

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**PARTİKÜL MADDE KONSANTRASYONU TAHMİNİNDE DERİN
ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI
İNCELENMESİ**

DOKTORA TEZİ

İpek AKSANGÜR

Çevre Mühendisliği Anabilim Dalı

TEMMUZ 2023

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**PARTİKÜL MADDE KONSANTRASYONU TAHMİNİNDE DERİN
ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI
İNCELENMESİ**

DOKTORA TEZİ

İpek AKSANGÜR

Çevre Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Beytullah EREN

Ortak Danışman: Dr. Öğr. Üyesi Caner ERDEN

TEMMUZ 2023

İpek AKSANGÜR tarafından hazırlanan “PARTİKÜL MADDE KONSANTRASYONU TAHMİNİNDE DERİN ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI İNCELENMESİ” adlı tez çalışması 24.07.2023 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Çevre Mühendisliği Anabilim Dalı Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :

Jüri Üyesi :

Jüri Üyesi :

Jüri Üyesi :

Jüri Üyesi :



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “PARTİKÜL MADDE KONSANTRASYONU TAHMİNİNDE DERİN ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI İNCELENMESİ” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(...../...../2023)

(imza)

İpek AKSANGÜR





Aileme



TEŞEKKÜR

Doktora eğitimim süresince yönlendirmeleri, katkıları, ilgi ve desteği için danışman hocam Doç. Dr. Beytullah EREN'e, bilimsel tecrübelerini benimle paylaşan ve yönlendirmeleri ile destek olan ortak danışman hocam Dr. Öğr. Üyesi Caner ERDEN'e, bu uzun süreçte desteklerini ve katkılarını her zaman hissettiğim aileme sonsuz teşekkürlerimi sunarım.



İpek AKSANGÜR



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1. GİRİŞ	1
1.1. Tezin Amaç ve Kapsamı	2
1.2. Tezin Yapısı	2
2. HAVA KİRLİLİĞİ	5
2.1. PM Kaynakları	6
2.2. Hava Kirliliğinin İnsan Sağlığına Etkisi	13
2.3. Yasal Mevzuat	16
3. LİTERATÜR ARAŞTIRMASI	19
3.1. Geleneksel Zaman Serisi Yöntemleri	19
3.2. Makine Öğrenmesi Yöntemleri	22
3.3. Derin Öğrenme Yöntemleri	23
3.4. Araştırma Hedefi ve Bilimsel Katkı	27
4. MATERYAL VE METOT	29
4.1. Çalışma Alanı	30
4.2. Hava Kirliliği Verileri	31
4.3. Meteorolojik Veriler	33
4.4. Veri Setine İstatistiksel Bakış	34
4.5. Modelleme Süreci	35
4.5.1. Veri ön işleme ve özellik seçimi (VÖİ-ÖS)	35
4.5.1.1. Eksik veri analizi	36
4.5.1.2. Aykırı veri analizi	37
4.5.1.3. Veri ölçeklendirme	38
4.5.1.4. Özellik seçimi	39
4.5.2. Zaman serisi çerçevesi oluşturma stratejisi	39
4.5.3. Derin öğrenme uygulamaları	40
4.5.3.1. Yinelemeli sinir ağlar (RNN)	41
4.5.3.2. Uzun kısa süreli bellek (LSTM)	42
4.5.3.3. Geçişli tekrarlayan birim (GRU)	46
4.6. Performans Değerlendirme İndeksleri	47
4.7. Kullanılan Programlama Dili Kütüphaneleri	48

5. ARAŞTIRMA BULGULARI	51
5.1. PM ₁₀ Tahmininde LSTM'nin Uygunluğu	51
5.1.1. Veri ön işleme ve özellik seçimi adımı sonuçları	51
5.1.2. LSTM tahmin sonuçları	55
5.2. PM _{2.5} Tahmininde DÖ Modellerinin Performanslarının Karşılaştırılması	67
5.2.1. Gizli katman sayısı ve katmandaki nöron sayısı seçimi.....	67
5.2.2. Farklı derin öğrenme algoritmalarının tahmin performans değerlendirme	69
5.2.3. Önerilen model genelleme kabiliyetinin değerlendirilmesi	74
6. TARTIŞMA VE SONUÇLAR	77
KAYNAKLAR.....	79
ÖZGEÇMİŞ.....	91



