



**YAPAY ZEKA TEKNİKLERİ İLE TÜRKİYE'DE
BÖLGESEL SALGIN RİSK ANALİZİ VE
HARİTALANDIRILMASI**

Beyzanur OKUDAN

**Danışman: Dr. Öğr. Üyesi Abdullah Ammar
KARCIOĞLU**

Yüksek Lisans Tezi

Yazılım Mühendisliği Ana Bilim Dalı

2025

(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAZILIM MÜHENDİSLİĞİ ANA BİLİM DALI

**YAPAY ZEKA TEKNİKLERİ İLE TÜRKİYE'DE BÖLGESEL SALGIN RİSK
ANALİZİ VE HARİTALANDIRILMASI**

(Regional Epidemic Risk Analysis and Mapping in Turkey with Artificial Intelligence
Techniques)

YÜKSEK LİSANS TEZİ

Beyzanur OKUDAN

Danışman: Dr. Öğr. Üyesi Abdullah Ammar KARCIOĞLU

Erzurum
Ocak, 2026



FEN BİLİMLERİ ENSTİTÜSÜ
Graduate School of Natural and
Applied Sciences

T.C.
ATATÜRK ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü Müdürlüğü
TEZ KABUL VE ONAY TUTANAĞI

**YAPAY ZEKA TEKNİKLERİ İLE TÜRKİYE'DE BÖLGESEL SALGIN RİSK ANALİZİ
VE HARİTALANDIRILMASI**

Dr. Öğr. Üyesi Abdullah Ammar KARCIOĞLU danışmanlığında, Beyzanur OKUDAN tarafından hazırlanan bu çalışma, 12/01/2026 tarihinde aşağıdaki jüri tarafından Yazılım Mühendisliği Anabilim Dalı Yazılım Mühendisliği Bilim Dalı'nda Yüksek Lisans tezi olarak **oybirliği / oy çokluğu (.../...)** ile kabul edilmiştir.

Jüri Başkanı: Dr. Öğr. Üyesi Merve YILDIRIM

Karadeniz Teknik Üniversitesi

Aslı ıslak imzalıdır

Danışman: Dr. Öğr. Üyesi Abdullah Ammar
KARCIOĞLU

Atatürk Üniversitesi

Aslı ıslak imzalıdır

Jüri Üyesi: Dr. Öğr. Üyesi Şeyma YÜCEL ALTAY

Atatürk Üniversitesi

Aslı ıslak imzalıdır

Enstitü Yönetim
Kurulunun/....../.... tarih
ve sayılı kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof.Dr. Alper NUHOĞLU

Enstitü Müdürü

Aslı ıslak imzalıdır

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildiriş, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Yüksek Lisans Tezi olarak Dr. Öğr. Üyesi Abdullah Ammar KARCIOĞLU danışmanlığında sunulan “YAPAY ZEKA TEKNİKLERİ İLE TÜRKİYE'DE BÖLGESEL SALGIN RİSK ANALİZİ VE HARİTALANDIRILMASI” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	%0	30
Kuramsal Temeller	%0	30
Materyal ve Metot	%2	35
Araştırma Bulguları ve Tartışma	%1	20
Sonuçlar ve Öneriler	%0	20
Tezin Geneli	%3	25

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.

Sunulan bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ettiğimizi beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
Beyzanur OKUDAN	Dr. Öğr. Üyesi Abdullah Ammar KARCIOĞLU
15.12.2025	15.12.2025
İmza: Aslı ıslak imzalıdır	İmza: Aslı ıslak imzalıdır

TEŞEKKÜR

Bu tez çalışmasının hazırlanması sürecinde bilimsel bir bakış açısı kazanmamda ve araştırma sürecinin her aşamasında değerli katkılarıyla bana yol gösteren danışmanım Dr. Öğr. Üyesi Abdullah Ammar Karcıoğlu'na teşekkür ederim.

Çalışmada kullanılan epidemik, demografik, coğrafi ve mobilite verilerinin derlenmesi sırasında açık veri politikaları kapsamında erişim sağlayan TÜİK, Google Cloud Platform, Wikipedia, Eurostat, UNData, WHO ve diğer açık veri sağlayıcılarına teşekkür ederim. Bu tezde kullanılan veri setlerinin oluşturulmasında açık kaynaklardan yararlanılmış olması, çalışmanın şeffaflık ve yeniden üretilebilirlik ilkelerine katkı sağlamıştır.

Tez süreci boyunca moral ve motivasyon kaynaklarım olan aileme ve yakın çevreme sabırları, anlayışları ve destekleri için minnettarım.

Bu çalışma, doğrudan ya da dolaylı katkı sağlayan tüm kişi ve kurumların desteğiyle ortaya çıkmıştır.

Beyzanur OKUDAN

ÖZET

YÜKSEK LİSANS TEZİ YAPAY ZEKA TEKNİKLERİ İLE TÜRKİYE'DE BÖLGESEL SALGIN RİSK ANALİZİ VE HARİTALANDIRILMASI

Beyzanur OKUDAN

Danışman: Dr. Öğr. Üyesi Abdullah Ammar KARCIOĞLU

Amaç: Bu tez çalışmasında, Türkiye ve komşu ülkelerinin bölgesel salgın risk seviyelerinin doğru, güvenilir ve yorumlanabilir bir şekilde tahmin edilmesi amacıyla makine öğrenimi yöntemleri ile çok-görünümlü derin öğrenme yaklaşımlarının bir arada kullanıldığı bütünlük bir model önerilmiştir. Salgın hastalıkların hızlı yayılımı, özellikle nüfus yoğunluğu yüksek bölgelerde sağlık yönetimini zorlaştırdığı için risk seviyelerinin veri odaklı olarak belirlenmesi kritik bir ihtiyaçtır. Bu doğrultuda, Türkiye ve komşu ülkelere ait demografik, coğrafi, sağlık altyapısı ve mobilite bilgilerinden oluşan çok kaynaklı bir veri seti de oluşturulmuştur.

Yöntem: İlk aşamada Türkiye’de bölgesel salgın risk ele alınmış olup çalışmaya Lojistik Regresyon, Rastgele Orman, Destek Vektör Makineleri, Naive Bayes ve XGBoost gibi makine öğrenimi modelleri uygulanmıştır. İkinci aşamada ise Türkiye’de ve komşu ülkelerinde bölgesel salgın risk ele alınmış olup TÜİK, Google Cloud Platform, Wikipedia, Eurostat, UNData, WHO gibi kaynaklardan çok-görünümlü çalışmaya özgü yeni bir veri seti oluşturulmuştur. Bu veri setindeki sınıf dengesizliğini gidermek amacıyla CGAN yöntemi kullanılarak sınıf koşullu sentetik örnekler üretilmiştir. Önerilen MV-RiskNet modeli, veriyi demografik, sağlık, coğrafi ve mobilite olmak üzere dört farklı görünüm altında işleyerek çoklu temsil öğrenmekte olup bu temsilleri dikkat tabanlı bir füzyon mekanizması ile birleştirmektedir.

Bulgular: Makine öğrenimi modellerinin karşılaştırılması sonucunda, XGBoost algoritması %98 doğruluk oranı ile en yüksek performansı göstermiştir. CGAN ile dengelenen veri setinde önerilen MV-RiskNet modeli ise %97,2 doğruluk ve %97,4 F1 skoru elde ederek derin öğrenme yöntemleri arasında en başarılı sonuçları üretmiştir. Elde edilen risk haritaları, yüksek risk seviyelerinin nüfus yoğunluğu yüksek bölgelerde ve ulaşım hareketliliğinin fazla olduğu alanlarda yoğunlaştığını göstermiştir. Bu bulgular, makine öğrenimi ve derin öğrenme yöntemlerinin salgın riskinin mekânsal ve istatistiksel analizinde etkili olduğunu gösterilmiştir.

Sonuç: Elde edilen sonuçlar, makine öğrenmesi yöntemleriyle yüksek doğruluk elde edilmesine rağmen çok değişkenli ve çok boyutlu veri setlerinde sınırlı kaldığı gözlemlenmiştir. Bu nedenle, CGAN yöntemi ile çok-görünümlü probleme özgü veri seti sınıfları dengelenmiş olup önerilen MV-RiskNet derin öğrenme modeli epidemik risk analizinde etkili, uygulanabilir ve güçlü bir yöntem sunduğunu göstermiştir. Önerilen model, bölgesel salgın yönetimi, erken uyarı sistemleri ve sağlık politikalarının geliştirilmesi için karar destek mekanizmalarında kullanılabilecek niteliktedir.

Anahtar Kelimeler: Salgın risk analizi, makine öğrenimi, çok-görünümlü derin öğrenme, CGAN, risk haritalandırma.

Ocak 2026, 66 sayfa

ABSTRACT

MASTER THESIS

REGIONAL EPIDEMIC RISK ANALYSIS AND MAPPING IN TURKEY WITH ARTIFICIAL INTELLIGENCE TECHNIQUES

Beyzanur OKUDAN

Supervisor: Assist. Prof. Dr. Abdullah Ammar KARCIOĞLU

Purpose: This thesis proposes an integrated model that combines machine learning methods with multi-view deep learning approaches to accurately, reliably, and interpretably estimate regional epidemic risk levels in Turkey and its neighboring countries. The rapid spread of infectious diseases, particularly in densely populated areas, makes health management challenging, making data-driven risk assessment a critical need. To this end, a multi-source dataset consisting of demographic, geographical, health infrastructure, and mobility information for Turkey and its neighboring countries has been created.

Methodology: In the first stage, regional epidemic risk in Turkey was addressed, and machine learning models such as Logistic Regression, Random Forest, Support Vector Machines, Naive Bayes, and XGBoost were applied to the study. In the second stage, regional epidemic risk in Turkey and neighboring countries was examined, and a new multi-view dataset specific to the study was created from sources such as TUIK, Google Cloud Platform, Wikipedia, Eurostat, UNData, and WHO. To address class imbalance in this dataset, class-conditional synthetic samples were generated using the CGAN method. The proposed MV-RiskNet model learns multiple representations by processing the data under four different views: demographic, health, geographic, and mobility, and combines these representations with an attention-based fusion mechanism.

Findings: A comparison of machine learning models revealed that the XGBoost algorithm demonstrated the highest performance with a 98% accuracy rate. The proposed MV-RiskNet model, balanced with CGAN, achieved the most successful results among deep learning methods, with 97.2% accuracy and a 97.4% F1 score. The resulting risk maps showed that high-risk levels were concentrated in densely populated areas and regions with high transportation mobility. These findings demonstrate the effectiveness of machine learning and deep learning methods in the spatial and statistical analysis of pandemic risk.

Conclusion: The results obtained showed that although high accuracy was achieved with machine learning methods, it remained limited in multivariate and multidimensional datasets. Therefore, balancing the multi-faceted problem-specific dataset classes with the CGAN method, the proposed MV-RiskNet deep learning model demonstrated that it offers an effective, applicable, and powerful method in epidemic risk analysis. The proposed model is suitable for use in decision support mechanisms for regional outbreak management, early warning systems, and the development of health policies.

Keywords: Epidemic risk analysis, machine learning, multi-view deep learning, CGAN, risk mapping.

December 2025, 66 pages

İÇİNDEKİLER

KABUL VE ONAYI TUTANAĞI.....	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER.....	vi
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ	ix
KISALTMALAR VE SİMGELER DİZİNİ	x
GİRİŞ.....	1
Salgın Risk Analizinin Önemi	1
Veri Tabanlı Epidemik Yaklaşımlar	1
Makine Öğrenimi ve Derin Öğrenmenin Rolü	2
Üretken Modellerle Epidemik Veri Setlerinde Sınıf Dengesizliğini Giderme	3
Araştırma Problemi	4
Çalışmanın Amacı.....	4
Çalışmanın Bilimsel ve Pratik Katkıları	5
Tezin Organizasyonu	6
KURAMSAL TEMELLER.....	7
Epidemik Modelleme ve Salgın Risk Kavramı.....	7
Mekansal Epidemi ve Coğrafi Risk Analizi	7
Veri Tabanlı Epidemik Yaklaşımlar	8
Makine Öğrenmesi Yöntemleri Kullanılarak Salgın Risk Analizi	9
Derin Öğrenme Modelleri Kullanılarak Salgın Risk Analizi.....	11
Çoklu Görünümlü Öğrenme Yaklaşımları	12
Üretken Modeller ve CGAN Yöntemi ile Veri Dengeleme.....	12
Literatür Değerlendirmesi ve Araştırma Boşlukları.....	13
MATERYAL VE YÖNTEM	14
1. Aşama: Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye’de Bölgesel Salgın Risk Analizi ve Haritalandırılması.....	14
Türkiye’de salgın risk veri setinin oluşturulması.....	14
Veri ön işleme ve hazırlık süreci.....	15

Veri ön işleme ve hazırlık süreci.....	16
Öznitelik Seçimi ve Öznitelik Öneminin Belirlenmesi.....	17
Makine Öğrenimi Algoritmalarının Uygulanması	18
Risk haritalandırma ve simülasyon süreci	19
2. Aşama: Derin Öğrenme Modelleri Kullanılarak Türkiye ve Komşu Ülkelerinin Bölgesel Salgın Risk Tahmini ve Haritalandırması	20
Türkiye ve komşu ülkelerinin epidemik risk veri seti.....	20
Çok-görünümlü veri yapısının oluşturulması	21
Çok-görünümlü veri hazırlık süreci	22
CGAN ile sınıf dengesizliğinin giderilmesi	24
Karşılaştırma amaçlı derin öğrenme modelleri	27
Önerilen Model: MV-RiskNet	28
Eğitim Süreci ve Değerlendirme Metrikleri.....	31
ARAŞTIRMA BULGULARI	34
1. Aşama: Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye’de Bölgesel Salgın Risk Analizi ve Haritalandırılması.....	34
Makine öğrenimi modellerinin genel sınıflandırma performansı	34
Makine Öğrenimi Modellerine Ait Karmaşıklık Matrisi	35
Makine Öğrenimi Modellerine Ait ROC–AUC Grafiği	36
Türkiye İl Bazlı Salgın Risk Haritalandırma Bulguları	37
2. Aşama: Derin Öğrenme Modelleri Kullanılarak Türkiye ve Komşu Ülkelerinin Bölgesel Salgın Risk Tahmini ve Haritalandırması	39
Derin Öğrenme Modellerinin Genel Performansı.....	39
Önerilen Modelin Sınıf Bazlı Sınıflandırma Performansı	40
CGAN Tabanlı Veri Artırmanın MV-RiskNet Performansına Etkisi.....	41
MV-RiskNet modelinin eğitim süreci ve öğrenme davranışları	42
Bölgesel risk dağılımı bulguları	44
SONUÇ VE ÖNERİLER.....	46
Makine Öğrenimi Temelli Bölgesel Risk Sınıflandırma Sonuçları	46
Çok-Görünümlü Derin Öğrenme Yaklaşımı ile Bölgesel Epidemik Risk Tahmini	47
Bölgesel Risk Dağılımının Mekânsal Değerlendirilmesi.....	47
Genel Değerlendirme	47
Sınırlılıklar ve Gelecek Çalışmalar	48
KAYNAKÇA	49
ÖZGEÇMİŞ.....	53

TABLÖLER DİZİNİ

Tablo 1. ML Algoritmalarının Sonuçları.....	35
Tablo 2. DL Algoritmalarının Sonuçları.....	40



ŞEKİLLER DİZİNİ

Şekil 1. Türkiye il bazlı salgın risk veri setinin yapısı ve değişken içerikleri.....	15
Şekil 2. Korelasyon matrisi	16
Şekil 3. Öznitelik önem sıralaması.....	18
Şekil 4. Veri setinin genel yapısı.....	21
Şekil 5. Çok-görünümlü önerilen modelin gösterimi	22
Şekil 6. Korelasyon matrisi	24
Şekil 7. CGAN tabanlı veri artırma sürecinin genel akış diyagramı.....	25
Şekil 8. CGAN ile sınıf koşullu sentetik örnek üretim süreci	26
Şekil 9. Çalışmada kullanılan karşılaştırma dl modellerinin mimari görünümleri	28
Şekil 10. Önerilen MV-RiskNet yüksek seviyeli mimarisi.....	30
Şekil 11. Önerilen MV-RiskNet detaylı blok diyagramı.....	31
Şekil 12. Karmaşıklık matrisleri.....	36
Şekil 13. ROC eğrisi sonuçları.....	37
Şekil 14. Türkiye il bazlı salgın risk haritası.....	38
Şekil 15. Türkiye salgın risk yoğunluğunu gösteren ısı haritası	38
Şekil 16. Karmaşıklık matrisleri.....	41
Şekil 17. ROC-AUC eğrileri	42
Şekil 18. Accuracy ve loss grafikleri	44
Şekil 19. Önerilen model ile tahmin edilen bölgesel salgın risk seviyelerinin haritalandırılması	45

KISALTMALAR VE SİMGELER DİZİNİ

Acc.	: Accuracy (Doğruluk)
AE-CLS	: Autoencoder–Classifier Model (Otokodlayıcı–Sınıflandırıcı Model)
Attn.	: Attention Mechanism (Dikkat Mekanizması)
AUC	: Area Under the Curve (Eğri Altında Kalan Alan)
CBS	: Coğrafi Bilgi Sistemleri
CGAN	: Conditional Generative Adversarial Network (Koşullu Üretken Çelişkili Ağlar)
CM	: Confusion Matrix (Karışıklık Matrisi)
CNN	: Convolutional Neural Network (Evrişimsel Sinir Ağı)
DL	: Deep Learning (Derin Öğrenme)
F1	: F1-Score (F1-Skoru)
FN	: False Negative (Yanlış Negatif)
FP	: False Positive (Yanlış Pozitif)
GCN	: Graph Convolutional Network (Graf Evrişimsel Ağ)
GIS	: Geographic Information Systems (Coğrafi Bilgi Sistemleri)
LR	: Logistic Regression (Lojistik Regresyon)
LSTM	: Long Short-Term Memory (Uzun Kısa Süreli Bellek)
ML	: Machine Learning (Makine Öğrenimi)
MLP	: Multilayer Perceptron (Çok Katmanlı Algılayıcı)
MV	: Multi-View (Çoklu Görünüm)
MV-RiskNet	: Multi-View Risk Network (Çoklu Görünüm Risk Ağı – Önerilen Model)
NB	: Naive Bayes (Naive Bayes)
OHE	: One-Hot Encoding (Tek-Sıcak Kodlama)
Prec.	: Precision (Kesinlik)
Rec.	: Recall (Duyarlılık)
RF	: Random Forest (Rastgele Orman)
RL	: Risk Level (Risk Düzeyi)
ROC	: Receiver Operating Characteristic (Alıcı İşletim Karakteristiği Eğrisi)
RS	: Risk Score (Risk Skoru)
SS	: StandardScaler (Standartlaştırıcı)

SEIR	: Susceptible–Exposed–Infected–Recovered (Duyarlı–Maruz Kalan–Enfekte–İyileşmiş)
SIR	: Susceptible–Infected–Recovered (Duyarlı–Enfekte–İyileşmiş)
SVM	: Support Vector Machines (Destek Vektör Makineleri)
TN	: True Negative (Doğru Negatif)
TP	: True Positive (Doğru Pozitif)
XGB	: XGBoost – Extreme Gradient Boosting (Aşırı Gradyan Artırma)



GİRİŞ

Salgın Risk Analizinin Önemi

Bulaşıcı hastalıkların bölgesel ölçekte yayılım dinamiklerinin anlaşılması ve risk düzeylerinin güvenilir biçimde tahmin edilmesi, günümüz halk sağlığı yönetiminin temel bileşenlerinden biri hâline gelmiştir. Artan nüfus yoğunluğu, hızlanan kentleşme süreçleri ve küresel ölçekte çeşitlenen insan hareketliliği, salgın hastalıkların yalnızca sağlık sistemlerini etkilememektedir aynı zamanda ekonomik yapı, sosyal yaşam, ulaşım ağları ve kamu güvenliği gibi birçok alanı eş zamanlı olarak etkilemesine neden olmaktadır. Bu çok boyutlu etki alanı, salgın riskinin yalnızca biyolojik parametreler üzerinden değerlendirilmesini yetersiz kılmakta, daha kapsamlı ve veri temelli yaklaşımları zorunlu hâle getirmektedir.

Covid-19 pandemisi, bölgesel hareketlilik ile hastalık yayılımı arasındaki güçlü ilişkiyi açık biçimde ortaya koymuş olup klasik epidemik modellerin karmaşık toplumsal ve mekânsal etkileşimleri tek başına açıklamakta sınırlı kaldığını göstermiştir. Bu süreçte, farklı veri kaynaklarının birlikte değerlendirilmesine dayanan analitik modellerin erken uyarı, risk izleme ve müdahale planlaması açısından kritik bir rol üstlendiği görülmüştür. Dolayısıyla salgın riskinin doğru biçimde tahmin edilebilmesi, yalnızca vaka sayılarının izlenmesiyle değil demografik yapı, coğrafi özellikler, sağlık altyapısı ve mobilite örüntülerinin bütüncül biçimde analiz edilmesiyle mümkün olmaktadır.

Türkiye ve komşu ülkeler, nüfus yapıları, ulaşım ağları, göç hareketlilikleri ve sosyoekonomik koşullar bakımından belirgin farklılıklar göstermektedir. Bu heterojen yapı, salgın riskinin hem ülke içi hem de ülkeler arası düzeyde farklılaşmasına neden olmakta ve bölgesel risk analizlerini daha karmaşık hâle getirmektedir. Bu nedenle, bölgesel salgın riskinin veri odaklı ve çok boyutlu modeller aracılığıyla ortaya konulması, sağlık kaynaklarının etkin kullanımı ve müdahale stratejilerinin doğru biçimde planlanması açısından büyük önem taşımaktadır.

Veri Tabanlı Epidemik Yaklaşımlar

Veri bilimi ve yapay zekâ alanındaki hızlı gelişmeler, salgın risk analizinde çok katmanlı ve heterojen veri yapılarının etkin biçimde işlenebilmesine olanak sağlamaktadır. Geleneksel epidemik modeller, bulaşma süreçlerinin temel çerçevesini açıklamakta önemli bir rol üstlenmekle birlikte nüfus hareketliliği, bölgesel sosyoekonomik farklılıklar ve sağlık

altyapısındaki eşitsizlikler gibi karmaşık ilişkileri yeterince temsil edememektedir. Bu sınırlılıklar, daha esnek ve veri odaklı yöntemlerin kullanılmasını gerekli kılmaktadır.

Makine öğrenimi ve derin öğrenme yaklaşımları, çok sayıda değişkeni aynı anda değerlendirebilme yetenekleri sayesinde epidemik risk analizinde güçlü bir alternatif sunmaktadır. Özellikle farklı tematik özelliklere sahip değişkenlerin birlikte ele alınabildiği bu yöntemler, bölgesel risk seviyelerinin sınıflandırılmasında yüksek performans elde edilmesini mümkün kılmaktadır. Ancak çok sayıda ve farklı yapısal özelliklere sahip veri kaynağının bulunduğu durumlarda, bu verilerin tek bir birleşik temsil altında modellenmesi önemli zorluklar doğurmaktadır.

Bu noktada çok-görünümlü veri yapıları, epidemik risk analizine önemli bir katkı sağlamaktadır. Çok-görünümlü yaklaşım, her bir veri grubunun kendi yapısal özelliklerini koruyarak ayrı öğrenme süreçleri altında ele alınmasına imkân tanımakta ve böylece temsil gücünü artırmaktadır. Demografik, coğrafi, sağlık altyapısı ve mobilite verilerinin ayrı görünümüler altında modellenmesi, epidemik riskin çok boyutlu doğasının daha gerçekçi biçimde yakalanmasını sağlamaktadır.

Makine Öğrenimi ve Derin Öğrenmenin Rolü

ML yöntemleri, çok değişkenli veri yapılarında sundukları güçlü genelleme yeteneği ve yüksek sınıflandırma performansı sayesinde epidemik risk analizlerinde yaygın olarak kullanılmaktadır. LR gibi doğrusal modeller, salgın riskini etkileyen temel faktörlerin incelenmesinde açıklayıcı bir başlangıç noktası sunarken RF, SVM ve özellikle XGB gibi daha gelişmiş algoritmalar doğrusal olmayan ilişkileri öğrenebilme kapasiteleri sayesinde daha yüksek doğruluk oranlarına ulaşabilmektedir. Literatürde XGB'nin yüksek boyutlu ve karmaşık veri setlerinde kararlı performans sergilediği sıklıkla rapor edilmekte, bu durum yöntemin epidemik risk sınıflandırmalarında güçlü bir alternatif olduğunu göstermektedir.

