüzyılın başında ortaya çıkan korona virüsler: 2002 yılında SARS, 2012 yılında MERS ve son olarak 2019 yılında Covid-19 virüsü ile küresel bir sağlık problemi haline gelmiştir. Covid-19 virüsü ilk olarak 2019 yılının Aralık ayında Çin'in Hubei eyaletine bağlı Wuhan şehrinde ortaya çıkmış olan virüs ağır akut solunumsal sorunlara yol açan bir RNA virüsüdür. İnsandan insana damlacık ve temas yoluyla bulaşabilen ve insanlarda ciddi hasara veya ölümcül sonuçlara neden olan bir virüs olduğu bilinmektedir (24).

Türkiye'de ilk olarak 11 Mart 2020 tarihinde görülen Covid-19 virüsü ilk vakanın ardından hızla tüm ülkeyi etkisi altına almış ve günümüzde de hala etkileri devam etmektedir. Sağlık Bakanlığı tarafından Türkiye'de virüsün yayılma hızı, hastanelerde yoğun bakım servislerinde tedavi gören, solunum cihazına bağlı olan, iyileşen hasta sayıları ile ölüm istatistikleri her gün periyodik olarak açıklanmaktadır (25).

Covid-19'un yaygın olarak görülen diğer solunum yolu enfeksiyonlarından ayrılan güvenilir bir klinik özelliği ve teşhis yöntemi yoktur. Yapılan bir çalışmada hastalığın ilk aşamasında en sık görülen semptomların %98'i ateş, %76'sı öksürük, %44'ü halsizlik veya yorgunluk ve atipik semptom olarak %28 balgam, %8 baş ağrısı, %5 hemoptizi, %3 ishal olduğu ve hastaların yaklaşık yarısında nefes darlığı gibi problemlerin yaşandığı görülmüştür (1-3). Yapılan çalışmalarda Covid-19 virüsünün etkisi normal hastalarda maksimum 15 gün kuluçka, 15 gün tedavi olmak üzere 1 aylık iyileşme sürecinden oluşmaktadır (4, 5).

Makine öğrenmesi öngörücü disiplinlerden biri olmasından dolayı mevcut verilerden hareketle tutarlı sonuçlar ortaya koymaktadır. Bu amaçla geliştirilen farklı algoritmalarından biri olan Random Forest yöntemi 2001 yılında Breiman tarafından öne sürülmüştür (22). Breiman'ın bu çalışması 1996 yılında kendisinin önerdiği Bagging yöntemi ile 1998'de Ho tarafından önerilen The Random Subspace tekniğinin bir kombi-nasyonu olarak oluşturmuştur (6,7). Ayrıca 1997'de tanımlanan, her düğüm için en iyi ayrımın rasgele bir seçim üzerinden belirlendiği belirtilen bir çalışmadan da etkilenmiştir (8). Hem regresyon hem de sınıflandırma problemlerine uygulanabilir olmasından dolayı popüler bir modelleme türüdür.

Random Forest yönteminin kullanılmış olduğu Rusya'da yapılmış bir çalışmada iklimsel faktörlerin virüsün yayılmasını etkilemiş olduğu bulunmuş. Araştırmaya göre mevsim sıcaklığı ve günlük sıcaklık aralığının COVID-19 iletiminin yoğunluğu bağlamında önemli olduğu rapor edilmiş. Çalışma bulgularına göre, bahar dönemindeki sıcaklık artışı nedeni ile COVID-19 bulaşma yoğunluğunun azalmasına neden olabileceği belirtilmiştir (13).

Endonezya'da yapılmış bir çalışmada, Nisan ve Eylül ayları arasında süren kurak mevsimde yüksek güneş ışığına

maruz kalma süresi ile iyileşme oranları arasında ilişki bulunduğu rapor edilmiş (23).

Çalışmanın temel hipotezi Türkiye'de Covid-19 verilerinin Random Forest yöntemi ile aylık olarak incelendiğinde pandeminin seyri ve mevsim normallerinde aylara göre değişen hava sıcaklıklarından nasıl etkilenmiş olduğunu ortaya koymaktır.