KISALTMALAR

AB	: Avrupa Birliđi
AÇA	: Avrupa Çevre Ajansı
ADAM	: Adaptive Moment Estimation
ARMA	: Autoregressive Moving Average
Bi- GRU	: Bi-Directional Gated Recurrent Unit
Bi-LSTM	: Bi-Directional Long Short Term Memory
CB-GRU	: Convolutional-Based Bidirectional Gated Recurrent Unit
CEEMD	: Complementary Ensemble Empirical Mood Decomposition
CLRTAP	: The Convention on Long-Range Transboundary Air Pollution
CNN	: Convolutional Neural Network
ÇİSİP	: Çevre İklim ve Sağlık için İşbirliđi Projesi
DFNN	: Deep Feed Forward Neural Network
DSÖ	: Dünya Sağlık Örgütü
DTR	: Decision Tree Regression
EMEP	: European Monitoring and Evaluation Programme
GBR	: Gradient Boosting Regressor
GRU	: Gated Recurrent Unit
HKDYY	: Hava Kalitesi Deđerlendirme ve Yönetimi Yönetmeliđi
IARC	: International Agency for Research on Cancer
LSTM	: Long-Short Term Memory
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MART	: Multiple Additive Regression Tree
MGM	: Meteoroloji Genel Müdürlüđü
MLP	: Multi-Layer Perceptron
MSE	: Mean Squared Error
NSW Health	: New South Wales Department of Health
OECD	: Organisation for Economic Co-operation and Development
ONEIROS	: Open-ended Neuro-Electronic Intelligent Robot Operating System

PAH	: Polisiklik Aromatik Hidrokarbon
PCR	: Principal Component Regression
PM	: Partikül Madde
ReLU	: Rectified Linear Unit
RF	: Random Forest
RMSE	: Root Mean Square Error
RNN	: Recurrent Neural Network
SD	: Standard Deviation
SİM	: Sürekli İzleme Merkezi
SVR	: Support Vector Regression
TL-LSTM	: Long Short Term Memory With Transfer Learning
TMMOB	: Türk Mühendis ve Mimar Odaları Birliği
TÜİK	: Türkiye İstatistik Kurumu
U.S. EPA	: U.S. Environmental Protection Agency
VÖİ-ÖS	: Veri Ön İşleme ve Özellik Seçimi
WHO	: World Health Organization
YZ	: Yapay Zeka
YSA	: Yapay Sinir Ağı

SİMGELER

PM_{2.5}	: Partikül madde 2,5 µm
PM₁₀	: Partikül madde 10 µm
p	: Anlamlılık düzeyi
r	: Korelasyon katsayısı
O₃	: Yer seviyesinde ozon
CO	: Karbon monoksit
NO	: Azot monoksit
Pb	: Kurşun
SO₂	: Kükürt dioksit
µm	: Mikrometre
SiO₂	: Silisyum dioksit
Al₂O₃	: Aluminyum [III] oksit
Fe₂O₃	: Demir [III] oksit
NO_x	: Azot oksit
As	: Arsenik
Cd	: Kadmiyum
Ni	: Nikel
mg/m³	: miligram/metreküp
m/sn	: metre/saniye
%	: yüzde
mbar	: milibar
m	: metre
R²	: determinasyon (belirlilik) katsayısı
Tg	: Teragram
°C	: Santigrat derece