Bununla birlikte, klasik ML modellerinin önemli bir sınırlılığı, heterojen veri kaynaklarını çoğunlukla tek bir birleşik temsil altında işlemesidir. Demografik, coğrafi, sağlık altyapısı ve mobilite gibi yapısal açıdan farklı veri gruplarının aynı özellik uzayına zorlanması, modellerin bu veri gruplarına özgü ilişkileri yeterince ayırtmasını güçleştirmekte ve temsil kaybına yol açabilmektedir. Bu durum, çok kaynaklı epidemik veri setlerinde ML tabanlı yaklaşımların her zaman optimum performans sağlayamamasına neden olmaktadır.

DL modelleri, bu sınırlılıkların aşılmasında önemli avantajlar sunmaktadır. CNN, LSTM, MLP, AE-CLS ve GCN gibi modeller, yüksek boyutlu veri yapıları içerisindeki karmaşık örüntüleri katmanlı biçimde öğrenebilmekte ve daha zengin temsiller

oluşturabilmektedir. Ancak bu modellerin büyük bir bölümü de ML yaklaşımlarına benzer şekilde tek bir veri görünümü üzerinden çalışmakta ve çok kaynaklı epidemik verilerin doğal çok-görünümlü yapısını doğrudan modelleyememektedir. Bu nedenle, farklı veri gruplarının kendilerine özgü yapısal özellikleri öğrenme sürecinde yeterince temsil edilememektedir.

Bu bağlamda literatürde, çok-görünümlü DL yaklaşımlarına yönelik ilgi giderek artmaktadır. Çok-görünümlü yapı, her bir veri grubunun kendi tematik bütünlüğünü koruyacak şekilde ayrı alt ağlarda işlenmesine olanak tanımakta ve görünüm bazlı temsillerin ortak bir öğrenme sürecinde bütünleştirilmesini sağlamaktadır. Bu yaklaşım, epidemik risk analizlerinde demografi, coğrafya, sağlık altyapısı ve mobilite verileri arasındaki karmaşık etkileşimlerin daha derin ve tutarlı biçimde modellenmesine imkân tanımakta olup klasik ML ve tek-görünüm DL yöntemleriyle ulaşılamayan bir temsil gücü sunmaktadır.

Üretken Modellerle Epidemik Veri Setlerinde Sınıf Dengesizliğini Giderme

Epidemik veri setlerinin önemli bir özelliği, çoğu zaman dengesiz sınıf dağılımına sahip olmalarıdır. Özellikle yüksek risk kategorisine ait gözlemlerin sınırlı sayıda olması, sınıflandırma modellerinin bu sınıfı yeterince öğrenememesine neden olmaktadır. Bu durum, modellerin düşük ve orta risk sınıflarına odaklanmasına, yüksek riskli bölgelerin ise hatalı veya eksik biçimde tahmin edilmesine yol açmaktadır.

Sınıf dengesizliği problemi, yalnızca performans ölçütlerini olumsuz etkilemekle kalmamakta olup aynı zamanda elde edilen risk haritalarının mekânsal tutarlılığını da zayıflatmaktadır. Bu nedenle, epidemik risk analizlerinde veri setlerinin dengelenmesi kritik bir gereklilik olarak öne çıkmaktadır. CGAN, sınıf koşullu sentetik örnekler üreterek bu sorunun çözümünde etkili bir yaklaşım sunmaktadır.

CGAN yöntemi, belirli bir sınıfa ait yeni ve gerçekçi örnekler üreterek veri dağılımını dengelemeyi amaçlayan üretici ve bu örneklerin gerçekliğini değerlendiren ayırt edici bileşenlerden oluşmaktadır. Eğitim süreci sonunda CGAN, özellikle nadir görülen yüksek risk sınıfına ait yapısal özellikleri temsil eden çeşitli ve tutarlı sentetik örnekler üretebilmektedir. Bu yaklaşım, sınıflandırma modellerinin öğrenme kapasitesini artırarak epidemik açıdan kritik örüntülerin daha sağlıklı biçimde yakalanmasını sağlamaktadır.

Bu tez çalışmasında CGAN yöntemi, önerilen MV-RiskNet modelinin eğitildiği veri setindeki sınıf dengesizliğini gidermek amacıyla kullanılmıştır. Yüksek risk sınıfına ait sentetik örneklerin üretilmesiyle modelin genelleme yeteneği güçlendirilmiş ve risk seviyelerinin daha güvenilir biçimde sınıflandırılması hedeflenmiştir.

Araştırma Problemi

Bu tez çalışması, çok kaynaklı ve heterojen yapıya sahip epidemik verilerin analizinde karşılaşılan temel metodolojik sınırlılıkları ele almaktadır. Mevcut yaklaşımlar, farklı veri gruplarının doğal çoklu görünüm yapısını yeterince dikkate almamakta ve sınıf dengesizliği gibi kritik veri problemleri nedeniyle özellikle yüksek risk sınıfına ilişkin güvenilir tahminler üretmekte zorlanmaktadır. Bu durum, yalnızca sınıflandırma performansını değil, aynı zamanda elde edilen risk çıktılarının mekânsal analizlerde tutarlılığını da olumsuz etkilemektedir.

Bu bağlamda, epidemik riskin hem istatistiksel hem de mekânsal boyutlarıyla güvenilir biçimde tahmin edilebilmesi için çoklu veri görünümlerini birlikte değerlendirebilen, sınıf dengesizliği problemini etkin biçimde gideren ve bölgesel farklılıkları yansıtabilen bütünleşik bir modelleme yaklaşımına ihtiyaç duyulmaktadır. Bu çalışma, söz konusu ihtiyaca yanıt verecek bir yöntemsel çerçeve geliştirmeyi hedeflemektedir.

Çalışmanın Amacı

Bu çalışmanın temel amacı, çok kaynaklı epidemik verilerin birlikte değerlendirilmesine dayalı olarak bölgesel salgın riskini güvenilir biçimde sınıflandırabilecek bütünleşik bir modelleme çerçevesi geliştirmektir. Bu kapsamda, makine öğrenimi ve derin öğrenme yaklaşımlarının epidemik risk tahminindeki etkinliği aşamalı olarak incelenmiş ve çoklu görünüm yapısını doğrudan öğrenebilen özgün bir derin öğrenme modeli önerilmiştir. Bu doğrultuda çalışmanın özel amaçları aşağıda listelenmiştir:

- Türkiye ve komşu ülkeleri kapsayan demografik, coğrafi, sağlık altyapısı ve mobilite verilerinden oluşan çok kaynaklı bir epidemik veri seti oluşturmak.
- Bölgesel salgın riskinin sınıflandırılmasında, farklı ML modellerinin performansını değerlendirmek ve temel karşılaştırma çerçevesini oluşturmak.
- Çok kaynaklı epidemik verilerin analizinde, farklı DL modellerinin temsil ve sınıflandırma kapasitelerini incelemek.
- Veri setinde mevcut olan sınıf dengesizliği problemini gidermek amacıyla CGAN yöntemi ile sentetik örnek üretimi uygulamak.
- Epidemik verilerin doğal çoklu görünüm yapısına uygun olarak tasarlanmış MV-RiskNet modelini geliştirmek ve eğitmek.
- Önerilen modelinin performansını, ML ve diğer DL modelleri ile karşılaştırmalı olarak değerlendirmek.

- Elde edilen risk tahminlerini coğrafi risk haritaları aracılığıyla görselleştirerek bölgesel salgın risk dağılımlarını yorumlamak.

Bu amaç doğrultusunda geliştirilen yaklaşımın, epidemik risk analizine hem yöntemsel hem de uygulamalı açıdan özgün katkılar sunması hedeflenmektedir.

Çalışmanın Bilimsel ve Pratik Katkıları

Bu tez çalışması, epidemik risk analizine hem yöntemsel hem de uygulamaya dönük açılardan çeşitli katkılar sunmaktadır. Çalışmanın bilimsel katkıları, öncelikle farklı yapısal özelliklere sahip epidemik veri kaynaklarının tek bir analiz çerçevesi altında bütünleştirilmesine dayanmaktadır. Bu kapsamda, demografik yapı, coğrafi özellikler, sağlık altyapısı ve insan hareketliliği verileri birlikte ele alınarak çok kaynaklı epidemik verilerin analizi için bütüncül bir yaklaşım ortaya konulmuştur. Ayrıca, epidemik veri setlerinde sıklıkla karşılaşılan sınıf dengesizliği problemi, koşullu üretken modellerden yararlanan CGAN yöntemi ile ele alınmış ve bu yaklaşımın epidemik risk sınıflandırma performansını artırmadaki etkisi gösterilmiştir.

Çalışmanın bir diğer önemli bilimsel katkısı, epidemik verilerin doğal çoklu görünüm yapısını doğrudan modelleyebilen çok-görünümlü bir derin öğrenme modelinin geliştirilmesidir. Önerilen model, farklı veri görünümlerini ayrı alt ağlarda işleyerek bu temsilleri dikkat tabanlı bir füzyon mekanizması ile birleştirmekte ve bu yönüyle literatürdeki tek-görünüm yaklaşımlarına kıyasla daha zengin bir temsil gücü sunmaktadır. Bununla birlikte, ML ve DL modelleri arasında gerçekleştirilen kapsamlı karşılaştırmalı analizler, önerilen yaklaşımın performansını nesnel biçimde ortaya koyarak literatüre karşılaştırmalı bir değerlendirme çerçevesi kazandırmaktadır.

Pratik katkılar açısından değerlendirildiğinde, bu çalışma bölgesel salgın riskinin coğrafi olarak görselleştirilmesine olanak tanıyan bir analiz altyapısı sunmaktadır. Üretilen risk haritaları, karar vericilerin risk kümelenmelerini ve bölgesel farklılıkları hızlı ve sezgisel biçimde değerlendirmesine imkân tanımaktadır. Ayrıca, geliştirilen model yapısı gerçek veriye dayalı erken uyarı ve karar destek sistemlerinde kullanılabilecek nitelikte olup, sağlık kaynaklarının bölgesel düzeyde planlanmasına ve salgın yönetimine veri temelli bir bakış açısı kazandırmaktadır.

Bu yönleriyle tez çalışması, hem akademik literatüre özgün yöntemsel katkılar sunmakta hem de sağlık yönetimi ve politika geliştirme süreçlerinde uygulanabilir bir çerçeve ortaya koymaktadır.

Tezin Organizasyonu

Bu tez toplam altı bölümden oluşmaktadır. Birinci bölümde çalışmanın genel çerçevesi sunularak araştırma probleminin tanımı, çalışmanın amacı ve tezin literatüre sağladığı katkılar açıklanmıştır. İkinci bölümde epidemik risk analizi, makine öğrenmesi yöntemleri, derin öğrenme modelleri, çoklu görünüm yaklaşımı ve dikkat tabanlı füzyon teknikleri gibi kuramsal temeller ile literatürde yer alan ilgili çalışmalar detaylı biçimde incelenmiştir. Üçüncü bölümde veri setlerinin yapısı, veri ön işleme süreçleri, sınıf dengesizliğini gidermek için kullanılan CGAN yöntemi, uygulanan makine öğrenimi algoritmaları ve önerilen MV-RiskNet modeli açıklanmaktadır. Dördüncü bölümde deneysel çalışmaların sonuçları sunulmuş olup makine öğrenimi ve derin öğrenme yöntemlerinin performansları karşılaştırılmış ve elde edilen risk seviyeleri coğrafi haritalar üzerinden görselleştirilmiştir. Beşinci bölümde bulgular literatür ışığında tartışılarak modelin güçlü yönleri ve sınırlılıkları değerlendirilmiştir. Son bölüm olan altıncı bölümde ise çalışmanın genel sonuçları özetlenmiş ve gelecekte yapılabilecek araştırmalar için önerilerde bulunulmuştur.

KURAMSAL TEMELLER

Epidemik Modelleme ve Salgın Risk Kavramı

Epidemik modelleme, bulaşıcı hastalıkların yayılım dinamiklerini anlamak, risk faktörlerini belirlemek ve geleceğe yönelik öngörülerde bulunmak amacıyla kullanılan temel yaklaşımlardan biridir. Bu modeller, salgınların ortaya çıkışı, yayılma hızı ve etki alanlarının analiz edilmesine olanak tanıyarak halk sağlığı politikalarının şekillendirilmesinde önemli bir rol üstlenmektedir. Bu kapsamda SIR ve SEIR gibi bölmeli modeller, bireyleri hastalığa duyarlı, enfekte ve iyileşmiş gruplara ayırarak salgının yayılım sürecini temsil eden yaygın epidemik çerçeveler olarak kullanılmaktadır. Özellikle Covid-19 bağlamında yapılan çalışmalar, bu tür klasik epidemik modellerin salgının temel yayılım dinamiklerini açıklamak amacıyla hâlen referans çerçeve olarak kullanıldığını göstermektedir (Kang vd., 2020).

Bununla birlikte, günümüz salgınlarının çok boyutlu yapısı, yalnızca bu tür teorik çerçevelerle açıklanamayacak ölçüde karmaşık dinamikler içermektedir. Covid-19 pandemisi, salgınların yalnızca biyolojik süreçlerden değil aynı zamanda nüfus yoğunluğu, insan hareketliliği, sosyoekonomik yapı ve sağlık sistem kapasitesi gibi çok sayıda bileşenin etkileşiminden etkilendiğini ortaya koymuştur. Salgının erken dönem bulaş dinamiklerini ele alan çalışmalar, yayılımın kısa sürede hızlanabildiğini ve erken aşamadaki yayılım karakteristiklerinin süreç üzerinde belirleyici olduğunu göstermektedir (Li vd., 2020). Ayrıca vaka tespiti ve bildirimle ilişkin eksikliklerin, epidemik değerlendirmelerde belirsizliği artırarak risk tahmin süreçlerini doğrudan etkileyebildiği vurgulanmaktadır (Pullano vd., 2021).

Bu bağlamda epidemik risk kavramı, yalnızca hastalığın varlığına değil, belirli bir bölgenin salgına karşı kırılganlığına ve müdahale kapasitesine bağlı olarak değişen çok boyutlu bir yapıya karşılık gelmektedir. Epidemik risk, bir bölgede salgının ortaya çıkma olasılığı, yayılma potansiyeli ve olası etkilerinin büyüklüğünü birlikte ele alan bütüncül bir değerlendirme alanıdır. Dolayısıyla epidemik risk analizi, farklı veri kaynaklarının birlikte değerlendirilmesini gerektiren karmaşık bir problem hâline gelmiştir.

Mekansal Epidemik ve Coğrafi Risk Analizi

Mekânsal epidemik, bulaşıcı hastalıkların coğrafi dağılımını ve mekânsal örüntülerini inceleyerek salgınların belirli bölgelerde neden yoğunlaştığını ve hangi mekânsal faktörlerin yayılım üzerinde etkili olduğunu ortaya koymayı amaçlayan bir araştırma alanıdır. Bu yaklaşım, epidemik riskin yalnızca zamansal değil, aynı zamanda mekânsal heterojenlik gösterdiği varsayımına dayanmakta ve bölgesel farklılıkların anlaşılmasını hedeflemektedir. Bu

kapsamda GIS, epidemik verilerin mekânsal olarak analiz edilmesi ve görselleştirilmesinde yaygın biçimde kullanılan temel araçlardan biridir (Franch-Pardo vd., 2020).

Covid-19 sürecinde gerçekleştirilen çok sayıda çalışma, salgının mekânsal yayılımında insan hareketliliği ve coğrafi yakınlığın belirleyici rol oynadığını göstermiştir. Çin’de yapılan analizler, şehirler arası nüfus hareketliliğinin Covid-19 vakalarının mekânsal-zamansal dağılımını doğrudan etkilediğini ortaya koymuştur (Jia vd., 2020). Benzer biçimde, hareketliliğin kısıtlanmasına yönelik önlemlerin salgının yayılım hızını anlamlı ölçüde etkilediği rapor edilmiştir (Kraemer vd., 2020). Bu bulgular, epidemik süreçlerin mekânsal bağlamdan bağımsız değerlendirilemeyeceğini açıkça ortaya koymaktadır.

Avrupa ölçeğinde gerçekleştirilen mekânsal analizler ise, sosyoekonomik yapı, yaşlı nüfus oranı ve sağlık hizmetlerine erişim gibi değişkenlerin bölgesel risk farklılıklarını açıklamada önemli olduğunu göstermektedir. Özellikle sosyo-demografik kırılganlığın yüksek olduğu bölgelerde Covid-19 kaynaklı mortalite oranlarının daha yüksek seyrettiği rapor edilmiştir (Sannigrahi vd., 2020). Bu tür çalışmalar, epidemik riskin yalnızca bulaşma mekanizmalarıyla değil, bölgesel kırılganlık göstergeleriyle birlikte ele alınması gerektiğini vurgulamaktadır.

GIS tabanlı mekânsal modelleme yaklaşımları, Covid-19 insidansının coğrafi düzlemde temsil edilmesi ve risk haritalarının oluşturulması açısından önemli katkılar sunmuştur. Küresel ve ulusal ölçekte yapılan çalışmalar, GIS tabanlı yöntemlerin epidemik riskin mekânsal dağılımını açıklamada etkili olduğunu göstermektedir (Mollalo vd., 2020; Leslie vd., 2021). Ayrıca mekânsal ve mekânsal-zamansal istatistik yöntemlerinin Covid-19 epidemisinde yaygın biçimde kullanıldığı, bu alanda yapılan sistematik derlemeler aracılığıyla ortaya konmuştur (Nazia vd., 2022).

Bununla birlikte, mekânsal epidemi literatüründeki çalışmaların önemli bir bölümü betimleyici analizlerle sınırlı kalmaktadır. Riskin tahmin edilmesi veya sınıflandırılması çoğu zaman ikincil düzeyde ele alınmaktadır. GIS tabanlı analizlerin, ML ve DL tabanlı sınıflandırma ve tahmin modelleriyle bütünleştirilmediği durumlarda, epidemik riskin dinamik ve çok boyutlu doğası tam olarak yansıtılamamaktadır. Bu durum, mekânsal analizlerin veri temelli öğrenme yaklaşımlarıyla entegre edildiği bütünsel çerçevelere olan ihtiyacı ortaya koymaktadır.

Veri Tabanlı Epidemik Yaklaşımlar

Covid-19 pandemisi ile epidemik risk analizinde veri tabanlı yaklaşımların kullanımı belirgin biçimde artmıştır. Geleneksel epidemik modellerin sınırlı varsayımlarına karşılık, veri

temelli yöntemler salgın süreçlerini çok sayıda değişken üzerinden ele alarak daha esnek ve genellenebilir analizler sunabilmektedir. Bu kapsamda demografik yapı, nüfus yoğunluğu, sağlık altyapısı, mobilite verileri, sosyoekonomik göstergeler ve coğrafi özellikler gibi farklı veri türleri epidemik riskin modellenmesinde yaygın biçimde kullanılmaktadır.

Literatürde yer alan çalışmalar, çok kaynaklı veri yapılarının birlikte kullanılmasıyla geliştirilen modellerin, yalnızca vaka sayısına veya zamansal verilere dayalı yaklaşımlara kıyasla daha başarılı sonuçlar ürettiğini göstermektedir. Özellikle sosyoekonomik göstergelerin ve çevresel faktörlerin modele dâhil edilmesi, bölgesel risk farklılıklarının daha doğru biçimde temsil edilmesini sağlamaktadır (Razavi-Termeh vd., 2021). Buna ek olarak mobilite verilerinin entegrasyonu, salgının mekânsal yayılım potansiyelinin değerlendirilmesinde önemli bir bileşen olarak öne çıkmaktadır (Ilin vd., 2021).

Bununla birlikte veri tabanlı epidemik yaklaşımların temel zorluklarından biri, kullanılan veri kaynaklarının heterojen yapıya sahip olmasıdır. Farklı veri türlerinin değişken ölçeklerde, farklı zamansal çözünürlüklerde ve farklı istatistiksel dağılımlarda sunulması, bu verilerin tek bir model yapısı altında birlikte ele alınmasını zorlaştırmaktadır. Bu durum, heterojen veri kaynaklarının doğrudan birleştirilmesi hâlinde bilgi kaybına ve model performansında düşüşe yol açabilmektedir.

Ayrıca veri tabanlı epidemik analizlerde veri kalitesi ve raporlama eksiklikleri önemli bir belirsizlik kaynağı oluşturmaktadır. Covid-19 sürecinde vaka tespiti ve bildirimle ilişkin eksikliklerin epidemik değerlendirmeleri doğrudan etkilediği ve risk tahminlerinde sapmalara neden olabildiği rapor edilmiştir (Pullano vd., 2021). Bu nedenle veri temelli epidemik modellerin, yalnızca yüksek doğruluk üretmeye odaklanmak yerine, veri yapısının özelliklerini ve sınırlılıklarını dikkate alacak biçimde tasarlanması gerekmektedir.

Bu çerçevede veri tabanlı epidemik risk analizi, farklı veri kaynaklarının etkin biçimde bütünleştirilmesini gerektiren bir temsil ve öğrenme problemi olarak ele alınmaktadır. Heterojen veri yapılarını dikkate alan makine öğrenmesi ve derin öğrenme tabanlı yaklaşımlar, epidemik riskin çok boyutlu doğasının modellenmesi açısından literatürde giderek daha fazla önem kazanmaktadır.

Makine Öğrenmesi Yöntemleri Kullanılarak Salgın Risk Analizi

ML yöntemleri, çok değişkenli ve karmaşık veri yapılarında örüntü öğrenme yetenekleri sayesinde epidemik risk analizi çalışmalarında yaygın biçimde kullanılmaktadır. Covid-19 pandemisi sürecinde denetimli öğrenme tabanlı algoritmalar, vaka sayısı tahmini, mortalite riski değerlendirmesi ve bölgesel risk sınıflandırması gibi farklı görevlerde uygulanmıştır.

Literatürde LR, RF, SVM ve XGB gibi algoritmaların epidemik veri kümeleri üzerinde anlamlı sınıflandırma performansları üretebildiği rapor edilmiştir (Rustam vd., 2020; Chumachenko vd., 2022).

Özellikle ağaç tabanlı yöntemler ve boosting yaklaşımları, heterojen özellik uzaylarında doğrusal olmayan ilişkileri modelleyebilme kabiliyetleri nedeniyle epidemik risk tahmini çalışmalarında öne çıkmaktadır. XGB algoritması, ölçeklenebilir yapısı ve aşırı öğrenmeye karşı görece dayanıklı mimarisi sayesinde farklı epidemik tahmin senaryolarında başarılı biçimde uygulanmıştır (Chen & Guestrin, 2016). Benzer şekilde RF tabanlı modellerin, çok sayıda değişken içeren epidemik veri setlerinde kararlı performans sergilediği gösterilmiştir (Chumachenko vd., 2022; Sun vd., 2024).

ML yöntemleri, yalnızca tahmin performansı açısından değil, aynı zamanda karar destek sistemleri için sundukları yorumlanabilirlik düzeyi nedeniyle de tercih edilmektedir. Özellikle bölgesel risk sınıflandırması bağlamında, ML tabanlı modellerin ürettiği risk skorları ve sınıf etiketleri, GIS tabanlı haritalandırma ile ele alındığında karar vericiler için pratik ve uygulanabilir çıktılar sunabilmektedir (Razavi-Termeh vd., 2021). Bu yönüyle ML yaklaşımları, epidemik riskin mekânsal boyutunu destekleyen önemli bir araç hâline gelmiştir.

Bununla birlikte ML tabanlı epidemik risk analizlerinin temel sınırlılığı, heterojen veri kaynaklarını çoğunlukla tek bir birleşik özellik uzayı altında ele almasıdır. Demografik yapı, mobilite göstergeleri, sağlık altyapısı ve coğrafi değişkenlerin doğrudan bir araya getirilmesi, her bir veri türünün yapısal özelliklerinin yeterince korunamamasına yol açabilmektedir. Bu durum, modelin epidemik riskin çok boyutlu doğasını tam olarak temsil edememesine ve genellenebilirliğinin sınırlandırılmasına neden olmaktadır.