Gereç ve Yöntem

Veri seti tanımı

Veri seti oluşturulurken temelde beş farklı öznelik ele alınmıştır. Bunlar: Sağlık Bakanlığının açıkladığı günlük vakalar, vefatlar, iyileşenler, testler ve ağır hasta sayılarıdır (21). Bu veriler Sağlık Bakanlığı tarafından oluşturulmuş veriler olup buradan elde edilmiştir. Her gün kaydedilen bilgiler, 11 Mart 2020 ile 12 Mart 2021 arasındaki zamanı içermektedir.

Sağlık Bakanlığı tarafından açıklanan hem laboratu-var onaylı hem de klinik olarak teşhis edilmiş vakalar kullanılmıştır. Onaylanmış vakalar veya ölümler kadar medyada yer almayan iyileşmiş vakaların, yapılan testlerin ve ağır hasta sayılarının da büyük bir önemi mevcuttur. Çünkü verilerin doğruluğu medyada oluşmuş veriler değil, hükümetin resmi organlarından açıklanan verilerdir.

Random forest sınıflama algoritması

Araştırmada veri madenciliği alt çalışma alanlarından olan makine öğrenmesinin Random Forest sınıflandırma algoritması kullanılmıştır.

Random Forest algoritması hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. Algoritma sınıflama için şu şekilde çalışır: Sınıflandırıcı girdi özelliğindeki vektörü alır, ormandaki her ağaçla sınıflandırır ve önermenin çoğunluğunu alan sınıf etiketini çıkarır. Bir regresyon durumunda sınıflandırıcının cevabı ormandaki tüm ağaçlar üzerindeki cevapların ortalamasıdır. Tüm ağaçlar aynı parametrelerle fakat farklı eğitim setleriyle eğitime alınmıştır. Eğitim-den geçen bütün ağaçların her bir düğümünde, bütün değişkenler en iyi bölünmeyi bulmaya değil bunların bir alt kümesini bulmak için kullanılır. Her düğümde yeni bir alt set oluşturulur. Rastgele ağaçlarda eğitim hatasının tahmini için ayrı bir test setine gerek yoktur. Çünkü hata eğitim esnasında tahmin edilmektedir (9).

Random Forest yönteminde, torbalama ve rastgele özellik seçimi birlikte kullanılır. Her yeni eğitim seti, orijinal eğitim setinden iadeli olarak çekilir. Ardından rastgele özellik seçimi kullanılarak yeni eğitim setinde bir ağaç yetiştirilir. Yetiştirilen ağaçlarda budama yapılmaz. Yapılan çalışmalar, özellik seçimi ölçütlerinin değil, budama yöntemlerinin seçiminin ağaç tabanlı sınıflandırıcıların performansını etkilediğini göstermektedir. Torbalamanın kullanılmamasının iki nedeni vardır: Birincisi, rastgele özellik seçimi kullanıldığında torbalama kullanımının doğruluğu arttırdığı görülmektedir. İkincisi, torba dışı hataların (OOB)

hesaplanmasıdır. Budamanın olmaması rastgele ormanları diğer karar ağacı yöntemlerinden daha uygun hale getirmiştir (10).

Model sınıflaması için performansı değerlendiren ve başarıyı ölçen birtakım ölçütler vardır. Bu ölçütlerin hesaplanmasında hata matrisi kullanılmaktadır. Hata matrisi, tahminlerin doğruluğu hakkında bilgi veren 2x2'lik bir matristir. Model performansını ölçmek için tahmin değerleri

ile gerçek değerleri karşılaştırarak bu matris oluşturulur. Bu matris Tablo 1'deki gibidir (11).

Sınıflandırma işlemini yapan sistemin pozitif sınıf etiketlerinin tahmin etmedeki etkililiği duyarlılık olarak tanımlanmaktadır. Duyarlılık, doğru sınıflandırılan pozitif örneklerin toplam pozitif örnek sayısına oranıdır (11).