TABLO LİSTESİ

Sayfa

Tablo 2.1. Hava kirleticileri.....	6
Tablo 2.2. Hava kirliliği yönetimi kapsamında bazı ulusal/uluslararası mevzuatlar	17
Tablo 2.3. PM ₁₀ için sınır değerleri	18
Tablo 2.4. PM _{2.5} için sınır değerleri.....	18
Tablo 4.1. Veri kaynakları olarak seçilen istasyonlar.	32
Tablo 4.2. BHKİİ veri seti istatistiksel dağılımı	34
Tablo 4.3. KHKİİ veri seti istatistiksel dağılımı	35
Tablo 5.1. VÖİ-ÖS öncesi ve sonrası her bir hava kirliliği ve meteorolojik parametrenin veri setlerine ait tanımlayıcı istatistikler.	54
Tablo 5.2. Geliştirilen LSTM modelleri karar bağı parametreleri.	55
Tablo 5.3. Geliştirilmiş LSTM modelleri tek yönlü ANOVA sonuçları.....	62
Tablo 5.4. Tukey ve Fisher test gruplama bilgileri	63
Tablo 5.5. PM _{2.5} tahmininde model performans sonuçları.	68
Tablo 5.6. Model#17 ait hiperparametre değerleri.	70
Tablo 5.7. Mevcut literatür ile karşılaştırma.	73
Tablo 5.8. Dokuz farklı istasyon model#17 genelleme kabiliyeti performansı (p<0,05).	74
Tablo 5.9. SD/RMSE performans dereceleri.....	75



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Partikül maddelerin sınıflandırılması	7
Şekil 2.2. Partikül madde dağılımı ve terimleri	8
Şekil 2.3. Antropojenik kaynakların sınıflandırılması	9
Şekil 2.4. Parçacık boyutları ve kaynakları	10
Şekil 2.5. Partikül madde oluşum aşamaları	11
Şekil 2.6. ABD’de PM kaynaklarına göre salınım miktarları.....	13
Şekil 2.7. Hava kirliliğinden etkilenen organlar ve hastalıklar.....	14
Şekil 3.1. Zaman serileri analiz yöntemleri	20
Şekil 3.2. 01/01/2019-30/11/2019 tarihleri arasında Kağıthane ilçesi için saatlik hava kalitesi ve meteorolojik veriler zaman serisi grafiği	22
Şekil 4.1. Tez araştırma konuları	29
Şekil 4.2. Araştırma konusu 1 akış şeması.	30
Şekil 4.3. Araştırma konusu 2 akış şeması.	30
Şekil 4.4. Modelleme süreci	35
Şekil 4.5. Zaman serisi çerçevesi oluşturma stratejisi	40
Şekil 4.6. Temel RNN yapısı	42
Şekil 4.7. LSTM yapısı a)birim b) döngü.....	43
Şekil 4.8. GRU yapısı	47
Şekil 4.9. AK-1 LSTM mimarisi	49
Şekil 5.1. Özellikler arası korelasyon ısı haritası.....	52
Şekil 5.2. Model tahmin performansı a) VÖİ-ÖS’siz LSTM b) VÖİ-ÖS’li LSTM. .	56
Şekil 5.3. PM ₁₀ için ölçülen ve tahmin edilen değerlerin karşılaştırılması a) VÖİ-ÖS’siz LSTM modelleri b)VÖİ-ÖS’li LSTM modelleri.....	58
Şekil 5.4. Model tahminleri ve ölçülen değerler arasındaki dağılım grafikleri a-c)VÖİ-ÖS’siz LSTM modelleri d-f)VÖİ-ÖS’li LSTM modelleri.	59
Şekil 5.5.VÖİ-ÖS’siz ve VÖİ-ÖS’li LSTM modellerinin RMSE, MAE ve R ² değerlerinin çubuk grafiği.	60
Şekil 5.6. En iyi LSTM modelinin ilgili hatalarının gözlemlenen PM ₁₀ değerlerine karşı çizilmiş dağılım grafiği.	60
Şekil 5.7. Çoklu bağımsız çalışmalarda önerilen LSTM modelleri için MAE performans değerlerinin kutu grafiği.	61
Şekil 5.8. Beşiktaş ve Kağıthane bölgeleri için VÖİ-ÖS’siz ve VÖİ-ÖS’li LSTM modellerinin RMSE, MAE ve R ² değerlerinin çubuk grafikleri.....	63
Şekil 5.9. Beşiktaş bölgesi için model tahmini ve ölçülen değerler arasındaki dağılım a-c) VÖİ-ÖS’siz LSTM modelleri d-f) VÖİ-ÖS’li LSTM modelleri.	65
Şekil 5.10. Kağıthane bölgesi için model tahminleri ile ölçülen değerler arasındaki dağılım grafikleri a-c)VÖİ-ÖS’siz LSTM modelleri d-f)VÖİ-ÖS’li LSTM modeller.....	66

