T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

SALGIN PARAMETRELERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TAHMİN EDİLMESİ

Adnan KEÇE

YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

DANIŞMAN
Dr. Öğr. Üyesi Faruk SERİN

TUNCELİ – 2021

T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

SALGIN PARAMETRELERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TAHMİN EDİLMESİ

Adnan KEÇE
(190080017)

YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

DANIŞMAN
Dr. Öğr. Üyesi Faruk SERİN

TUNCELİ – 2021

T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

SALGIN PARAMETRELERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TAHMİN EDİLMESİ

Adnan KEÇE
YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

Bu tez 11/06/2021 tarihinde aşağıdaki jüri üyeleri tarafından **oybirliği** ile kabul edilmiştir.

İmza	**İmza**	**İmza**
Doç. Dr. Muhammet GÜL (Munzur Üniversitesi)	Dr. Öğr. Üyesi Faruk SERİN (Gaziantep Üniversitesi)	Dr. Öğr. Üyesi Bülent HAZNEDAR (Gaziantep Üniversitesi)
BAŞKAN	**DANIŞMAN**	**ÜYE**

Bu tez, Enstitümüz Mühendislik Yönetimi Anabilim Dalı'nda hazırlanmıştır.

Doç. Dr. Murat KORUNUR
Enstitü Müdürü
İmza ve Mühür

NOT: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı "Fikir ve Sanat Eserleri Kanunu'ndaki hükümlere tabidir.

11/06/2021

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

İmza
Adnan KEÇE

Danışman
Dr. Öğr. Üyesi Faruk SERİN

TEŞEKKÜR

Bu çalışmanın yürütülmesi sırasında desteğini esirgemeyen danışmanım Dr. Öğr. Üyesi Faruk Serin'e, yoğun çalışmalarım sırasında sabır gösterdikleri ve bana katlandıkları için eşim Şeyma'ya, oğlum Hüseyin'e ve desteklerinden dolayı arkadaşım Yiğit'e teşekkür ederim.

Adnan KEÇE
TUNCELİ-2021

İÇİNDEKİLER

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ I
TEŞEKKÜR II
İÇİNDEKİLER III
ŞEKİLLER LİSTESİ IV
TABLOLAR LİSTESİ V
SEMBOLLER LİSTESİ VI
KISALTMALAR LİSTESİ VII
ÖZET VIII
ABSTRACT IX
1. GİRİŞ 1
1.1. Literatür Taraması 3
1.2. Tezin Katkısı ve Organizasyonu 7
2. MATERYAL VE METOT 8
2.1. Verilerin Toplanması ve Düzenlenmesi 9
2.2. Kullanılan Python Kütüphane Modülleri 9
2.3. Salgın Modelleme Yöntemleri 11
2.3.1. SI/SIS modeli 11
2.3.2. SIR/SIRS modeli 12
2.3.3. SEIR modeli 13
2.4. Eğri Uyarlama Yöntemleri 14
2.4.1. L-BFGS-B yöntemi 15
2.4.2. Powell yöntemi 16
2.4.3. CG yöntemi 16
2.4.4. COBYLA yöntemi 17
2.4.5. SLSQP yöntemi 17
2.5. Destek Vektör Makinesi (DVM) 18
2.5.1. Doğrusal DVM 19
2.5.2. Doğrusal olmayan DVM 20
2.6. LSTM 21
2.7. Performans Ölçütü 23
3. BULGULAR 25
3.1. SEIR Modeli Tahmin Sonuçları 25
3.2. Makine Öğrenmesi Yöntemleri Tahmin Edilmesi 28
3.2.1. DVM yöntemi ile tahmin sonuçları 29
3.2.2. LSTM yöntemi tahmin sonuçları 30
4. SONUÇ VE ÖNERİLER 34
4.1. Öneriler 34
4.2. Kısıtlamalar ve Zorluklar 35
4.3. Potansiyel Gelecek Çalışmalar 35
5. KAYNAKLAR 36
ÖZGEÇMİŞ 41

ŞEKİLLER LİSTESİ

Şekil 2.1. Akış şeması 9
Şekil 2.2. SI modeli 12
Şekil 2.3. SIR modeli 12
Şekil 2.4. SEIR modeli 13
Şekil 2.5. İki sınıflı bir problem için hiper-düzlemler ve destek vektörleri 19
Şekil 2.6. Doğrusal olarak ayrılabilen veri setleri için hiper-düzlemin belirlenmesi 20
Şekil 2.7. Doğrusal ayrılamayan veri seti ve bunlar için hiper-düzlemin belirlenmesi 21
Şekil 2.8. Standart Bir RNN 'de ki Yinelenen Modül 22
Şekil 2.9. Bir LSTM'deki Yinelenen Modül 22
Şekil 2.10. Şekil 2.9'da kullanılan şekillerin tanımı 23
Şekil 3.1. Mısır (a) ve Güney Kore'ye (b) ait "Duyarlı Kişi Sayısının" SEIR modeli ile tahmin sonuçları 26
Şekil 3.2. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" SEIR modeli ile tahmin sonuçları 27
Şekil 3.3. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" SEIR modeli ile tahmin sonuçları 27
Şekil 3.4. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" SEIR modeli ile tahmin sonuçları 28
Şekil 3.5. Mısır (a) ve Güney Kore'ye (b) ait "Duyarlı Kişi Sayısının" DVM yöntemi ile tahmin sonuçları 29
Şekil 3.6. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" DVM yöntemi ile tahmin sonuçları 29
Şekil 3.7. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" DVM yöntemi ile tahmin sonuçları 30
Şekil 3.8. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" DVM yöntemi ile tahmin sonuçları 30
Şekil 3.9. Oluşturulan LSTM modeli 31
Şekil 3.10. Mısır (a) ve Güney Kore'ye (b) ait "Duyarlı Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları 31
Şekil 3.11. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları 32
Şekil 3.12. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları 32
Şekil 3.13. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları 32

TABLOLAR LİSTESİ

Tablo 3.1. Optimizasyon algoritmalarına ait sonuçların MAPE değerleri 26
Tablo 3.2. Parametrelerin geçiş oranları ve R0 değerleri 26
Tablo 3.3. Mısır'a ait sonuçlar 33
Tablo 3.4. Güney Kore'ye ait sonuçlar 33

SEMBOLLER LİSTESİ

f	: Kriterler
β	: Enfeksiyon oranını
γ	: Ortalama gecikme süresi oranı
δ	: Ölüm oranı
κ	: İyileşme hızı
k	: Hessian matrisi yaklaşma adım katsayısı
∇	: Gradyan sembolü
x_i	: Veri kümesi noktaları
y_i	: Sınıf etiketleri
w	: Ağırlık vektörü
i	: Kriter sıra sayısı
b	: Eğilim değeri
ξ	: Yapay değişken

KISALTMALAR LİSTESİ

ANFIS	: Adaptive Neuro-Fuzzy Inference System
CG	: Conjugate Gradient Method
COBYLA	: Constrained Optimization by Linear Approximation
DVM	: Destek Vektör Makineleri
LSTM	: Long Short Term Memory
L-BFGS-B	: Limitted**-**Broydon-Fletcher-Goldfarb-Shanno-Bound
MLP	: Multi-Layer Perceptron
MAPE	: Mean Absolute Percentage Error
SI	: Suspect-Infectious
SIS	: Suspect-Infectious-Suspect
SIR	: Suspect-Infectious-Recovery
SIRS	: Suspect-Infectious-Recovery-Suspect
SEIR	: Suspect-Exposed-Recovery-Suspect
SLSQP	: Sequential Least Squares Programming

ÖZET

Tarih boyunca farklı zamanlarda, çeşitli sebeplerle salgın hastalıklar meydana gelmiştir. Salgın hastalıklar can kaybı, sosyal ve ekonomik yapı bozulması gibi türlü sonuçlar doğurabilmektedirler. Bu nedenle, ülkeler salgının etkisini azaltmak için bir takım önlemler alırlar. Salgınla mücadelede alınan önlemlerin, potansiyel maliyetlerini ve faydalarını ölçmek için salgın modelleme yöntemleri ve optimizasyon algoritmaları kullanılmaktadır. Bu tez çalışmasında da günümüzde dünyayı etkisi altına almış olan COVID-19 salgınının modellenmesi ve parametrelerinin tahmin edilmesine yönelik çalışmalar yapıldı. Çalışmalarda Mısır ve Güney Kore'deki COVID-19 salgınına ait veriler kullanıldı. Deneysel uygulamalarda, ilk önce salgın modelleme yöntemi olarak yaygın bir kullanıma sahip olan SEIR modeli üzerinden, L-BFGS-B, Powell, CG, COBYLA ve SLSQP yöntemleri ile optimizasyonlar gerçekleştirildi. Ardından, hesaplanan R0 değeri, ölüm oranı ve salgın parametreleri arasındaki geçiş oranları gibi değerler kullanılarak tahminler yapıldı. Aynı veri setleri kullanılarak makine öğrenmesi yöntemlerinden LSTM ve DVM yöntemleriyle de tahminler yapıldı. SEIR, LSTM ve DVM yöntemlerinden elde edilen sonuçların hem MAPE metriği kullanılarak tahmin başarıları karşılaştırıldı hem de grafiklerle desteklenerek sonuçları değerlendirildi. Yapılan 24 farklı tahmin sonucunun %79'unun MAPE değeri 10'nun altında çıkmıştır.

Anahtar Kelimeler: COVID-19, makine öğrenmesi, SEIR, LSTM, DVM

ABSTRACT

Prediction of epidemic parameters using machine learning methods

Epidemics have occurred at different times throughout history for various reasons. Epidemics can cause various consequences such as loss of life, deterioration of the social and economic structure. Therefore, countries take some measures to reduce the impact of the epidemic. Epidemic modelling methods and optimization algorithms are used to evaluate the potential costs and benefits of the measures taken in the fight against the epidemic. In this thesis, studies were carried out to model the COVID-19 epidemic, which has affected the world today, and to predict its parameters. Data from the COVID-19 outbreak in Egypt and South Korea were used in the experiments. First, optimizations were carried out using the L-BFGS-B, Powell, CG, COBYLA, and SLSQP methods on the SEIR model, which is widely used as an epidemic modelling method. Then, predictions were made using calculated values such as the R0 value, mortality rate, and transition rates between epidemic parameters. Using the same datasets, predictions were also made using the machine learning methods LSTM and DVM. The prediction successes of the results obtained from SEIR, LSTM, and DVM methods were compared using the MAPE metric, and the results were evaluated with graphics. The MAPE value of 79% of the 24 different prediction results was below 10.

Keywords: COVID-19, machine learning, SEIR, LSTM, SVM

1. GİRİŞ

Sağlık, insan yaşamının sürdürülmesinde, yaşam kalitesinin yükseltilmesinde ve korunmasında özel bir öneme sahiptir. Sağlığın korunması ve geliştirilmesi kişinin öncelikle kendi sağlığını koruması ve sağlık bilincini geliştirmesi ile mümkündür. Ancak bazı durumlarda, salgın hastalıklar da sağlığımızı tehdit edebilmektedir. Madhav ve ark. (2017) salgın hastalıkları "Salgınlar, geniş bir coğrafi alanda hastalık ve ölüm oranını büyük oranda arttıran, önemli ekonomik, sosyal ve politik aksamalara neden olan büyük ölçekli bulaşıcı hastalıklardır" şeklinde tanımlamaktadırlar.

Aslan (2020) salgın hastalıklar üzerine yaptığı çalışmada salgını, kısa bir süre içinde belirli bir popülasyondaki çok sayıda insana hızla yayılması olarak tanımlamış ve salgın hastalıkları yayılma türlerine göre şu şekilde isimlendirmiştir. İlk olarak, ortaya çıktığı popülasyonda normalden fazla gözlemlenmesi halinde epidemi olarak tanımlanır. Eğer salgın geniş bir coğrafi bölgeye yayılmışsa endemik olarak tanımlanır. Son olarak, hastalık geniş coğrafyalara hatta tüm dünyaya yayılan, insanlar veya hayvanlarda hastalık ve ölümlere yol açıyorsa pandemi olarak tanımlanmaktadır.

1984 yılında Birleşmiş Milletlere bağlı olarak kurulmuş, Cenevre merkezli Dünya Sağlık Örgütü (DSÖ), toplum sağlığıyla ilgili uluslararası çalışmalar yapmaktadır. DSÖ salgın hastalıkları, yeni bir hastalığın dünya çapında yayılması olarak tanımlamaktadır. Aynı zamanda DSÖ kaynakları yaklaşık 20 çeşit pandemik ve epidemik hastalık tanımlamıştır. Bunlardan bazıları geçmişten günümüze Dünya'yı etkisi altına alan H1N1, SARS ve COVID gibi gribal salgınlarıdır (URL-1, 2020).

En yaygın salgın hastalık tiplerinden biri olan grip, insanlık tarihi boyunca birçok salgının temelinde yatan bulaşıcı ve akut solunum yolu hastalıklarına neden olan viral enfeksiyonlardır. Grip virüslerinin, hem tekrarlayan hem de hızla yayılan çoğu yaş grubunu etkileyebilen daha ciddi pandemilere neden olma yetenekleri de bulunmaktadır. Salgınların ve pandemilerin boyutu ve bunların göreceli etkileri, virüsün antijenik (hastalıklı hücre) varyasyonunun kapsamı arasındaki etkileşimi yansıtır. Grip virüsleri, genellikle yaşlı ve akciğer hastalığı olanlarda ölüme neden olabilmektedir (Cox ve Subbarao, 1999).

Grip virüsleri arasında salgın nedeniyle yaşanan ölüm oranı en yüksek olanlardan biriside koronavirüsleridir. Koronavirüsleri, biyolojik açıdan farklı birçok hayvan sınıfına etki edebilecek derecede çeşitliliğe sahiptir. Bu derece etkili olan korona tipi virüslerinin

insanlara etki etmesi sonucunda, enfekte olan insanlarda hafif ya da şiddetli solunum yolu enfeksiyon belirtileri gözlemlenebilmektedir. Yakın tarihte örnek olarak 2002 ve 2012 yıllarında insanlarda şiddetli akut solunum sendromu koronavirüsü (SARS-CoV) ve Orta Doğu solunum sendromu koronavirüsü (MERS-CoV) ortaya çıkarak ölümcül solunum yolu hastalıklarına neden olmuştur (Hu, 2020). Aralık 2019'da, Çin'in Hubei eyaletinin başkenti Wuhan'da zoonotik kökenli yeni bir koronavirüsü (SARS-CoV-2), atipik bir pnömoni salgını meydana gelmiştir (Q. Li ve ark., 2020). 30 Ocak 2020'de DSÖ "Uluslararası Öneme Sahip Halk Sağlığı Acil Durumu" ilan etmiştir (URL-2, 2020). 12 Şubat 2020'de ise yeni koronavirüsünün neden olduğu hastalığa "Koronavirüs Hastalığı 2019" (COVID-19) adını vermiştir (URL-3, 2020). Dünyada hızla yayılmaya başlayan bu salgın hastalık Türkiye'de ilk kez 10 Mart 2020 tarihinde tespit edildiği duyurulmuştur. 11 Mart 2020 tarihinde ise salgın DSÖ tarafından pandemi olarak ilan edilmiştir (URL-4, 2020).