Ayrıca epidemik risk sınıflandırması problemleri çoğu zaman sınıf dengesizliği içermektedir. Özellikle yüksek risk sınıfının düşük sayıda örnekle temsil edilmesi, ML modellerinin bu kritik sınıfa ilişkin örüntüleri öğrenmesini zorlaştırmakta ve Rec. ile F1 gibi performans ölçütlerinde düşüşe neden olabilmektedir. Bu bağlamda yalnızca Acc. değerine dayalı değerlendirmelerin yanıltıcı olabileceğini Prec., Rec., F1 ve ROC–AUC gibi ölçütlerin birlikte ele alınmasının daha sağlıklı bir performans değerlendirmesi sunduğu literatürde vurgulanmaktadır (Lever vd., 2016; Tharwat, 2021; Chicco & Jurman, 2020).

Bu sınırlılıklar, epidemik risk analizi bağlamında ML yöntemlerinin tek başına yeterli olmadığını, heterojen veri kaynaklarını yapısal olarak ayırabilen ve daha güçlü temsil öğrenme yeteneklerine sahip modellere ihtiyaç duyulduğunu göstermektedir. Bu ihtiyaç, bir sonraki aşamada ele alınacak olan derin öğrenme ve çoklu görünümlü öğrenme yaklaşımlarının temel motivasyonunu oluşturmaktadır.

Derin Öğrenme Modelleri Kullanılarak Salgın Risk Analizi

DL yöntemleri, yüksek boyutlu ve karmaşık veri setlerinde otomatik özellik öğrenme yetenekleri sayesinde epidemik süreçlerin modellenmesinde önemli bir araştırma alanı hâline gelmiştir. Covid-19 pandemisi sürecinde özellikle zaman serisi tabanlı tahmin problemlerinde LSTM ve GRU gibi tekrarlayan sinir ağı mimarilerinin yaygın biçimde kullanıldığı görülmektedir. Bu modeller, salgının zamansal dinamiklerini öğrenebilme kabiliyetleri sayesinde kısa vadeli vaka tahmini ve eğilim analizi çalışmalarında başarılı sonuçlar üretmiştir (Ribeiro vd., 2020; Zeroual vd., 2020).

Literatürde yer alan DL tabanlı epidemik çalışmaların önemli bir bölümü, tek değişkenli veya sınırlı sayıda girdi içeren zaman serileri üzerinden yürütülmüştür. Bu yaklaşımlar, salgının zamansal seyrini modellemede etkili olmakla birlikte, epidemik riskin mekânsal ve sosyo-demografik belirleyicilerini çoğu zaman model dışında bırakmaktadır. Benzer şekilde, CNN tabanlı modellerin uzamsal örüntüleri veya karmaşık veri yapılarındaki ilişkileri öğrenmede güçlü olduğu gösterilmiş olsa da bu tür yaklaşımlar genellikle epidemik risk sınıflandırması bağlamında çok kaynaklı veri entegrasyonuna odaklanmamaktadır.

Derin öğrenme yöntemlerinin bir diğer önemli avantajı, manuel özellik mühendisliği ihtiyacını azaltarak veriden doğrudan temsil öğrenebilmesidir. Ancak bu avantaj, heterojen veri kaynaklarının tek bir temsil altında birleştirilmesi durumunda sınırlı kalabilmektedir. Demografik göstergeler, mobilite verileri, sağlık altyapısı ve coğrafi değişkenler gibi farklı veri türlerinin yapısal özellikleri göz ardı edildiğinde, DL modellerinin temsil gücü epidemik riskin çok boyutlu doğasını tam olarak yansıtamamaktadır. Bu durum, DL tabanlı modellerin tek görünüm üzerinden öğrenme yaptığı senaryolarda genellenebilirliğinin azalmasına yol açmaktadır. Ayrıca DL modellerinin epidemik risk gibi yüksek etkili alanlarda kullanımı, model güvenilirliği ve dayanıklılığı açısından da tartışılmaktadır (Goodfellow vd., 2014; Madry vd., 2017). Bu tür çalışmalar, DL modellerinin yüksek temsil gücüne rağmen, uygun mimari tasarım ve veri yapısına duyarlı öğrenme stratejileriyle desteklenmediği sürece sınırlılıklar barındırabileceğini ortaya koymaktadır.

Bu bağlamda literatür, epidemik risk analizi için DL yöntemlerinin tek başına yeterli olmadığını, heterojen veri kaynaklarını ayrı temsiller altında ele alabilen ve bu temsilleri bütünleştirebilen daha gelişmiş öğrenme çerçevelerine ihtiyaç duyulduğunu göstermektedir. Bu ihtiyaç, bir sonraki bölümde ele alınacak olan çoklu görünümlü öğrenme ve veri füzyonu yaklaşımlarının temel motivasyonunu oluşturmaktadır.

Çoklu Görünümlü Öğrenme Yaklaşımları

Çoklu görünümlü öğrenme, aynı probleme ait farklı veri kaynaklarının ayrı temsil alanlarında öğrenilmesini ve bu temsillerin daha sonra bütünleştirilmesini amaçlayan bir öğrenme yaklaşımıdır. Bu yaklaşım, heterojen veri kaynaklarının tek bir özellik uzayına indirgenmesi sonucu ortaya çıkabilecek bilgi kaybını azaltmayı hedeflemektedir.

Literatürde multimodal ve çoklu görünümlü öğrenme yaklaşımlarının, farklı veri türlerinin birlikte kullanıldığı problemlerde tek görünüm tabanlı modellere kıyasla daha güçlü temsiller üretebildiği gösterilmiştir (Ramachandram & Taylor, 2017; Baltrušaitis vd., 2019; Yan vd., 2021). Özellikle derin öğrenme tabanlı mimarilerde, her bir veri görünümü için ayrı alt ağların kullanılması, veri kaynaklarının yapısal özelliklerinin korunmasına olanak tanımaktadır (Gao vd., 2020; Zhao vd., 2024).

Epidemik risk analizi bağlamında değerlendirildiğinde ise demografik yapı, mobilite, sağlık altyapısı ve coğrafi göstergeler doğal olarak farklı veri görünümelerini temsil etmektedir. Buna rağmen literatürdeki epidemik risk çalışmalarının büyük bir bölümü bu veri kaynaklarını tek bir birleşik yapı altında ele almakta veya yalnızca belirli bir veri görünümüne odaklanmaktadır. Bu durum, epidemik riskin çok boyutlu yapısının yeterince temsil edilememesine yol açmaktadır.

Bu nedenle çoklu görünümlü öğrenme yaklaşımları, epidemik risk analizinde heterojen veri kaynaklarının bütüncül ve yapısal olarak tutarlı biçimde modellenenebilmesi için uygun bir çerçeve sunmaktadır.

Üretken Modeller ve CGAN Yöntemi ile Veri Dengeleme

Epidemik risk analizi çalışmalarında sık karşılaşılan temel problemlerden biri sınıf dengesizliğidir. Özellikle yüksek risk sınıfının düşük sayıda örnekle temsil edilmesi, sınıflandırma modellerinin bu kritik sınıfa ilişkin örüntüleri öğrenmesini zorlaştırmakta ve Rec. ile F1 gibi performans ölçütlerinde düşüşe neden olmaktadır. Bu durum, erken uyarı ve risk önceliklendirme sistemlerinin güvenilirliğini doğrudan etkilemektedir.

Bu problemi gidermek amacıyla literatürde örnekleme ve veri artırma yaklaşımları önerilmiş olup özellikle GAN ve CGAN tabanlı üretken modellerin, düşük temsil edilen sınıflar için gerçekçi sentetik örnekler üretebildiği gösterilmiştir. CGAN tabanlı veri artırma yöntemlerinin tıbbi sınıflandırma ve Covid-19 tahminleme çalışmalarında sınıflandırma performansını artırdığı rapor edilmiştir (Guo vd., 2023). Buna karşın epidemik risk seviyelerinin çok kaynaklı veri yapısı üzerinden sınıflandırıldığı senaryolarda CGAN kullanımının sınırlı kaldığı görülmektedir.

Bu durum, sınıf dengesizliği problemini doğrudan hedefleyen ve epidemik risk bağlamına uyarlanmış üretken model tabanlı yaklaşımlara duyulan ihtiyacı ortaya koymaktadır.

Literatür Değerlendirmesi ve Araştırma Boşlukları

Epidemik risk analizine yönelik literatür incelendiğinde, çalışmaların genel olarak mekânsal analizler, makine öğrenmesi ve derin öğrenme tabanlı tahmin modelleri ile mobilite odaklı yaklaşımlar etrafında yoğunlaştığı görülmektedir. GIS tabanlı mekânsal çalışmalar, salgının coğrafi heterojenliğini ve bölgesel risk farklılıklarını ortaya koymada önemli katkılar sunmuştur. Bununla birlikte bu çalışmalar çoğunlukla betimleyici analizlerle sınırlı kalmış ve elde edilen bulguların veri temelli sınıflandırma modelleriyle bütünleştirilmesi sınırlı düzeyde gerçekleştirilmiştir.

ML ve DL tabanlı çalışmalar, epidemik riskin tahmini ve sınıflandırılması açısından yüksek doğruluk değerleri üretmiş olsa da bu yaklaşımların büyük bir bölümü heterojen veri kaynaklarını tek bir özellik uzayı altında ele almaktadır. Demografik yapı, mobilite, sağlık altyapısı ve coğrafi göstergeler gibi farklı veri türlerinin yapısal özellikleri yeterince ayrıştırılmadan birlikte kullanılması, epidemik riskin çok boyutlu doğasının tam olarak temsil edilememesine yol açmaktadır. Bu durum, çoklu veri görünümelerini dikkate alan öğrenme yaklaşımlarına olan ihtiyacı açıkça ortaya koymaktadır.

Literatürde öne çıkan bir diğer önemli sınırlılık, epidemik veri setlerinde yaygın olarak karşılaşılan sınıf dengesizliği problemi. Yüksek risk sınıfının düşük sayıda örnekle temsil edilmesi, sınıflandırma modellerinin bu kritik sınıfa ilişkin performansını olumsuz etkilemektedir. Üretken modellerin sınıf dengesizliğini azaltmada etkili olduğu farklı alanlarda gösterilmiş olsa da bu yaklaşımların epidemik risk seviyelerinin tahminine sistematik biçimde entegre edildiği çalışmalar sınırlı sayıdadır.

Son olarak, epidemik risk tahmininden elde edilen çıktıların mekânsal bağlam ile değerlendirilmesi literatürde yeterince ele alınmamıştır. GIS tabanlı mekânsal analizler ve öğrenme tabanlı sınıflandırma modellerinin bütünleşik biçimde kullanıldığı çerçeveler, karar destek süreçleri açısından önemli potansiyel taşımasına rağmen sınırlı düzeyde kalmıştır.

Bu çalışma, literatürde belirlenen söz konusu boşlukları bütüncül bir yaklaşımla ele almayı hedeflemektedir. Bu kapsamda önerilen CGAN destekli MV-RiskNet modeli, çoklu veri görünümelerini DL tabanlı bir mimari altında bütünleştirerek, sınıf dengesizliğini CGAN tabanlı veri dengeleme ile azaltmakta ve elde edilen risk tahminlerini mekânsal risk haritalarıyla birlikte değerlendirerek literatürdeki temel metodolojik boşluklara doğrudan yanıt vermektedir.

MATERYAL VE YÖNTEM

Bu çalışmada iki farklı yöntemsel yaklaşım uygulanmıştır. Birinci aşamada, Türkiye'nin 81 ili için ML tabanlı risk sınıflandırması gerçekleştirilmiş ve elde edilen sonuçlar GIS kullanılarak mekânsal olarak haritalandırılmıştır. İkinci aşamada ise, çoklu veri görünümüne dayalı ve DL temelli bir mimariye sahip olan MV-RiskNet modeli kullanılarak Türkiye ve komşu ülkeleri kapsayan bölgesel epidemik risk analizi yapılmıştır. Bölümün devamında her iki aşamaya ait veri yapıları, modelleme süreçleri ve değerlendirme adımları ayrıntılı biçimde açıklanmaktadır.

1. Aşama: Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye'de Bölgesel Salgın Risk Analizi ve Haritalandırılması

Türkiye'de salgın risk veri setinin oluşturulması

Bu çalışmanın birinci aşamasında, Türkiye'deki illerin epidemik risk seviyelerinin belirlenmesi amacıyla ML tabanlı analizlerde kullanılmak üzere çok boyutlu bir veri seti oluşturulmuştur. Analiz birimi olarak Türkiye'nin 81 ili ele alınmış olup her il için epidemik riski etkileyebileceği düşünülen demografik, coğrafi, sağlık altyapısı, ulaşım ve ekonomik göstergeler derlenmiştir. Kullanılan verilerin büyük bölümü 2022 yılına ait olup, farklı güvenilir açık veri kaynaklarından temin edilmiştir. Bu çalışma kapsamında oluşturulan veri seti, araştırmaya özgü olarak derlenmiş olup şu an için açık erişime sunulmamıştır.

Demografik veriler (nüfus, nüfus yoğunluğu, yüzölçümü ve il-ilçe idari bilgileri) TÜİK kaynaklarından elde edilmiştir. Sağlık altyapısına ilişkin göstergeler (hastane sayısı ve sağlık hizmetlerine erişimi temsil eden değişkenler) TÜİK ve ilgili resmî istatistik kaynaklarından derlenmiştir. Ekonomik göstergeler ve ulaşım altyapısına ilişkin veriler yine TÜİK verileri ile açık kaynaklı resmî istatistik platformlarından temin edilmiştir.

Veri setinin coğrafi bileşenini oluşturan enlem ve boylam bilgileri, il ve ilçe merkezlerini temsil edecek şekilde Google Cloud Console ve Wikipedia gibi açık kaynaklardan elde edilmiştir. Bu coğrafi bilgiler, hem epidemik risk seviyelerinin mekânsal dağılımının analiz edilmesinde hem de sonuçların GIS ile haritalandırılmasında temel girdi olarak kullanılmıştır.

Oluşturulan veri setinde her satır bir ili temsil etmekte, sütunlar ise hastane sayısı, ekonomik göstergeler, ulaşım kapasitesi, nüfus, yüzölçümü, nüfus yoğunluğu, enlem-boylam bilgileri ve merkeze uzaklık gibi değişkenlerden oluşmaktadır. Buna ek olarak, sağlık ve nüfus

verilerinden türetilen hastane başına düşen nüfus ve ekonomi–nüfus oranı gibi bileşik öznitelikler hesaplanarak veri setine dâhil edilmiştir. Bu değişkenler, illerin epidemik açıdan taşıdığı yapısal risk faktörlerini daha doğru biçimde temsil edebilmek amacıyla kullanılmıştır. Veri setinin genel yapısı ve içerdiği değişkenler Şekil 1’de gösterilmiştir.

il	il_kod	Hastane	Ekonomi	Ulasim	Nufus	Yuzolcumu	Nufus Yogunlugu	Latitude	Longitude		
adana	0	90.0	56.0	1794.0	2237940.0	14045.0	159.3	36.9914194	35.3308285		
adiyaman	1	41.0	36.0	985.0	635169.0	7614.0	83.4	39.750545	35.837347		
afyonkarahisar	2	72.0	47.0	2141.0	744179.0	14230.0	52.3	38.7568852	30.5387038		
aksaray	4	44.0	57.0	960.0	429069.0	7617.0	56.3	38.368626	34.0297		
amasya	5	38.0	45.0	803.0	335494.0	5520.0	60.8	40.656455	35.837347		
ankara	6	239.0	101.0	3000.0	5663322.0	25437.0	222.5	39.9333625	32.8507119		
antalya	7	173.0	81.0	2217.0	2562271.0	20177.0	12				
							Merkez_Uzaklik	Ekonomi_Nufus_Orani	Hastane_Basina_Nufus	Risk Skoru	Risk Seviyesi
							3.842099154666449	2.5023012234465625e-05	24866.0	33.02116919199013	Orta
							2.9832562088486507	5.667782905022128e-05	15491.926829268292	9.411155828095348	Düşük
							2.602154959842604	6.315684801640466e-05	10335.819444444443	24.59582850050132	Düşük
							1.953821299678147	0.0001328457660655	9751.568181818182	31.176744211832187	Düşük
							3.0641785505472803	0.0001341305656733	8828.78947368421	22.054552919163907	Düşük
							5.5568516270985015e-05	1.7834055700876622e-05	23695.907949790795	72.62954555612271	Yüksek

Şekil 1. Türkiye il bazlı salgın risk veri setinin yapısı ve değişken içerikleri

Her il için oluşturulan öznitelikler kullanılarak bir Risk Skoru hesaplanmış ve bu skorlar düşük, orta ve yüksek olmak üzere üç farklı Risk Seviyesi altında sınıflandırılmıştır. Elde edilen risk seviyeleri, ML modellerinin eğitiminde hedef değişken olarak kullanılmış ve aynı zamanda sonuçların mekânsal olarak yorumlanabilmesine olanak sağlamıştır.

Sonuç olarak oluşturulan bu veri seti, Türkiye’nin il düzeyindeki epidemik risk dinamiklerini bütüncül biçimde temsil eden bir yapı sunmakta olup, çalışmanın birinci aşamasında gerçekleştirilen tüm sınıflandırma ve mekânsal analizlerin temelini oluşturmaktadır.

Veri ön işleme ve hazırlık süreci

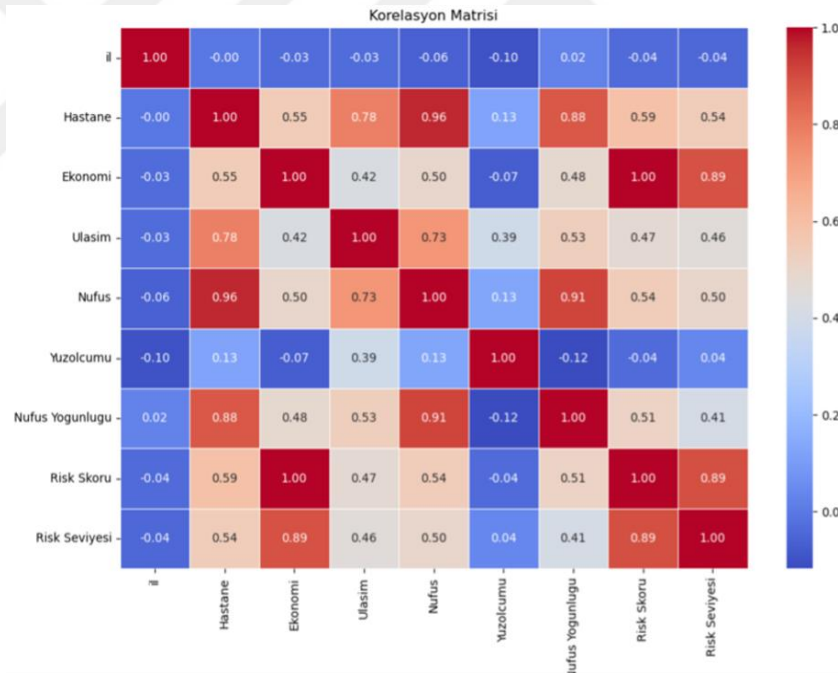
Bu çalışmada kullanılan veri seti, ML algoritmalarının gerektirdiği bütünlük ve tutarlılığı sağlamak amacıyla kapsamlı bir ön işleme sürecinden geçirilmiştir. İlk aşamada farklı kaynaklardan derlenen veri tabloları birleştirilmiş olup eksik, tutarsız veya aykırı değer içeren gözlemler analiz edilmiştir. Yapılan inceleme sonucunda eksik veri oranının sınırlı düzeyde olduğu görülmüş ve bu eksiklikler, benzer demografik ve coğrafi özelliklere sahip illerin ortalama değerleri kullanılarak tamamlanmıştır. Bu yaklaşım, veri setinin homojenliğini artırarak model tahminlerinde oluşabilecek yanlışlıkların azaltılmasına katkı sağlamıştır.

Ön işleme sürecinin devamında epidemik riskin yapısal belirleyicilerini daha iyi temsil edebilmek amacıyla bazı türetilmiş öznitelikler oluşturulmuştur. Bu kapsamda, sağlık altyapısı ve nüfus verileri kullanılarak hastane başına düşen nüfus, ekonomik göstergeler ile nüfus

büyüklüğü dikkate alınarak ekonomi–nüfus oranı gibi bileşik değişkenler hesaplanmış ve veri setine dâhil edilmiştir. Bu öznetelikler, iller arasındaki göreceli kırılma farklılıklarının daha açık biçimde modellenmesini sağlamaktadır.

Tüm sayısal değişkenler, farklı ölçeklerdeki değerlerin model üzerindeki etkisini dengelemek amacıyla SS yöntemi kullanılarak normalize edilmiştir. Normalizasyon işlemi sonucunda her değişkenin ortalaması sıfır, standart sapması bir olacak şekilde dönüştürülmüştür. Ayrıca hedef değişken olan Risk Seviyesi, düşük, orta ve yüksek olmak üzere üç sınıfa ayrılmış ve sırasıyla 0, 1 ve 2 değerleriyle etiketlenmiştir. Böylece problem, çok sınıflı sınıflandırma yapısına uygun hâle getirilmiştir.

Veri hazırlık sürecinin son aşamasında, değişkenler arasındaki ilişkileri incelemek amacıyla korelasyon analizi gerçekleştirilmiştir. Korelasyon matrisi, bazı demografik, sağlık altyapısı ve ekonomik değişkenlerin risk seviyesi ile anlamlı ilişkiler sergilediğini göstermiştir. Elde edilen bu bulgular, bir sonraki aşamada gerçekleştirilen öznetelik seçimi ve modelleme sürecine yön vermek amacıyla kullanılmıştır. Korelasyon matrisi Şekil 2’de gösterilmiştir.



Şekil 2. Korelasyon matrisi

Veri ön işleme ve hazırlık süreci

Ön işleme sürecinin ardından her bir il için epidemik riskin nicel olarak temsil edilebilmesi amacıyla bileşik bir RS tanımlanmıştır. Bu skor, normalize edilmiş özneteliklerin doğrusal birleşimine dayanmaktadır. Öncelikle, veri setinde yer alan her bir sayısal değişken x_j , z-skor normalizasyonu kullanılarak standartlaştırılmıştır. Denklem 1’de formül verilmiştir ve

burada x_{ij} , i . ile ait j . değişkenin ham değerini, μ_j ve σ_j sırasıyla j . değişkenin ortalama ve standart sapmasını, x'_{ij} ise normalize edilmiş değeri ifade etmektedir.

Normalize edilmiş öznitelikler kullanılarak her bir il için Risk Skoru ise Denklem 2'de gösterilmiştir. Bu denklemde RS_i , i . ilin epidemik risk skorunu, n , risk hesaplamasında kullanılan toplam öznitelik sayısını temsil etmektedir.

Bu yaklaşımda tüm özniteliklerin risk üzerindeki katkısının eşit olduğu varsayılmıştır. Böylece herhangi bir değişkene öznel veya veri dışı bir ağırlık atanmadan, epidemik risk skoru tamamen veri temelli ve şeffaf bir biçimde elde edilmiştir. Elde edilen sürekli Risk Skoru değerleri, risk seviyelerinin sınıflandırılabilmesi amacıyla yüzdelik dilimlere dayalı bir eşikleme yöntemi kullanılarak üç sınıfa ayrılmıştır ve burada Q_{33} ve Q_{66} , Risk Skoru dağılımının sırasıyla %33 ve %66 yüzdelik değerlerini göstermektedir.

$$Risk_Seviyesi = \begin{cases} \text{Düşük Risk,} & RS_i \leq Q_{33} \\ \text{Orta Risk,} & Q_{33} < RS_i \leq Q_{66} \\ \text{Yüksek Risk,} & RS_i > Q_{66} \end{cases}$$

Bu çalışmada risk seviyesi, doğrudan vaka sayısını temsil etmemekte olup bir bölgenin salgına karşı yapısal kırılganlığını ve potansiyel risk düzeyini ifade etmektedir. Risk seviyesi, demografik yoğunluk, sağlık altyapısı kapasitesi, ekonomik göstergeler, coğrafi konum ve mobilite örüntülerinin birlikte değerlendirilmesiyle elde edilen Risk Skoru üzerinden tanımlanmaktadır.