Şekil 5.11. LSTM+LSTM modeli kullanılarak tahmin edilen ve ölçülen $PM_{2.5}$ değerlerinin karşılaştırılması a)KHKİİ eğitim veri seti ve b)KHKİİ test veri seti ($p<0,05$).	70
Şekil 5.12. Model#17 tahmin edilen ve ölçülen $PM_{2.5}$ değerlerinin dağılım grafikleri a) KHKİİ eğitim veri seti ve b) KHKİİ test veri seti ($p < 0,05$).	71
Şekil 5.13. Tüm modellerin keman grafikleri a)Model#1-Model#6 b)Model#7-Model#12 c)Model#13-Model#18 d) Model#19-Model#24.....	72
Şekil 5.14. Model#17 genelleme kabiliyeti dağılım grafikleri (Aksaray, Avcılar, Beşiktaş, Kadıköy, Kartal, Maslak, Sultangazi, Tuzla, Ümraniye)	76



PARTİKÜL MADDE KONSANTRASYONU TAHMİNİNDE DERİN ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI İNCELENMESİ

ÖZET

Hava kirliliği, insan sağlığını etkileyen en önde gelen faktörlerden biridir. Kötü hava kalitesi, hem fiziksel hem de zihinsel açıdan insan sağlığına ve yaşam kalitesine olumsuz etkiler yaratmaktadır. Özellikle Partiküler Madde (PM) hava kirliliğinin ciddi sağlık sorunlarına yol açtığı iyi bilinmektedir. Bu nedenle, yoğun nüfuslu şehirlerde hava kalitesi yönetimi, hava kirliliğinin olumsuz etkilerini azaltmak veya engellemek adına vazgeçilmez bir konudur. Partiküler Madde (PM_{2.5} ve PM₁₀) kirleticilerinin kirlilik seviyelerini tahmin etmek için güvenilir modeller geliştirmek, karar vericiler için hayati bir araç olacaktır. Bu çalışma, PM kirliliği tahmini için yüksek doğruluklu yapay zeka tabanlı tahmin modelleri geliştirmeyi amaçlamaktadır. Bu hedef doğrultusunda, hava kalitesi izleme istasyonlarından elde edilen geniş veri setleri üzerinde derin öğrenme algoritmalarıyla çalışılmıştır.

Bu çalışma iki ayrı araştırma konusunu ele almaktadır. İlk araştırma konusu, PM₁₀ kirlilik tahmini için farklı Uzun-Kısa Süreli Bellek (LSTM) algoritmalarının uygunluğunun değerlendirilmesi ve buna ek olarak Veri Ön İşleme ve Özellik Seçimi (VÖİ-ÖS) sürecinin LSTM modellerinin tahmin doğruluğu üzerindeki etkisinin incelenmesidir. Bu amacı gerçekleştirmek için Yalın (Vanilla), Çift Yönlü (Bi-Directional) ve Yığılı (Stacked) olmak üzere üç farklı LSTM modeli geliştirilmiştir. Tüm tahmin modellerinin sonuçları RMSE, MAE ve R² performans endeksleri kullanılarak karşılaştırılmıştır. Deney sonuçları, önerilen LSTM modellerinin VÖİ-ÖS işlemi uygulandığında daha yüksek tahmin performansı sergilediğini göstermiştir ve bu modellerin saatlik PM₁₀ konsantrasyonlarını tahmin etmek için kullanılabileceğini ortaya koymuştur. Ek olarak, önerilen VÖİ-ÖS'li LSTM modellerinin farklı bölgelerden elde edilen veri setleri için de oldukça iyi bir tahmin performansı gösterdiği görülmüştür.