Tablo 1: Hata Matrisi

		Tahmin Edilen Sınıf	
		Pozitif (P)	Negatif (N)
Gerçek Sınıf	Doğru (T)	TP (Doğru tahmin edilen pozitif sınıf)	FN (Yanlış Tahmin Edilen Negatif Sınıf)
	Yanlış (F)	FP (Yanlış Tahmin Edilen Pozitif Sınıf)	TN (Doğru Tahmin Edilen Negatif Sınıf)

Tablo 2: Türkiye için aylara göre Performans Ölçütleri

Aylar	Kesinlik	Duyarlılık	F Ölçüsü	AUC
Mart 2020	0,99	0,99	0,99	0,99
Nisan 2020	0,99	0,99	0,99	0,99
Mayıs 2020	0,99	0,88	0,93	0,99
Haziran 2020	0,80	0,57	0,67	0,98
Temmuz 2020	0,57	0,99	0,73	0,99
Ağustos 2020	0,99	0,99	0,99	0,99
Eylül 2020	0,50	0,33	0,40	0,99
Ekim 2020	0,71	0,83	0,77	0,99
Kasım 2020	0,99	0,99	0,99	0,99
Aralık 2020	0,99	0,99	0,99	0,99
Ocak 2021	0,99	0,99	0,99	0,99
Şubat 2021	0,86	0,99	0,92	0,99
Mart 2021	0,99	0,83	0,91	0,98
Model	0,90	0,89	0,89	0,99

Tablo 3: Değişken Önem Değerleri

	Doğruluk Önemi	Sınıf ayırma önemi
Günlük Hasta Sayısı	0.058	0.536
Ağır Hasta Sayısı	0.364	0.293
Günlük Test Sayısı	0.362	0.272
Günlük Vefat Sayısı	0.223	0.250
Günlük İyileşen Sayısı	0.541	0.008

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

Sınıflandırma işlemini yapan sistemin negatif sınıf etiketlerinin tahmin etmedeki etkililiği özgüllük şeklinde tanımlanmaktadır. Özgüllük, doğru sınıflandırılan negatif örneklerin toplam negatif örnek sayısına oranıdır (12).

$$\text{Özgüllük} = \frac{TN}{TN + FP}$$

Doğruluk oranı: Sınıflandırma modellerini değerlendirmek için de bir metrik yöntemdir. Doğruluk oranı testi ile yapılan sınıflandırma modelinin ne kadar verimli olduğunun da tespiti yapılmış olur. Doğruluk oranı yüzde olarak ne kadar yüksek çıkarsa sınıflandırma modelinin de o kadar iyi çalıştığı anlaşılır (11).

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik oranı: Doğru tahmin edilen pozitif sınıf değerinin, bütün pozitif olarak tahmin edilen sınıf değerlerine oranıdır (11).

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

F1 ölçüsü: Duyarlılık ile kesinlik ölçütlerini beraber değerlendirilebilmek için kullanılan F-ölçütü, bu iki ölçütün harmonik ortalamasıdır (11).

$$F \text{ ölçüsü} = 2 \times \frac{\text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}}$$

Modelimizin doğruluğunu tespit etmek için ROC (Receiver Operating Characteristic Curve- Alıcı İşletim Karakteristik Eğrisi) eğrisi kullanılır bu eğri oluşturulurken eğrinin altında kalan alanı belirlemek için AUC (Area Under Curve) değeri hesaplanır. Bu değer modelin sınıfları ne kadar başarılı ayırt edebildiğini anlatır. AUC-ROC değeri 1'e ne kadar yakınsa modelin performansı o kadar iyi denilebilir. Bu değer alabileceği en düşük değer ise 0,5'tir (11).

ROC eğrisi çizilirken y ekseninde duyarlılık, x ekseninde ise (1- Özgüllük) değerleri kesişim noktası belirlenerek grafik çizilir (11).

Yukarıda verilen değerler tüm sınıflar için belirlenerek sonuçlar tüm sınıflara göre değerlendirilir. Bu değerler sayesinde model ile ilgili değerlendirme yapılacaktır.

Random Forest sınıflama algoritmasında değişkenler için doğruluktan ortalama düşüş (mean decrease in accuracy) ve

düğüm saflığında toplam artış (total increase in node purity) olarak iki farklı değer hesaplanmaktadır. Bu değerler 1'e yaklaştıkça değişken önemi artmakta azaldıkça değişken önemi azalmaktadır. Yapılan tahmin ile en iyi sınıflamayı yapmasında en çok yardımcı olan değişkenin belirlenmesi sağlanır. Bunun sonucunda değişkenin model için kullanılabilirliği incelenir (9).