Yeni bir grip pandemisinin olumsuz etkilerini hafifletmek için stratejilerin geliştirilmesi artık küresel bir halk sağlığı önceliğidir. İnfluenza önleme ve sınırlama stratejileri; antiviral, aşı ve farmasötik olmayan (vaka izolasyonu, ev karantina, okul veya işyeri kapatma, seyahat kısıtlamaları) tedbirlerin geniş kategorileri altında değerlendirilebilir (Ferguson ve ark., 2006).

Salgınların müdahale stratejilerini, alınan tedbirleri, potansiyel maliyetlerini ve faydalarını ölçmek için bazı matematiksel modeller ve optimizasyon algoritmaları kullanılmaktadır. Salgın hastalıklarda modelleme yapılırken birkaç faktör üzerinde durulur. Bu faktörlerden bir tanesi "eğriyi tahmin etme" durumu yani salgında ölen ya da hastalanan kişilere ait sayısal veriler kullanılarak ileriye yönelik tahminlerin yapılmasıdır. Bir başka faktör de karantina, kısıtlama gibi salgının yayılmasını önleme ve kontrol altına alma amaçlı önlemlerdir. Bu da enfekte olan hastalara ait eğrileri istenilen seviyede tutmaya yöneliktir. Yani "eğriyi düzleştirme" durumu vardır. Bununla birlikte, çok az model anahtar soruyu yanıtlamaya çalışır. Kontrol önlemleri nasıl optimal bir şekilde uygulanır ve gevşetilir? Burada "eğriyi optimize etme" durumundan bahsedilir. Çünkü gerçekten de karantina gibi güçlü kontrol önlemlerinin çok büyük olumsuz toplumsal ve ekonomik etkileri olabilmektedir.

Salgın hastalıkları tahmin etmek için kullanılan matematiksel modellerden bazıları bölmeli modellerdir. Bölmeli modellerin başında, temel parametreler dediğimiz "sağlıklı", "hasta", "iyileşmiş/ölmüş" kişilerin hesaplamasını yapan SIR modeli gelmektedir. SIR

modeline ek olarak salgına yönelik farklı tahmin ve hesaplamalar için model geliştirilebilmektedir. Geliştirilen modellerden biri de SEIR modelidir. SEIR modeli salgının yayılmasına ilişkin bilgileri tanımlamak için daha uygun ve kapsamlıdır (Zhao ve ark., 2018).

SEIR modeli, SIR modeline ek olarak hastalığa maruz kalan bireyleri modele dahil etmektedir. Hastalığa maruz kalan bireylerin oluşturduğu sınıf, hastalığın gizli olduğunun bilindiği ancak yayılıp yayılmadığının tam olarak bilinmediği durumlardaki kişilerde oluşan bir sınıfı tanımlar. Bu nedenle, COVID-19 gibi kuluçka süresi olan salgınların modellemesinde SEIR modelinin kullanılması durumunda daha kapsamlı sonuçlar elde edilebilecektir.

Bu çalışmada, Güney Kore ve Mısır'daki COVID-19 salgınının doğrulanmış vaka verileri kullanıldı. Bu veriler hem düzenli hem de erişimi kolay olmasından dolayı tercih edildi. Elde edilen salgın verileri kullanılarak SEIR modeli oluşturuldu ve optimizasyon yöntemleri ile modele ait parametreler hesaplandı. Bununla beraber, literatürde sıklıkla bahsedilen makine öğrenmesi yöntemlerinden Uzun Kısa Süreli Bellek (LSTM) ve Destek Vektör Makinesi (DVM) ile veriler üzerinden tahminler gerçekleştirildi. Son olarak optimizasyon yöntemlerinin ve makine öğrenmesi tahmin yöntemlerinin sonuçları kıyaslanarak değerlendirildi ve önerilerde bulunuldu.

1.1. Literatür Taraması

Roques ve ark., (2020) tarafından yapılan çalışmada, gözlemlenemeyen vakaları tanımlayan SIR modeli, veri toplama sürecini tanımlayan olasılıksal model ve istatistiksel çıkarımla birleştirilerek Fransa'daki COVID-19 salgınında ölen ve hastalığa yakalanmış gerçek insan sayısını tahmin etmeye çalışılmıştır. Çalışmalarında tahmin edilen vaka sayısının, gözlemlenen vaka sayısından daha fazla olduğunu raporlanmıştır.

Li ve ark., (2014) SIRS modelini kullanarak yapmış oldukları çalışmada toplumda hastalığa yakalanmayan kesimin durumunu koruyabilmesinin, nüfus içerisindeki vaka artışının dengeye yaklaşmasına bağlı olduğunu belirtmişlerdir.

Jansen ve Twizell (2002) yapmış oldukları çalışmada, enfeksiyon oranının sabit tutulduğu SEIR modelini, tek noktalı yineleme fonksiyonlarının bir sistemi şeklinde, birinci dereceden sonlu farklar yöntemi kullanarak çözüldüğünü ve denklemlerinin kritik

noktalarıyla özdeş olan iki sabit nokta ile aynı kararlılık özelliklerine sahip olduklarını göstermişlerdir.

Peng ve ark., (2020) yapmış oldukları çalışmada SEIR modelinin hem Çin'de hem de diğer ülkelerde COVID-19 salgınını karakterize etmek ve salgın süreçlerini simüle etmek için geliştirilen yaygın bir model olduğunu ifade etmiştir.

Teles, SEIR modelini SARS-COV-2 pandemisinin Portekiz'deki aktif ve hastanede yatan vakaların zaman yörüngelerini analiz etmek için kullandı. Ülkeye yayılan virüsün zaman evriminin yeterince modellendiğini ve modelin değiştirilebilir parametrelere sahip olduğunu belirtmiştir. Uygulanan yöntem ile aktif vakaların ne kadarının kontrol altına alındığını, hastaneye yatırılan hasta sayısını, yoğun bakıma alınan hasta sayısını ve ihtiyaç duyulan vantilatör sayısını analiz edilmiştir (Teles, 2020).

Godio ve ark., (2020) Particle Swarm Optimization (PSO) yöntemiyle İtalya'nın salgından en çok etkilenen bölgelerinden (Lombardia, Veneto ve Piemonte) elde edilen verilere SEIR modelinde girdi olarak kullanmışlardır. Modele ait tahmin güvenilirliğini artırmak için stokastik bir yaklaşımla çok sayıda parametre uyarlanmasını önermişlerdir.

Schneckenreither ve ark., (2008) SIR modeli ile analizi yapılabilen salgınların; Lattice Gas Cellular Automata (LGCA), Stochastic Cellular Automata (SCA) ve Partial Differential Equations (PDE) gibi farklı yöntemlerle hibrit modellerinin oluşturulabileceğini ortaya koymuşlardır.

O'Regan ve ark., (2010) çeşitli SIR ve SIRS modellerini kullanarak salgın hastalıklar biliminde kullanılacak yeni bir Lyapunov fonksiyonu oluşturmuşlardır. Bu fonksiyonlar ile kullanılacak sistemlerin yerel bazdaki denge durumunun küresel bazdaki istikrarı destekleyeceğini belirtmişlerdir.

Vargas-De-León yapmış olduğu çalışmada kalıcı bağışıklık ve geçici bağışıklık sağlayan hastalıkların SIR ve SIRS gibi epidemiyolojik sistemler ile modellenebileceğini ifade etmiştir (Vargas-De-León, 2011).

Attaluri ve ark., (2009) H1N1 virüsünü karakterize etmek ve evrimsel kökenlerini tanımlamak için biyoinformatik analiz yapmışlardır. Filogenetik analiz, karar ağacı analizi ve destek vektör makinesi (DVM) analizi için üç sekans grubu (insan, domuz ve insan / domuz) kullanılarak pandemik viral yükün domuz kaynaklı olduğu bulgusunun güçlü bir şekilde desteklediğini bulmuşlardır.

N Roy ve ark., (2020) Katkı modelini (Additive Regression Model) kullanarak bazı ülkelerin onaylanmış vaka sayılarıyla birlikte iyileşen ve ölen hastaların karşılaştırılmasını

yapmışlardır. Karşılaştırma sonucunda tahminlerde bulunarak tüm dünyadaki SARS-COVID-19 salgınının ekonomik değerlendirmesini açıklamışlardır.

Vaishya ve ark., (2020) yapmış oldukları çalışmada yapay zekâ uygulamalarının doktorlara COVID-19 semptomlarını tanımlamaları için yardımcı olduğunu, hastaların önceki verilerini analiz ederek ölüm riskini azaltabileceğini ve hastalık kontrolünün yüksek doğrulukla tahmin edilebileceğini ifade etmişlerdir.

COVID-19 pandemisinin hastalarda görülen ölüm riski, kritik hastalığa geçmeden önce potansiyel olarak tahmin edilebilmektedir. Yuan ve ark., (2020) XGBoost makine öğrenme algoritmasına dayalı bir tahmin modeli oluşturarak hem ölüm riskinin tahmin edebilmesine hem de kritik vakaların tanınmasına klinik bir yol sunabileceğini ifade etmişlerdir. Erken teşhis ve müdahale ile doktorlara yardımcı olunabileceğini, böylece ölüm oranının potansiyel olarak azaltılabileceğini belirtmişlerdir.

Ardabili ve ark., (2020) yapmış oldukları çalışmada COVID-19 salgını tahmin etmek için makine öğrenmesi ve yumuşak hesaplama modellerinin karşılaştırmalı bir analizini yapmışlardır. Makine öğrenmesi modellerinden MLP ve ANFIS yöntemlerinin sonuçlarının, uzun vadeli tahmin için yüksek bir genelleme yeteneğine sahip olduğunu belirtmişlerdir.

Hu ve ark., (2020) yapmış oldukları çalışmada gelecekteki COVID-19 vaka sayısını tahmin etmek için Modified Auto Encoders yöntemi kullanmış ve yapay zekâ temelli yöntemlerinin doğruluğunun yüksek olduğunu ifade etmişlerdir.

Z. Yang ve ark. (2020) LSTM modelini bulaşma olasılığı, kuluçka hızı, iyileşme veya ölüm olasılığı gibi COVID-19 epidemiyolojik parametrelerini içeren bir veri setiyle eğittiler. Doğrulanmış vakaların gerçek sayısı ile LSTM tarafından tahmin edilen vaka sayılarına ait eğriler arasında dikkate değer bir uyum olduğunu keşfettiler.

Wang ve ark., (2020) yapmış oldukları çalışmada LSTM modeli kullanarak elde ettikleri tahmin sonuçlarının, rapor edilen günlük pozitif vakalarla oldukça tutarlı olduğunu, böylece önerilen LSTM yöntemi ile salgın eğilimini doğru bir şekilde analiz edebileceğini ifade etmişlerdir.

Said ve ark., (2020) çok değişkenli zaman serilerine uygulanan Çift Yönlü Uzun Kısa Süreli Bellek (Bi-LSTM) modelini kullanarak kümülatif COVID-19 vakalarının sayısını tahmin etmek için uygulamışlardır. Katar'daki hastalık salgınını inceleyerek önerilen tekniğin son çıkan tahmin yaklaşımlarından daha iyi performans gösterdiğini ve

tedbirlere ek olarak birden fazla ülkenin verilerinin kullanılması, günlük kümülatif COVID-19 vakalarının tahmininin doğruluğunu arttırdığını belirtmişlerdir.

Singh ve ark., (2020) DVM modelini kullanarak ölen, iyileşen ve doğrulanan vakaları gerçek zamanlı tahmin etmeyi amaçlamışlardır. Yapmış oldukları çalışmada COVID-19 ölüm oranının, doğrulanmış vaka sayısıyla pozitif yönde ilişkili olduğunu göstermişlerdir.

Rahimi ve ark., (2020) yapmış oldukları çalışmada tüm dünyada etkisini gösteren korona virüs salgınının epidemiyolojik parametrelerini analiz etmek için SIR, SEIRQ ve makine öğrenmesi yöntemlerini birlikte kullanmışlardır. Salgında parametre tahminleri için çeşitli algoritmalar bulunmasına rağmen, tüm durumlar için optimal olacak tek bir algoritma olmadığını belirtmişlerdir.

Gupta ve ark., (2020) COVID-19 hastalığının Hindistan'daki yayılım değişimini analiz etmek ve tahminde bulunmak için SEIR ve regresyon tekniklerini kullanılarak R0 (bulaştırma oranı) tahminini yaptılar. SEIR modelinin regresyon tekniklerinden daha iyi sonuçlar verdiğini belirtmişlerdir.

Feng ve ark., (2020) Wuhan'daki salgın eğilimini analiz etmek için SEIR modelini ve Wuhan dışındaki bölgelerdeki salgın eğilimini analiz etmek için Makine öğrenmesi modellerinden Deep Neural Network (DNN) ve Recurrent Neural Network (RNN) modellerini kullanmışlardır. Yapmış oldukları çalışmada, kullanılan SEIR ve Makine Öğrenmesi modelleri, COVID-19 salgınının zirve noktalarını ve boyutlarını tahmin etmede etkili olduğunu belirtmişlerdir. Ayrıca alınan tedbirlerin kaldırılması sonucu Wuhan dışındaki bölgelerde salgının iki katına çıkacağını ifade etmişlerdir.

Y. Yang ve ark., (2020) COVID-19'un salgın dinamiklerini incelemek ve Çin'deki enfekte nüfusu tahmin etmek için ilk olarak LSTM modeli kullanmışlardır. Elde ettikleri sonuçların kısa vadeli tahmin için doğru olduğunu ancak uzun vadeli evrim için doğru olmadığını belirtmişlerdir. Bu sebeple epidemik yayılma sürecini açıklamak için yaygın olarak kabul edilen SEIR modelini kullanmışlardır. Salgının yayılma sürecindeki dinamiklerini ve popülasyonların eğilimini doğru bir şekilde tahmin edebilmek için SEIR modelini önermişlerdir.