Bu sınıflandırma yöntemi sayesinde risk seviyeleri, veri setinin kendi dağılımına dayalı olarak belirlenmiş; sınıflar arasında dengeli bir örnek dağılımı sağlanmıştır. Tanımlanan risk seviyeleri, makine öğrenimi modellerinin hedef değişkeni olarak kullanılmış ve mekânsal risk analizlerinin temelini oluşturmuştur.

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (1)$$

$$RS_i = \frac{1}{n} \sum_{j=1}^n x'_{ij} \quad (2)$$

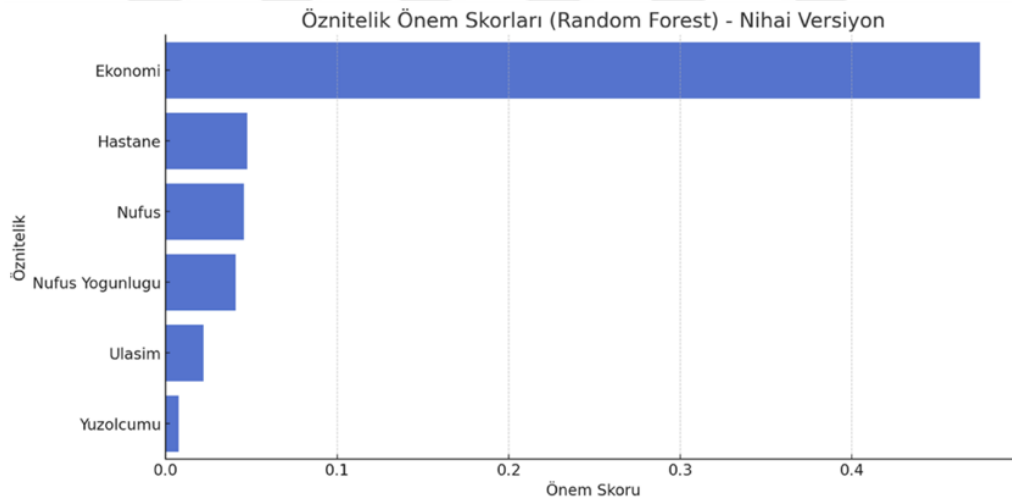
Öznitelik Seçimi ve Öznitelik Öneminin Belirlenmesi

Veri ön işleme sürecinin ardından, sınıflandırma performansına katkısı sınırlı olan değişkenlerin elenmesi ve modelin genellenebilirliğinin artırılması amacıyla öznitelik seçimi gerçekleştirilmiştir. Bu aşamada, epidemik risk sınıflandırması üzerinde etkili olan değişkenlerin göreceli önem düzeylerinin belirlenmesi hedeflenmiştir.

Öznitelik önemlerinin hesaplanmasında, ağaç tabanlı yapısı sayesinde değişkenler arasındaki doğrusal olmayan ilişkileri modelleyebilen ve öznitelik katkılarını doğrudan ölçebilen RF algoritmasından yararlanılmıştır. RF modeli kullanılarak her bir değişkenin epidemik risk sınıflandırmasına olan katkısı hesaplanmış ve bu katkılar önem skorlarına göre sıralanmıştır.

Elde edilen sonuçlar, ekonomik göstergelerin epidemik risk sınıflandırması üzerinde açık ara en yüksek etkiye sahip olduğunu göstermektedir. Ekonomik değişkenleri sırasıyla sağlık altyapısı (hastane sayısı), nüfus, nüfus yoğunluğu ve ulaşım göstergeleri takip etmektedir. Buna karşılık yüzölçümü değişkeninin sınıflandırma üzerindeki katkısının oldukça düşük olduğu belirlenmiştir. Bu bulgular, epidemik riskin yalnızca coğrafi büyüklükten ziyade sosyoekonomik yapı ve sağlık kapasitesi ile daha güçlü biçimde ilişkili olduğunu ortaya koymaktadır.

Önem skoru düşük olan değişkenler analiz dışında bırakılarak model karmaşıklığı azaltılmış, aşırı öğrenme riski düşürülmüş ve daha kararlı bir sınıflandırma yapısı elde edilmiştir. Nihai öznitelik seti, ML modellerinin eğitiminde kullanılmıştır. RF algoritmasına dayalı öznitelik önem sıralaması Şekil 3'te gösterilmiştir.



Şekil 3. Öznitelik önem sıralaması

Makine Öğrenimi Algoritmalarının Uygulanması

Öznitelik seçimi sürecinin tamamlanmasının ardından, Türkiye'deki illerin epidemik risk seviyelerini sınıflandırmak amacıyla farklı öğrenme prensiplerine sahip beş ML algoritması uygulanmıştır. Bu algoritmalar LR, NB, SVM, RF ve XGB modelleridir. Seçilen algoritmalar, epidemik veri setlerinde sık karşılaşılan doğrusal ve doğrusal olmayan ilişkileri farklı açılardan modelleyebilmeleri ve literatürde yaygın biçimde kullanılmaları nedeniyle tercih edilmiştir.

LR modeli, epidemik riskin temel belirleyicilerini doğrusal bir karar fonksiyonu üzerinden tahmin eden referans bir sınıflandırıcı olarak kullanılmıştır. NB algoritması, öznitelikler arasında koşullu bağımsızlık varsayımına dayanan olasılıksal bir yaklaşım sunmakta ve özellikle sınırlı veri koşullarında hızlı ve kararlı sonuçlar üretebilmektedir. SVM modeli, sınıflar arasındaki ayırıcı sınırı maksimize etmeyi amaçlayan optimizasyon tabanlı bir yöntem olup, yüksek boyutlu özellik uzaylarında güçlü genelleme performansı sergilemektedir.

RF modeli, çok sayıda karar ağacından oluşan topluluk yapısı sayesinde değişkenler arasındaki karmaşık ve doğrusal olmayan ilişkileri etkili biçimde öğrenebilmektedir. XGB algoritması ise gradyan artırma yaklaşımının optimize edilmiş bir sürümü olup, düzenleme mekanizmaları sayesinde aşırı öğrenmeyi azaltarak yüksek doğruluk ve genelleme kapasitesi sunmaktadır.

Tüm modeller, aynı eğitim–doğrulama–test veri bölünmesi altında eğitilmiş ve karşılaştırılmıştır. Böylece modeller arası performans farklarının veri bölünmesinden kaynaklanabilecek etkilerden arındırılarak nesnel biçimde değerlendirilmesi sağlanmıştır. Model eğitim sürecinde, öznitelik seçimi ve ölçekleme adımları yalnızca eğitim verisi üzerinde gerçekleştirilmiş; doğrulama ve test kümelerine aynı dönüşümler uygulanarak veri sızıntısı önlenmiştir.

Model performanslarının değerlendirilmesinde çok sınıflı sınıflandırma problemleri için yaygın olarak kullanılan Acc., Prec., Rec., F1 ve ROC–AUC ölçütleri dikkate alınmıştır. Özellikle sınıf dengesizliği içeren epidemik risk problemlerinde tek başına Acc. değerinin yanıltıcı olabileceği göz önünde bulundurularak, sınıf bazlı performans ölçütleri birlikte değerlendirilmiştir.

Uygulanan ML modellerine ait ayrıntılı performans sonuçları ve karşılaştırmalı değerlendirmeler, tezin Bulgular bölümünde verilmiştir.

Risk haritalandırma ve simülasyon süreci

ML modelleri kullanılarak elde edilen epidemik risk sınıflandırma çıktıları, sonuçların mekânsal bağlamda değerlendirilebilmesi amacıyla GIS tabanlı bir analiz sürecine dâhil edilmiştir. Bu kapsamda, her bir il için tahmin edilen Risk Seviyesi değerleri, ilgili coğrafi konum bilgileri (enlem–boylam) ile eşleştirilmiştir.

Mekânsal analiz sürecinde, düşük, orta ve yüksek risk sınıflarının coğrafi dağılımlarını temsil edebilmek amacıyla il bazlı tematik haritalar oluşturulmuştur. Haritalandırma aşamasında, sınıflandırma sonuçları renk temelli bir gösterim ile kodlanmıştır böylece epidemik

risk seviyelerinin ülke genelindeki dağılım örüntüleri görsel olarak analiz edilebilir hâle getirilmiştir.

Sınıflandırma sonuçlarının mekânsal yoğunluğunu ve olası kümelenme eğilimlerini incelemek amacıyla, risk seviyelerine dayalı mekânsal yoğunluk analizleri gerçekleştirilmiştir. Bu analizler, yüksek risk sınıfına ait bölgelerin coğrafi olarak yakınlık gösterip göstermediğinin değerlendirilmesine ve epidemik riskin mekânsal süreklilik özelliklerinin incelenmesine olanak sağlamıştır.

Elde edilen mekânsal analiz çıktıları, epidemik riskin yalnızca il bazlı özniteliklere değil, aynı zamanda mekânsal komşuluk ilişkilerine ve bölgesel etkileşimlere bağlı olarak şekillendiğini değerlendirebilmek amacıyla kullanılmıştır. Bu süreç, ML tabanlı sınıflandırma sonuçlarının coğrafi bağlam ile bütünleştirilmesini sağlayarak, karar destek ve bölgesel önceliklendirme analizleri için temel bir altyapı sunmaktadır.

2. Aşama: Derin Öğrenme Modelleri Kullanılarak Türkiye ve Komşu Ülkelerinin Bölgesel Salgın Risk Tahmini ve Haritalandırması

Türkiye ve komşu ülkelerinin epidemik risk veri seti

Çalışmanın ikinci aşamasında, bölgesel epidemik riskin çok boyutlu ve sınır-ötesi yapısını temsil edebilmek amacıyla Türkiye ve komşu ülkeleri kapsayan çok kaynaklı bir veri seti oluşturulmuştur. Bu veri seti, epidemik riskin belirleyicileri olarak kabul edilen demografik yapı, ekonomik göstergeler, sağlık altyapısı, coğrafi konum ve insan hareketliliği bileşenlerini birlikte içerecek şekilde tasarlanmıştır. Çalışma kapsamına Türkiye'nin yanı sıra Yunanistan, Bulgaristan, Gürcistan, Ermenistan, İran ve Irak dâhil edilmiştir. Bu ülkelerin seçilmesinin temel nedeni, Türkiye ile olan kara sınırları ve mekânsal sürekliliklerinin sınır-ötesi epidemik risk dinamiklerinin değerlendirilmesine olanak sağlamasıdır.

Oluşturulan veri setinde her satır bir idarî bölgeyi temsil etmektedir ve sütunlar ise ilgili bölgenin epidemik riskini etkileyebilecek nicel göstergelerden oluşmaktadır. Ham veriler TÜİK, Eurostat, WHO, UNData ve çeşitli uluslararası açık veri portallarından derlenmiştir. Ülkeler arasındaki idarî sınıflandırma ve veri tanım farklılıkları, gerekli durumlarda güvenilir açık kaynaklar (ör. Wikipedia) kullanılarak uyumlaştırılmıştır. Coğrafi konum bilgileri (enlem–boylam), her idarî birimin merkezini temsil edecek şekilde Google Maps / Geocoding servisleri aracılığıyla elde edilmiştir.

Veri seti, çok-görünümlü öğrenme yaklaşımına uygun olarak dört ana tematik görünüm altında organize edilmiştir. Demografik görünüm, nüfus büyüklüğü, nüfus yoğunluğu ve ekonomik göstergeleri; sağlık görünümü, hastane sayısı ve hastane başına düşen nüfus gibi

sağlık kapasitesini temsil eden değişkenleri; coğrafi görünüm, bölgesel alan bilgisi ile enlem-boylam koordinatlarını; mobilite görünümü ise ulaşım ve erişilebilirliği temsil eden göstergeleri içermektedir. Bu tematik ayrışma, heterojen veri kaynaklarının yapısal özelliklerinin korunarak modellenmesini mümkün kılmaktadır

Veri setinin hedef değişkeni olan araştırmacılar, düşük, orta ve yüksek olmak üzere üç sınıf altında tanımlanmış ve sırasıyla 0, 1 ve 2 değerleri ile etiketlenmiştir. Sınıf dağılımı incelendiğinde, özellikle yüksek risk sınıfının diğer sınıflara kıyasla daha düşük örnek sayısına sahip olduğu görülmüş; bu dengesiz yapı, ilerleyen aşamada CGAN tabanlı veri artırma yöntemi ile ele alınmıştır. Veri setinin genel yapısı ve görünüm bazı değişken grupları Şekil 4'te gösterilmektedir.

City	Area_km2	Population	Population_Density	Economic_EUR	Economic_Per_Capita	Hospitals	Population_Per_Hospital	Latitude	Longitude	Mobility_Index	Country	Risk_Score	Risk_Level	Country_English
Pazardzhik	4456.92	275548.0	51.56	6725.669902884434	0.0244090681220129	12.0	22962.33333333333	42.16558369218918	24.152004797200973	78.0	Bulgaristan	0.403468931934367	high	Bulgaria
Ararat	2090.0	260367.0	119.1	4823.245470771534	0.0185247956567903	19.0	13703.526315789471	40.0053589	44.7315157	2.0	Ermenistan	0.1455892823131937	low	Armenia
Tavush	2704.0	128609.0	42.5	3515.2518350645287	0.0273328603368701	17.142857142857142	7502.191666666667	40.8275136	45.1847828	1.0	Ermenistan	-0.033468400014101	low	Armenia
Aragatsotn	2773.0	132925.0	46.5	3310.9651797000406	0.0249085211939066	13.0	10225.0	40.3827655	44.114727	11.0	Ermenistan	0.042296674183827	medium	Armenia
Imereti	6475.0	463100.0	83.0	4211.058826940583	0.0090931954803294	92.0	5033.695652173913	42.18756	42.9521554	78.0	Gürcistan	-0.083246537909417	low	gucistan
Samegrelo-Zemo Svaneti	7440.0	209300.0	45.0	4520.205567997884	0.0151025912729631	38.0	7876.315789473684	42.6629046	42.2328587	13.0	Gürcistan	-0.013791688288107	medium	gucistan
Al-Anbar	53478.0	1771656.0	29.1	55.315478851424885	3.122247143431055e-05	118.0	15014.033998305084	32.7889443	41.6093983	17.0	Irak	0.2638073126563303	high	Iraq
Karbala	1944.0	1218732.0	548.6	54.5640084458273	4.477186817596264e-05	267.0	4564.539325842697	32.5984057	44.0157931	22.0	Irak	-0.0553962574243276	low	Iraq
Sulaymaniyah	6573.0	2053305.0	285.8	68.18275901534355	3.320634733531723e-05	429.0	4786.258741258741	35.5570501	45.4425963	56.0	Irak	-0.0041425669265316	medium	Iraq
Tehran	18814.0	705.2	705.2	127481565.51332957	180773.6323217945	2345.0	0.300724946695096	35.6892523	51.3896004	552.0	Iran	-0.2516286178881455	low	Iran
kahramanmaraş	14725.0	1157940.0	78.6	6819.0	0.0058889061609409	71.0	16309.014084507042	39.750545	35.837347	1296.0	Türkiye	0.2750619644582845	high	Turkey
karabük	4142.0	248458.0	60.0	7141.0	0.0287412761915494	43.0	5778.093023255814	39.750545	35.837347	708.0	Türkiye	-0.0696552297702486	low	Turkey
kırklareli	6550.0	366383.0	55.9	11398.0	0.031112202924421	37.0	9901.702702702703	39.750545	35.837347	950.0	Türkiye	0.0523398376798044	medium	Turkey
Atina	36172.0	2664776.0	7366.98	34050.72375109694	0.0127780810661372	166.0	16052.867466879518	37.98853705100004	23.74205028088212	75.0	Yunanistan	0.3473164607581622	high	Greece
Kafalonya	90439.0	39488.0	4396.0	17682.642827630243	0.447797883600469	7.0	5641.142857142857	38.26711130350006	20.53845319532181	9.0	Yunanistan	-0.0881202689817254	low	Greece
Böotya	2951.62	131085.0	4441.0	17503.82023659861	0.1335303065690095	12.0	10923.75	38.36181979450007	23.14396055853965	41.0	Yunanistan	0.0655144875377845	medium	Greece

Şekil 4. Veri setinin genel yapısı

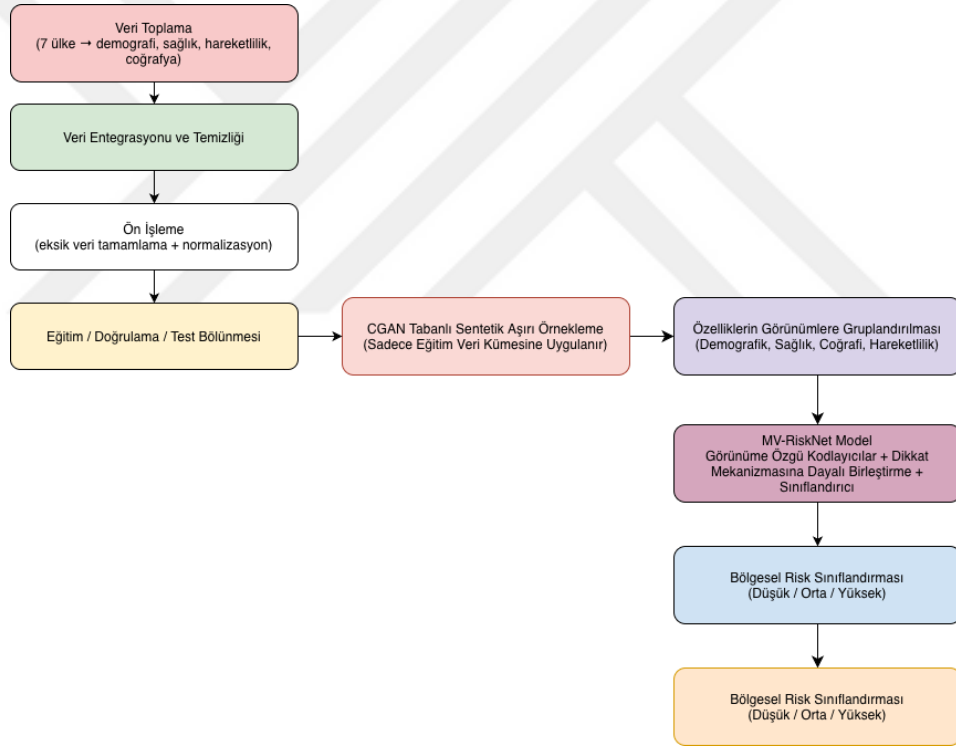
Çok-görünümlü veri yapısının oluşturulması

Bu çalışmanın ikinci aşamasında kullanılan veri seti, epidemik riskin çok boyutlu yapısını temsil edebilmek amacıyla çok-görünümlü bir yaklaşım çerçevesinde yapılandırılmıştır. Epidemik riskin yalnızca tek bir veri kaynağına bağlı olmadığına demografik özellikler, sağlık altyapısı, coğrafi konum ve hareketlilik gibi farklı tematik bileşenlerin etkileşimi sonucunda şekillendiği kabul edilmiştir. Bu doğrultuda, veri setinde yer alan değişkenler tematik bütünlükleri korunacak biçimde dört ayrı görünüm altında organize edilmiştir.

Demografik görünüm, bölgesel nüfus büyüklüğü, nüfus yoğunluğu, ekonomik düzeyi temsil eden gösterge ve nüfus ile ekonomik kapasite arasındaki ilişkiyi yansıtan bileşik değişken olmak üzere dört değişkenden oluşmaktadır. Bu görünüm, epidemik riskin sosyo-demografik belirleyicilerini temsil etmektedir. Sağlık görünümü, sağlık hizmetlerinin kapasitesini yansıtan hastane sayısı ve sağlık altyapısının nüfus üzerindeki yükünü ifade eden Population_Per_Hospital değişkenlerini içermektedir. Bu yapı, bölgelerin salgına müdahale edebilme kapasitesini modelleme sürecine dâhil etmeyi amaçlamaktadır.

Coğrafi görünüm, idarî bölgelerin fiziksel ve mekânsal özelliklerini temsil eden yüzölçümü ile konumsal bilgiyi yansıtan enlem ve boylam değişkenlerinden oluşmaktadır. Bu görünüm, epidemik riskin mekânsal dağılımını ve bölgesel etkileşimleri dikkate almayı mümkün kılmaktadır. Mobilite görünümü ise bölgeler arası ulaşım ve hareketlilik potansiyelini temsil eden Transportation değişkenini içermekte olup, salgının yayılım hızını etkileyebilecek insan hareketliliği boyutunu modele dâhil etmektedir.

Bu tematik ayrışma sayesinde her veri grubu, modelleme sürecinde bağımsız bir bilgi kaynağı olarak ele alınabilmekte; heterojen veri yapılarının tek bir birleşik özellik uzayına zorlanması sonucu oluşabilecek bilgi kaybı önlenmektedir. Çok-görünümlü veri organizasyonu, çalışmada önerilen modelin temelini oluşturmakta ve her bir görünüm için görünüm-özel temsillerin öğrenilmesine olanak tanımaktadır. Veri setinde yer alan değişkenlerin görünüm bazlı tematik organizasyonu Şekil 5'te gösterilmektedir.



Şekil 5. Çok-görünümlü önerilen modelin gösterimi

Çok-görünümlü veri hazırlık süreci

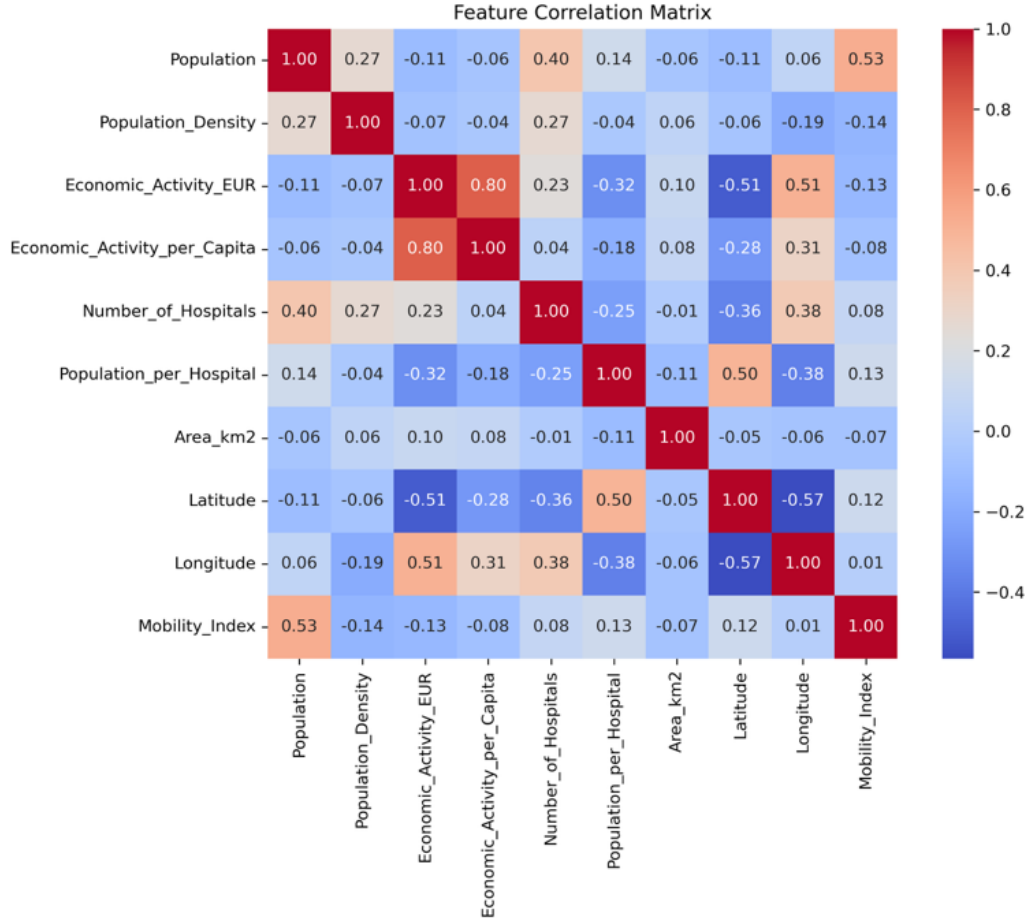
Çok-görünümlü veri yapısının oluşturulmasının ardından, verilerin önerilen modele uygun biçimde işlenebilmesi amacıyla kapsamlı bir veri hazırlık süreci uygulanmıştır. Bu sürecin temel hedefi, farklı ülkelerden ve farklı istatistiksel yapılardan derlenen değişkenlerin tutarlı, karşılaştırılabilir ve çok-görünümlü öğrenmeye uygun hâle getirilmesidir.