Çalışmanın ikinci araştırma konusu olarak PM_{2.5} tahmininde Uzun-Kısa Süreli Bellek (LSTM), Yinelemeli Sinir Ağı (RNN) ve Geçişli Tekrarlayan Birimler (GRU) olmak üzere üç tür derin öğrenme algoritmasının tahmin performanslarının karşılaştırılmıştır. Tüm tahmin modelleri sonuçları MSE, RMSE, MAE, MAPE ve R² performans indeksleri kullanılarak kıyaslanmıştır. Deney sonuçları, LSTM+LSTM modelinin, diğer derin öğrenme algoritmalarına kıyasla eğitim ve test veri setleri için anlamlılık düzeyinde ($p < 0,05$) 0,98 ve 0,97'lik bir R² ile önemli ölçüde daha üstün bir performans gösterdiğini ortaya koymuştur. Ayrıca, önerilen modelin genelleme kabiliyetini değerlendirmek için İstanbul'un dokuz farklı semtindeki istasyonlardan alınan bir yıllık veriler ile model genelleme çalışmaları yapılmıştır. Sonuçlar önerilen LSTM+LSTM modelinin tüm istasyonlar için 0,90 ($p < 0,05$) ve üzeri R² doğruluk oranı ile iyi bir genelleme yeteneğine sahip ve doğrusal ve durağan olmayan çok boyutlu zaman serisi verileri için kullanılabilir olduğunu göstermiştir. Ayrıca sonuçlar

literatürdeki diğer çalışmalarla karşılaştırılarak, önerilen LSTM+LSTM modelinin $PM_{2.5}$ konsantrasyonlarını tahmin etmede daha iyi bir performansa sahip olduğu görülmüştür.



A COMPARATIVE ANALYSIS OF DEEP LEARNING APPROACHES FOR THE PREDICTION OF PARTICULATE MATTER CONCENTRATION

SUMMARY

Air pollution is one of the primary factors affecting human health. Poor air quality has adverse physical and mental effects on human health and quality of life. In particular, it is known that particulate matter (PM) air pollution leads to serious health problems. Therefore, air quality management is essential in densely populated cities to reduce or prevent the adverse effects of air pollution. Developing reliable models for predicting particulate matter pollution levels (PM_{2.5} and PM₁₀) will be essential for decision-makers. This study aims to develop high-accuracy artificial intelligence-based prediction models for PM pollution. Deep learning algorithms were run extensive data obtained from air quality monitoring stations, which are environmental facilities.

The study consists of two primary research areas. The first research subject involves an assessment of various Long-Short Term Memory (LSTM) algorithms' applicability in predicting PM₁₀ pollution levels while concurrently exploring the impact of the Data Preprocessing and Feature Selection (DPFS) process on the predictive accuracy of the LSTM models. The DP-FS phase encompasses multiple steps, including the analysis of missing data, identification of outlier data, data scaling, and feature selection. In the subsequent phases of the study, three distinct LSTM models are developed: Vanilla, Bi-Directional, and Stacked. The performance of each prediction model is evaluated using metrics such as Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

The datasets employed in the study consist of both air pollutant parameters including PM₁₀ ($\mu\text{g}/\text{m}^3$), SO₂ ($\mu\text{g}/\text{m}^3$), NO ($\mu\text{g}/\text{m}^3$), NO₂ ($\mu\text{g}/\text{m}^3$), NO_x ($\mu\text{g}/\text{m}^3$), O₃ ($\mu\text{g}/\text{m}^3$), and CO ($\mu\text{g}/\text{m}^3$) and meteorological parameters including atmospheric temperature ($^{\circ}\text{C}$), wind direction, wind speed (m/s), relative humidity (%) and atmospheric pressure (mbar). Specifically, the air pollution dataset originates from the Başakşehir Air Quality Monitoring Station (BAQMS). Additionally, the robustness and generalization capability of the model have been evaluated using datasets obtained from the Beşiktaş Air Quality Monitoring Station (BeAQMS) and the Kağıthane Air Quality Monitoring Station (KAQMS).

The decision parameters for all the developed LSTM models were set as follows: activation function: tanh, dropout rate: 0.2, epoch:50, batch size: 72, loss function: MAE and optimizer: Adam were chosen as decision parameters. Regarding the LSTM layers, the Vanilla LSTM model utilized a single LSTM layer with 50 units, the Bidirectional LSTM model employed a single BI-LSTM layer with 50 units, and the Stacked LSTM model consisted of two hidden layers, each with 50 units. The outcomes of the comprehensive analyses of all models revealed that the MAE values without DP-FS were approximately 0.1, whereas the MAE values after the DP-FS stage were around 0.06.

Furthermore, an assessment was conducted on the residual errors between the predicted and observed PM₁₀ values, revealing that the highest error values corresponded to the highest observed PM₁₀ values.