1.2. Tezin Katkısı ve Organizasyonu

Bu çalışmanın katkıları şöyle sıralanabilir:

- Salgın hastalıkların modellenmesinde kullanılabilecek modellere değinilmiştir.
- Bu modellerde kullanılan parametreler ve değişkenler açıklanmıştır.
- Bu parametrelerin optimizasyonu için kullanılabilecek eğri uyarlama yöntemleri ve bu uyarlama yöntemlerinin nasıl kullanılabileceği açıklanmıştır.
- Salgın hastalık parametrelerinin makine öğrenmesi yöntemleri ile nasıl tahmin edilebileceği açıklanmıştır.
- Makine öğrenmesi yöntemleri ile salgın modelleme yöntemlerinin tahmin başarıları karşılaştırılmıştır.

Bu çalışmadaki yöntemler kullanılarak:

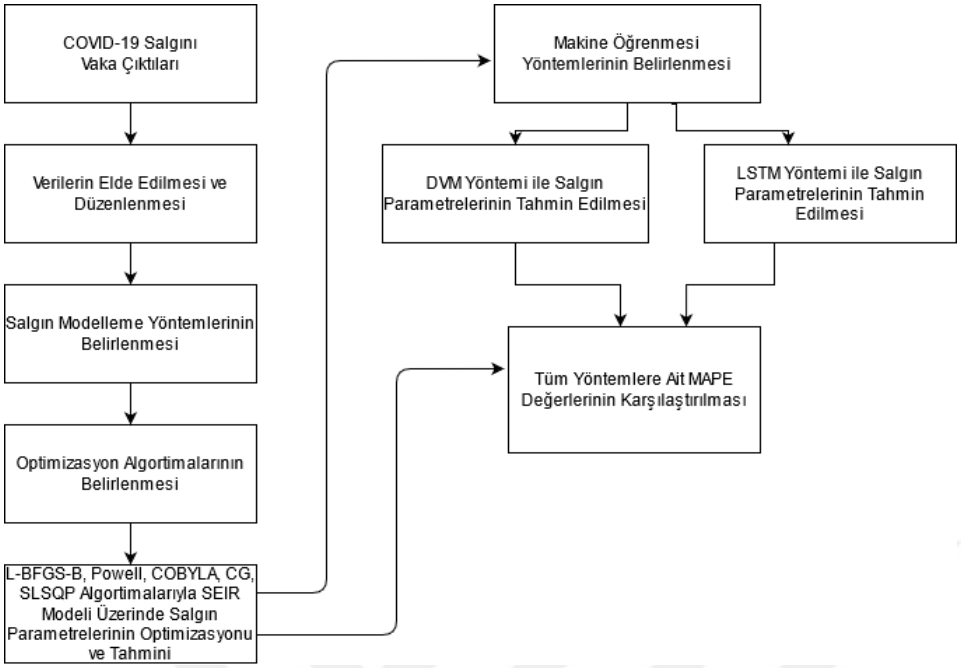
- Salgının seyri hakkında yetkili kuruluşların bilgi edinebilmesi,
- Sağlık kuruluşlarının ve hükümetlerin salgının ilerleyişini durdurmak veya azaltmak için karantina tedbirlerinin ne zaman alabileceklerini,
- COVID-19 gibi salgın hastalıkların ya da etkileşim halindeki olayların etkisinin araştırılması ve tahmin edilebilmesi için SIR, SEIR gibi modellerin yanında makine öğrenmesi yöntemlerinin de nasıl kullanabileceği konularında kolaylık sağlayacaktır.

Bu tez dokümanı şöyle düzenlenmiştir: Birinci bölümde giriş yapılıp literatür ve tezin katkısından bahsedilmiştir. İkinci bölümde ise önerilen sistemi gerçekleştirmek için gerekli olan materyal ve metotlar açıklanmıştır. Üçüncü bölümde, önerilen sistemi gerçeklemek ve test etmek için yapılan örnek uygulamaya ve çıktılarına yer verilmiştir. Son bölümde sırasıyla sonuçlar, öneriler, kısıtlamalar ve potansiyel gelecek çalışmalar hakkında bilgiler sunulmuştur.

2. MATERYAL VE METOT

Bu çalışmada, metot temel olarak dört aşamada ele alınmıştır. İlk olarak kullanılacak ham verinin tedarik edilmesi ve girdi olarak kullanılabilecek hale getirilmesi için yapılan işlemler açıklanmıştır. İkinci aşamada SEIR modelinde kullanılan parametrelere ait optimizasyon işlemleri yapılmıştır. Üçüncü aşamada salgın parametrelerinin makine öğrenmesi yardımıyla tahmini gerçekleştirilmiştir. Son aşamada ise iki farklı modelden elde edilen çıktılara ait veriler karşılaştırılmaktadır. Bu çalışmaya ait akış şeması Şekil 2.1.'de gösterilmektedir. Aşağıda her bir aşamada kullanımına ihtiyaç duyulan materyal ve metotlar detaylı olarak açıklanmıştır.

Matematiksel modellerin bilgisayar ortamlarında hesaplanması aşamasında en önemli adımlardan biri uygun programlama dilinin belirlenmesidir. Programlama dilinin güncel, yaygın, ulaşılabilir kaynaklarının olması ve amaca yönelik geliştirilmiş olması tercih edilir. Yazılım geliştirme süreçlerinde kullanılan ve kaynak kod yönetim sistemi olan Github platformu, en çok aranan ve kullanılan dillerin oluşturduğu popüler programlama dillerine ait liste yayınlamıştır. Buna göre popülerliği en yüksek 5 dil sırası ile Python, Java, Javascript, C#, C/C++ olduğunu belirmiştir (Carbonnelle, 2020). Program geliştirme için sadece programlama dili değil aynı zamanda üzerinde çalışılan platformda önemlidir. Stackoverflow (www.stackoverflow.com) internet sitesi tarafından yapılan anket sonuçlarına göre, Linux işletim sistemi %53 ile program geliştiricilerin en çok kullanmış olduğu platformu seçilmiştir (URL-5, 2019). Bu nedenlerden ötürü yapılan çalışmada, işletim sistemi olarak Linux (Ubuntu 18.04), programlama dili olarak Python programlama dili ve ilgili kütüphaneleri kullanılmıştır.



Şekil 2.1. Akış şeması

2.1. Verilerin Toplanması ve Düzenlenmesi

Yapılan çalışmada kullanılan veri seti, John Hopkins üniversitesi tarafından GitHub platformu üzerinden günlük olarak güncellenen Mısır ve G. Kore'deki COVID-19 salgınına ait verilerinden elde edildi (URL-6, 2021). Mısır'daki 317 günlük ve G. Kore'deki 341 günlük salgın verisi içerisinde; "Doğrulanmış Vaka Sayısı", "Ölen Hasta Sayısı", "İyileşen Hasta Sayısı", "Tarih" ve "Ülke Nüfusu" bilgileri ayıklanarak veri seti oluşturuldu.

2.2. Kullanılan Python Kütüphane Modülleri

Kütüphane modülleri, program yazma aşamasında ihtiyaç duyulabilen bazı işlevleri yerine getirmeyi sağlayan, içerisinde yapılış amacına uygun olan sınıfları, fonksiyonları, nitelikleri vb. barındıran ve ihtiyaç duyulmayan kısımların çıkartılması sonucu hızlı, okunabilirliği yüksek bir yapıda olan ve bu sayede kaynakları etkin kullanan hazır araçlardır.

Python, açık kaynak kodlu ve yaygın olarak kullanılan bir programlama dili olması nedeniyle çok sayıda hazır kütüphaneye sahiptir. Bu sayede yazılan programlarda tekrar eden komut sayıları azalmakta ve kütüphaneler tekrar tekrar kullanılabilmektedir.

Bu çalışmada kullanılan başlıca kütüphane modülleri:

Numpy kütüphanesi, bilimsel hesaplamaları hızlı şekilde yapan, matrisler ile çalışma imkânı sağlayan, ayrık Fourier dönüşümü, trigonometrik, üstel fonksiyon, logaritmik gibi ileri matematiksel işlemler yapmak için kullanılan Python programlama dilinde yazılmış bir matematik kütüphanesidir.

Pandas kütüphanesi, verileri işlemek ve analiz etmek için Python programlama dilinde yazılmış olan bir kütüphanedir. Bu kütüphane temel olarak zaman serileri ve sayısal tabloları işlemek için bir veri yapısı oluşturur ve işlemler bu veri yapısı üzerinde gerçekleştirmektedir.

Matplotlib ve Plotly kütüphaneleri, Python programlama dilinde oluşturulmuş bir veri görselleştirme kütüphanesidir.

Scipy kütüphanesi, Python'un Numpy kütüphanesine dayanan matematiksel algoritmalar ve kullanışlı işlevler koleksiyonudur. Kullanıcıya verileri işlemek ve görselleştirmek için üst düzey komutlar ve sınıflar sağlayarak etkileşimli Python projeleri gerçekleştirmek için önemli katkı sunar. Scipy ile etkileşimli yapılan Python çalışmaları MATLAB, IDL, Octave, R-Lab ve SciLab gibi sistemlerle rekabet eden bir veri işleme ve sistem prototipleme ortamı haline gelir. Scipy kütüphanesine ait çok sayıda optimizasyon yöntemi vardır. Bu tez çalışmasında L-BFGS-B, Powell, CG, COBYLA ve SLSQP yöntemleri kullanılmıştır.

Scikits.learn veya sklearn olarak da bilinen Scikit-learn kütüphanesi Python programlama dili üzerinde kullanılan ücretsiz bir kütüphanedir. Destek vektör makineleri, rastgele orman, gradyan artırma, K-Means ve DBSCAN dahil olmak üzere çeşitli sınıflandırma, regresyon ve kümeleme algoritmalarına sahiptir. Python'un Numpy ve Scipy kütüphaneleri ile birlikte çalışmak üzere tasarlanmıştır.

Math kütüphanesi modülü, Python ile yapılan karmaşık matematiksel hesaplamalar için kullanabileceğimiz Python'daki bazı yaygın matematik işlevlerine ve sabitlerine erişim sağlayarak hesaplama yapar. Kütüphane yerleşik bir Python modülüdür, bu nedenle onu kullanmak için herhangi bir kurulum gerekmez.

Keras kütüphanesi, Python'da programlanmış bir sinir ağı kütüphanesidir. Keras TensorFlow, Microsoft Cognitive Toolkit, R, Theano veya PlaidML ile birlikte

çalışabilmektedir. İçerisinde birçok modül barındırmaktadır. Bu tez çalışmasında "Aktivasyon (Activations) Modülü", "Katmanlar (Layers) Modülü", "Modeller (Models) Modülü", "Optimize Ediciler (Optimizers) Modulü" ve "Hata Hesaplama (Losses) Modülü" kullanıldı.

2.3. Salgın Modelleme Yöntemleri

Salgın hastalıkları modellemek için en sık kullanılan yöntemlerin başında bölmeli modeller gelmektedir. Bölmeli modeller, bulaşıcı hastalıkların matematiksel modellemesini basitleştirir. Bu modellemelerde popülasyon, model içerisindeki bölümlere atanır. Örneğin SIR modeli S, I ve R, (Duyarlı, Bulaşıcı ve İyileştirilmiş) olarak gruplandırılır. İnsanlar sağlık durumlarına göre bölümlere atanır ve insanların gruplar/bölmeler arasında geçişleri olabilmektedir. Etiketlerin sırası genellikle bölmeler arasındaki akış modellerini gösterir.

Modeller çoğunlukla (deterministik olan) sıradan diferansiyel denklemlerle çalıştırılır, ancak daha gerçekçi bir analiz çok daha karmaşık olan stokastik (rastgele) bir çerçeve ile de kullanılabilir.

Modeller bir hastalığın ne kadar yayıldığını, enfekte olan toplam sayıyı, salgının süresi ve üreme katsayısı gibi çeşitli parametreleri tahmin etmeye çalışırlar. Bu tür modeller, farklı halk sağlığı müdahalelerinin salgının sonucunu nasıl etkileyebileceğini gösterebilir. Örneğin, belirli bir popülasyonda sınırlı sayıda aşı yapmak için en etkili tekniğin hangisi olduğuna yönelik tahmin yapılabilir.

2.3.1. SI/SIS Modeli

Bu salgın modelinde bireyler duyarlı ve bulaşıcı olarak sınıflandırılır. Şekil 2.2'de gösterildiği gibi hastalığa maruz kalan bireylerden enfekte olan yani hastalanan bireylere doğru bir akış söz konusudur. Bu modelde şayet bağışıklık olmamış olsaydı bireyler S-I-S döngüsünde devam edeceklerdi. Bu durumda Denklem (2.1) duyarlı kişilerin, Denklem (2.2) hastalanan kişilerin sayısını hesaplamaktadır. Popülasyondaki toplam değişim Denklem (2.3)'te gösterildiği gibi hesaplanır (Vargas-De-León, 2011).

Şekil 2.2. SI modeli

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{2.1}$$

$$\frac{dI(t)}{dt} = \beta S(t) \tag{2.2}$$

Denklem (2.1) ve (2.2)'de geçen (S) parametresi duyarlı vakaların hedef zaman geçmişlerini, (I) parametresi bulaşıcı vakaların hedef zaman geçmişlerini, (β) enfeksiyon oranını ifade eder (Saeedian ve ark., 2017).

2.3.2. SIR/SIRS Modeli

SIR modelinde bireyler üç sınıfa ayrılmıştır. Bunlar duyarlı, bulaşıcı ve iyileşmiş olan sınıflardır. Bu sınıflar modelleme yapılırken sırasıyla S, I ve R ile ifade edilir. Bu modelde akış S sınıfından I sınıfına oradan da R sınıfına doğru ilerler. Hastalığın tekrar bulaşma durumuna göre modelleme yapılmak istenirse SIRS modeli kullanılabilir. SIRS modelinde ise SIR modelindeki akışa ek olarak akış R sınıfından S sınıfına geri döner. SIR modeli aşağıda belirtilen Denklem (2.4), (2.5) ve (2.6)'da belirtildiği gibi hesaplanmaktadır (Saeedian ve ark., 2017). Standart SIR modeli akışı Şekil 2.3'te gösterilmiştir.