İlk aşamada veri setindeki eksik gözlemler analiz edilmiş ve eksik değerlerin özellikle sağlık altyapısı ve mobilite göstergelerinde yoğunlaştığı belirlenmiştir. Eksik veriler, ilgili bölgenin demografik ve mekânsal benzerlikleri dikkate alınarak uygun istatistiksel imputasyon yöntemleriyle tamamlanmıştır. Bu yaklaşım, görünüm bazlı veri bütünlüğünün korunmasını sağlamış ve model eğitiminde yapay sapmaların oluşmasını engellemiştir.

Ön işleme sürecinin devamında, farklı ölçeklerdeki değişkenlerin model üzerindeki etkisini dengelemek amacıyla her bir görünüm altında yer alan sürekli değişkenlere z-skor normalizasyonu uygulanmıştır. Normalizasyon parametreleri yalnızca eğitim verisi üzerinden hesaplanmış ve aynı dönüşüm doğrulama ile test kümelerine uygulanarak veri sızıntısı riski ortadan kaldırılmıştır. Bu işlem, encoder katmanlarının her görünüm için daha kararlı ve ayrıştırılabilir temsiller öğrenmesine katkı sağlamıştır.

Veri seti, çok-görünümlü yapının tutarlılığı korunacak şekilde eğitim, doğrulama ve test alt kümelerine ayrılmıştır. Bu bölme sırasında her bölgeye ait demografik, sağlık, coğrafi ve mobilite görünümünün tamamı aynı alt kümeye karşılık gelecek biçimde eşlenmiştir böylece örnek bütünlüğü bozulmadan modelleme sürecine aktarılmıştır.

Son aşamada, görünüm bazlı değişkenler arasındaki ilişkileri değerlendirmek amacıyla korelasyon analizi gerçekleştirilmiştir. Elde edilen korelasyon matrisi, özellikle demografik ve ekonomik değişkenlerin epidemik riskle anlamlı ilişkiler sergilediğini göstermiştir. Bu analiz, veri yapısının çok boyutlu karakterini desteklemekte ve görünüm bazlı temsil öğreniminin uygunluğunu ortaya koymaktadır. Korelasyon analizine ait görsel çıktı tezde Şekil 6'da gösterilmektedir.



Şekil 6. Korelasyon matrisi

CGAN ile sınıf dengesizliğinin giderilmesi

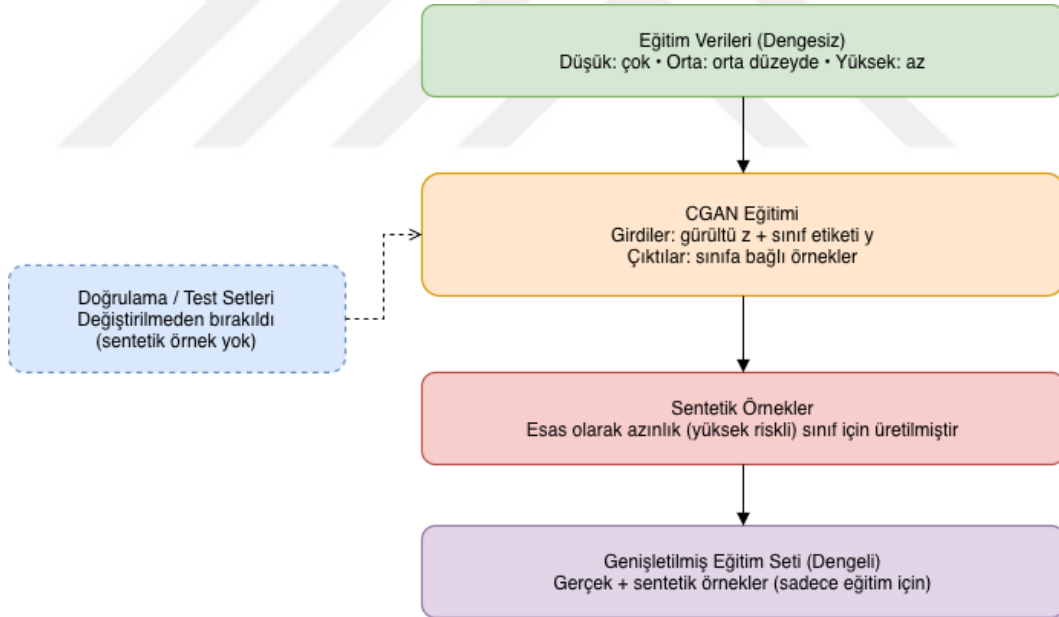
CGAN tabanlı veri artırma uygulanmadan önce veri setinde yüksek risk sınıfı, toplam örneklerin yaklaşık %10–15’lik bir kısmını oluşturmakta olup belirgin bir sınıf dengesizliği mevcuttur. Bu dengesizlik nedeniyle CGAN öncesi MV-RiskNet modelinin yüksek risk sınıfına ait sınıflandırma performansı görece düşük kalmıştır. CGAN uygulanmadan önce elde edilen sonuçlarda, yüksek risk sınıfı için Precision %91,2, Recall %88,6 ve F1 skoru %89,9 olarak ölçülmüştür.

CGAN yöntemi ile sınıf koşullu sentetik örnekler üretilerek veri seti dengelendikten sonra, yüksek risk sınıfına ait örnek oranı artırılmış ve sınıflar arası dağılım daha homojen bir yapıya kavuşmuştur. CGAN sonrası elde edilen sonuçlarda, yüksek risk sınıfı için Precision %96,1, Recall %97,8 ve F1 skoru %96,9 seviyesine yükselmiştir. Bu iyileşme, özellikle yanlış negatif tahminlerin (yüksek riskli bölgelerin düşük risk olarak sınıflandırılması) belirgin biçimde azaldığını göstermektedir.

Elde edilen bu bulgular, CGAN tabanlı veri artırma yönteminin epidemik açıdan kritik olan yüksek risk sınıfının daha etkin biçimde öğrenilmesine katkı sağladığını ve MV-RiskNet modelinin genellenebilirliğini güçlendirdiğini ortaya koymaktadır.

Çok-görünümlü veri setinde hedef değişken olan Risk_Level sınıflarının dağılımı incelendiğinde, özellikle yüksek risk sınıfının diğer sınıflara kıyasla daha az sayıda örnekle temsil edildiği görülmüştür. Bu durum, sınıflandırma modellerinin karar sınırlarını olumsuz etkileyerek yüksek risk sınıfının ayırt edici biçimde öğrenilememesine ve model performansının düşmesine yol açabilmektedir. Bu nedenle çalışmada, sınıf dengesizliği problemini doğrudan ele alan koşullu üretici karşıt ağlar CGAN tabanlı veri artırma yaklaşımı kullanılmıştır.

CGAN tabanlı veri artırma süreci, sınıf koşullu sentetik örneklerin üretilmesi ve bu örneklerin yalnızca eğitim veri kümesine eklenmesi olmak üzere iki temel aşamadan oluşmaktadır. Bu sürecin genel işleyişi Şekil 7’de gösterilmektedir. Bu yapı, yalnızca az temsil edilen sınıflar için sentetik veri üretmekte; doğrulama ve test alt kümeleri sentetik veriden bağımsız tutulmaktadır. Böylece model değerlendirme sürecinin tarafsızlığı korunmaktadır.



Şekil 7. CGAN tabanlı veri artırma sürecinin genel akış diyagramı

CGAN yöntemi, üretici ve ayırt edici olmak üzere iki ağıdan oluşmakta olup klasik GAN yapısından farklı olarak veri üretim sürecini sınıf etiketlerine koşullandırmaktadır. Bu çalışmada CGAN yöntemi, epidemik risk seviyelerini temsil eden sınıf etiketi c ile rastgele gürültü vektörü z 'yi birlikte girdi olarak alacak şekilde yapılandırılmıştır. Üretici ağ, bu girdileri kullanarak ilgili risk sınıfına ait sentetik örnekler üretmektedir. Üretici ağın hesaplama süreci Denklem 3'te ifade edilmiştir. Burada $[z; c]$, gürültü vektörü ile sınıf etiketinin

birleştirilmiş girdisini; W ve b terimleri ağ ağırlıklarını ve bias değerlerini; $\phi(\cdot)$ aktivasyon fonksiyonunu ve $\sigma(\cdot)$ çıkış aktivasyonunu temsil etmektedir.

Ayırt edici ağ ise, gerçek veya sentetik bir örneğin ilgili sınıfa ait gerçek bir gözlem olup olmadığını değerlendirmektedir. Ayırt edici ağın çıktısı Denklem 4’te gösterilmiştir. CGAN’ın eğitimi, üretici ve ayırt edici ağlar arasında tanımlanan klasik min–max oyun çerçevesinde gerçekleştirilmektedir. Ayırt edici (Denklem 5) ve üretici ağlara (Denklem 6) ait kayıp fonksiyonları sırasıyla tanımlanmıştır.

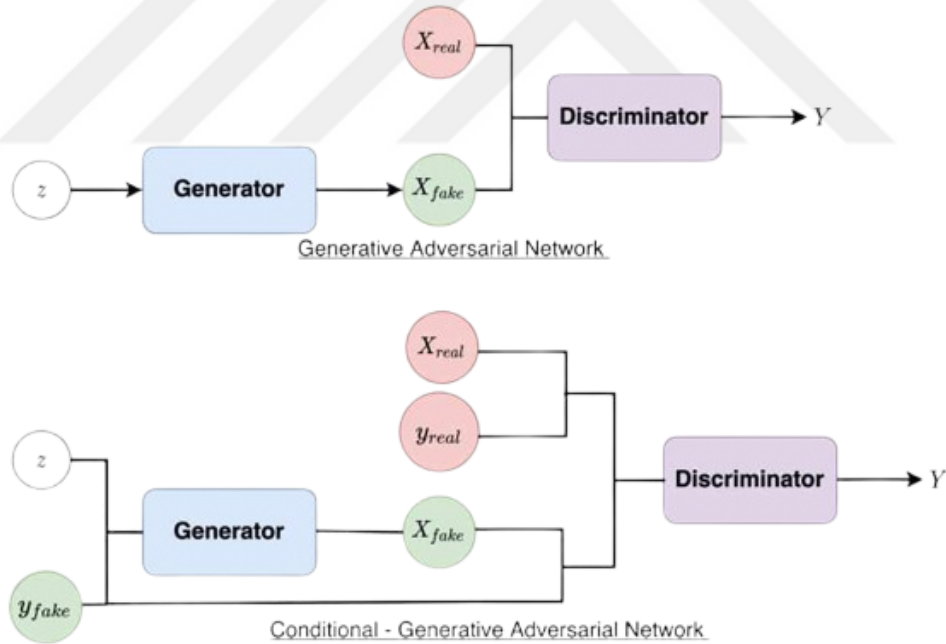
$$G(z, c) = \sigma(W_2 \phi(W_1[z; c] + b_1) + b_2) \quad (3)$$

$$D(x, c) = \text{sigmoid}(W_d[x; c] + b_d) \quad (4)$$

$$L_D = -\mathbb{E}_{x,c}[\log D(x, c)] - \mathbb{E}_{z,c}[\log(1 - D(G(z, c), c))] \quad (5)$$

$$L_G = -\mathbb{E}_{z,c}[\log D(G(z, c), c)] \quad (6)$$

Bu karşıt öğrenme süreci sonucunda CGAN yöntemi, her bir risk sınıfına özgü istatistiksel dağılımları öğrenerek özellikle az temsil edilen yüksek risk sınıfına ait gerçekçi ve çeşitli sentetik örnekler üretebilmektedir. Sınıf koşullu sentetik örnek üretim mekanizması Şekil 8’de şematik olarak gösterilmektedir.



Şekil 8. CGAN ile sınıf koşullu sentetik örnek üretim süreci

Üretilen sentetik örnekler, ters normalizasyon işlemi uygulanarak orijinal veri ölçeğine geri dönüştürülmüş ve yalnızca eğitim veri kümesine dâhil edilmiştir. Doğrulama ve test kümelerine herhangi bir sentetik örnek eklenmemiştir. Ayrıca, çok-görünümlü veri yapısının bütünlüğünün korunabilmesi amacıyla CGAN ile üretilen her sentetik örnek, demografik, sağlık, coğrafi ve mobilite görünümünün tamamı için tutarlı biçimde oluşturulmuştur. Bu

yaklaşım, sınıf dağılımının dengelenmesinin yanı sıra çok-görünümlü temsiller arasındaki yapısal tutarlılığın korunmasını da sağlamıştır.

Karşılaştırma amaçlı derin öğrenme modelleri

Önerilen modelinin performansını nesnel biçimde değerlendirebilmek amacıyla, literatürde yaygın olarak kullanılan ve farklı öğrenme prensiplerini temsil eden çeşitli DL mimarileri karşılaştırma modelleri olarak ele alınmıştır. Bu modeller, veri işleme yaklaşımları ve temsil öğrenme kapasiteleri bakımından önerilen model ile kıyaslanarak, önerilen mimarinin sağladığı katkının yalnızca veri setinden değil, mimari tasarımdan kaynaklandığının ortaya konulması hedeflenmiştir.

Bu kapsamda ilk olarak MLP modeli kullanılmıştır. MLP, tüm girdilerin tek bir birleşik özellik vektörü hâlinde işlendiği temel bir tam bağlantılı ağ yapısına sahiptir. Doğrusal olmayan ilişkileri öğrenebilmesine rağmen, çok-görünümlü veri setinin tematik ayrışmasını doğrudan temsil edememektedir. Bu nedenle MLP, tek-görünüm tabanlı derin öğrenme yaklaşımları için bir referans model olarak değerlendirilmiştir.

CNN modeli, bir boyutlu evrişim katmanları aracılığıyla öznitelikler arasındaki yerel örüntüleri öğrenebilen bir mimaridir. CNN'ler düzenli yapıya sahip verilerde güçlü performans sergilemekle birlikte, bu çalışmada kullanılan çok-görünümlü ve tablo tabanlı veri yapısında mekânsal komşuluk ilişkileri açık biçimde tanımlı olmadığından, CNN modeli karşılaştırma amaçlı bir referans olarak kullanılmıştır.

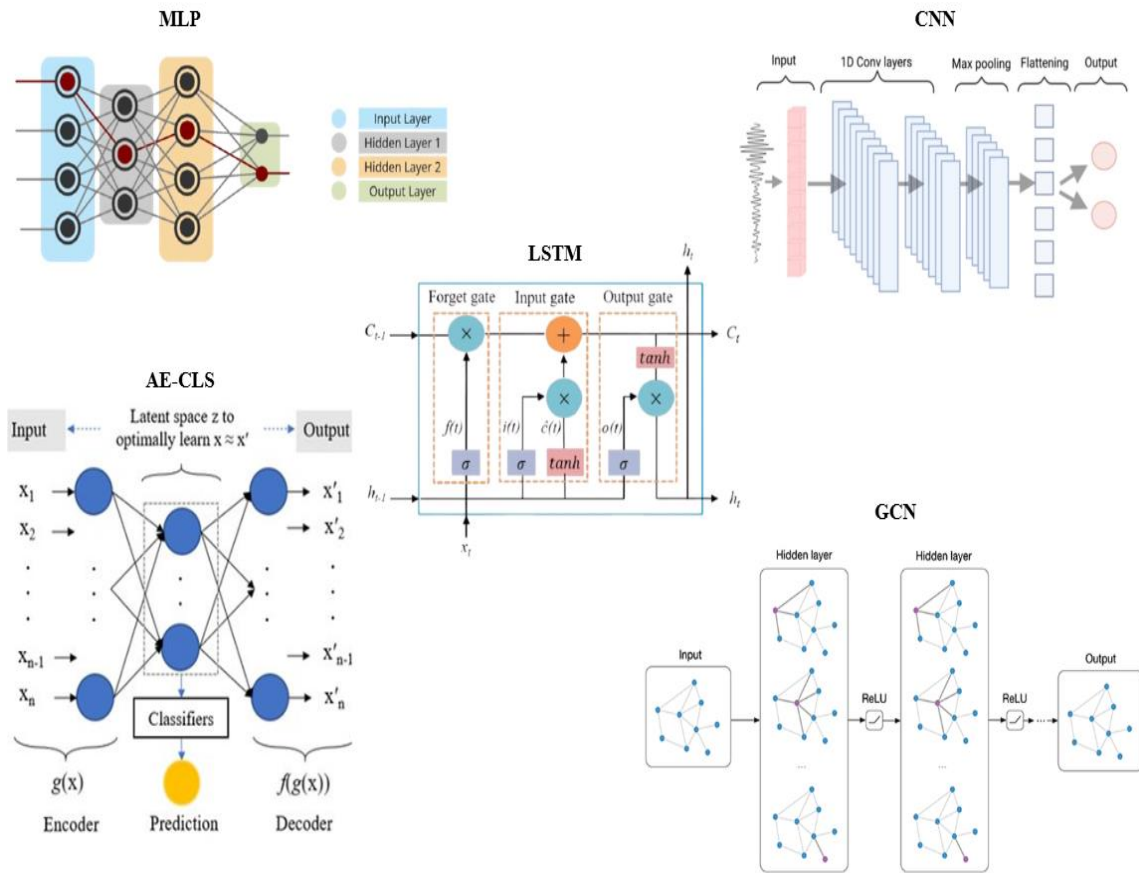
LSTM ağı, ardışık bağımlılıkları öğrenebilme kapasitesine sahip kapı mekanizmaları ile çalışan bir tekrarlayan sinir ağı mimarisidir. LSTM modelleri özellikle zaman serisi verilerinde başarılı sonuçlar üretmektedir. Bu çalışmada kullanılan veriler statik nitelikte olmasına rağmen, LSTM modeli derin öğrenme tabanlı ardışık yapıların performansını değerlendirmek amacıyla karşılaştırma setine dâhil edilmiştir.

AE-CLS modeli, encoder-latent uzay-decoder yapısı ile düşük boyutlu temsiller öğrenirken eş zamanlı olarak sınıflandırma çıktısı üreten bir mimaridir. AE-CLS, temsil öğrenme açısından güçlü bir yapı sunmasına rağmen, tüm veri görünümlerini tek bir latent uzayda birleştirmesi nedeniyle çok-görünümlü veri setinin tematik özelliklerini ayrıştırarak öğrenememektedir.

Son olarak, GCN modeli karşılaştırma amaçlı olarak kullanılmıştır. GCN, grafik yapılar üzerinde tanımlanan düğümler ve kenarlar aracılığıyla komşuluk ilişkilerini öğrenebilen bir mimaridir. Coğrafi yakınlık temelli ilişkilerin modellenmesi açısından avantajlı olmakla birlikte, demografik, sağlık ve mobilite gibi tematik olarak farklı veri görünümlerinin tek bir

grafik yapısı altında birleştirilmesi yapısal bilgi kaybına yol açabilmektedir. Bu nedenle GCN, önerilen modelin çoklu görünüm temelli yaklaşımının sağladığı avantajları ortaya koymak açısından önemli bir karşılaştırma modeli olarak değerlendirilmiştir.

Karşılaştırma amacıyla kullanılan tüm derin öğrenme modelleri, önerilen model ile aynı veri bölünmeleri, aynı değerlendirme metrikleri ve benzer eğitim protokolleri altında eğitilmiştir. Böylece elde edilen performans farklarının yalnızca model mimarisinden kaynaklanması sağlanmış ve karşılaştırmalar adil biçimde gerçekleştirilmiştir. Karşılaştırma modellerine ait mimari özetler tezde Şekil 9’da gösterilmektedir.



Şekil 9. Çalışmada kullanılan karşılaştırma dl modellerinin mimari görünüşleri

Önerilen Model: MV-RiskNet

Bu çalışmada önerilen MV-RiskNet, epidemik riskin çok boyutlu yapısını etkili biçimde temsil edebilmek amacıyla çoklu veri görünümünü derin öğrenme tabanlı bir mimari altında bütünleştiren bir sınıflandırma modelidir. Model, demografik, sağlık, coğrafi ve mobilite görünümünün her birini ayrı alt ağlar aracılığıyla işleyerek, bu görünümlere özgü temsilleri öğrenmekte ve elde edilen temsilleri ortak bir karar mekanizması altında birleştirmektedir. Bu yaklaşım, heterojen veri kaynaklarının tek bir birleşik özellik vektörüne zorlanması sonucu oluşabilecek bilgi kaybını önlemeyi amaçlamaktadır.

Önerilen MV-RiskNet modelinde demografik, sağlık, coğrafi ve mobilite verileri ayrı encoder ağları tarafından işlenmektedir. Her encoder, kendi veri grubuna özgü yapısal özellikleri öğrenmeye yönelik çok katmanlı tam bağlantılı ağlardan oluşmakta ve giriş verisini daha düşük boyutlu fakat bilgi açısından zengin gizil temsillere dönüştürmektedir. Bu yapı, heterojen veri kaynaklarının tek bir özellik uzayına zorlanmasıyla oluşabilecek bilgi kaybını önlemektedir.

Encoder ağlarından elde edilen görünüm bazlı temsiller, dikkat (attention) tabanlı bir füzyon mekanizması aracılığıyla bütünleştirilmektedir. Attention mekanizması, her bir veri görünümünün epidemik risk tahminine olan göreceli katkısını öğrenilebilir ağırlıklar ile belirlemekte ve modele esnek bir entegrasyon yeteneği kazandırmaktadır. Böylece farklı bölgelerde demografik, mobilite veya sağlık altyapısı verileri risk tahminine farklı düzeylerde katkı sunabilmektedir.

Her bir veri görünümü için tanımlanan girdi özellik kümesi X_v , ilgili görünüm için tasarlanan alt ağ tarafından işlenerek görünüm-özel bir latent temsil h_v elde edilmektedir. Bu işlem Denklem 6’da ifade edilmiş olup burada $f_v(\cdot)$, ilgili veri görünümüne ait çok katmanlı tam bağlantılı ağ yapısını temsil etmektedir. Bu yapı sayesinde her bir görünüm, kendi istatistiksel ve yapısal özellikleri doğrultusunda bağımsız biçimde öğrenilmektedir.

Encoder ağları tarafından üretilen görünüm-özel latent temsiller, füzyon katmanında birleştirilmektedir. Bu çalışmada füzyon işlemi, her bir görünümün modele katkısını temsil eden öğrenilebilir ağırlıklar aracılığıyla gerçekleştirilmiştir. Birleştirilmiş temsil Denklem 7’de tanımlanmıştır. Burada α_v , her bir veri görünümünün epidemik risk üzerindeki göreceli önemini temsil eden ağırlık katsayısını ifade etmektedir. Bu yapı, modelin farklı veri görünümülerinden gelen bilgileri dengeli ve yorumlanabilir biçimde bütünleştirmesine olanak tanımaktadır. Elde edilen birleşik temsil h , tam bağlantılı katmanlardan oluşan sınıflandırıcı yapıya aktarılmakta ve epidemik risk seviyesi düşük, orta ve yüksek olmak üzere üç sınıfta tahmin edilmektedir. Model çıktısı softmax fonksiyonu kullanılarak Denklem 8’de olduğu gibi elde edilmektedir.