Veri analizi

Random Forest sınıflandırma algoritması için maksimum 100 ağaç oluşturulmuş ve bu 100 ağacın %50'si eğitim seti için kullanılmıştır. Bu eğitim setinde kullanılacak ağaçların her biri bir tahmin üreterek algoritmayı eğitmiştir. Eğitilen algoritma sonucunda tahminler oluşturulmuştur. Tahminler veri setindeki sonuç çıktılarının %20'si kullanılarak test edilmiştir. Buradan yola çıkarak veri setinde kullanılan değişkenler bağımsız olsa da test seti ile eğitim seti birbirinden bağımsız değildir. Bu ilişkiden yararlanarak doğruluk oranları oluşturulmuştur. Test seti ile eğitim setinin kesiştiği optimum doğruluk oranı bize modelimizin doğruluk oranını (Out Of Bag-OOB) ve en verimli ağacı belirlememizi sağlamıştır. Eğitilen verinin ise %20'si doğrulama için kullanılmıştır. Elde edilen sonuçlar doğrultusunda modelin doğruluk eğrisi çizilmiştir. Modelin performansını değerlendirmek için Kesinlik, Duyarlılık, F Ölçüsü ve AUC-ROC performans ölçütleri kullanılmış ve bu ölçütler yardımı ile ROC eğrisi çizilmiştir. Son olarak modelde kullanılan değişkenlerin önemleri değerlendirilmiştir. Model hakkında genel bir değerlendirme yapabilmek için karar sınır matrisleri kullanılmıştır. Bu matristeki her hücre kesişen iki sınıf hakkında değerlendirme yapması sağlanmıştır.

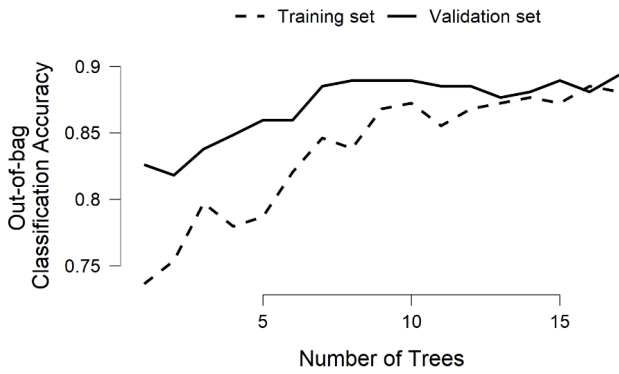
Analizimiz Jasp (JASP Team (2020). JASP (Version 0.14) [Computer software]) programı kullanılarak gerçekleştirildi. Jasp programında analizin nasıl yapılacağı eklerde görsel olarak verilmiştir.

Bulgular

Şekil 1'de araştırmada değişkenler arasındaki ilişkiler incelendiğinde en yüksek ilişkinin günlük test sayısı ile ağır hasta sayısı arasında %86, daha sonra günlük hasta sayısı ile ağır hasta sayısı arasında %80 ilişki bulunmuştur. En düşük ilişki ise %21 ile günlük hasta sayısı ile günlük iyileşen sayısı arasındadır. İncelenen tüm değişkenler arasında istatistiksel olarak anlamlı ilişkiler vardır.

Random Forest yöntemi ile oluşturulan modelde maksimum 100 ağaç kullanılmıştır. Bu ağaçlardan en verimli ağacın 17. ağaç olduğu tespit edilmiştir. Eğitilen model test aşamasında maksimum %89, doğrulama aşamasında ise maksimum %93 doğruluğa ulaşmıştır. Modelin doğruluğu (OOB) ise %83 olarak bulunmuştur. Modelin doğruluk grafiği Grafik 1'de verilmiştir.

Modelin performansını değerlendirmek için Tablo 2' de kesinlik, duyarlılık, F ölçüsü ve AUC-ROC performans ölçütleri verilmiştir.