Şekil 2.3. SIR modeli

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{2.4}$$

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t) \tag{2.5}$$

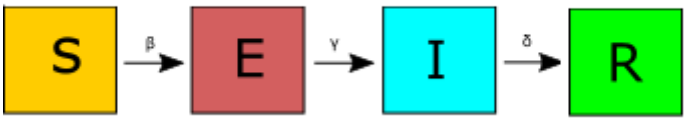
$$\frac{dR}{dt} = \gamma I(t) \tag{2.6}$$

Denklem (2.4), (2.5) ve (2,6)'da geçen (S) parametresi duyarlı vakaların hedef zaman geçmişlerini, (I) parametresi bulaşıcı vakaların hedef zaman geçmişlerini, (R) iyileşen/ölen vakaların hedef zaman geçmişlerini, (β) enfeksiyon oranını ve (γ) bağışıklık kazanmayanların oranını ifade eder.

2.3.3. SEIR Modeli

SEIR modeli bir salgının zaman geçmişini simüle eder. Model bir salgının geçmiş verilerini kullanarak dört farklı koşul arasında etkileşimli olarak modelleme yapar. Bu koşullar: Duyarlı (S), Maruz Kalan (E), Hastalanmış (I) ve İyileşmiş/Ölmüş ya da bağışıklık kazanmış (R) şeklinde bölümlere ayrılır (Godio ve ark., 2020).

Bölmeli modelin ilk kısmı olan duyarlı kişiler salgının birinci gününde toplam nüfusun tamamına eşittir ve henüz hastalıkla tanışmamışlardır. İkinci bölüm de duyarlı iken hastalığa maruz kalan kişilerden oluşmaktadır. Burada hastalık kuluçka dönemindedir. Üçüncü kısımda kesin olarak hastalanmış kişiler vardır. Bunlar hastalığa maruz kaldıktan sonra kuluçka dönemi sona ermiş semptomu olan kişilerden oluşmaktadır. Son olarak hasta olan kişilerin ölüm ya da iyileşme sonucu döngüden çıkan kişiler vardır. Salgın döneminde nüfusun dinamik transferi Şekil 2.4'teki diyagramda gösterilmiştir.



Şekil 2.4. SEIR modeli

SEIR modelindeki hesaplamalar Denklem (2.7), (2.8), (2.9) ve (2.10)'daki formüller kullanılarak yapılmaktadır.

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N} \tag{2.7}$$

$$\frac{dE(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma E(t) \tag{2.8}$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t) \tag{2.9}$$

$$\frac{dR(t)}{dt} = \delta(t) I(t) \tag{2.10}$$

Denklem (2.7)'de Duyarlı kişi sayısına, Denklem (2.8)'de Maruz kalan kişi sayısına, Denklem (2.9)'da hastalananların sayısına ve Denklem (2.10)'da ölen/iyileşen hasta sayısına ait geçişlerin t zamanındaki formülü verilmiştir.

Bu denklemlerde geçen terimler şu şekilde ifade edilmektedir: (S) parametresi duyarlı vakaların hedef zaman geçmişlerini, (E) parametresi maruz kalan vakaların hedef zaman geçmişleri, (I) parametresi bulaşıcı vakaların hedef zaman geçmişlerini, (R) parametresi ölen/iyileşen vakaların hedef zaman geçmişlerini, (β) enfeksiyon oranını, (γ) ortalama gecikme süresinin tersini ve (δ) ölüm oranına bağlı kullanılan katsayıları ifade eder (Godio ve ark., 2020).

2.4. Eğri Uydurma Yöntemleri

Eğri uyarlama doğrusal veya doğrusal olmayan bir $f(x)$ model fonksiyonu oluşturmaya çalışan veri analiz yöntemidir. Optimizasyon ise matematiksel bir fonksiyonun optimumunu, yani en düşük ve en yüksek değerlerini belirleyen işlemler olarak adlandırılır. Optimizasyon, eğri uydurma yöntemlerinden makine öğrenmesi yöntemlerine kadar birçok alanda ve model üzerinde kullanılabilir (Zielesny, 2011).

2.4.1. L-BFGS-B Yöntemi

L-BFGS-B metodunu anlamak için öncelikle BFGS ve L-BFGS metotlarını bilmek gerekir. Türev tabanlı arama yöntemi olan BFGS metodu diğer optimizasyon metotlarından farklı olarak başlangıç adımında belirlenen metrik matrisin Hessian matrisini her adımda kendisine yakınsaması ve çözüm adımında eşitlemesidir (Cengiz, 2013).

Denklem (2.11)'de BFGS Metodunun arama yöntemi belirtilmiştir. Burada B_k, *f*'nin ters Hessian matrisinin *k*'ninci yaklaşımıdır. Bu k'inci adım yaklaşımı Denklem (2.12)'deki BFGS formülü ile hesaplanır.

$$d = -B_k^{-1} f\nabla(x_k) \tag{2.11}$$

$$B_{k+1} = \left(I - \frac{s_k \gamma_k^T}{\gamma_k^T s_k}\right) B_k \left(I - \frac{\gamma_k s_k^T}{\gamma_k^T s_k}\right) + \frac{s_k s_k^T}{\gamma_k^T s_k} \tag{2.12}$$

Denklem (2.13)'te $\gamma_k = f\nabla(x_{k+1}) - (x_k)$ ve $s_k = (x_{k+1}) - (x_k)$ değerlerine eşittir (Nocedal ve Wright, 2006).

L-BFGS yöntemi sınırlandırılmış BFGS yöntemi olarak bilinir. Bu algoritma, kapalı bir Hessian yaklaşımını tersine güncellemek için yalnızca sınırlı miktarda bilgisayar belleği kullanarak BFGS'yi yaklaştırır. Yoğun bir n × n matrisi depolamak yerine, L-BFGS, m'nin önceden seçilen küçük bir sayı olduğu son m yinelemelerin kaydını tutar.

L-BFGS-B, L-BFGS'nin bir uzantısıdır. B, Bound-aries'in dahil edilmesini ifade eder. L-BFGS-B, L-BFGS'nin üstüne iki ekstra adım gerektirir. Birincisi sorunun boyutluluğunu azaltan gradyan projeksiyonu adı verilen bir adımdır. Soruna bağlı olarak, gradyan projeksiyonu, problemin ilk boyutluluğunu ve yineleme sayısını ve çalışma süresini optimum düzeyde azaltarak ve sınırlarında olan değişkenleri ortadan kaldırarak potansiyel olarak birçok yinelemeden tasarruf edilebilir (Henao ve ark., 2014).

L-BFGS-B yöntemi bir objektif fonksiyonu en aza indiren yinelemeli bir algoritmadır. Denklem (2.13)'te verilen L-BFGS-B algoritmasında x bazı kısıtlamalara bağlıdır. Denklem de k'inci yinelemede, amaç fonksiyonu, x_k noktasındaki ikinci dereceden bir modelle yaklaşık olarak belirlenir:

$$m_k(x) = f(x_k) + g_k^T(x - x_k) + \frac{1}{2}x - x_k{}^T B_k(x - x_k) \quad (2.13)$$

Burada, g_k, x_k noktasındaki gradyan ve B_k , x_k noktasındaki Hessian matrisine yaklaşan sınırlı hafızalı BFGS matrisidir (Fei ve ark., 2014).

L-BFGS-B algoritması, değişkenler üzerinde sınırlı kısıtlamalar (diğer bir deyişle bağlı kısıtlamaları) kullanır. Her bir değişken için sırasıyla alt ve üst sınırlar belirlenir. Yöntem, her adımda sabit ve serbest değişkenleri tanımlayarak (basit bir gradyan yöntemi kullanarak) ve ardından daha yüksek doğruluk elde etmek için yalnızca serbest değişkenler üzerinde L-BFGS yöntemini tekrarlayarak çalışır.

L-BFGS-B ve Nelder-Mead gibi bazı optimizasyon algoritmaları SIR ve SEIR modellerinin parametrelerini optimize etmek ve tahmin yeteneklerini geliştirmek için önerilmektedir (Rahimi ve ark., 2020).

2.4.2. Powell Yöntemi

Powell yöntemi, bir fonksiyonun yerel minimumunu bulmak için Michael J. D. Powell tarafından önerilen bir algoritmadır. Fonksiyonun türevlenebilir olması gerekmez ve türev alınmaz. Eşlenik gradyan ile herhangi bir matrisin depolanmasını gerektirmediğinden, çok sayıda değişkenin fonksiyonlarını en aza indirmek için özellikle yararlıdır. (Powell, 1977)

Bu yöntem, BFGS yönteminden farklı olarak son iterasyonda denklem (2.14)'e göre güncellenmektedir.

$$B_{k+1} = (I - \rho_k \gamma_k s_k^T) B_k (I - \rho_k s_k \gamma_k^T) + \rho_k \gamma_k \gamma_k^T \quad (2.14)$$

$$\rho_k = \frac{1}{\gamma_k^T s_k} \quad (2.15)$$

Denklem (2.14)'te verilen ρ_k değeri Denklem (2.15) ile hesaplanmaktadır (Ding, ve ark., 2020).

2.4.3. CG Yöntemi

Magnus Hestenes ve Eduard Stiefel tarafından geliştirilmiş Eşlenik Gradyan yöntemi, matematikte belirli doğrusal denklem sistemlerinin, yani matrisi pozitif-tanımlı

olanların sayısal çözümü için kullanılan bir algoritmadır. Ayrıca optimizasyon problemlerini çözmek için de kullanılabilir.

Denklem (2.16)'da $p(n)$ arama işlemi sırasında maliyet fonksiyonunu, $g(n)$ artık hata ve $\beta(n)$ ise yeni aramalardaki ortoganlığı korumaktadır (Göğüsdere ve ark., 2002).

$$p(n+1) = g(n) + \beta(n)p(n) \tag{2.16}$$

2.4.4. COBYLA Yöntemi

COBYLA, doğrusal olmayan eşitsizlik kısıtlamalarını yönetebilen gradyan içermeyen bir optimizasyon algoritmasıdır. Bu nedenle, Powell basitliklerin köşelerinde enterpolasyon yaparak amaç ve kısıtlama fonksiyonlarına doğrusal polinom yaklaşımları oluşturan COBYLA algortimasını 1994 yılında geliştirmiştir. Denklem (2.17)'de verilen kısıtlanmamış COBYLA algoritmasında x amaç fonksiyonunu, x_0 amaç fonksiyonunun en küçük değerini ve L doğrusal polinom değerini ifade etmektedir (Powell, 2007).

$$x = x_0 - (\Delta/\|\nabla L\|)\nabla L \tag{2.17}$$

2.4.5. SLSQP Yöntemi

SLSQP yöntemi genellikle amaç fonksiyonunun ve kısıtların sürekli türevlenebilir olduğu durumlarda kullanılır. SLSQP algoritmasının kullanıldığı matematiksel problemlerde, her adım iki alt problemi çözer. Denklem (2.18), (2.19) ve (2.20)'de programlamadaki doğrusal olmayan kısıtların işleyiş denklemi gösterilmiştir.

$$g\,(x)\ =\ 0, j\ =\ 1,\ldots,k\,, \tag{2.18}$$

$$g\,(x)\ \geq\ 0, j\ =\ k\ +\ 1,\ldots,m, \tag{2.19}$$

$$x\,l\ \leq\ x\ \leq\ x\,u\,, \tag{2.20}$$

Burada m eşitlik ve eşitsizlik sınırlamalarının sayısını ve k eşitlik sınırlamalarının sayısını ifade etmektedir (Gupta ve Gupta, 2018).

2.5. Destek Vektör Makinesi

Destek Vektör Makinesi (DVM) başlangıçta Vapnik ve arkadaşları tarafından tanıtılan ve daha sonraları başka bir araştırmacı tarafından genişletilen nispeten yeni bir öğrenme algoritmasıdır (Furey ve ark., 2000). Sınıflandırma, veri kümesinin girdi alanındaki doğrusal veya doğrusal olmayan bir ayırma yüzeyi ile gerçekleştirilir. Algoritma, milyonlarca nokta ile problemleri doğru bir şekilde çözebilir ve problemin tipik olarak küçük (100) boyutlu giriş alanı için yaygın olarak bulunan bir doğrusal denklem çözücüden daha karmaşık bir şey gerektirmez (Mangasarian ve Musicant, 2000). Vapnik'in hassas olmayan yaklaşımını kullanarak, tahmin edilen fonksiyon etrafında simetrik olarak minimum yarıçaplı esnek bir kanal oluşturulur, bu şekilde hata hesaplaması sonucunda eşik dışında kalan değerler ihmal edilir. Ayrıca regresyon, genelleme ve yüksek tahmin yeteneğine sahiptir (Awad ve Khanna, 2015). DVM veri setinin doğrusal olarak ayrılıp ayrılamama durumuna göre temel olarak ikiye ayrılmaktadır.

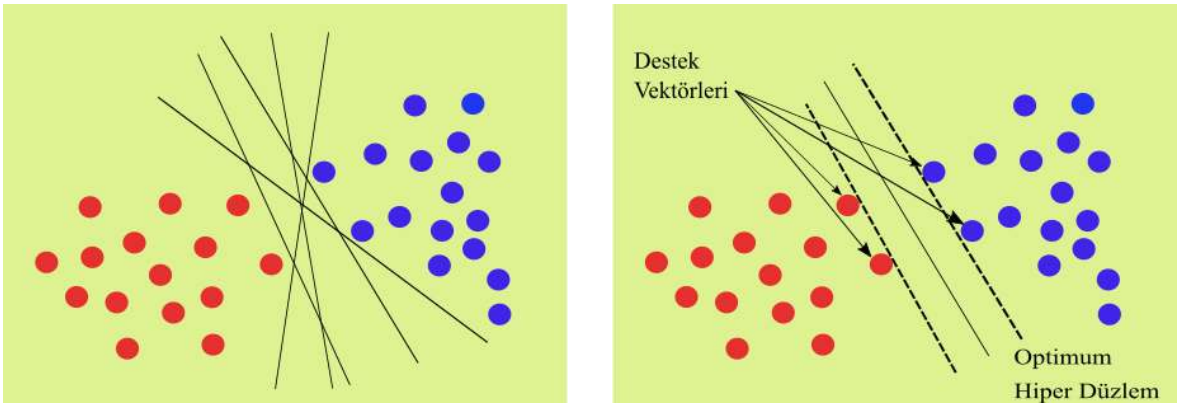
2.5.1. Doğrusal DVM

DVM ile sınıflandırmada genel olarak {-1, +1} şeklinde sınıf etiketleri ile gösterilen iki sınıfa ait örneklerin, eğitim verisi ile elde edilen bir karar fonksiyonu yardımıyla birbirinden ayrılması hedeflenir. Hiper-düzlem karar fonksiyonunu kullanarak eğitim verisini en iyi şekilde ayırır (Kavzoğlu ve Çölkesen, 2010). Doğrusal olarak ayrılabilen iki sınıflı bir sınıflandırma probleminde DVM'nin eğitimi için k sayıda örnekten oluşan eğitim verisinin $\{ x_i , y_i \}, i = 1, ... , k$ olduğu kabul edilirse, optimum hiper düzleme ait eşitsizlikler Denklem (2.21) ve (2.22)'deki gibi olur.