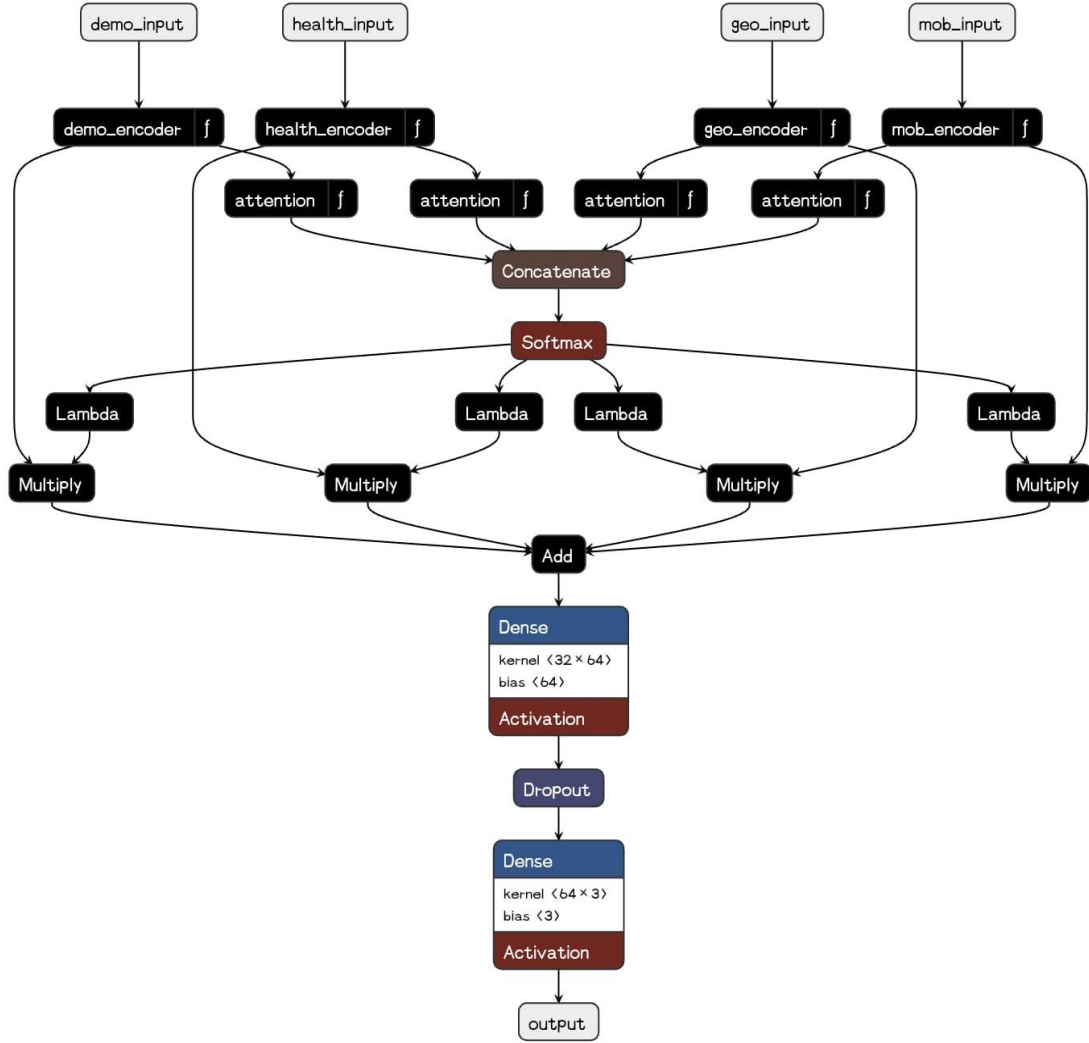
Modelin eğitimi sırasında çok sınıflı çapraz entropi kayıp fonksiyonu kullanılmış olup, sınıf dengesizliğinin etkisini azaltmak amacıyla sınıf ağırlıkları kayıp fonksiyonuna entegre edilmiştir. Parametre güncellemeleri Adam optimizasyon algoritması ile gerçekleştirilmiş; aşırı öğrenmenin önlenmesi amacıyla Dropout ve hafif L2 düzenleme teknikleri uygulanmıştır.

$$h_v = f_v(X_v) \quad (6)$$

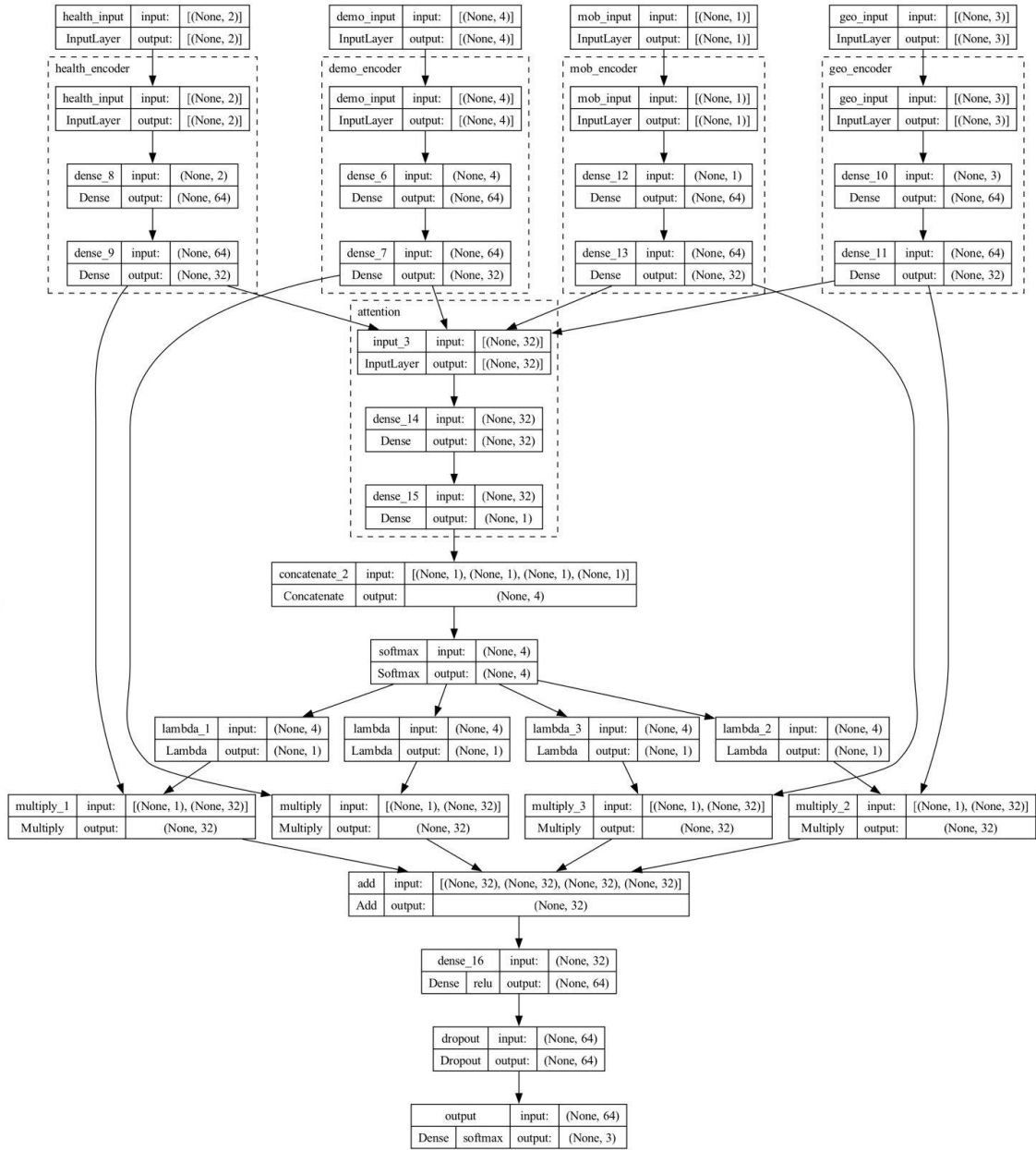
$$h = \sum_{v=1}^V \alpha_v h_v \quad (7)$$

$$\hat{y} = \text{softmax}(Wh + b) \quad (8)$$

Önerilen MV-RiskNet modelinin yüksek seviyeli mimari yapısı Şekil 10’da görünüm-özel encoderlar ile füzyon ve sınıflandırma aşamalarını içeren ayrıntılı blok diyagramı ise Şekil 11’de gösterilmektedir.



Şekil 10. Önerilen MV-RiskNet yüksek seviyeli mimarisi



Şekil 11. Önerilen MV-RiskNet detaylı blok diyagramı

Eğitim Süreci ve Değerlendirme Metrikleri

Önerilen MV-RiskNet modelinin eğitim süreci, çok-görünümlü veri yapısının bütünlüğünü koruyacak şekilde yapılandırılmıştır. Veri seti eğitim, doğrulama ve test olmak üzere üç alt kümeye ayrılmış olup bu ayrım tüm görünüm için aynı örnek indeksleri kullanılarak gerçekleştirilmiştir. Böylece her bölgeye ait demografik, coğrafi, sağlık ve mobilite görünimleri veri kümesi boyunca tutarlı bir eşleşme ile temsil edilmiştir. Eğitim verisi üzerinde CGAN yöntemi uygulanmış, doğrulama ve test kümeleri ise yalnızca gerçek gözlemlerden oluşturulmuştur. Bu yaklaşım, model değerlendirmesinin tarafsız kalmasını ve gerçek performansın yansıtılmasını sağlamaktadır.

Modelin öğrenme sürecinde çok sınıflı çapraz entropi kayıp fonksiyonu kullanılmıştır. Risk sınıflarındaki doğal dengesizliği azaltmak amacıyla, eğitimde kullanılan kayıp fonksiyonuna sınıf ağırlıkları entegre edilmiştir. Optimizasyon işlemi Adam algoritması ile yürütülmüş olup öğrenme oranı doğrulama kaybındaki iyileşmelere bağlı olarak gerektiğinde azaltılmıştır. Aşırı öğrenmenin önlenmesi amacıyla eğitim sürecinde erken durdurma mekanizması uygulanmış, belirli sayıda epoch boyunca doğrulama kaybında gelişme görülmemesi durumunda eğitim otomatik olarak sonlandırılmıştır. Ayrıca sınıflandırıcı katmanda Dropout ve hafif L2 düzenleme teknikleri kullanılarak modelin genelleme kabiliyeti artırılmıştır.

Batch size, epoch sayısı ve encoder katman boyutları, doğrulama performansı göz önünde bulundurularak deneysel olarak belirlenmiştir. MV-RiskNet'in çoklu encoder yapısı nedeniyle her veri görünümü için ayrı ağırlıklar optimize edilmiş, füzyon katmanından itibaren ağ ortak bir temsil uzayında öğrenme gerçekleştirmiştir. Bu yaklaşım, görünüm-özel temsil öğrenimi ile bütüncül karar mekanizması arasındaki optimizasyon dengesini korumuştur.

Model performansının değerlendirilmesinde, çok sınıflı sınıflandırma problemleri için yaygın olarak kullanılan Doğruluk, Kesinlik, Duyarlılık, F1-Skoru ve ROC-AUC metrikleri birlikte kullanılmıştır. Doğruluk, modelin genel sınıflandırma başarısını ölçen temel bir metriktir ve doğru sınıflandırılan örneklerin toplam örnek sayısına oranı olarak hesaplanmaktadır (Denklem 9). Bununla birlikte, sınıf dengesizliği içeren yapılarda doğruluk metriği tek başına yanıltıcı olabilmektedir. Bu nedenle sınıf bazlı performansı daha doğru değerlendirebilmek amacıyla kesinlik ve duyarlılık metrikleri de dikkate alınmıştır. Kesinlik (Denklem 10), model tarafından pozitif olarak tahmin edilen örneklerin ne kadarının gerçekten pozitif olduğunu ifade ederken, duyarlılık (Denklem 11) gerçek pozitif örneklerin ne kadarının doğru biçimde yakalandığını göstermektedir. Kesinlik ve duyarlılık arasındaki dengeyi ölçen F1-Skoru (Denklem 12), özellikle sınıf dengesizliği bulunan veri setlerinde model performansını daha güvenilir biçimde yansıtan bir ölçüttür.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (11)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

Buna ek olarak, sınıflar arası ayırt edicilik kapasitesini değerlendirmek amacıyla ROC-AUC analizi uygulanmıştır. Çok sınıflı yapı nedeniyle her bir risk seviyesi, diğer sınıflara karşı ayrı bir ikili sınıflandırma problemi olarak ele alınmış ve her sınıf için ayrı ROC eğrileri

oluşturulmuştur. Elde edilen AUC değerlerinin ortalaması, modelin genel çok sınıflı ayırt edicilik kapasitesini yansıtan tamamlayıcı bir performans göstergesi olarak değerlendirilmiştir.

Önerilen modelin değerlendirilmesi, tüm karşılaştırma modellerinin aynı veri bölünmeleri ve aynı değerlendirme metrikleri altında eğitilmesiyle gerçekleştirilmiştir. Böylece modelin performans artışının yalnızca mimari tasarımdan kaynaklandığı garanti altına alınmıştır. Bu eğitim ve değerlendirme stratejisi, önerilen modelin hem çoklu görünüm yapısını etkili biçimde öğrenebildiğini hem de sınıf dengesizliğine rağmen özellikle yüksek risk kategorisinde güçlü bir ayırım performansı sergilediğini ortaya koymaktadır.

Bu çalışmada kullanılan makine öğrenimi ve derin öğrenme modellerine ait temel eğitim ayarları ve hiperparametreler, literatürde yaygın olarak kullanılan değerler ve deneysel gözlemler doğrultusunda belirlenmiştir. Modellerin eğitiminde, aşırı öğrenmenin önlenmesi ve genellebilirliğin artırılması amacıyla öğrenme oranı, epoch sayısı ve optimizasyon yöntemi gibi temel parametreler dikkate alınmıştır.

Makine öğrenimi modellerinde, algoritmaların varsayılan parametre ayarları korunmuş; performansı doğrudan etkileyen temel parametreler üzerinde sınırlı ve kontrollü düzenlemeler yapılmıştır. Derin öğrenme tabanlı MV-RiskNet ve CGAN modellerinde ise eğitim süreci boyunca optimizasyon yöntemi, öğrenme oranı ve eğitim süresi gibi parametreler izlenmiş ve doğrulama kümesi performansına göre değerlendirilmiştir.

Bu yaklaşım ile deneysel kurulumun yeniden üretilebilirliği desteklenmiş, aynı zamanda modellerin karşılaştırılabilir ve tutarlı biçimde değerlendirilmesi sağlanmıştır.

ARAŞTIRMA BULGULARI

Bu bölümde çalışmanın yöntem kısmında tanımlanan iki aşamalı analiz sürecine karşılık gelen deneysel bulgular sunulmaktadır. Bulgular, yöntemsel akışla tutarlı olacak şekilde iki ana alt başlık altında ele alınmıştır. İlk aşamada, Türkiye'nin 81 ili için ML tabanlı epidemik risk sınıflandırması ve bu sınıflandırmaya dayalı mekânsal haritalandırma sonuçları değerlendirilmiştir. İkinci aşamada ise, çoklu veri görünümüne dayalı DL mimarisi olan MV-RiskNet modeli kullanılarak Türkiye ve komşu ülkeleri kapsayan bölgesel epidemik risk tahminine ilişkin bulgular incelenmiştir.

1. Aşama: Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye'de Bölgesel Salgın Risk Analizi ve Haritalandırılması

Makine öğrenimi modellerinin genel sınıflandırma performansı

Çalışmanın birinci aşamasında, Türkiye'nin 81 iline ait epidemik risk seviyelerinin sınıflandırılması amacıyla LR, NB, SVM, RF ve XGB algoritmaları uygulanmıştır. Tüm modeller, aynı öznitelik seti, aynı veri bölünmesi ve aynı ön işleme adımları kullanılarak eğitilmiş ve test edilmiştir. Böylece performans karşılaştırmalarının yöntemsel olarak tutarlı olması sağlanmıştır.

Modellerin performansları doğruluk, kesinlik, duyarlılık ve F1-skoru ölçütleri üzerinden değerlendirilmiş olup elde edilen sonuçlar Tablo 1'de gösterilmiştir. Sonuçlar incelendiğinde, XGB modelinin %98 doğruluk ve %97,78 F1-skoru ile en yüksek genel sınıflandırma başarısını elde ettiği görülmektedir. Bu durum, XGB modelinin gradyan artırma temelli yapısı sayesinde epidemik risk veri setinde yer alan doğrusal olmayan ilişkileri ve değişken etkileşimlerini etkin biçimde öğrenebilmesiyle açıklanabilir.

RF modeli %96 doğruluk ve dengeli kesinlik-duyarlılık değerleri ile ikinci en yüksek performansı sergilemiştir. Karar ağaçlarının topluluk yapısı içerisinde birleştirilmesi, modelin genelleme yeteneğini artırmış ve farklı risk sınıfları arasında daha kararlı tahminler üretmesini sağlamıştır. SVM modeli de yüksek duyarlılık değeri (%97,78) ile özellikle riskli sınıfların yakalanması açısından güçlü bir performans ortaya koymuştur.

Buna karşılık, LR ve NB modelleri diğer algoritmalara kıyasla daha düşük performans değerleri üretmiştir. LR'nin doğrusal karar fonksiyonuna dayalı yapısı, epidemik riskin demografik, ekonomik ve sağlık altyapısı gibi çok boyutlu bileşenlerden oluşan karmaşık yapısını yeterince temsil edememiştir. NB algoritmasının koşullu bağımsızlık varsayımı ise

değişkenler arasındaki güçlü ilişkilerin bulunduğu bu veri setinde sınıflandırma başarısını sınırlamıştır.

Genel değerlendirme olarak, ağaç tabanlı ve topluluk öğrenme yaklaşımlarının epidemik risk sınıflandırmasında daha yüksek ve dengeli performans göstermiştir. Bununla birlikte, tüm makine öğrenmesi modellerinde yüksek risk sınıfına ilişkin performansın görece sınırlı kalması, veri setindeki sınıf dengesizliğinin sınıflandırma başarısı üzerindeki etkisini ortaya koymaktadır. Bu bulgu, çalışmanın ikinci aşamasında ele alınan derin öğrenme ve veri artırma temelli yaklaşımların gerekliliğini desteklemektedir.

Tablo 1. ML Algoritmalarının Sonuçları

Algoritma	Doğruluk (%)	Kesinlik	Duyarlılık	F1-Skoru
RF	96.0	0.833	0.963	0.869
NB	92.0	0.796	0.941	0.840
XGB	98.0	0.976	0.980	0.980
LR	88.0	0.736	0.904	0.763
SVM	96.0	0.797	0.978	0.971

XGBoost algoritmasının yüksek performans göstermesinin temel nedenleri arasında doğrusal olmayan ilişkileri etkin biçimde modelleyebilmesi, düzenlileştirme mekanizmaları sayesinde aşırı öğrenmeye karşı dayanıklı olması ve heterojen veri yapılarında güçlü genelleme yeteneği yer almaktadır. Bu özellikler, epidemik risk analizinde kullanılan çok boyutlu veri yapıları için XGBoost'u güçlü bir yöntem hâline getirmektedir.

Makine Öğrenimi Modellerine Ait Karmaşıklık Matrisi

Makine öğrenmesi algoritmalarının sınıf bazındaki sınıflandırma performansları, her model için oluşturulan karışıklık matrisleri üzerinden incelenmiştir. LR, NB, SVM, RF ve XGB modellerine ait karışıklık matrisleri Şekil 12'te gösterilmektedir.

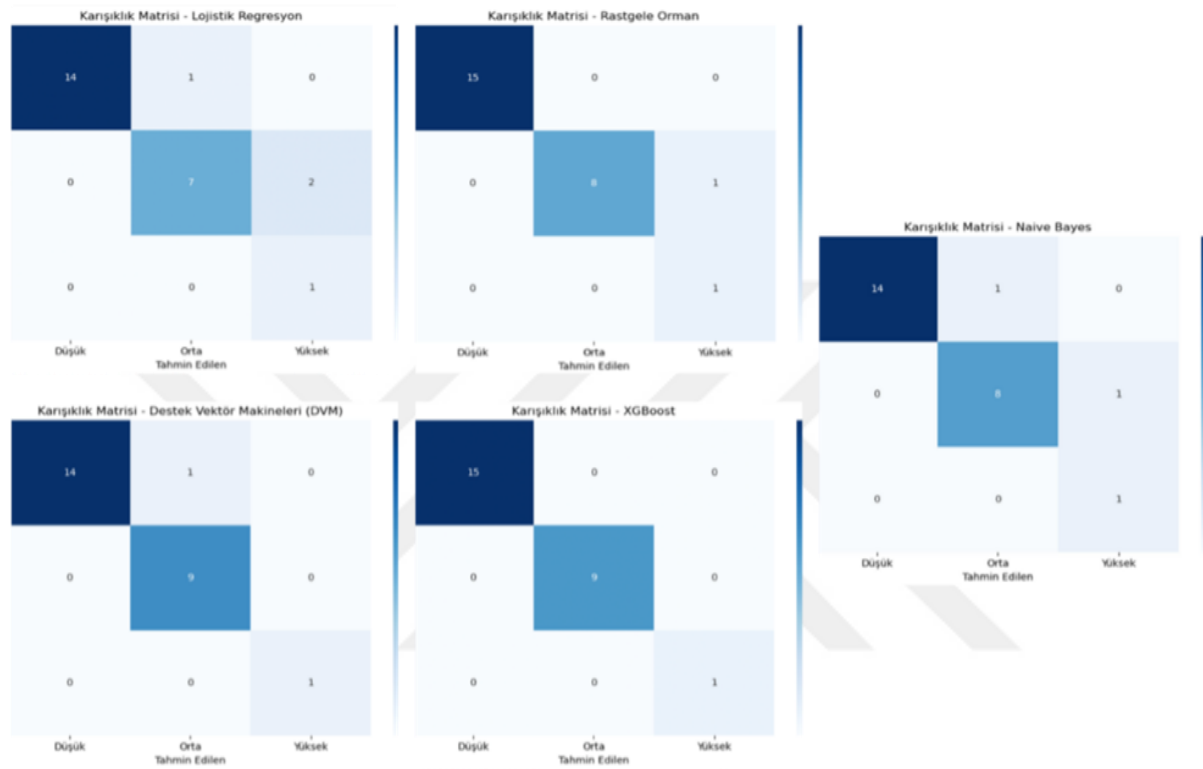
Tüm modeller düşük risk sınıfında yüksek doğru sınıflandırma başarısı göstermiştir. LR, NB ve SVM modelleri düşük risk sınıfına ait 15 örnekten 14'ünü doğru sınıflandırırken, RF ve XGB modelleri düşük risk sınıfındaki tüm örnekleri doğru tahmin etmiştir.

Orta risk sınıfında LR modeli 7 doğru tahmin üretmiş, NB modeli 8 doğru tahmin elde etmiştir. SVM, RF ve XGB modelleri orta risk sınıfında daha yüksek doğru sınıflandırma oranlarına ulaşmıştır. Yanlış sınıflandırmalar ağırlıklı olarak orta risk sınıfının yüksek risk sınıfı ile karıştırılması şeklinde gerçekleşmiştir.

Yüksek risk sınıfı, tüm makine öğrenmesi modelleri için en düşük performansın gözlemlendiği sınıf olmuştur. İncelenen tüm modeller yüksek risk sınıfına ait örneklerden yalnızca

birini doğru sınıflandırabilmiştir. Yüksek risk örneklerinin büyük bölümü orta risk sınıfı olarak tahmin edilmiştir.

Karmaşıklık matrisi sonuçları, ML modellerinin düşük ve orta risk sınıflarında başarılı olduğunu, ancak yüksek risk sınıfının sınıflandırılmasında belirgin sınırlamalar bulunduğunu göstermektedir. Bu durum, yüksek risk sınıfının veri setinde düşük temsil edilmesi ve sınıf sınırlarının yeterince ayrışmaması ile ilişkilidir.



Şekil 12. Karmaşıklık matrisleri

Makine Öğrenimi Modellerine Ait ROC-AUC Grafiği

ML modellerinin sınıflar arasındaki ayırt edicilik kapasiteleri, ROC eğrileri ve ROC-AUC değerleri kullanılarak değerlendirilmiştir. Çok sınıflı yapı nedeniyle ROC eğrileri her bir risk seviyesi için ayrı ayrı hesaplanmış ve sonuçlar Şekil 13'te gösterilmiştir.