Grafik 1: Model Doğruluk Grafiği

Kesinlik değeri modelin %90 doğru sınıflandırıldığını göstermektedir. Duyarlılık oranı ise eğitim veri setinin %89 ile gerçek sonuçları temsil ettiğini göstermektedir. F ölçüsü ise kesinlik be duyarlılık değerlerinin harmonik ortalaması ile bulunduğu ve metriklerin doğruluğunu göstermektedir. Metrikler %89 doğruluğa sahiptir. Eğrinin altında kalan alan (AUC-ROC) %99 olarak bulunmuştur. Modelin performansının aylara göre değerlendirilmesi sonucunda elde edilen metrikler Tablo 2' de detaylı olarak verilmiştir.

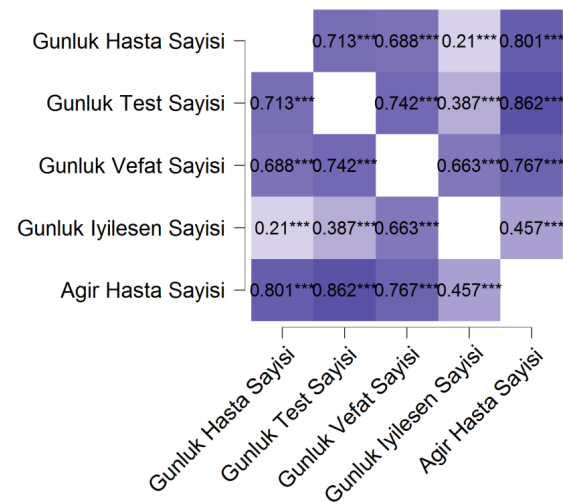
Şekil 2' de çalışmadaki aylara ait ROC eğrileri çizilmiştir. Tablo 2' de verilen AUC-ROC değerleri ROC eğrisinin altında kalan alanı belirtmektedir. Buradan yola çıkarak Random Forest sınıflandırma yönteminin aylara göre çok iyi bir sınıflandırma yaptığı görülmektedir.

Tablo 3'e göre doğrulukdaki ortalama düşüş, her bir özelliğin Random Forest modelinin doğruluğu üzerindeki etkisini ölçmek için değerlendirilmiştir. Bunun sonucunda Random Forest modelinin doğruluğu için en önemli değişken günlük iyileşen sayısı olarak belirlenmiştir. Sınıf ayırma önemi ise hangi değişkenin sınıfları belirlemede daha önemli olduğunu göstermektedir. Bunun sonucunda sınıf belirlemedeki en önemli değişken günlük iyileşen sayısı olarak belirlenmiştir.

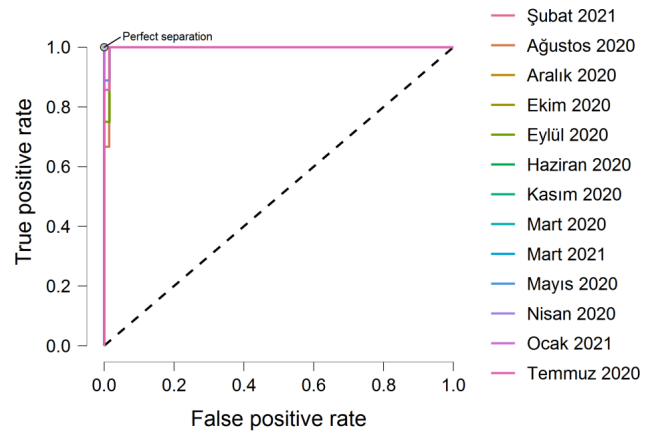
Şekil 3' e göre günlük hasta sayısı ile günlük test sayısı birlikte incelendiğinde; günlük test sayısı en çok Ocak 2021, Şubat 2021, Mart 2021 aylarında, günlük hasta sayısı en çok Ekim 2020, Kasım 2020, Aralık 2020 aylarında artmıştır.

Günlük hasta sayısı ile günlük vefat sayısı birlikte incelendiğinde; günlük test sayısı en çok Ocak 2021, Şubat 2021, Mart 2021 aylarında, günlük vefat sayısı ise en çok Ekim 2020, Kasım 2020, Aralık 2020 aylarında Nisan 2020, Mayıs 2020 aylarında da artmıştır.