Also, multiple experiments with thirty random numbers were carried out to conduct a more comprehensive analysis of the difference between the proposed LSTM models and the MAE values of the model. According to the results, the lowest median value was observed in the Bi-Directional LSTM model with 0.05031, stacked LSTM with 0.05037, and Vanilla LSTM with 0.05053.

Experimental results demonstrate that the proposed LSTM models exhibit high prediction performance when the DP-FS process is applied, indicating their usability in predicting hourly PM₁₀ concentrations. The proposed DPFS-based LSTM models also show excellent prediction performance for datasets from different regions.

The secondary research focus of this study entails a comparative examination of predictive performances across three distinct deep learning algorithms: Long-Short Term Memory (LSTM), Recurrent Neural Network (RNN), and Gated Recurrent Units (GRU), specifically in the context of PM_{2.5} prediction. The performance of all forecasting models is evaluated using performance metrics, including Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R²).

The study's datasets comprise air pollutant variables such as PM_{2.5} (µg/m³), SO₂ (µg/m³), NO (µg/m³), NO₂ (µg/m³), NO_x (µg/m³), O₃ (µg/m³) and meteorological variables including atmospheric temperature (°C), wind speed (m/s), relative humidity (%), and air pressure (mbar). The air quality dataset originates from the Kağıthane Air Quality Monitoring Station (KAQMS).

The research topic involves two stages. In the initial stage, the models are executed to determine the optimal dimensions of hidden layers (1-2-4) and the quantity of neurons (5-10-20-30-40). Based on the outcomes, the LSTM model (model#9) exhibiting two hidden layers and 30 neurons in each layer emerges as the most effective, displaying superior performance metrics for both training and testing phases, including MSE = 24.95 and 25.20, RMSE = 4.771 and 4.818, MAE = 3.099 and 3.129, MAPE = 0.13 and 0.132, and R² = 0.976 and 0.966, respectively (p < 0.05).

The second stage consists of evaluating the prediction performances of different deep learning algorithms. In this part of the study, a series of ten repetitions were conducted for hybrid algorithms, namely GRU + GRU, LSTM+LSTM, RNN + RNN, GRU + LSTM, GRU + RNN, LSTM+RNN, LSTM+GRU, RNN + GRU, and RNN + LSTM, in order to achieve comprehensive evaluation outcomes. The LSTM+LSTM model (model#17) demonstrated superior performance outcomes in terms of training and testing, boasting MSE values of 24.60 and 24.84, RMSE values of 4.739 and 4.785, MAE values of 3.056 and 3.086, MAPE values of 0.129 and 0.13, and R² values of 0.98 and 0.97, respectively (p < 0.05).

Regarding the LSTM+LSTM model, decision parameters such as activation function (relu), dropout rate (0), epoch (50), batch size (32), loss function (MAE), optimizer (Adam), hidden layer number (2), each layer neuron number (30), learning rate (0.01), and train size (0.8) were carefully selected.

In summary, the experimental findings highlight a significant superiority of the LSTM+LSTM model over alternative deep learning algorithms with R^2 values of 0.98 and 0.97 for the training and test sets, respectively, at a significance level ($p < 0.05$).

Furthermore, to assess the model's generalization capability, comprehensive studies were conducted employing one year's worth of data collected from monitoring stations situated across nine distinct districts of Istanbul (Aksaray, Avcılar, Beşiktaş, Kadıköy, Kartal, Maslak, Sultangazi, Tuzla, and Ümraniye).

Performance evaluation studies were carried out to evaluate the model's generalization ability. The outcomes of the generalization assessment revealed the following R^2 values for the respective regions: Aksaray: 0.932, Avcılar: 0.961, Beşiktaş: 0.913, Kadıköy: 0.963, Kartal: 0.911, Maslak: 0.923, Sultangazi: 0.914, Tuzla: 0.904, and Ümraniye: 0.912. These outcomes were also assessed based on the SD/RMSE performance metrics for each region: Aksaray: 2.932, Avcılar: 2.262, Beşiktaş: 2.355, Kadıköy: 2.552, Kartal: 3.843, Maslak: 2.388, Sultangazi: 2.307, Tuzla: 2.617, and Ümraniye: 2.536. The findings underscore the district of Kartal achieving a "Very good" performance rating, while the other districts received a "Good" performance rating.