Amaç: min $\frac{1}{2}\left|\left|w\right|\right|^2$ (2.21)

Kısıtlamalar: $y_i(wx_i + b) - 1 \geq 0 \; ve \; y_i \in \{-1, 1\}$ (2.22)

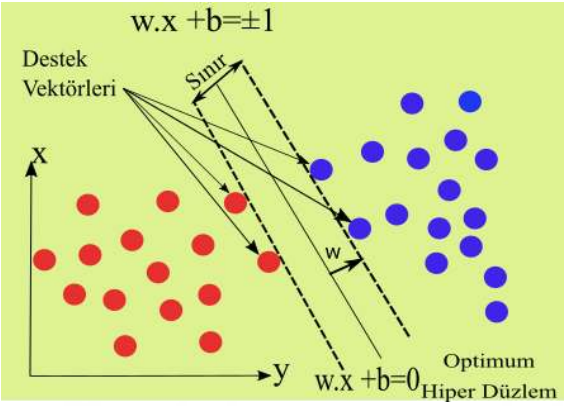
Burada x veri kümesindeki noktaları, $y \in \{-1, 1\}$ ise sınıf etiketlerini, w ağırlık vektörünü ve b eğilim değerini göstermektedir (Osuna ve ark., 1997).



Şekil 2.5. İki sınıflı bir problem için hiper-düzlemler ve destek vektörleri

Şekil 2.5'te gösterildiği üzere iki sınıflı verileri birbirinden ayırabilen birçok hiper-düzlem çizilebilir. Ancak DVM'nin amacı kendisine en yakın noktalar arasındaki uzaklığı maksimuma çıkaran hiper-düzlemi bulabilmektir. Şekil 2.5'te görüldüğü üzere sınırı maksimuma çıkararak en uygun ayrımı yapan hiper-düzleme optimum hiper-düzlem ve sınır genişliğini sınırlandıran noktalar ise destek vektörleri olarak adlandırılır. Optimum

hiper-düzlemin belirlenebilmesi için bu düzleme paralel ve sınırlarını oluşturacak iki hiper-düzlemin belirlenmesi gerekir (Şekil 2.6) (Kavzoğlu ve Çölkesen, 2010).



Şekil 2.6. Doğrusal olarak ayrılabilen veri setleri için hiper-düzlemin belirlenmesi

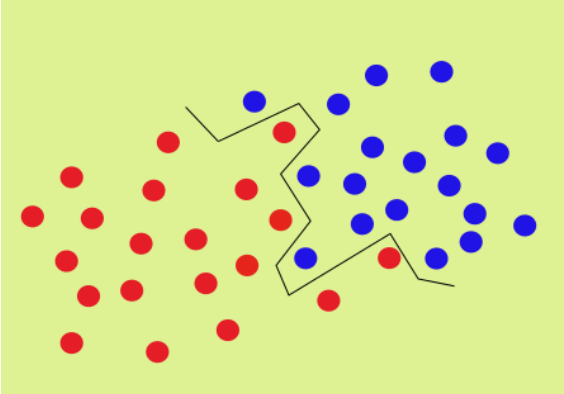
2.5.2. Doğrusal Olmayan DVM

Problem çözümünde veriler her zaman doğrusal olarak ayrılmayabiliyor (Şekil 2.7). Bu tür durumlarda eğitim verilerinin bir kısmının optimum hiper-düzlemin diğer tarafında kalmasından kaynaklanan problem pozitif bir yapay değişkenin (ξ_i) tanımlanması ile çözülür (Şekil 2.7) (Kavzoğlu ve Çölkesen, 2010). Denklem (2.23)'te amaç fonksiyonu, Denklem (2.24) ve (2.25)'te fonksiyona ait kısıtlar belirtilmiştir (Osuna ve ark., 1997).

$$\frac{1}{2}||w||^2 + C(\sum_{i=1}^{t} \xi_i)^k \tag{2.23}$$

$$y_i(wx_i + b) - 1 \geq 1 - \xi_i \qquad i = 1 \ldots t \tag{2.24}$$

$$\xi_i \geq 0 \qquad i = 1 \ldots t \tag{2.25}$$



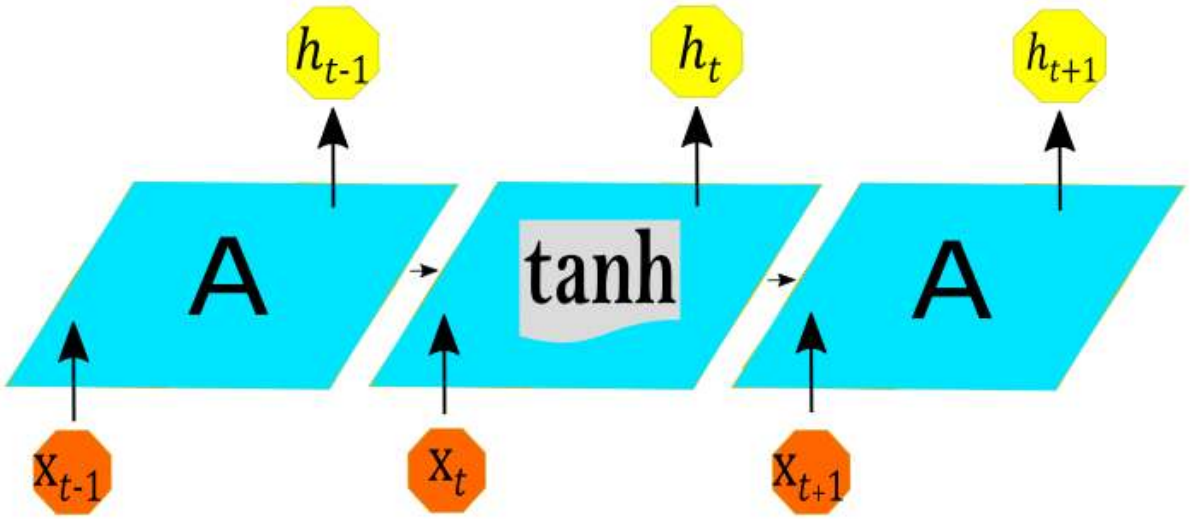
Şekil 2.7. Doğrusal ayrılamayan veri seti

2.6. LSTM

LTSM ilk olarak 1997 yılında Hochreiter ve Schmidhuber tarafından işlendi ve 3 yıl sonra Flex Gers ekibi tarafından geliştirildi (Gers ve ark., 2002). LSTM katmanı, bellek blokları olarak bilinen bir dizi tekrarlayan bağlı bloktan oluşur. Bu bloklar, dijital bir bilgisayarda bellek yongalarının farklı bir versiyonu olarak düşünülebilir. Her biri, hücreler için yazma, okuma ve sıfırlama işlemlerinin sürekli analoglarını sağlayan bir veya daha fazla tekrarlayan bağlı bellek hücresi ve üç çarpımsal birim (giriş, çıkış ve unut kapılar) içerir (Graves ve Schmidhuber, 2005).

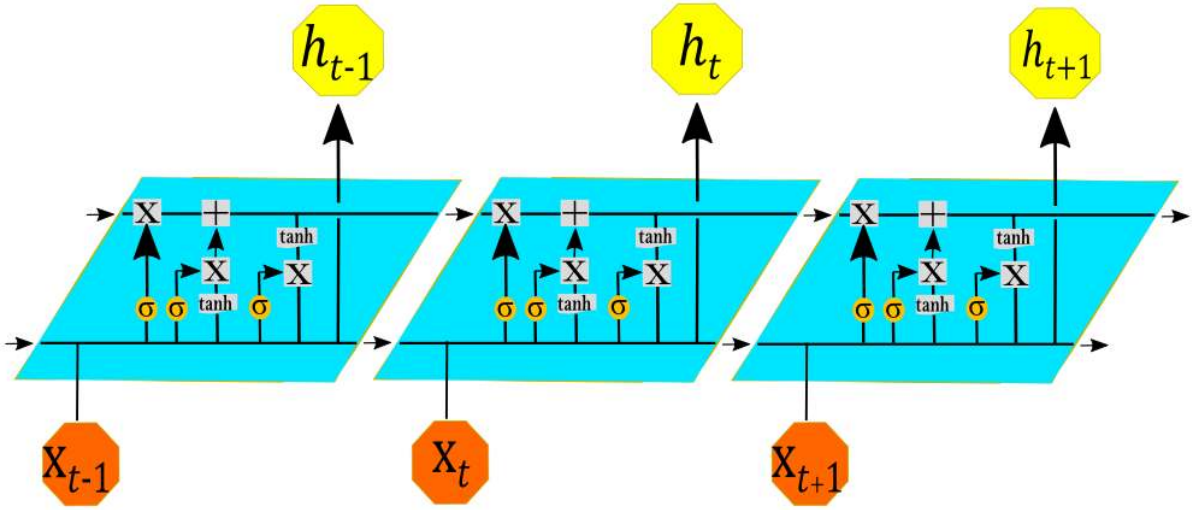
Bazı büyük teknoloji şirketleri de bu algoritmayı konuşma tanıma ve çeviri uygulamaları için kullanıyor. LSTM, tekrarlayan sinir ağı sınıfındandır, bu nedenle hücrelerinin genelde lojistik fonksiyonları olan bir aktivasyon fonksiyonu vardır. Genellikle, LSTM Mimarisi bir bellek hücresi, bir giriş kapısı, bir çıkış kapısı ve bir unutma kapısından oluşur. Bellek hücreleri girişlerini bir süre saklar. Giriş kapıları hücreye yeni değerlerin akışını kontrol eder, unutma kapıları bellekte hangi değerlerin saklanması gerektiğine karar verir ve çıkış kapıları LSTM ünitesinin çıkış aktivasyonunu hesaplamak için bellekte hangi değerlerin kullanılması gerektiğini kontrol eder. Bu yöntem çok popüler bir yapay zekâ tekniğidir ve uygulamaları robot kontrolü, zaman serisi tahminleri, insan aksiyonu tanıma, semantik ayrıştırma vb. birçok farklı alanda görülebilir.

Tüm tekrarlayan sinir ağları (RNN), bir sinir ağının yinelenen modül zincirinin formuna sahiptir. Standart RNN'lerde, bu yinelenen modül, tek bir tanh katmanı gibi çok basit bir yapıya sahip olacaktır (Şekil 2.8).

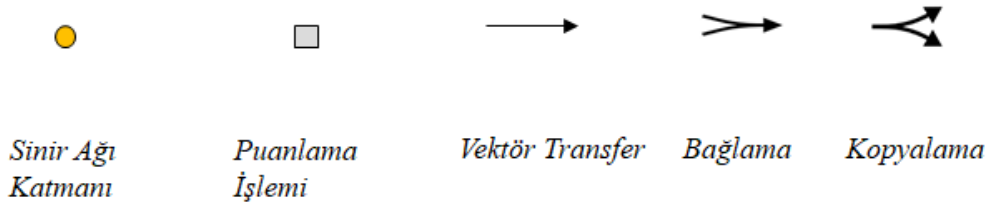


Şekil 2.8. Standart bir RNN'deki yinelenen modül

LSTM'ler de bu zincir benzeri yapıya sahiptir, fakat tekrar eden modül farklı bir yapıya sahiptir. Tek bir sinir ağı katmanına sahip olmak yerine, çok özel bir şekilde etkileşime giren dört tane sinir ağı katmanı var (Şekil 2.9).



Şekil 2.9. Bir LSTM'deki yinelenen modül



Şekil 2.10. Şekil 2.9'da kullanılan şekillerin tanımı

Şekil 2.9'da her satır bir düğümün çıktısından diğerlerinin girişlerine kadar bir vektörü taşır. Turuncu daireler, vektör eklenmesi gibi noktalara yönelik işlemleri temsil ederken, mavi kutular sinir ağı katmanlarını öğrenir. Birleştiren satırlar birleşmeyi belirtirken, bir satır damgası içeriğinin kopyalandığını ve kopyaların farklı konumlara gittiğini belirtir (URL-7, 2015).

LTSM'de bellek hücrelerinin gizli durumu Denklem (2.26), (2.27), (2.28), (2.29) ve (2.32)'de ki formüllerle hesaplanır:

$$i_t = \sigma(W_{ix}x_t + W_{hi}h_{t-1} + W_{ic}\ o\ c_{t-1} + b_i) \quad (2.26)$$

$$f_t = \sigma(W_{fx}x_t + W_{hf}h_{t-1} + W_{fc}\ o\ c_{t-1} + b_f) \quad (2.27)$$

$$c_t = f_t\ o\ c_{t-1} + i_t o\ tanh(W_{cx}x_t + W_{ch}h_{t-1} + W_{hc}c_{t-1} + b_i) \quad (2.28)$$

$$o_t = \sigma(W_{ox}x_t + W_{ho}h_{t-1} + W_{oc}c_{t-1} + b_o) \quad (2.29)$$

$$h_t = o_t\ o\ tanh\ (c_t) \quad (2.30)$$

Burada i_t giriş kapısını, f_t unut kapısını, o_t çıkış kapısını, W ilgili kapıdaki ağırlığı, σ standart sigmoid değerini, önceki LSTM bloğunun çıkışı (zaman damgası-1'de olduğunda), x_t geçerli zaman damgası girişini, c_t hücre belleğindeki zaman damgasını, h_t son durum kontrolünü ve *b* sapma değerini ifade etmektedir (Shi ve ark., 2015).

2.7. Performans Ölçütü

Bu tez çalışmasında COVID-19 pandemisinin tahmin modellemesinde kullanılan yöntemlerden elde edilen veriler ile gerçek veriler arasındaki uyumu hesaplamak için "Ortalama Mutlak Yüzde Hata" (Mean Absolute Percentag Error-MAPE) performans ölçütü kullanılmıştır. MAPE formülü Denklem (2.31)'de belirtilmiştir.