Günlük iyileşen sayısı ile günlük test sayısı birlikte incelendiğinde; Aralık 2020'de hem günlük test sayısı hem de günlük iyileşen sayısı artmıştır. Ocak 2021 ve Şubat 2021'de artış daha çok günlük test sayısında olmuştur. Diğer tüm aylarda günlük iyileşen sayısında ciddi bir artış olmamıştır.



Şekil 1: Değişkenler arasındaki ilişkiler

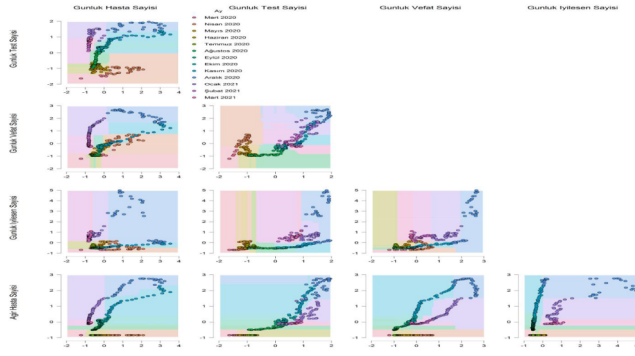


Şekil 2: Random Forest Yöntemi için ROC Eğrileri

Ağır hasta sayısı ile günlük iyileşen sayısı birlikte incelendiğinde; Eylül 2020, Ekim 2020, Kasım 2020' de ağır hasta sayısı en yüksek seviyeye ulaşmıştır. Aralık 2020' de ise hem ağır hasta sayısı hem de günlük iyileşen sayısı yüksektir. Ocak 2021 ve Şubat 2021' de ağır hasta sayısı daha yüksek seviyededir.

Tartışma

Covid-19 salgını, önemli küresel sağlık ve ekonomik zorluklara neden olmuştur. Bu çalışmada Türkiye Covid-19 verileri Random Forest sınıflama algoritması ile aylık olarak sınıflayıp virüsün aylara göre durumları değerlendirilmiştir. Çalışmada hava sıcaklıkları değişken olarak ele alınmamıştır. Hava sıcaklıklarının mevsim normallerinde seyrettiği düşünüldüğünde aylara göre temel alınarak incelenmiştir. Sonuç olarak oluşturduğumuz modelin doğruluğu %83 olarak bulunmuştur. Kesinlik, duyarlılık, F ölçüsü ve AUC-ROC performans ölçütleri değerlendirilmiş ve bu değerlendirme sonucunda kesinlik oranı %90, duyarlılık oranı %89, F ölçüsü %89 ve eğrinin altında kalan alan (AUC-ROC) %99



Şekil 3: Türkiye için aylara göre karar sınır matrisi

olarak bulunmuştur. Değişken önemleri açısından modelin doğruluğu için en önemli değişken günlük iyileşen sayısı iken sınıf belirlemedeki en önemli değişken günlük iyileşen sayısı olduğu belirlenmiştir.

Buradan yol çıkarak Covid-19 modelinin aylık olarak çok yüksek derecede sınıflandırılma oranına (%83) sahip olduğu belirlenmiştir. Bu oranın yüksek olması, verilerin aylık olarak farklı etkilere sahip olduğunu göstermektedir. Ayrıca Rusya'da yapılan bir çalışmada Rusya'daki bölgelerdeki biyoklimatik değişkenler ve iklim değişkenleri ile Covid-19 vakaları arasındaki bağlantıları bulunmuş. Bu çalışmada mevsimsel durumların Covid-19 üzerinde anlamlı bir etkisinin olduğu tespit edilmiş. Ayrıca nemli bölgelerde sıcaklık ve rüzgâr hızı ile Covid-19 yoğunluğu arasında pozitif bir ilişki olduğunu tespit edilmiş. Mevsimsel olaylar aylara göre değişiklik gösterdiği için bu sonuçlar bu çalışmadaki bulguları desteklemektedir (13). Öte yandan, yapılan diğer bir çalışmada altı bölgede, günlük sıcaklık aralığı, nem ve sıcaklığın tümü, Covid-19'un yayılması üzerinde olumlu bir etkiye sahip olduğu belirlenmiş. İklim ilişkilerinin sonuçları önceki çalışmalarla tutarlıdır (14). Ek olarak, Covid-19 vakalarının ve ölümlerinin sıcaklık dalgalanmasından etkilendiğini ve daha düşük sıcaklıkla önemli bir ilişkinin bulunduğu tespit edilmiş (15, 16, 17). Sıcaklığın Covid-19 vakalarının üzerindeki etkisini inceleyen çalışmalar da vardır (18, 19). Bazı çalışmalar hem sıcaklığın hem de soğukun etkilerinin solunum hastalıklarını olumsuz etkileyebileceği kanısındadır. Ayrıca 30 doğu Asya şehrinde artan günlük sıcaklık aralığı riskinin daha yüksek kardiyovasküler hastalık riski ile bağlantılı olduğu bulunmuş (19, 20). Bu sonuçlara göre değerlendirilmesi bize yeni bir bakış açısı sağlayacaktır.