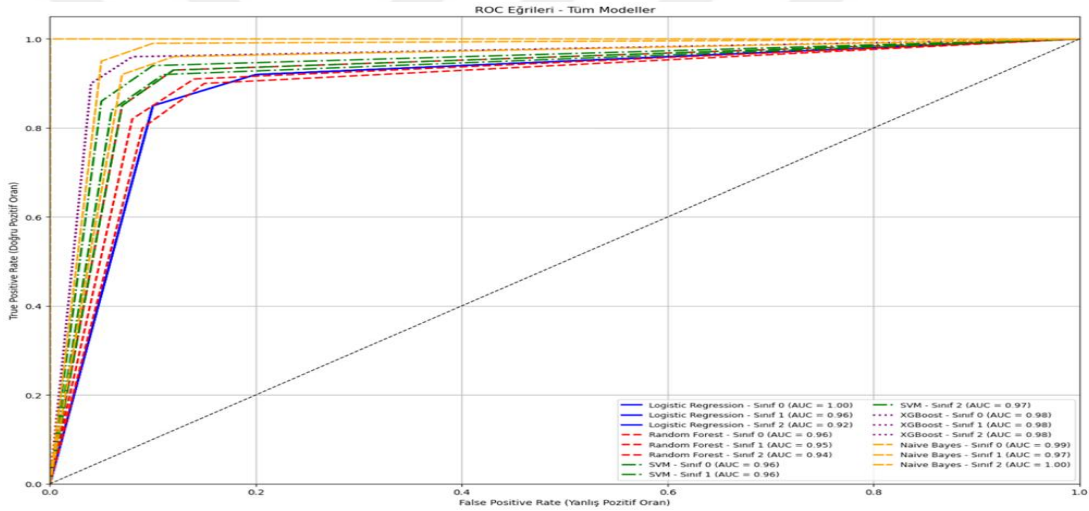
ROC eğrileri incelendiğinde, düşük risk sınıfı için tüm modellerin yüksek ayırt edicilik performansı sergilediği görülmektedir. Eğrilerin büyük ölçüde sol üst köşeye yakın seyretmesi, düşük risk sınıfına ait örneklerin özellik uzayında diğer sınıflardan net biçimde ayrıştığını göstermektedir. Bu durum, karmaşıklık matrisi sonuçları ile tutarlıdır.

Orta risk sınıfı için ROC eğrileri daha dengeli bir dağılım sergilemektedir. Modellerin AUC değerleri düşük risk sınıfına kıyasla daha düşük olmakla birlikte, sınıflar arasında ayırt

edici bilgi taşıyan bir performans ortaya koymaktadır. Bu sonuç, orta risk sınıfının düşük ve yüksek risk sınıfları ile kısmen örtüşen bir yapıya sahip olduğunu göstermektedir.

Yüksek risk sınıfı için ROC eğrilerinin diğer sınıflara kıyasla daha sınırlı bir ayırt edicilik sunduğu görülmektedir. Eğrilerin diyagonale daha yakın seyretmesi, modellerin yüksek risk sınıfını ayırt etme kapasitesinin düşük olduğunu göstermektedir. Bu bulgu, yüksek risk sınıfına ait örnek sayısının görece az olması ve sınıf içi varyasyonun yüksekliği ile ilişkilidir.

Genel olarak ROC–AUC analizi, ML modellerinin düşük risk sınıfında güçlü, orta risk sınıfında kabul edilebilir, yüksek risk sınıfında ise sınırlı bir ayırıştırma performansı sergilediğini ortaya koymaktadır. Elde edilen sonuçlar, yalnızca doğruluk temelli değerlendirmelerin yeterli olmadığını ve özellikle yüksek risk sınıfının güvenilir biçimde tahmin edilebilmesi için sınıf dengesizliğini ve çok boyutlu veri yapısını ele alan ileri modelleme yaklaşımlarına ihtiyaç duyulduğunu göstermektedir.



Şekil 13. ROC eğrisi sonuçları

Türkiye İl Bazlı Salgın Risk Haritalandırma Bulguları

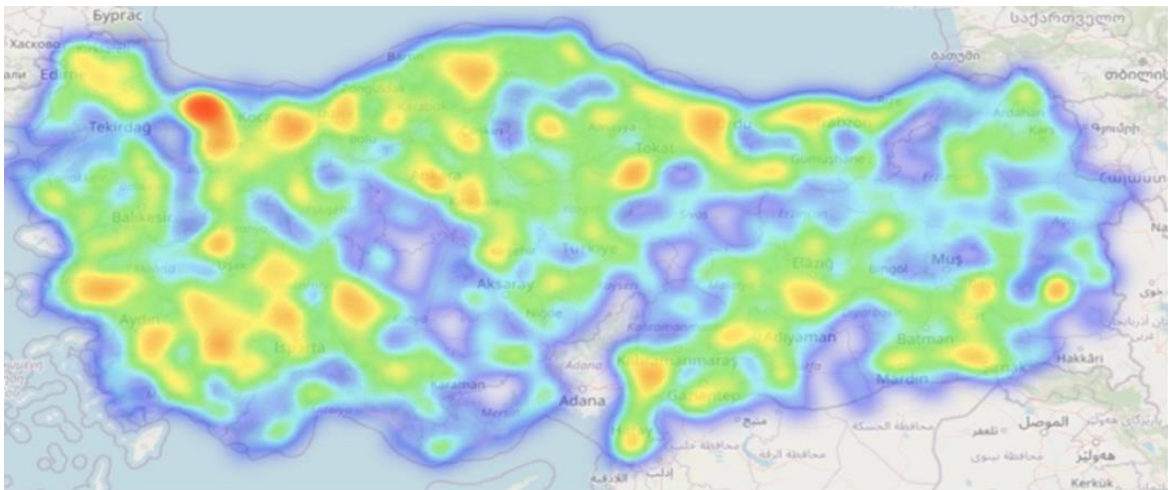
ML modelleri kullanılarak elde edilen epidemik risk sınıflandırma sonuçlarının mekânsal dağılımı, Türkiye il bazlı risk haritaları üzerinden değerlendirilmiştir. Şekil 14'te gösterilen il bazlı salgın risk haritası, düşük, orta ve yüksek risk sınıflarının ülke genelinde homojen bir dağılım sergilemediğini ve belirli bölgelerde yoğunlaştığını göstermektedir.



Şekil 14. Türkiye il bazlı salgın risk haritası

Harita çıktıları incelendiğinde, orta ve yüksek risk sınıflarının ağırlıklı olarak nüfus yoğunluğunun, ekonomik faaliyetlerin ve ulaşım ağlarının görece yoğun olduğu bölgelerde konumlandığı görülmektedir. Buna karşılık düşük risk sınıfına ait illerin, nüfus yoğunluğu ve bölgesel hareketliliğin daha sınırlı olduğu alanlarda daha yaygın olduğu gözlemlenmektedir.

Sınıflandırma sonuçlarının mekânsal yoğunluğunu ortaya koymak amacıyla oluşturulan ısı haritası Şekil 15’de verilmiştir. Isı haritası sonuçları, yüksek risk sınıfına ait illerin coğrafi olarak kümelenme eğilimi gösterdiğini ve risk yoğunluğunun belirli bölgelerde arttığını ortaya koymaktadır. Bu kümelenmeler, epidemik riskin yalnızca il bazlı özelliklerle değil, mekânsal komşuluk ilişkileri ve bölgesel etkileşimler ile değerlendirilmesi gerektiğine işaret etmektedir.



Şekil 15. Türkiye salgın risk yoğunluğunu gösteren ısı haritası

Genel olarak il bazlı risk haritaları, makine öğrenmesi tabanlı sınıflandırma çıktılarının mekânsal bağlamda tutarlı örüntüler ürettiğini göstermektedir. Bununla birlikte, yüksek risk sınıfına ait belirsizliklerin devam etmesi, birinci aşama kapsamında elde edilen haritaların

özellikle kritik risk bölgelerinin belirlenmesinde dikkatli ve destekleyici bir karar aracı olarak değerlendirilmesi gerektiğini ortaya koymaktadır.

2. Aşama: Derin Öğrenme Modelleri Kullanılarak Türkiye ve Komşu Ülkelerinin Bölgesel Salgın Risk Tahmini ve Haritalandırması

Derin Öğrenme Modellerinin Genel Performansı

Çalışmanın ikinci aşamasında, önerilen modelin performansını değerlendirebilmek amacıyla literatürde yaygın olarak kullanılan çeşitli derin öğrenme mimarileri karşılaştırma modelleri olarak ele alınmıştır. Bu kapsamda MLP, CNN, LSTM, AE-CLS ve GCN modelleri kullanılmıştır. Tüm modeller, aynı eğitim-doğrulama-test bölünmesi, aynı normalizasyon işlemleri ve aynı değerlendirme metrikleri altında eğitilmiştir.

Karşılaştırma modellerine ait doğruluk, kesinlik, duyarlılık ve F1-skoru değerleri Tablo 2’de gösterilmiştir. Tablo 2 incelendiğinde, CNN ve AE-CLS modellerinin %0.917 doğruluk ve yaklaşık %0,92 F1-skoru değerleri ile klasik DL mimarileri arasında en yüksek performansı sergilediği görülmektedir. Bu sonuçlar, söz konusu modellerin doğrusal olmayan ilişkileri öğrenebilme kapasitesinin epidemik risk verisindeki karmaşık örüntülerin temsil edilmesine katkı sağladığını göstermektedir.

MLP modeli Tablo 2’de %0.889 doğruluk ve %0.889 F1-skoru değerleri ile orta düzeyde bir performans ortaya koymuştur. Tüm özniteliklerin tek bir birleşik özellik vektörü hâlinde işlendiği bu mimaride, çoklu veri görünümüne ait tematik ayrışmanın korunamaması sınıflandırma başarısını sınırlamıştır. LSTM modeli Tablo 2’de görüldüğü üzere %0.778 doğruluk değeri üretmiş olup, bu durum verinin statik yapısı nedeniyle ardışık bağımlılık temelli mimarinin yeterli katkıyı sağlayamamasıyla ilişkilidir.

Sonuçlara göre, GCN modeli %0.667 doğruluk ile en düşük performansı sergilemiştir. GCN mimarisi coğrafi komşuluk ilişkilerini öğrenmede avantajlı olmakla birlikte, demografik, sağlık ve mobilite gibi tematik olarak farklı veri görünümlerinin tek bir grafik yapısı altında temsil edilmesi bilgi kaybına neden olmuş ve sınıflandırma başarısını sınırlandırmıştır.

Genel olarak Tablo 2’deki karşılaştırmalı sonuçlar, klasik DL modellerinin epidemik risk sınıflandırmasında belirli bir başarı sağladığını, ancak çok kaynaklı ve heterojen veri yapısının tematik ayrışmasını koruyamadıklarını göstermektedir. Özellikle yüksek risk sınıfına ilişkin performans sınırlılıkları, tek-görünüm temelli derin öğrenme yaklaşımlarının yetersiz kaldığını ortaya koymakta ve çoklu veri görünümüne dayalı bir mimarinin gerekliliğini desteklemektedir.

Tablo 2. DL Algoritmalarının Sonuçları

Algoritma	Doğruluk (%)	Kesinlik	Duyarlılık	F1-Skoru
MLP	0.889	0.889	0.897	0.889
AE-CLS	0.917	0.922	0.923	0.919
CNN	0.917	0.916	0.923	0.918
LSTM	0.778	0.787	0.795	0.772
GCN	0.667	0.693	0.692	0.668
MV-RİSKNET (ÖNERİLEN)	0.944	0.946	0.949	0.944
CGAN'LI MV-RİSKNET	0.972	0.976	0.974	0.974

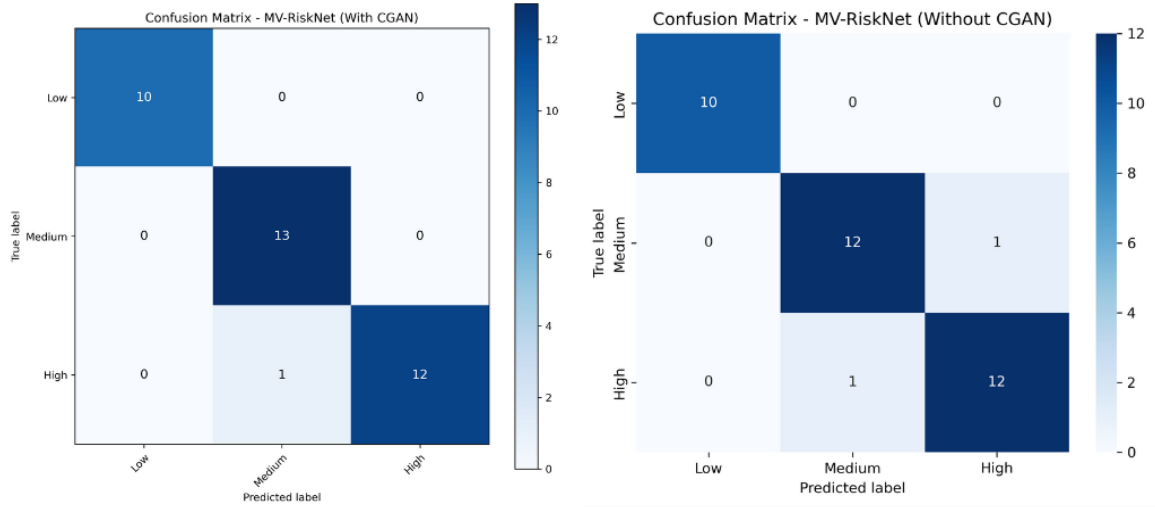
Önerilen Modelin Sınıf Bazlı Sınıflandırma Performansı

Önerilen modelin sınıf bazlı sınıflandırma performansı, CGAN uygulanmadan elde edilen karmaşıklık matrisi üzerinden değerlendirilmiştir. Modele ait karmaşıklık matrisi Şekil 16'da gösterilmiştir.

Karmaşıklık matrisi sonuçları incelendiğinde, düşük risk sınıfına ait tüm örneklerin doğru biçimde sınıflandırıldığı görülmektedir. Düşük risk sınıfında herhangi bir yanlış sınıflandırma gözlenmemiştir. Bu durum, düşük risk sınıfının çoklu veri görünümleri altında tutarlı bir şekilde temsil edilebildiğini göstermektedir.

Orta risk sınıfında, örneklerin büyük çoğunluğu doğru biçimde sınıflandırılmış olup yalnızca bir orta risk örneğinin yüksek risk sınıfı olarak tahmin edildiği görülmektedir. Benzer şekilde, yüksek risk sınıfında da bir örneğin orta risk sınıfı olarak sınıflandırıldığı tespit edilmiştir. Bu karşılıklı karışma, orta ve yüksek risk sınıfları arasındaki tematik örtüşmenin bir göstergesi olarak değerlendirilmektedir.

Genel olarak karmaşıklık matrisi analizi, önerilen modelinin risk sınıfları arasında yüksek doğrulukla ayırım yapabildiğini, ancak CGAN uygulanmadığı durumda orta ve yüksek risk sınıfları arasında sınırlı düzeyde bir karışma olduğunu ortaya koymaktadır. Bu bulgu, sınıf dengesizliğinin ve özellikle yüksek risk sınıfındaki örnek çeşitliliğinin model performansı üzerindeki etkisinin tamamen ortadan kalkmadığını göstermektedir.



Şekil 16. Karmaşıklık matrisleri

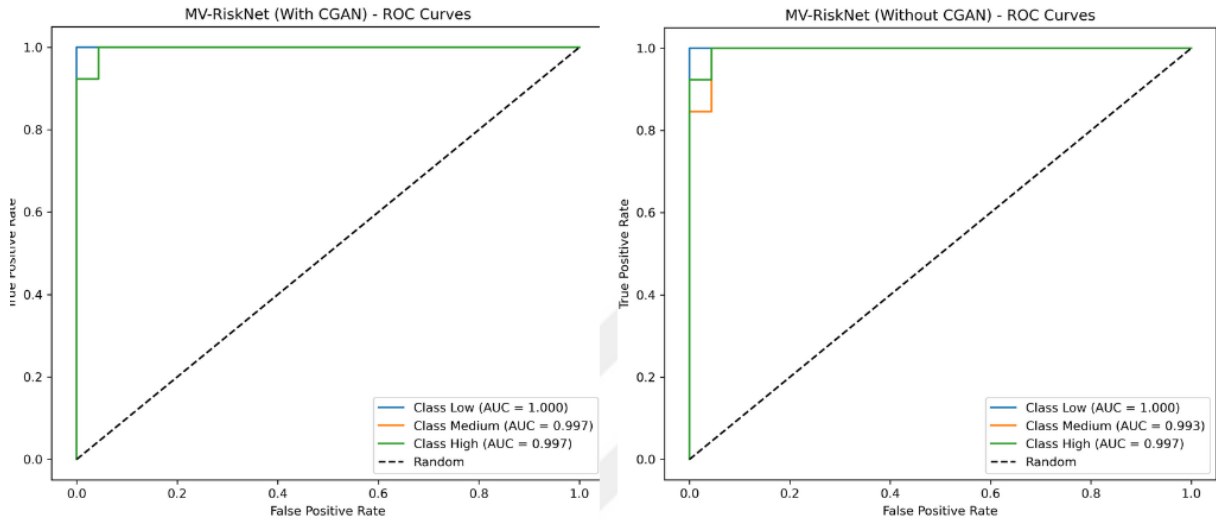
CGAN Tabanlı Veri Artırmanın MV-RiskNet Performansına Etkisi

Bu aşamada, veri setinde gözlenen sınıf dengesizliğinin önerilen modelinin sınıflandırma performansı üzerindeki etkisini incelemek amacıyla, model CGAN uygulanmadan ve CGAN tabanlı veri artırma yöntemi kullanılarak ayrı ayrı eğitilmiştir. Elde edilen sonuçlar, sınıf bazlı değerlendirme yapılabilmesi için karmaşıklık matrisleri ve sınıflar arası ayırt edicilik kapasitesinin incelenebilmesi için ROC-AUC eğrileri üzerinden karşılaştırmalı olarak analiz edilmiştir.

CGAN uygulanmadan ve CGAN uygulanarak eğitilen MV-RiskNet modeline ait karmaşıklık matrisleri Şekil 17’de gösterilmektedir. Şekil 17 incelendiğinde, CGAN uygulanmadan eğitilen modelde düşük risk sınıfına ait tüm örneklerin doğru biçimde sınıflandırıldığı, orta ve yüksek risk sınıfları arasında ise sınırlı düzeyde karşılıklı yanlış sınıflandırma gerçekleştiği görülmektedir. Özellikle bir orta risk örneğinin yüksek risk sınıfı olarak, bir yüksek risk örneğinin ise orta risk sınıfı olarak tahmin edilmesi, bu iki sınıf arasındaki tematik örtüşmeye ve yüksek risk sınıfındaki örnek çeşitliliğinin sınırlı olmasına işaret etmektedir.

CGAN tabanlı veri artırma yöntemi uygulandıktan sonra elde edilen karmaşıklık matrisi sonuçları, sınıf bazlı ayrımın daha belirgin hâle geldiğini göstermektedir. CGAN destekli MV-RiskNet modelinde düşük ve orta risk sınıflarına ait tüm örneklerin doğru biçimde sınıflandırıldığı, yüksek risk sınıfında ise yanlış sınıflandırma sayısının azaldığı görülmektedir. Bu durum, CGAN ile üretilen sentetik örneklerin özellikle az temsil edilen sınıflara ait veri çeşitliliğini artırarak modelin karar sınırlarını daha dengeli biçimde öğrenmesine katkı sağladığını göstermektedir.

CGAN tabanlı veri artırma yaklaşımının MV-RiskNet modelinin sınıflar arası ayırt edicilik kapasitesi üzerindeki etkisi ayrıca ROC–AUC analizi ile değerlendirilmiştir. CGAN uygulanmış ve uygulanmamış MV-RiskNet modellerine ait ROC–AUC eğrileri Şekil 17’de gösterilmektedir. Şekil 17 incelendiğinde, CGAN uygulanmadan eğitilen modelde düşük risk sınıfı için yüksek bir ayırt edicilik sağlandığı, buna karşın orta ve yüksek risk sınıflarında ROC eğrilerinin diyagonale daha yakın seyrettiği görülmektedir.



Şekil 17. ROC-AUC eğrileri

CGAN tabanlı veri artırma sonrası elde edilen ROC–AUC eğrileri ise tüm risk sınıflarında ayırttırma performansının arttığını göstermektedir. Özellikle orta ve yüksek risk sınıflarına ait ROC eğrilerinin diyagonalden daha uzak bir seyir izlemesi ve AUC değerlerinin artması, CGAN’ın sınıf dengesizliği problemini azaltarak MV-RiskNet modelinin sınıf ayırım kapasitesini güçlendirdiğini ortaya koymaktadır. Bu bulgular, Şekil 16’da gösterilen karmaşıklık matrisi sonuçları ile tutarlıdır.

Genel olarak Şekil 16 ve Şekil 17 birlikte değerlendirildiğinde, CGAN tabanlı veri artırma yaklaşımının önerilen modelin hem sınıf bazlı hem de global sınıflandırma performansı üzerinde olumlu bir etkiye sahip olduğu görülmektedir. Özellikle az temsil edilen risk sınıflarına ait örnek çeşitliliğinin artırılması, modelin daha dengeli ve kararlı karar sınırları öğrenmesine olanak sağlamıştır.

MV-RiskNet modelinin eğitim süreci ve öğrenme davranışları

Önerilen MV-RiskNet modelinin öğrenme davranışı, eğitim ve doğrulama sürecinde elde edilen doğruluk ve kayıp değerleri üzerinden incelenmiştir. Model, çoklu veri görünümüne dayalı yapının tutarlılığı korunacak şekilde eğitim, doğrulama ve test alt kümelerine ayrılmış

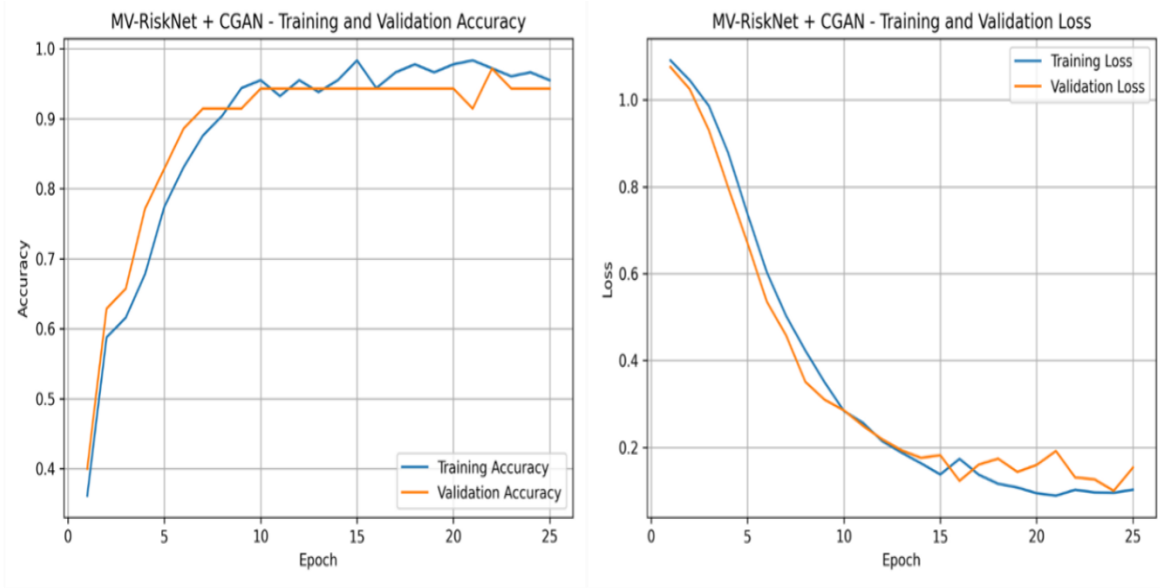
veri seti kullanılarak eğitilmiştir. Eğitim süreci boyunca modelin yakınsama davranışı ve aşırı uyum eğilimi analiz edilmiştir.

Eğitim sürecine ait doğruluk ve kayıp eğrileri Şekil 18’de gösterilmiştir. Şekil 18 incelendiğinde, MV-RiskNet modelinin ilk epoch’lardan itibaren kararlı bir öğrenme davranışı sergilediği görülmektedir. Eğitim doğruluğunun hızlı biçimde artması ve doğrulama doğruluğunun eğitim eğrisi ile paralel bir seyir izlemesi, modelin çoklu veri görünümünden anlamlı temsiller öğrenebildiğini göstermektedir.

Kayıp fonksiyonu değerleri değerlendirildiğinde, eğitim ve doğrulama kayıplarının epoch ilerledikçe düzenli biçimde azaldığı gözlemlenmiştir. Doğrulama kaybında ani artışlar veya dalgalanmalar görülmemesi, modelin aşırı uyum göstermeden genelleştirilebilir bir çözüm uzayına yakınsadığını ortaya koymaktadır. Bu durum, görünüm-özel encoder yapılarının ve düzenlileştirme stratejilerinin model eğitimine olumlu katkı sağladığını göstermektedir.