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}\right) * 100 \quad (2.31)$$

3. BULGULAR

Bu çalışmada salgın parametreleri SEIR modeli üzerinde 5 farklı optimizasyon algoritması kullanılarak verilerin optimizasyonu ve tahmini yapıldı. Ayrıca SEIR modeli üzerinde salgına ait ortalama sabitler (DSÖ kaynakları) verilerek ölüm oranı ve R0 değerleri hesaplandı. Elde edilen oranlar ve sabitler kullanılarak her bir gün için parametrelerin değerleri ve geçiş oranları hesaplandı. SEIR, LSTM ve DVM yöntemleri ile gerçekleştirilen hesaplamalar neticesinde elde edilen sonuçlar aşağıda gösterilmektedir. Ayrıca tahmin sonuçları ve MAPE değerleri ayrı ayrı grafiklerle desteklenerek açıklamasıyla verildi.

3.1.SEIR Modeli ile Tahmin Sonuçları

SEIR modelinde salgın verilerinin optimizasyonunu yapabilmek için bazı sabitlerin değerlerini önceden belirlemek gerekir. Bu çalışmadaki SEIR Modeli başlangıç parametreleri DSÖ tarafından yayınlanan çalışmalardan alındı. Parametreler arasındaki geçiş oranları hesaplamak için R0, ortalama kuluçka süresi ve enfeksiyon süresi aşağıda belirtildiği şekilde alındı (Roda ve ark., 2020).

Bulaştırma Oranı (R0) = 3,9 (kişi)

Kuluçka Süresi = 2,9 (gün)

Enfeksiyon Süresi = 5,2 (gün)

Optimizasyonu yapılacak olan veri setinin %70'i eğitim için %30'u test için ayrıldı. Veri seti içerisinde yer alan "Doğrulanmış Vaka Sayısı", "Ölen Hasta Sayısı", "İyileşen Hasta Sayısı" ve "Ülke Nüfusu" bilgileri modele tanıtıldı. Salgın parametreleri arasındaki geçiş oranlarını bulmak için R0 ve ölüm oranları kullanıldı. Geçiş oranları kullanılarak salgın parametrelerinin günlük değerleri hesaplandı. SEIR modeli üzerinde "Vaka" ve "Ölüm" sayılarının optimizasyonu için 5 farklı optimizasyon algoritmasına ait MAPE oranları Tablo 3.1'de gösterildi.

Tablo 3.1. Optimizasyon algoritmaları kullanılarak elde edilen sonuçların MAPE değerleri

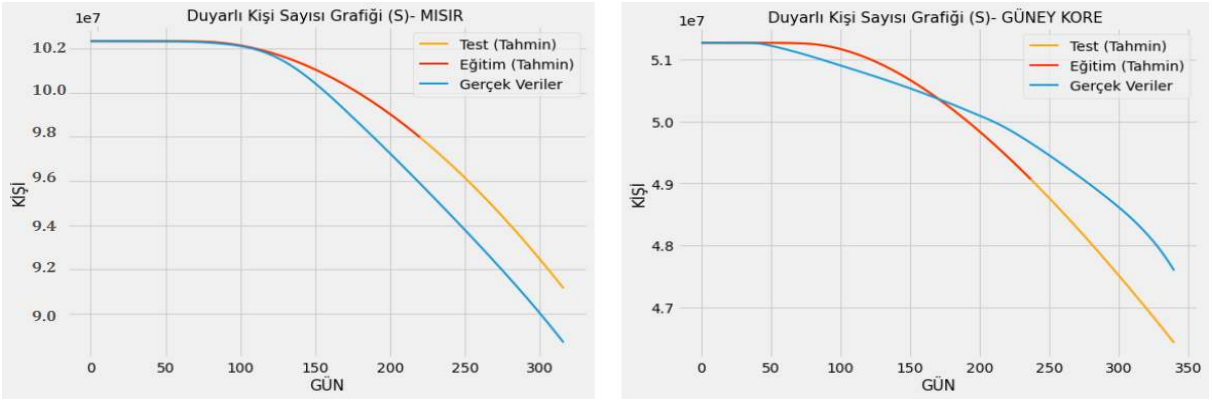
Ülke	MISIR		GÜNEY KORE	
Yöntem	Vaka	Ölüm	Vaka	Ölüm
L-BFGS-B	1.294	0.438	16.991	17.176
Powell	1.294	0.438	16.991	17.177
SLSQP	1.285	0.444	17.009	17.172
CG	1.294	0.438	16.991	17.177
COBYLA	4.822	3.868	18.187	17.698

Tablo 3.1'de görüldüğü üzere tüm yöntemler kabul edilebilir sonuçlar üretmiştir. SEIR modeli üzerinde parametreler arasındaki geçiş oranlarının hesaplanmasında L-BFGS-B yöntemine ait veriler aşağıda sunulmuştur. İlgili alanda daha önceden SEIQR modeli üzerinde kullanılan optimizasyon algoritmaları ile elde edilen sonuçlar yapmış olduğumuz çalışmadaki sonuçları destekler niteliktedir (Rahimi ve ark., 2020). Hesaplanan R0 ve geçiş oranları Tablo 3.2'de gösterildi.

Tablo 3.2. Parametrelerin geçiş oranları ve R0 değeri

Ülke/Parametre	S>E	E>I	I>R	R0
Mısır	0,679	0,340	0,050	1,970
Güney Kore	0,651	0,344	0,016	1,890

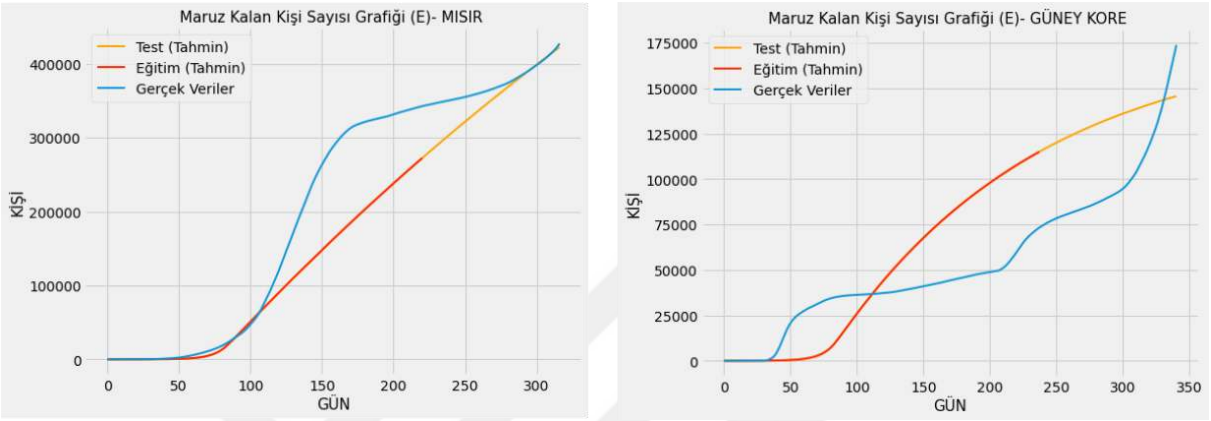
Veri setindeki S, E, I ve R parametrelerine ait değerlerinin tahmin sonuçları, Şekil (3.1), (3.2), (3.3) ve (3.4)'teki grafiklerde gösterildi.



(a) (b)

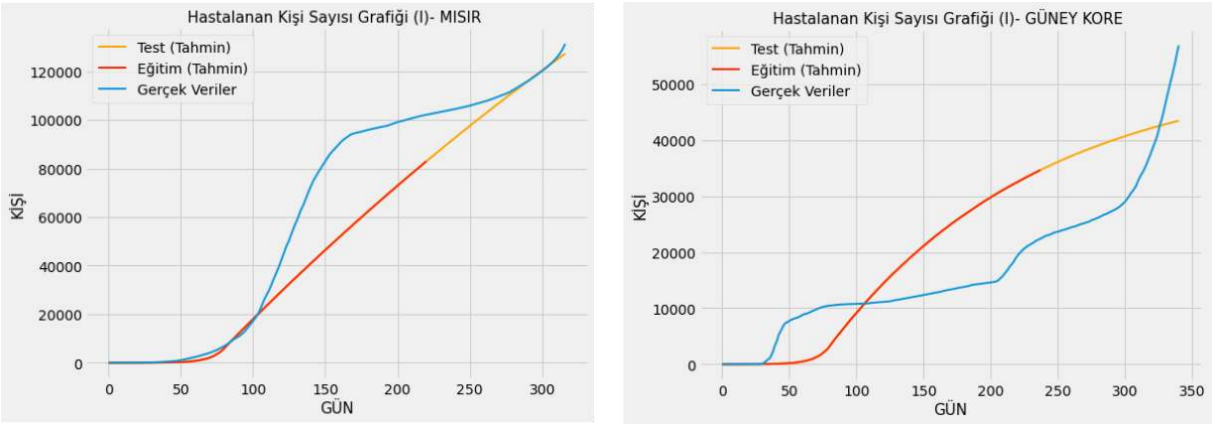
Şekil 3.1. Mısır (a) ve Güney Kore'ye (b) ait "Duyarlı Kişi Sayısının" SEIR modeli ile tahmin sonuçları

Şekil 3.1 (a) grafiğinde Mısır'daki duyarlı kişi sayısının SEIR modeli ile tahmin etmede yaklaşık ilk 100 gün tahmini ve gerçek veriler çok yakın olmuş olsa da 150 günden sonra gerçek ve tahmini veriler arasındaki farkın açıldığı görülmektedir. Şekil 3.1 (b) grafiğinde ise ilk 150 gün gerçek ve tahmini veriler birbirine çok yakın olmuş olsa da sonraki günlerde tahmini veriler ile gerçek veriler arasındaki farkın açıldığı görülmektedir.



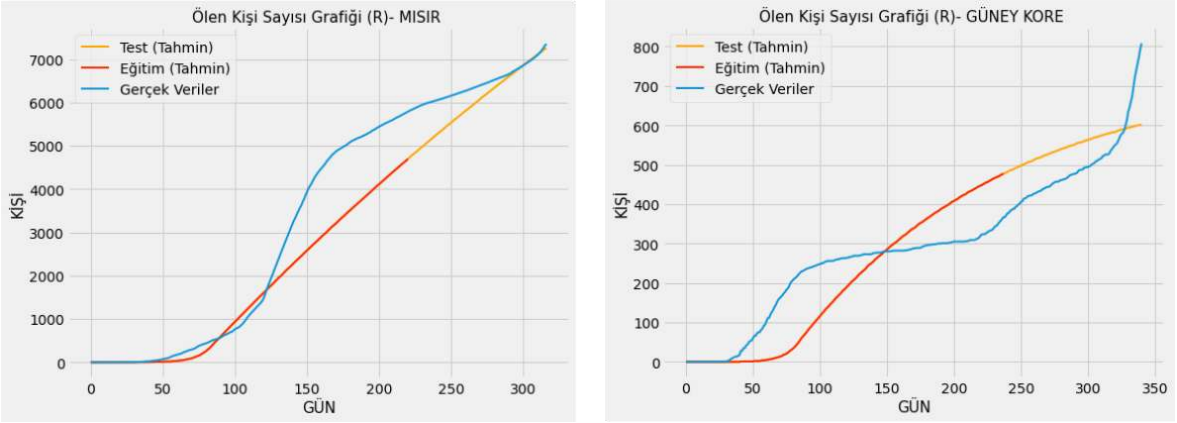
(a) (b)

Şekil 3.2. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" SEIR modeli ile tahmin sonuçları



(a) (b)

Şekil 3.3. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" SEIR modeli ile tahmin sonuçları



(a) (b)

Şekil 3.4. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" SEIR modeli ile tahmin sonuçları

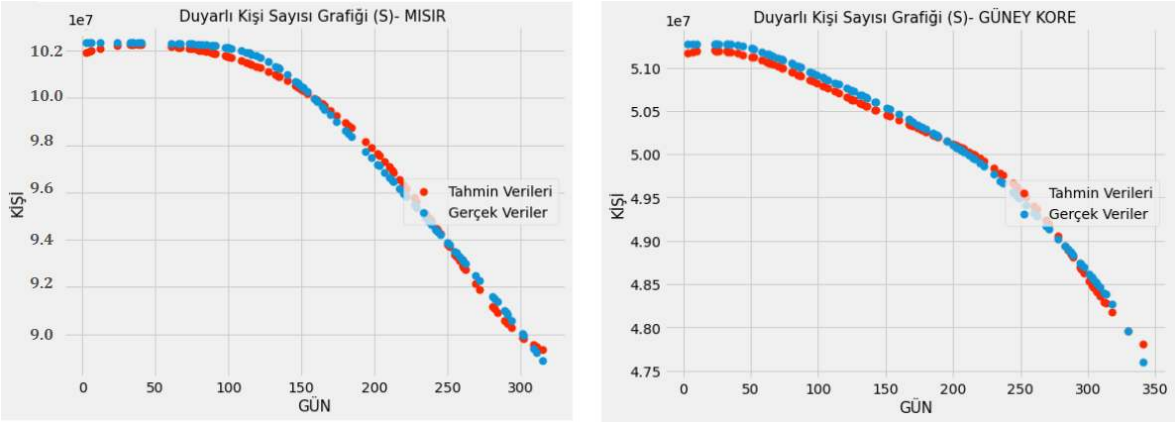
Şekil 3.2, 3.3, 3.4 (a) ve (b) grafiklerinde Mısır'daki ve Güney Kore'deki salgında maruz kalanlar, hastalananlar ve ölen kişilere ait tahmini ve geçek verilere ait grafiklerin birbirlerine benzerlik gösterdiği görülmektedir. Şekil 3.2, 3.3, 3.4 (a) grafiklerinde kişi sayılarını ilk 5 ayın sonun da pik yaptığı ancak tahmini verilerin daha doğrusal bir şekilde ilerlediği ve son iki aylık süreçte gerçek veriler ile tahmini verilere ait eğrilerin birbirine çok yaklaştığı görülmektedir. Şekil 3.2, 3.3, 3.4 (b) grafiklerinde ise Güney Kore'deki gerçek veriler ile tahmini verilere ait eğrilerin ilk 100 gün birbirlerine çok yakın olmuş olsa da sonraki süreçte tahmini verilerin parabolik arttığı ancak gerçek verilerin artışında ise dalgalanmaların olduğu görülmektedir. Son 1 aylık süreçte gerçek verilerin pik yaptığı ancak tahmini verilerin azalış eğilimine geçtiği görülmektedir.