Sonuç

Bu araştırmalar ışığında tüm solunum yolu rahatsızlıklarında olduğu gibi Covid-19 virüsünün de mevsimler ve sıcaklıklar ile arasında anlamlı ilişkiler vardır. Aylara göre virüsün etkilerinin farklı olduğu çalışmamız sonucunda bulunmuştur. Random Forest sınıflama algoritması bu değerlendirmeyi yapmada oldukça kullanışlı bir makine öğrenme algoritmasıdır.

Artan Covid-19 vakalarına ve ölümlerine hızlı yanıt vermek için iklim temelli, hava sıcaklıklarının göz önüne alınması ve erken uyarı sistemlerinin geliştirilmesi Covid-19 bulaşma riskini azalmak için önemli bir yol olabilir. Bu ilişkilerin daha iyi anlaşılması, Covid-19 pandemisini kontrol etmek ve önlemek için temel bir unsur olan daha iyi iklime duyarlı müdahaleler yoluyla hızla büyüyen vakaları ve ölümleri kontrol etmek için kritik öneme sahiptir.

Çıkar çatışması: yok

Finansal destek: yok

Yazar katkıları:

Motivasyon / Konsept: ZÖ, MD

Çalışma Tasarımı: ZÖ

Kontrol / Gözetim: ZÖ, MD

Veri Toplanması ve/veya İşlemesi: MD

Analiz ve/veya Yorum: MD

Literatür inceleme: ZÖ, MD

Makalenin Yazılması: ZÖ

Eleştirel İnceleme: ZÖ, MD

KAYNAKLAR

1. Wu D, Wu T, Liu Q, Yang Z. The Sars-Cov-2 outbreak: what we know, *International Journal of Infectious Diseases*. 2020 March; 94:44-48. doi: 10.1016/j.ijid.2020.03.004.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020 January; 395:10223:497-506. doi:10.1016/S0140-6736(20)30183-5.
3. Uğraş-Dikmen A, Kına HM, Özkan S, İlhan MN. Covid-19 epidemiyolojisi: Pandemiden ne öğrendik. *Journal of Biotechnology and Strategic Health Research*. 2020 April; 4: 29-36. doi: 10.34084/bshr.715153.
4. Atalay S, Ersan G. Covid-19 Tedavisi. *Tepecik Eđit. ve Arařt. Hast. Dergisi*. 2020 Temmuz; 30: 126-134. doi:10.5222/terh.2020.48030.
5. Sağlık Bakanlığı. Covid-19 Bilgilendirme Platformu İnkübasyon süresi [İnternet]. Türkiye Cumhuriyeti Sağlık Bakanlığı; 2020 Temmuz [Eriřim 2021 Mart 13]. Eriřim adresi: <https://covid19.saglik.gov.tr/TR-66474/inkubasyon-suresi.html>
6. Breiman L. Bagging predictors. *Machine learning*. 1996 August; 24.2: 123-140.
7. Ho TK. The random subspace method for constructing decision forests, *IEEE transactions on pattern analysis and machine intelligence*. 1998 August; 20.8: 832-844.
8. Amit Y, Geman D, Wilder K. Joint induction of shape features and tree classifiers. *IEEE transactions on pattern analysis and machine intelligence*. 1997 November; 19.11: 1300-1305.
9. Demirsöz M, Özel Z, Yonar H et al. 2020 Classification of Covid-19 Case Numbers for G8 Countries and Turkey by Random Forest Method, *Medicres-2020 7th International Multidisciplinary Congress On Good Medical, Research Conference Abstracts and Proceedings Book*. pp;52-53.
10. Palczewska A, Palczewski J, Robinson RM, Neagu D. Interpreting random forest classification models using a feature contribution method, In: *Integration of reusable systems*. Springer, Cham. 2014 February; 193-218.
11. Karlı ÖB. Makine öğrenme yöntemleri ile karaciđer hastalığının teşhisi [Yüksek lisans tezi]. Ağrı İbrahim Çeçen Üniversitesi; 2019.