Overall, the results indicate that the proposed LSTM+LSTM model demonstrates a noteworthy ability for generalization, maintaining an R^2 accuracy rate of 0.90 ($p < 0.05$) or higher across all stations, making it suitable for non-linear, non-stationary multidimensional time series data. Additionally, comparing the results with other studies, the proposed LSTM+LSTM model performs better predicting $PM_{2.5}$ concentrations.

The results obtained from this study will yield several benefits. Firstly, establishing more accurate and reliable prediction models related to air pollution will represent a significant step in urban areas for a better understanding of air quality management and its impacts on human health. These prediction models can provide a solid foundation for monitoring and assessing air quality to assist authorities in protecting and enhancing public health.

Secondly, predicting particulate matter pollution levels such as $PM_{2.5}$ and PM_{10} can empower air quality managers to anticipate pollution events and take appropriate measures in advance. This can lead to more effective planning and implementation of emergency responses, contributing to improved air quality for the health and safety of urban residents.

Furthermore, this study highlights the development of a model that can predict PM pollution levels in cases where air quality monitoring stations are lacking or inadequate, especially in the sampled regions. This could aid in better planning measures to mitigate the effects of air pollution and optimize resource utilization.

Lastly, by emphasizing the potential of deep learning algorithms in air pollution prediction, this study offers a new perspective on how these algorithms can address various environmental issues. This may encourage and stimulate future similar studies in environmental engineering and air quality management.



1. GİRİŞ

Hava kirliliği, modern toplumların karşı karşıya kaldığı önemli bir çevresel sorundur ve insan sağlığı üzerinde ciddi etkilere sahiptir. Kötü hava kalitesi, çeşitli sağlık sorunlarına neden olabilen zararlı kirleticilerin geniş bir yelpazesi içerir. Özellikle havada süzülen ince parçacıklar olarak adlandırılan partikül madde (PM), solunum yolu hastalıkları, kardiyovasküler sorunlar ve diğer sağlık endişeleriyle ilişkilendirilen önemli bir kirlilik kaynağını oluşturur.

Son zamanlarda, bilimsel bir araştırma PM_{2.5} kaynaklı kirliliğin ölü doğumlara katkısını açıkça ortaya koymuş ve Dünya Sağlık Örgütü'nün (DSÖ) hava kalitesi hedeflerine uyulmasıyla önemli ölçüde ölü doğumların önlenebileceği vurgulanmıştır (Xue ve ark, 2022). DSÖ'nün verilerine göre, dünya nüfusunun neredeyse tamamı (%99), DSÖ sınırlarını aşan ve yüksek seviyelerde kirletici madde içeren hava solumakta olup, düşük ve orta gelirli ülkeler en fazla etkilenenler arasında yer almaktadır (World Health Organization [WHO], 2022). DSÖ tarafından belirlenen standartların karşılanması durumunda Türkiye'de yaklaşık 31-45 bin ölümün önlenebileceği çalışmalarda belirtilmektedir (Çevre İklim ve Sağlık İçin İşbirliği Projesi [ÇİSİP], 2022). Bunun yanı sıra, Güzel ve Özer (2022) tarafından yapılan bir çalışma, Türkiye'de hava kirliliği nedeniyle minimum 25,845 milyon TL, maksimum 52,492 milyon TL sağlık maliyetinin oluştuğunu göstermektedir (Güzel ve Özer, 2022).

Tüm bu sebeplerden dolayı, hava kirliliği seviyelerinin öngörülmesi ve bu konuda etkili tedbirler alınması, toplum sağlığı açısından son derece kritik bir öneme sahiptir. Hava kirliliği tahmini, çevre bilimleri ve halk sağlığı için hayati bir meseledir. Önceden hava kirliliği düzeylerini tahmin etmek, halk sağlığını korumak ve hava kalitesi yönetim stratejilerini geliştirmek için büyük bir stratejik öneme sahiptir. Bu nedenle, hava kirliliği tahmin modellerinin geliştirilmesi ve iyileştirilmesi, araştırmacılar ve karar alıcılar için öncelikli bir görev haline gelmiştir.