3.2. Makine Öğrenmesi Yöntemleri ile Tahmin Sonuçları

SEIR modeli üzerinde optimize edilerek hesaplanan salgın parametreleri veri setine eklendi. Veri seti her parametre birbirinden bağımsız olacak şekilde hazırlandı. Parametrelerin tahmin edilmesi aşamasında makine öğrenmesi metotlarından LSTM ve DVM yöntemleri kullanıldı. Veri setinin %70'i eğitim için %30'u test için ayrıldı. Kullanılan makine öğrenmesi yöntemlerinde salgın parametreleri birbirinden bağımsız olarak eğitildi. Eğitim sonucu tahmini verilerle ile gerçek veriler arasındaki performans ölçütü olarak MAPE değerleri hesaplandı.

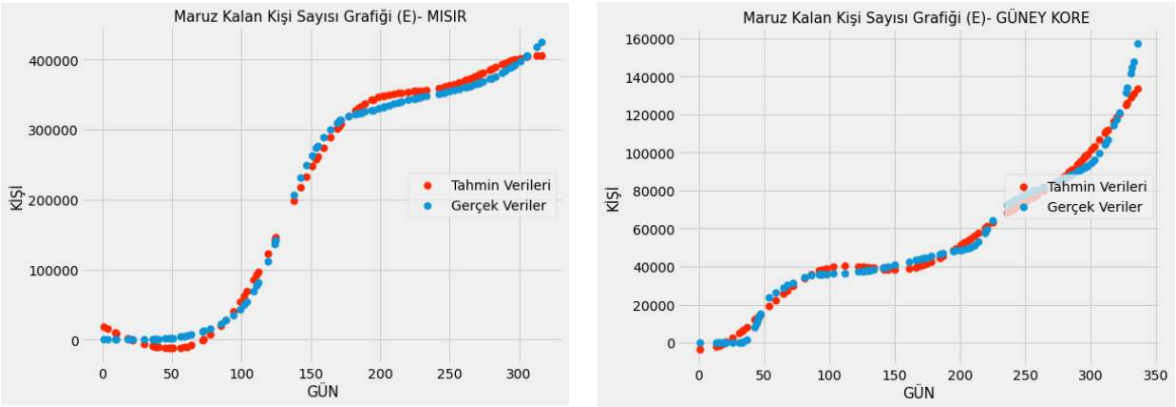
3.2.1. DVM Yöntemi ile Tahmin Sonuçları

Salgın parametrelerinin DVM yöntemi kullanılarak tahmin edilmesinde C kullanılan Kernel fonksiyonu, veri setinin küçük olmasından dolayı RBF tercih edildi. Kullanılan veri setine ait tahmin sonuçları Şekil (3.5), (3.6), (3.7) ve (3.8)'de gösterildi.



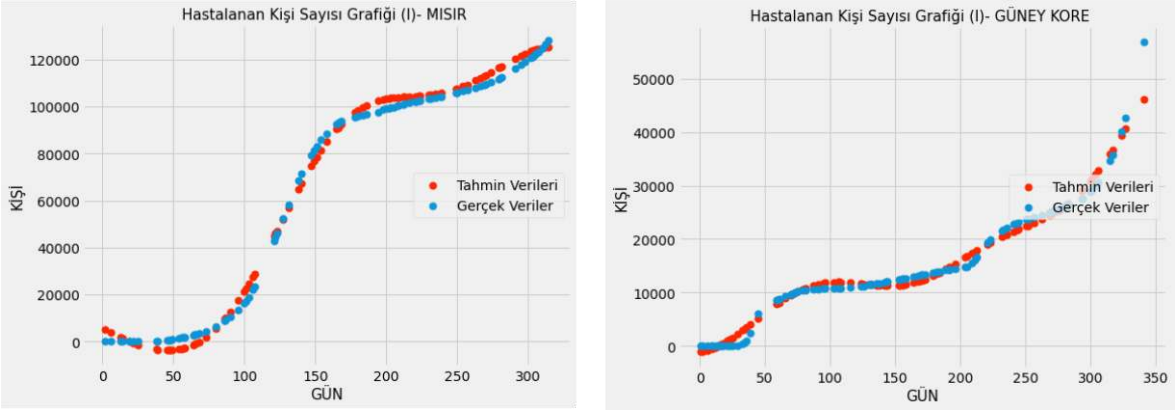
(a) (b)

Şekil 3.5. Mısır (a) ve Güney Kore'ye (b) ait "Duyarlı Kişi Sayısının" DVM yöntemi ile tahmin sonuçları



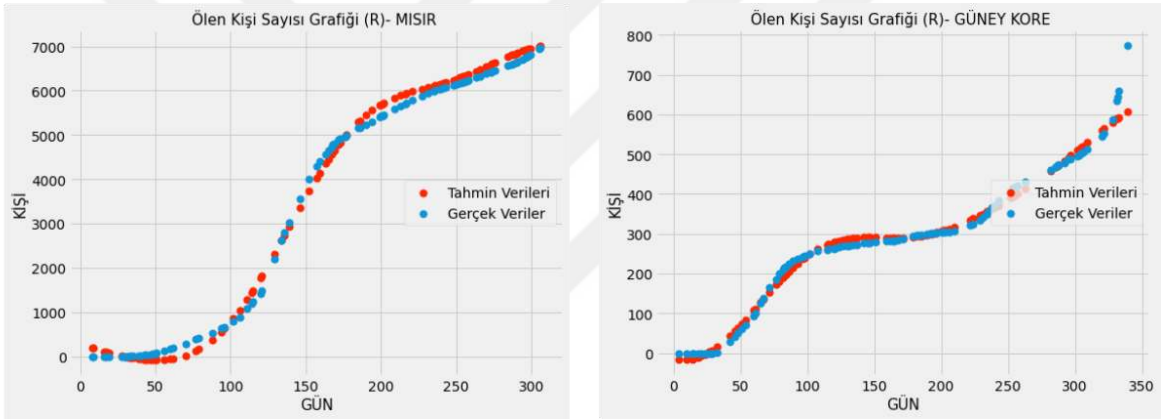
(a) (b)

Şekil 3.6. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" DVM yöntemi ile tahmin sonuçları



(a) (b)

Şekil 3.7. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" DVM yöntemi ile tahmin sonuçları



(a) (b)

Şekil 3.8. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" DVM yöntemi ile tahmin sonuçları

Şekil 3.5, 3.6, 3.7 ve 3.8 (a) ve (b) grafiklerinde Mısır'daki ve Güney Kore'deki Duyarlı, Maruz Kalan, Hasta ve Ölen Kişi sayısının DVM modeli ile tahmin etmede salgının ilk günlerinden itibaren gerçek veriler ile tahmini verilere ait eğrilerin birbirlerine yakın olduğu görülmektedir.

3.2.2. LSTM Yöntemi Tahmin Sonuçları

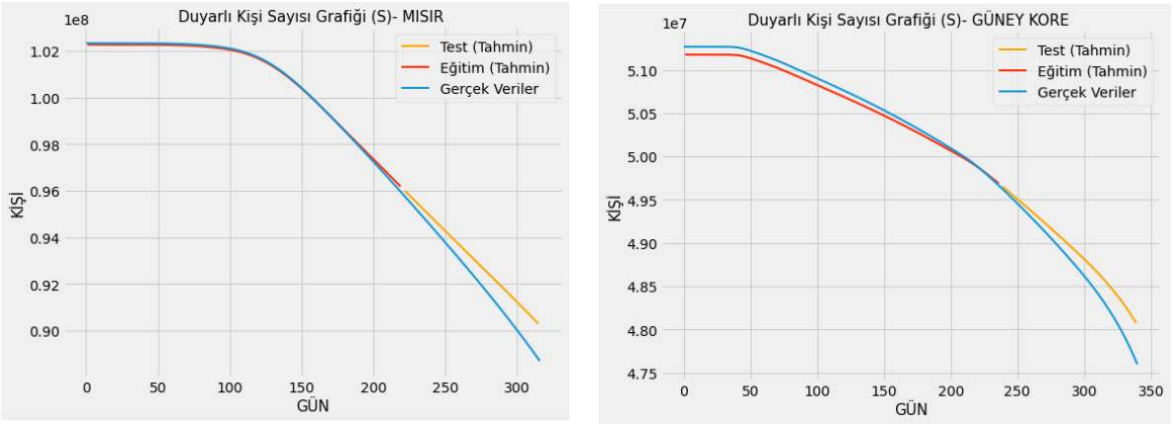
LSTM yöntemi ile elde edilen deneysel sonuçlar için TensorFlow Kütüphanesi kullanılmıştır. Oluşturulan modelin çıkış katmanı "Dense", optimizer "adam" ve

aktivasyon fonksiyonu “relu” kullanıldı. Modelin hiperparametreleri sezgisel olarak ayarlandı. Oluşturulan LSTM modeli Şekil 3.9’da gösterildi.

```
Model: "sequential_18"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_21 (LSTM)               (None, 100)               40800
_________________________________________________________________
dropout_8 (Dropout)          (None, 100)               0
_________________________________________________________________
dense_19 (Dense)             (None, 1)                 101
=================================================================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
```

Şekil 3.9. Oluşturulan LSTM modeli

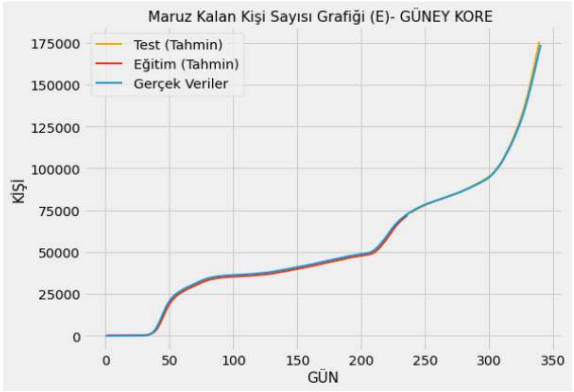
LSTM yöntemi ile yapılan Mısır ve Güney Kore’ye ait tahmin sonuçları Şekil (3.10), (3.11), (3.12) ve (3.13)’te gösterildi.



(a) (b)

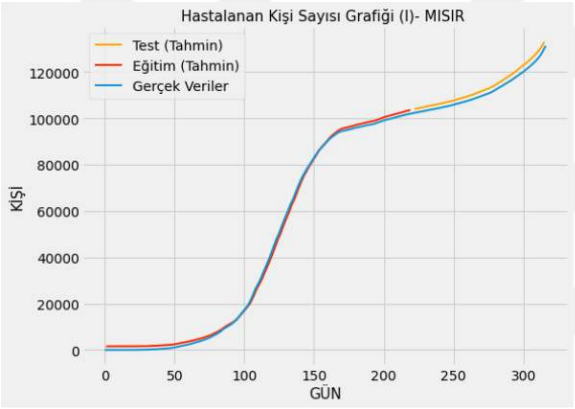
Şekil 3.10. Mısır (a) ve Güney Kore’ye (b) ait “Duyarlı Kişi Sayısının” LSTM yöntemi ile tahmin sonuçları

(a)

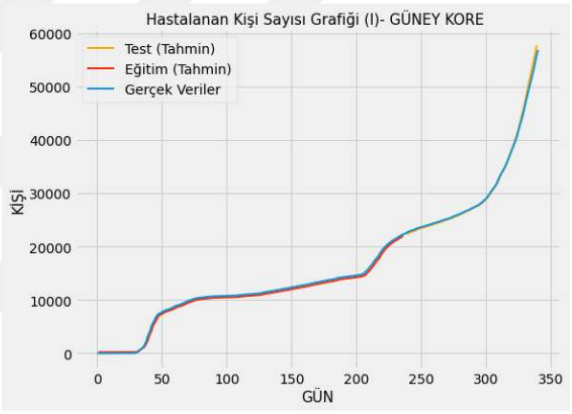


(b)

Şekil 3.11. Mısır (a) ve Güney Kore'ye (b) ait "Maruz Kalan Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları

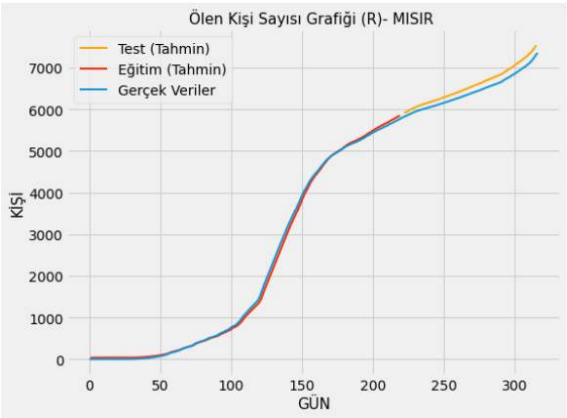


(a)

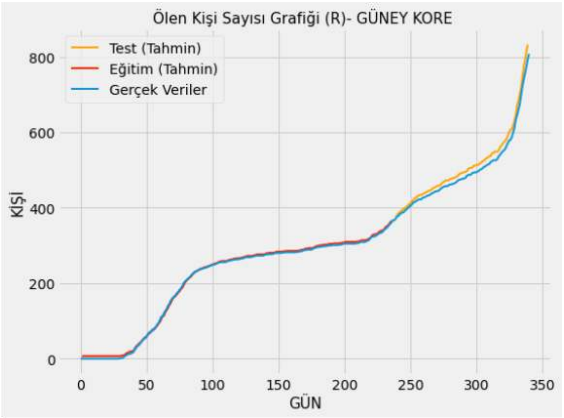


(b)

Şekil 3.12. Mısır (a) ve Güney Kore'ye (b) ait "Hastalanan Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları



(a)



(b)

Şekil 3.13. Mısır (a) ve Güney Kore'ye (b) ait "Ölen Kişi Sayısının" LSTM yöntemi ile tahmin sonuçları

Şekil 3.10, 3.11, 3.12 ve 3.13 (a) ve (b) grafiklerinde Mısır'daki ve Güney Kore'deki Duyarlı, Maruz Kalan, Hasta ve Ölen Kişi sayısının LSTM modeli ile tahmin etmede salgının ilk günlerinden itibaren gerçek veriler ile tahmini verilere ait eğrilerin birbirlerine yakın olduğu görülmektedir.

Tez çalışmasında SEIR, LSTM ve DVM yöntemleri ile Mısır ve Güney Kore'ye ait salgın parametrelerinin tahmini sonucu oluşan MAPE oranları Tablo 3.3 ve Tablo 3.4'te gösterildi.