12. Khan SUR, Islam MA, Aleem M, Iqbal M.A. Temporal specificity-based text classification for information retrieval, *Turkish Journal of Electrical Engineering & Computer Sciences*. 2018 November; 26.6: 2915-2926. doi:10.3906/elk-1711-136.
13. Pramanik M, Udmale P, Bisht P et al. Climatic factors influence the spread of Covid-19 in Russia, *International journal of environmental health research*. 2020 July; 1-15. doi:10.1080/09603123.2020.1793921.
14. Pramanik M, Chowdhury K, Rana MJ, et al. Climatic influence on the magnitude of Covid-19 outbreak: a stochastic model-based global analysis, *Medrxiv*. 2020 October; 1-17 doi:10.1101/2020.06.02.20120501.
15. Fallah GG, Mayvaneh F. Effect of air temperature and universal thermal climate index on respiratory diseases mortality in Mashhad, Iran. *Arch Iran Med*. 2016 September; 19(9):618-624. doi:10.161909/AIM.004.
16. Gomez-Acebo I, Llorca J, Dierssen T. Cold-related mortality due to cardiovascular diseases, respiratory diseases and cancer: a case-crossover study, *Public Health*. 2013 March; 127(3):252-258. doi:10.1016/j.puhe.2012.12.014.
17. Dadbakhsh M, Khanjani N, Bahrapour A, Haghighi PS. Deaths from respiratory diseases and temperature in Shiraz, *Int J Biometeorol*. 2017 July; 61(2):239-246. doi:10.1007/s00484-016-1206-z.
18. Xie J, Zhu Y. Association between ambient temperature and Covid-19 infection in 122 cities from China. *Sci Total Environ*. 2020 July; 724:138201. doi:10.1016/j.scitotenv.2020.138704.
19. Li M, Zhou M, Yang J et al. Temperature, temperature extremes, and cause-specific respiratory mortality in China: a multi-city time series analysis, *Air Qual Atmos Health*. 2019 January; 12(5):539-548. doi:10.1007/s11869-019-00670-3.
20. Kim J, Shin J, Lim Y et al. Comprehensive approach to understand the association between diurnal temperature range and mortality in East Asia, *Sci Total Environ*. 2016 January; 539:313-321. doi:10.1016/j.scitotenv. 2015.08. 134.
21. Tekindal MA, Yonar H, Yonar A, et al. Analyzing Covid-19 outbreak for Turkey and Eight Country with Curve Estimation Models, Box-Jenkins (ARIMA), Brown Linear Exponential Smoothing Method, Autoregressive Distributed Lag (ARDL) and SEIR Models *Eurasian J Vet Sci, Covid-19 Special Issue*. 2020 November; 142-155.
22. Breiman L. Random forests. *Machine learning*. 2001 October; 45(1), 5-32.
23. Asyary A, Veruswati M. Sunlight exposure increased Covid-19 recovery rates: A study in the central pandemic area of Indonesia, *Science of The Total Environment*. 2020 August; 729, 139016. doi:10.1016/j.scitotenv.2020.139016.
24. World Health Organization. Coronavirus disease (Covid-19) pandemic [İnternet]. [Eriřim 2021 Mart 13]. Eriřim adresi: <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov>
25. Sağlık Bakanlığı. Covid-19 Bilgilendirme Platformu [İnternet]. Türkiye Cumhuriyeti Sağlık Bakanlığı; [Eriřim 2021 Mart 13]. Eriřim adresi: <https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html>