CGAN tabanlı veri artırma yöntemi kullanılarak eğitilen modelde, doğrulama kaybının daha pürüzsüz bir seyir izlediği ve doğrulama doğruluğunun daha erken aşamada dengelendiği görülmektedir. Bu durum, sentetik örneklerle artırılan eğitim veri setinin özellikle az temsil edilen sınıflar için öğrenme sürecini desteklediğini ve karar sınırlarının daha kararlı biçimde öğrenilmesine katkı sağladığını göstermektedir.

Genel olarak eğitim süreci analizi, MV-RiskNet modelinin çoklu veri görünümüne dayalı mimarisinin kararlı bir öğrenme davranışı sergilediğini ve sınıf dengesizliğine rağmen istikrarlı bir yakınsama sağladığını ortaya koymaktadır. Elde edilen bulgular, modelin yalnızca sınıflandırma performansı açısından değil, aynı zamanda öğrenme sürecinin güvenilirliği ve genelleme kapasitesi açısından da epidemik risk analizi için uygun bir yapı sunduğunu göstermektedir.



Şekil 18. Accuracy ve loss grafikleri

Modelin genellenabilirliğini değerlendirmek amacıyla eğitim ve test kümelerine ait performans metrikleri karşılaştırılmıştır. Eğitim ve test doğrulukları arasındaki farkın sınırlı olduğu, ayrıca eğitim ve doğrulama kayıp eğrilerinin benzer bir azalma eğilimi gösterdiği gözlemlenmiştir. Bu durum, önerilen modelin eğitim verisine aşırı öğrenme göstermediğini ve farklı veri örnekleri üzerinde tutarlı performans sergilediğini ortaya koymaktadır. Ayrıca erken durdurma ve düzenleme stratejileri sayesinde aşırı öğrenme eğilimi kontrol altına alınmıştır.

Bölgesel risk dağılımı bulguları

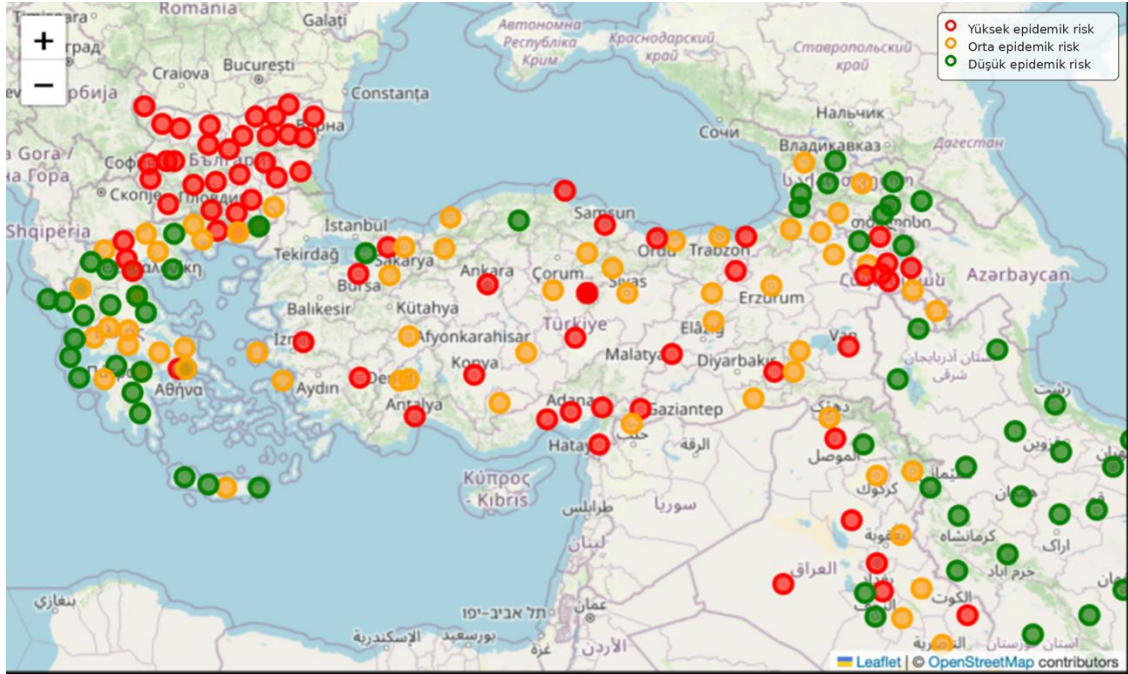
Önerilen MV-RiskNet modeli kullanılarak tahmin edilen epidemik risk seviyelerinin bölgesel dağılımı, Türkiye ve komşu ülkeleri kapsayan mekânsal risk haritaları üzerinden analiz edilmiştir. Model tarafından üretilen düşük, orta ve yüksek risk sınıflarına ait bölgesel dağılım sonuçları Şekil 19’da gösterilmektedir.

Bölgesel risk haritaları incelendiğinde, epidemik riskin homojen bir dağılım sergilemediği ve belirli coğrafi bölgelerde yoğunlaştığı görülmektedir. Orta ve yüksek risk sınıflarının, nüfus yoğunluğu, ekonomik etkileşim ve bölgesel hareketliliğin yüksek olduğu alanlarda daha baskın olduğu gözlemlenmiştir. Buna karşılık düşük risk sınıfının, nüfus yoğunluğu ve mobilite düzeyinin görece düşük olduğu bölgelerde yoğunlaştığı tespit edilmiştir.

Türkiye özelinde değerlendirildiğinde, yüksek risk sınıfının özellikle büyükşehir merkezleri, ulaşım koridorları ve sınır geçişlerine yakın bölgelerde kümелendiği görülmektedir. Bu durum, MV-RiskNet modelinin mobilite ve coğrafi görünümünden elde edilen bilgileri etkin biçimde kullanarak bölgesel etkileşimleri risk tahminine yansıtabildiğini göstermektedir.

Komşu ülkeler incelendiğinde, bazı sınır bölgelerinde orta ve yüksek risk sınıflarının süreklilik gösterdiği dikkat çekmektedir. Bu bulgu, epidemik riskin yalnızca ülke içi dinamiklerle değil, sınır ötesi hareketlilik ve bölgesel etkileşimlerle birlikte değerlendirilmesi gerektiğine işaret etmektedir. Özellikle sınır bölgelerinde gözlenen risk geçişleri, modelin çoklu veri görünümüne dayalı yapısının mekânsal süreklilikleri yakalayabildiğini ortaya koymaktadır.

Genel olarak bölgesel risk dağılımı bulguları, MV-RiskNet modelinin epidemik riskin mekânsal örüntülerini tutarlı ve epidemik açıdan anlamlı biçimde yansıttığını göstermektedir. Elde edilen haritalar, modelin yalnızca sınıflandırma performansı açısından değil, aynı zamanda bölgesel risk izleme ve erken uyarı sistemleri için kullanılabilir coğrafi çıktılar üretebildiğini ortaya koymaktadır.



Şekil 19. Önerilen model ile tahmin edilen bölgesel salgın risk seviyelerinin haritalandırılması

SONUÇ VE ÖNERİLER

Bu tez çalışmasında, Türkiye’de bölgesel salgın risklerinin belirlenmesi, sınıflandırılması ve mekânsal olarak haritalandırılması amacıyla iki aşamalı bir analitik çerçeve geliştirilmiştir. Çalışmanın ilk aşamasında, Türkiye’nin 81 ili için geleneksel makine öğrenimi algoritmaları kullanılarak il bazlı salgın risk seviyeleri tahmin edilmiştir. İkinci aşamada ise, Türkiye ve komşu ülkeleri kapsayan daha geniş bir coğrafyada epidemik risk tahmini gerçekleştirmek üzere çoklu veri görünümüne ve dikkat temelli füzyon mekanizmasına sahip MV-RiskNet modeli önerilmiştir. Her iki aşama, farklı veri yapıları ve modelleme yaklaşımları altında kapsamlı biçimde değerlendirilmiştir.

Makine Öğrenimi Temelli Bölgesel Risk Sınıflandırma Sonuçları

Çalışmanın birinci aşamasında oluşturulan veri seti demografik, coğrafi, ekonomik ve sağlık altyapısına ilişkin göstergeleri içeren bütünleşik bir yapı sunmaktadır. Bu veri seti üzerinde LR, NB, RF, SVM ve XGB algoritmaları karşılaştırmalı olarak değerlendirilmiştir. Elde edilen sonuçlar, XGBoost algoritmasının %98 doğruluk oranı ile en yüksek sınıflandırma performansını sergilediğini göstermektedir. RF ve SVM modellerinin %96 doğruluk ile güçlü ve dengeli sonuçlar üretmesi, özellikle demografik ve sağlık altyapısı değişkenlerinin risk sınıflandırmasında belirleyici rol oynadığını ortaya koymaktadır.

Birinci aşama bulguları, Türkiye genelinde salgın risk seviyelerinin homojen bir dağılım sergilemediğini göstermiştir. İstanbul, Ankara, İzmir, Bursa, Gaziantep ve Diyarbakır gibi nüfus yoğunluğu ve ulaşım erişilebilirliği yüksek illerde orta ve yüksek risk seviyelerinin baskın olduğu buna karşılık İç Anadolu’nun kırsal kesimleri, Doğu Anadolu’nun düşük nüfus yoğunluklu bölgeleri ve Karadeniz’in dağlık alanlarında düşük risk seviyelerinin daha yaygın olduğu gözlenmiştir. Bu sonuçlar, salgın riskinin yalnızca nüfus büyüklüğü ile değil, ekonomik hareketlilik, sağlık altyapısı kapasitesi, ulaşım ağları ve mekânsal bağlantılar gibi çok boyutlu faktörlerin ortak etkisiyle şekillendiğini ortaya koymaktadır.

Ayrıca gerçekleştirilen mekânsal analizler ve simülasyonlar, yüksek riskli illerden komşu bölgelere doğru salgın riskinin mekânsal süreklilik gösterdiğini ortaya koymuştur. Bu bulgu, coğrafi komşuluk ilişkilerinin ve bölgesel etkileşimlerin salgın riskinin yayılım dinamiklerinde önemli bir rol oynadığını göstermektedir.

Çok-Görünümlü Derin Öğrenme Yaklaşımı ile Bölgesel Epidemik Risk Tahmini

Tezin ikinci aşamasında, epidemik riskin çok boyutlu yapısını daha etkin biçimde temsil edebilmek amacıyla MV-RiskNet modeli geliştirilmiştir. Model, demografik, sağlık, coğrafi ve mobilite görünümelerini ayrı alt ağlar aracılığıyla işleyerek her bir görünüm için bağımsız temsiller öğrenmekte olup bu temsilleri dikkat temelli bir füzyon mekanizması ile bütünleştirmektedir. Veri setinde gözlenen sınıf dengesizliğini gidermek amacıyla CGAN tabanlı sentetik veri üretimi uygulanmış ve özellikle yüksek risk sınıfına ait örnek çeşitliliği artırılmıştır.

Deneyisel sonuçlar, MV-RiskNet modelinin %97,2 doğruluk ve %97,4 F1-skoru değerlerine ulaştığını göstermektedir. Elde edilen ROC–AUC değerlerinin tüm risk sınıfları için 0.997 ve üzerinde olması, modelin sınıflar arası ayrıştırma kapasitesinin oldukça yüksek olduğunu ortaya koymaktadır. Bu performans, hem temel derin öğrenme mimarileri (MLP, CNN, LSTM, AE-CLS) hem de grafik tabanlı yaklaşımlara kıyasla belirgin bir üstünlük sunmaktadır. Dikkat mekanizmasının özellikle demografik ve coğrafi görünümlere daha yüksek ağırlık atadığını mobilite görünümünün ise yüksek risk sınıfının ayrıştırılmasında tamamlayıcı bir rol üstlendiği gözlenmiştir.

Bölgesel Risk Dağılımının Mekânsal Değerlendirilmesi

Her iki aşamada elde edilen sonuçların mekânsal olarak değerlendirilmesi, salgın riskinin belirli bölgelerde kümelenme eğilimi gösterdiğini ortaya koymaktadır. İl bazlı analizlerde yüksek risk seviyelerinin büyükşehirler ve ulaşım ağlarının yoğun olduğu bölgelerde yoğunlaştığı görülürken, MV-RiskNet modeli tarafından üretilen bölgesel risk haritaları bu örüntünün uluslararası sınırları aşarak daha geniş bir coğrafyada süreklilik gösterdiğini ortaya koymuştur.

Özellikle Türkiye'nin doğu ve güney sınırlarına yakın bölgelerde yüksek risk seviyelerinin belirgin biçimde yoğunlaştığı gözlemlenirken Balkanlar ve Kafkasya bölgelerinde ise daha heterojen bir risk dağılımı sergilendiği gözlenmiştir. Mobilite verilerinin modele dâhil edilmesiyle, risk seviyelerinin ticaret merkezleri, yoğun ulaşım aksları ve sınır geçiş noktaları boyunca arttığı belirlenmiştir. Bu sonuçlar, model çıktılarının mekânsal olarak tutarlı olduğunu ve epidemik riskin sınır-ötesi bir olgu olarak ele alınması gerektiğini göstermektedir.

Genel Değerlendirme

Bu tezde geliştirilen iki aşamalı modelleme yaklaşımı, bölgesel salgın riskinin çok katmanlı ve çok boyutlu bir yapı sergilediğini açık biçimde ortaya koymaktadır. ML tabanlı modeller, il düzeyinde risk sınıflandırması için güçlü ve açıklanabilir sonuçlar sunarken MV-

RiskNet modeli çoklu veri görünümü, dikkat mekanizması ve sentetik veri üretimi sayesinde daha geniş coğrafi ölçekte daha derin ve dengeli temsiller üretmiştir.

Elde edilen bulgular, epidemik riskin demografik, ekonomik, coğrafi ve mobilite temelli faktörlerin etkileşimiyle şekillendiğini ve bölgesel hareketliliğin salgın dinamiklerinde kritik bir rol oynadığını göstermektedir. Bu yönüyle çalışma hem ulusal hem de uluslararası ölçekte epidemik risk izleme, bölgesel önceliklendirme ve karar destek sistemleri için ölçeklenebilir bir metodolojik çerçeve sunmaktadır.

Sınırlılıklar ve Gelecek Çalışmalar

Bu çalışmanın bazı sınırlılıkları bulunmaktadır. Kullanılan veri setlerinin büyük bölümü statik veya yıllık nitelikte olduğundan, salgın riskinin zamansal değişimi gerçek zamanlı olarak modellenememiştir. Farklı ülkelerden derlenen verilerin kurumsal ve istatistiksel standartlar açısından heterojen olması, veri tutarlılığında kısmi sınırlılıklara yol açabilmektedir. CGAN tabanlı sentetik veri üretimi sınıf dengesizliğini azaltmış olsa da gerçek veri çeşitliliğini tam olarak yansıtamayabilir.

Gelecek çalışmalar, daha yüksek zamansal çözünürlüğe sahip verilerin kullanılmasıyla salgın riskinin dinamik yapısının modellenmesine odaklanabilir. Spatio-temporal derin öğrenme modellerinin (ConvLSTM, Transformer tabanlı modeller, zamansal GNN'ler) entegrasyonu, risk tahminlerinin doğruluğunu artırabilir. Mekânsal çözünürlüğün ilçe ve mahalle düzeyine indirilmesi, mikro ölçekte risk farklılıklarının belirlenmesine olanak sağlayacaktır. Ayrıca model çıktılarının gerçek zamanlı karar destek sistemlerine entegre edilmesi, erken uyarı ve müdahale mekanizmalarının geliştirilmesi açısından önemli bir potansiyel sunmaktadır.

- Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. DOI: <https://doi.org/10.1109/TPAMI.2018.2798607>
- Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). DOI: <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, 1–13. DOI: <https://doi.org/10.1186/s12864-019-6413-7>
- Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission using LSTM. *Chaos, Solitons & Fractals*, 135, 109864. DOI: <https://doi.org/10.1016/j.chaos.2020.109864>
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. DOI: <https://doi.org/10.1126/science.aba9757>
- Chumachenko, D., Meniailov, I., Bazilevych, K., Chumachenko, T., & Yakovlev, S. (2022). Investigation of statistical machine learning models for COVID-19 epidemic process simulation: Random forest, K-nearest neighbors, gradient boosting. *Computation*, 10(6), 86. DOI: <https://doi.org/10.3390/computation10060086>
- Çakmak, M. A., Kurt, M. E., & Çakmak, C. (2022). Makine Öğrenmesi Algoritmaları ile Covid-19 Hastalarının Mortalite Risklerinin Hesaplanması. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 13(35), 994–1011. DOI: <https://doi.org/10.21076/vizyoner.1074212>
- Dabbagh, R., & Yousefi, S. (2019). A hybrid decision-making approach based on FCM and MOORA for occupational health and safety risk analysis. *Journal of Safety Research*, 71, 111–123. DOI: <https://doi.org/10.1016/j.jsr.2019.09.021>
- Edward, J., & Biddle, D. J. (2017). Using geographic information systems (GIS) to examine barriers to healthcare access for Hispanic and Latino immigrants in the US south. *Journal of Racial and Ethnic Health Disparities*, 4, 297–307. DOI: <https://doi.org/10.1007/s40615-016-0229-9>
- Franch-Pardo, I., et al. (2020). Spatial analysis and GIS in the study of COVID-19: A review. *Science of the Total Environment*, 739, 140033. DOI: <https://doi.org/10.1016/j.scitotenv.2020.140033>
- Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5), 829–864. DOI: https://doi.org/10.1162/neco_a_01273
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*. DOI: <https://doi.org/10.48550/arXiv.1412.6572>
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 235. DOI: <https://doi.org/10.3390/info15040235>

- Guo, Z., Lin, Q., & Meng, X. (2023). A comparative study on deep learning models for COVID-19 forecast. *Healthcare*, 11(17), 2400. DOI: <https://doi.org/10.3390/healthcare11172400>
- Gündüz, M. Ş., & Işık, G. (2023). A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models. *Journal of Real-Time Image Processing*, 20(5). DOI: <https://doi.org/10.1007/s11554-023-01276-w>
- Gündüz, M. Ş., & Işık, G. (2023). A new YOLO-based method for social distancing from real-time videos. *Neural Computing and Applications*, 35, 15261–15271. DOI: <https://doi.org/10.1007/s00521-023-08556-3>
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4), 599–653. DOI: <https://doi.org/10.1137/S0036144500371907>
- Ilin, C., Annan-Phan, S., Tai, X. H., Mehra, S., Hsiang, S., & Blumenstock, J. E. (2021). Public mobility data enables COVID-19 forecasting and management at local and global scales. *Scientific Reports*, 11, 13531. DOI: <https://doi.org/10.1038/s41598-021-92892-8>
- Jia, J. S., et al. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 583, 389–394. DOI: <https://doi.org/10.1038/s41586-020-2284-y>
- Kang, D., Choi, H., Kim, J.-H., & Choi, J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, 94, 96–102. DOI: <https://doi.org/10.1016/j.ijid.2020.03.076>
- Karacıoğlu, A. A., Tanışman, S., & Bulut, H. (2021). Türkiye'de COVID-19 Bulaşısının ARIMA Modeli ve LSTM Ağı Kullanılarak Zaman Serisi Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (32), 288–297. DOI: <https://doi.org/10.31590/ejosat.1039394>
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115(772), 700–721. DOI: <https://doi.org/10.1098/rspa.1927.0118>
- Khalid, J., Chuanmin, M., Altaf, F., Shafqat, M. M., Khan, S. K., & Ashraf, M. U. (2024). AI-driven risk management and sustainable decision-making: Role of perceived environmental responsibility. *Sustainability*, 16(16), 6799. DOI: <https://doi.org/10.3390/su16166799>
- Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497. DOI: <https://doi.org/10.1126/science.abb4218>
- Leslie, T. F., Delamater, P. L., Yang, Y., & Wang, J. (2021). Spatial modeling of COVID-19 incidence in the United States: A spatial regression approach. *Spatial and Spatio-temporal Epidemiology*, 39, 100442. DOI: <https://doi.org/10.1016/j.sste.2021.100442>
- Lever, J., Krzywinski, M., & Altman, N. (2016). Points of significance: Classification evaluation. *Nature Methods*, 13(8), 603–604. DOI: <https://doi.org/10.1038/nmeth.3945>
- Li, Q., Guan, X., Wu, P., Wang, X., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine*. DOI: <https://doi.org/10.1056/NEJMoa2001316>
- Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*. DOI: <https://doi.org/10.48550/arXiv.1706.06083>
- Miyazaki, Y., Kawakami, M., Kondo, K., Hirabe, A., Kamimoto, T., Akimoto, T., ... & Tsuji, T. (2024). Logistic regression analysis and machine learning for predicting post-stroke

- gait independence: A retrospective study. *Scientific Reports*, 14(1), 21273. DOI: <https://doi.org/10.1038/s41598-024-72206-4>
- Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence. *Science of the Total Environment*, 728, 138884. DOI: <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). Spatial modeling of COVID-19 incidence rates using GIS-based methods: A global review. *International Journal of Environmental Research and Public Health*, 17(12), 4204. DOI: <https://doi.org/10.3390/ijerph17124204>
- Naim, A. (2023). Application of machine learning techniques to identify the business financial risks. *Academy of Marketing Studies Journal*, 27(S5), 1–11.
- Nazia, N., Butt, Z. A., Bedard, M. L., Tang, W. C., Sehar, H., & Law, J. (2022). Methods used in the spatial and spatiotemporal analysis of COVID-19 epidemiology: A systematic review. *International Journal of Environmental Research and Public Health*, 19(14), 8267. DOI: <https://doi.org/10.3390/ijerph19148267>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Powers, D., & Ailab. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2, 2229–3981. DOI: <https://doi.org/10.9735/2229-3981>
- Pullano, G., Di Domenico, L., Sabbatini, C. E., Valdano, E., Turbett, S. E., & Colizza, V. (2021). Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature*, 590, 134–139. DOI: <https://doi.org/10.1038/s41586-020-03095-6>
- Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6), 96–108. DOI: <https://doi.org/10.1109/MSP.2017.2738401>
- Razavi-Termeh, S. V., Samadi, R., Heidarzadeh, N., & Choubin, B. (2021). COVID-19 risk mapping with considering socio-economic criteria using machine learning algorithms. *International Journal of Environmental Research and Public Health*, 18(18), 9657. DOI: <https://doi.org/10.3390/ijerph18189657>
- Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. S. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, 135, 109853. DOI: <https://doi.org/10.1016/j.chaos.2020.109853>
- Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W., & Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8, 101489–101499. DOI: <https://doi.org/10.1109/ACCESS.2020.2997311>
- Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, 62, 102418. DOI: <https://doi.org/10.1016/j.scs.2020.102418>
- Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India–USA comparative case study. *Chaos, Solitons & Fractals*, 140, 110227. DOI: <https://doi.org/10.1016/j.chaos.2020.110227>

- Stahlschmidt, S. R., & Gerlach, J. (2022). Multimodal deep learning for biomedical data fusion: A systematic review. *Briefings in Bioinformatics*, 23(2), bbab569. DOI: <https://doi.org/10.1093/bib/bbab569>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. DOI: <https://doi.org/10.1016/j.aci.2018.08.003>
- Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. *International Journal of Environmental Research and Public Health*, 17(11), 3903. DOI: <https://doi.org/10.3390/ijerph17113903>
- Yan, X., Xie, H., Zhang, Y., Zhang, C., & Sun, Z. (2021). Deep multi-view learning methods: A review. *Neurocomputing*, 448, 106–129. DOI: <https://doi.org/10.1016/j.neucom.2021.03.090>
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, 140, 110121. DOI: <https://doi.org/10.1016/j.chaos.2020.110121>
- Zhao, F., Cai, Z., & Li, J. (2024). Deep multimodal data fusion: A survey. *ACM Computing Surveys*. DOI: <https://doi.org/10.1145/3649447>

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı:	Beyzanur OKUDAN
Uyruğu:	
E-mail:	
Eğitim	
Lise:	İsmail Karataş Anadolu Sağlık Meslek Lisesi (2017)
Lisans:	Atatürk Üniversitesi, Mühendislik Fakültesi Bilgisayar Mühendisliği (2024)
Yüksek lisans:	Atatürk Üniversitesi, Yazılım Mühendisliği
Yabancı Dil Bilgisi	
Tezden Üretilmiş Yayınlar	
1. Karcıoğlu, A. A., & Okudan, B. (2025). Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye’de Bölgesel Salgın Risk Analizi ve Haritalandırılması. Journal of the Institute of Science and Technology, 15(4), 1178-1192. https://doi.org/10.21597/jist.1683256	