Tablo 3.3. Mısır'a ait sonuçlar

Yöntem/Parametreler	S	E	I	R
SEIR	2,570	6,100	5,270	6,550
DVM	0.002	12,500	5,880	0,430
LSTM	0.280	0,930	1,180	0,510

Tablo 3.4. Güney Kore'ye ait sonuçlar

Yöntem/Parametreler	S	E	I	R
SEIR	2,010	39,030	37,540	15,950
DVM	0,015	31,500	6,640	0.250
LSTM	0,110	0,860	0,450	1,540

Mısır'daki duyarlı ve ölen kişilerin sayısını tahmin etmede en düşük MAPE değeri DVM yöntemi ile elde edilirken, maruz kalan ve hastalanan kişilerin sayısını tahmin etmede en düşük MAPE değerini LSTM yöntemi vermiştir.

Güney Kore'deki maruz kalan ve hastalanan kişilerin sayısını tahmin etmede en düşük MAPE değeri LSTM yöntemi ile elde edilirken, duyarlı ve ölen kişilerin sayısını tahmin etmede en düşük MAPE değeri DVM yöntemi ile elde edilmiştir.

4. SONUÇ VE ÖNERİLER

Günümüzde birçok alanda kullanılan tahmin teorileri farklı yöntemlerle hesaplanabilmektedir. Salgın hastalıkların modellenmesinde tercih edilen modelin verilere iyi uyum göstermesi yani modelin öngörü başarısının yüksek olması önemlidir. Salgın ya da benzer şekilde etkileşimli olayların boyutlarını tahmin etmek için SIR ve SEIR gibi salgın modelleme teknikleri sıklıkla tercih edilmektedir. Bu modeller ile her ne kadar salgın boyutunu, yapılmış olan müdahalenin etkisini ve ölen hasta sayısının tahmin etme gibi konularda yaklaşık sonuçları elde etmek mümkün olsa da makine öğrenmesi yöntemleriyle de başarılı sonuçlar elde edilmektedir. İlgili alanda önceden yapılmış olan makine öğrenmesi metotlarının başarısı, yapmış olduğumuz çalışmadaki sonuçları destekler niteliktedir (Wang ve ark., (2020), Singh ve ark., (2020)). Makine öğrenmesi yöntemlerinde kullanılan algoritmaların kendi kendisini eğitmesi elde edilecek sonuçlara olan güveninin artmasında önemli rol oynamaktadır.

Literatürdeki çalışmalar zaman serilerini tek bir yöntemle tahmin etmek yerine, veri setindeki farklı fonksiyonel ilişkileri modelleyebilen birden fazla yöntemi birleştirerek tahminler yapmanın daha etkin sonuçlar verdiğini göstermektedir. Mevcut çalışmalara ek bir bakış açısı getirerek yapmış olduğumuz deneysel çalışmalar salgın hastalıklara ait parametrelerin tahmin edilmesinde MAPE oranları oldukça başarılı sonuçlar vermiştir.

Kültürel, sosyal ve ekonomik değişkenlerden dolayı verinin örüntü yapısı ülkeden ülkeye değişmektedir. Bu sebeple yapılacak olan çalışmalarda ülkelere ve tahmin edilecek değere uygun modeller tercih edilmelidir.

4.1. Öneriler

İnsan hayatının söz konusu olduğu salgın hastalıkların önlenmesinde ve tedavi sürecinde matematiksel modellemeler oldukça önem arz etmektedir. Yetkili kuruluşlar mevcut veriler ile yapılan tahmin sonuçlarına yönelik strateji geliştirirler. Bu nedenle yetkili sağlık kuruluşlarının gerçeğe yakın tahmin sonuçları elde edebilmesi için kamuoyuna açıklanan salgın verilerinin doğru bir şekilde paylaşılması önerilmektedir.

Bu tez çalışmasında örnek olarak seçilen iki ülkenin salgın parametreleri yöntemlere göre farklı sonuçlar vermiştir. Bu sebeple salgın hastalıkların seyrine yönelik

yapılan tahminlerde yetkili kurumlarca değerlendirilmeden önce birkaç farklı yöntemle karşılaştırılması önerilmektedir.

4.2. Kısıtlamalar ve Zorluklar

Sosyolojik ve ekonomik faktörlere sebep olan hastalıklara ait veriler ne kadar sağlıklı olursa tahmin başarısı da o denli yüksek olur. Bazı ülkeler de toplumsal baskılardan ve ekonomik sıkıntılardan dolayı salgına ait eksik veriler kamuoyuna açıklanıyor. Bu da tahmin modellerinde gerçek sonuca ulaşmada olumsuzluk teşkil etmektedir.

4.3. Potansiyel Gelecek Çalışmalar

Daha başarılı tahmin sonuçları elde etmek için, yeni matematiksel modeller geliştirilebilir. Mevcut ve geliştirilecek olan modeller için diğer optimizasyon yöntemleri de kullanılabilir. Zaman serisi tahmin yöntemleri ve diğer makine öğrenmesi yöntemleri de tercih edilebilir. Ayrıca tüm bu önerilerden hibrit yapılar da üretilebilir.

5. KAYNAKLAR

Ardabili, S. F., ve ark., 2020. COVID-19 Outbreak Prediction with Machine Learning. *Algorithms*, 13:10-249.

Aslan, R., 2020. Tarihten günümüze epidemiler, pandemiler ve Covid-19. *Ayrıntı Dergisi*, 8(85): 35-41.

Attaluri, P. K., ve ark., 2010. Applying machine learning techniques to classify H1N1 viral strains occurring in 2009 flu pandemic. *2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology,* Montreal, Canada , 2-5 Mayıs.

Awad, M., Khanna, R., 2015. Support Vector Regression., *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* pp. 67-80, eds. Awad, M., Khanna, R., Apress.

Cengiz, M., 2013. Numerik ve parçacık sürü optimizasyonu yöntemleri birleştirilerek kurgulanmış yeni bir optimizasyon algoritması, *Yüksek Lisans Tezi,* Pamukkale Üniversitesi, Denizli, 47s.

Cox, N. J., Subbarao, K., 1999. Influenza. *The Lancet*, 354(9186):1277-1282.

Carbonnelle, P., 2020. PYPL PopularitY of Programming Language, https://pypl.github.io/PYPL.html, Nisan 2020.

Fei, Y., ve ark., 2014. Parallel L-BFGS-B algorithm on GPU. *Computers & Graphics*, 40:1-9.

Feng, S., ve ark., 2020. Prediction of the COVID-19 Epidemic Trends Based on SEIR and AI Models. *Plos One,*16(1):1-15.

Ferguson, N. M., ve ark., 2006. Strategies for mitigating an influenza pandemic. *Nature* , 442:448–452.

Furey, T. S., ve ark., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906-914.

Gers, F. A., ve ark., 2002. Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research,* 3:115-143.

Godio, A., Pace, F., Vergnano, A., 2020. SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence. *Int. J. Environ. Res. Public Health*, 17(10):1-19

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6):602-610.

Göğüsdere, Ş., ve ark., 2002. Gram-Schmidt Eşlenik Yön ve Eşlenik Gradyan Algoritmalarının Adaptif Antenlere Uygulanması., *Eleco International Conference on Electrical and Electronics Eng. (ELECO)*, Bursa, Türkiye, 18-22 Aralık.

Gupta, M., Gupta, B., 2018. An Ensemble Model for Breast Cancer Prediction Using Sequential Least Squares Programming Method (SLSQP). *2018 Eleventh International Conference on Contemporary Computing (IC3)*, Noida, Hindistan, 2-4 Ağustos.

Gupta, R., ve ark., 2020. SEIR and Regression Model based COVID-19 outbreak predictions in India, *Medrxiv.*

Henao, W., 2014. An L-BFGS-B-NS optimizer for non-smooth functions. *Yüksek Lisans Tezi*, New York University, New York 32s.

Hu, B., Guo, ve ark., 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology*, 19(3):141–154.

Hu, Z., ve ark., 2020. Artificial Intelligence Forecasting of Covid-19 in China. arXiv:2002.07112

Jansen, H., Twizell, E. H., 2002. An unconditionally convergent discretization of the SEIR model. *Mathematics and Computers in Simulation*, 58(2):147-158.

Kavzoğlu, T., Çölkesen, İ., 2010. Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. *Harita Dergisi,*144(7),73-82.

Li, C.-H., ve ark., 2014. Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks. *Communications in Nonlinear Science and Numerical Simulation*, 19(4):1042-1054.

Li, Q., ve ark., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199-1207.

Liu, D. C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, *45*(1-3):503-528.

Madhav, N., ve ark., 2017. Pandemics: Risks, Impacts, and Mitigation. *Disease Control Priorities: Improving Health and Reducing Poverty* (3rd bs), pp315-345, eds, D. T. Jamison, H. Gelband, S. Horton, P. Jha, R. Laxminarayan, C. N. Mock, & R. Nugent The International Bank for Reconstruction and Development.

Mangasarian, O. L., Musicant, D. R., 2000. Active support vector machine classification. *Proceedings of the 13th International Conference on Neural Information Processing Systems,* Cambridge, United States, s. 556–562

N Roy, A., ve ark., 2020. Prediction and Spread Visualization of Covid-19 Pandemic Using Machine Learning, *Preprints*

Nocedal, J., Wright, S. J., 2006. Numerical optimization. Springer, New York 664s.

O'Regan, S. M., ve ark., 2010. Lyapunov functions for SIR and SIRS epidemic models. *Applied Mathematics Letters*, 23(4):446-448.

Osuna, E. E., ve ark., 1997. Training support vector machines: an application to face detection. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* San Juan, USA, 17-19 June*,* s. 130-136

Peng, L., ve ark., 2020. Epidemic analysis of COVID-19 in China by dynamical modeling. *medRxiv*

Powell, M. J. D., 2007. A view of algorithms for optimization without derivatives. *Mathematics Today,* 43(5)

Powell, M. J. D., 1977. Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12(1):241-254.

Rahimi, I., ve ark., 2020. Analysis and prediction of COVID-19 using SIR, SEIR, and machinel earning models: Australia, Italy, and UK cases. *Information*, 12(3):109

Roda, W. C., ve ark., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 5:271-281.

Roques, L., ve ark., 2020. Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology*, 9(5):97.

Said, A. B., ve ark., 2021. Predicting COVID-19 cases using Bidirectional LSTM on multivariate time series. *Environ Sci Pollut Res* .

Saeedian, M., ve ark., 2017. Memory effects on epidemic evolution: The susceptible-infected-recovered epidemic model. *Phys. Rev. E,* 95(2): 022409.

Schneckenreither, G., ve ark., 2008. Modelling SIR-type epidemics by ODEs, PDEs, difference equations and cellular automata – A comparative study. *Simulation Modelling Practice and Theory*, 16(8):1014-1023.

Singh, V., ve ark., 2020. Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(8):1583-1597

Shi, X., ve ark., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, Cambridge, United States, s. 802–810

Teles, P., 2020. A time-dependent SEIR model to analyse the evolution of the SARS-CoV-2 epidemic outbreak in Portugal. *Bull World Health Organ,* E-pub: 7 April 2020.

Vaishya, R., ve ark., 2020. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):337-339.

Vargas-De-León, C., 2011. On the global stability of SIS, SIR and SIRS epidemic models with standard incidence. *Chaos, Solitons & Fractals*, 44(12):1106-1110.

Wang, P., ve ark., 2020. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons & Fractals*, 140:110214.

Yang, Y., ve ark., 2020. Prediction of COVID-19 spread via LSTM and the deterministic SEIR model. *2020 39th Chinese Control Conference (CCC)*, Shenyang, China, s. 782-785.

Yang, Z., ve ark., 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12(3):165-174.

Ding, Y., ve ark., 2010. Investigation of quasi-Newton methods for unconstrained optimization. *International Journal of Computer Application,* 29:48-58

Yan, L., ve ark., 2020. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: A machine learning-based prognostic model with clinical data in Wuhan. *medRxiv*

Zhao, D., ve ark., 2018. An extended SEIR model considering homepage effect for the information propagation of online social networks. *Physica A: Statistical Mechanics and Its Applications*, 512:1019-1031.

Zielesny, A. 2011. Curve Fitting. *From Curve Fitting to Machine Learning: An Illustrative Guide to Scientific Data Analysis and Computational Intelligence,* eds, Zielesny, A., s. 53-147

URL-1, 2020. http://www.emro.who.int/pandemic-epidemic-diseases/outbreaks/index.html .Outbreaks.

URL-2, 2020. https://www.who.int/news/item/30-01-2020-statement-on-the-second meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019 nCoV).

URL-3, 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks at-the-media-briefing-on-2019-ncov-on-11-february-2020. WHO Director-General's remarks at the media Briefing on 2019-nCoV on 11 Şubat 2020.

URL-4, 2020. https://www.who.int/director-general/speeches/detail/who-director-general-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 11 Mart 2020.

URL-5, 2019. https://insights.stackoverflow.com/survey/2019/?utm_source=social-share&utm_medium= social &utm_campaign=dev-survey-2019. Developer Survey Result.

URL-6, 2020. https://github.com/CSSEGISandData. COVID-19. 25 Aralık 2020.

URL-7, 2015. https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf. Understanding LSTM Networks. Ağustos 2015.

ÖZGEÇMİŞ

Adı Soyadı: Adnan KEÇE
Yabancı Dil: İngilizce
Doğum Yeri ve Yılı: Erciş /1989
E-Posta: adnan.kece@adalet.gov.tr, bmdnankece@gmail.com
Eğitim ve Mesleki Geçmişi:

- 2017-Devam, Mühendis, Adli Tıp Kurumu, Adli Bilişim İhtisas Dairesi
- 2013-2017, Bilgisayar İşletmeni, Munzur Üniversitesi, Bilgi İşlem Daire Başkanlığı
- 2014-2016, Munzur Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü
- 2012-2013, TÜBİTAK –Teknoloji Geliştirme Bölgesi / Yazılım Geliştiricisi
- 2008-2012, Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü
- 2003-2007, Kazım Karabekir Anadolu Öğretmen Lisesi

Yayınları:

- 2019, Bildiri, Assessment of the Improvability of Production Capacity with Industrial Image Processing for Enterprises in TR1 Region of Turkey, 5th International Regional Development Conference, Malatya
- 2021, Makale, Hybrid Time Series Forecasting Methods for Travel Time Prediction, Physica A: Statistical Mechanics and its Applications, https://doi.org/10.1016/j.physa.2021.126134