Ayrıca, hava kirliliği parametrelerinde olası değişiklikleri tespit etmek ve çevresel felaketlerin yol açabileceği etkileri minimize etmek için yüksek doğruluklu tahmin değerleri sağlayan modellere ve teknolojilere olan gereksinim önem arz etmektedir.

1.1. Tezin Amaç ve Kapsamı

Günümüzde, partikül madde kaynaklı hava kirliliği önemli bir çevre ve insan sağlığı için ciddi bir sorundur. Bu tez çalışması, özellikle PM_{10} ve $PM_{2.5}$ gibi sağlık üzerinde büyük etkisi olan partikül madde türlerinin derin öğrenme algoritmaları kullanılarak tahmin edilmesine yönelik detaylı analizleri içermektedir. Tez çalışması; derin öğrenme algoritmaları olan LSTM (Uzun Kısa Süreli Bellek), GRU (Geçişli Tekrarlayan Birim) ve RNN (Yinelemeli Sinir Ağları) algoritmalarının performanslarını karşılaştırarak en uygun modeli belirlemeye yönelik çalışmalar sunmaktadır.

Çalışma, partikül madde konsantrasyonlarının tahmininde yapay zeka tabanlı modellerin uygulamalarını inceleyerek geniş bir literatür analizi gerçekleştirmiştir. İstanbul ilinde bulunan farklı hava kalitesi istasyonlarından elde edilen büyük veri setleri üzerinde çalışılmıştır. Bu veri setleri, Veri Ön İşleme ve Özellik Seçimi (VÖİ-ÖS) adımlarıyla model öncesinde daha anlamlı hale getirilmiş ve partikül madde tahmininde LSTM algoritmalarının doğruluğuna etkisi araştırılmıştır. VÖİ-ÖS aşamasında, eksik veri analizi, aykırı veri analizi, ölçeklendirme ve özellik seçimi gibi adımlar uygulanmıştır. Ayrıca, başarılı modelin farklı bölgelerdeki veri setleri ile genelleme yeteneği incelenerek uygulanabilirliği değerlendirilmiştir. Tezin amacı, partikül madde (PM) konsantrasyonları tahmininde VÖİ-ÖS yöntemlerinin etkisini değerlendirmek ve farklı derin öğrenme algoritmalarının performansını karşılaştırmaktır.

1.2. Tezin Yapısı

Bu çalışma altı bölümden oluşmaktadır. Giriş bölümünde, konu, amaç ve kapsam tanımlanmıştır. İkinci bölüm, hava kirliliği ve ilgili yasal düzenlemeler hakkında bilgi sunmaktadır. Üçüncü bölümde, literatür özeti ile çalışmanın bilimsel hedefleri ve katkısı aktarılmaktadır. Dördüncü bölümde, çalışma alanı, veri seti yapısı, veri ön işleme adımları ve derin öğrenme uygulamaları ayrıntılı olarak açıklanmaktadır.

Beşinci bölüm, iki ayrı araştırma konusunun sonuçlarını içermektedir. Bu bölümde, PM_{10} kirleticisi için LSTM modelinin uygunluğu ve $PM_{2.5}$ tahmini için farklı derin öğrenme modellerinin performans karşılaştırması sunulmaktadır. Son bölüm olan altıncı bölümde, çalışmanın sınırlamaları, gelecekte yapılacak araştırma önerileri ve sonuçlar detaylı bir şekilde açıklanmaktadır.





2. HAVA KİRLİLİĞİ

Temiz hava, sağlıklı yaşamın sürdürülebilmesi için temel bir gereksinimdir. Atmosfer, yaklaşık olarak %78 azot, % 21 oksijen ve % 1 diğer küçük gazlardan oluşur. Bu gazların yanı sıra, havada “aerosoller” olarak adlandırılan küçük parçacıklar da mevcuttur (İklimBu, 2019).

Hava kirliliği kavramına dair farklı tanımlar literatürde yer almaktadır. Dünya Sağlık Örgütü'ne (DSÖ) göre, hava kirliliği, “atmosferin doğal özelliklerini kimyasal, fiziksel veya biyolojik etkenlerle değiştirerek iç veya dış mekanları kirleten durum” olarak tanımlanmaktadır (World Health Organization [WHO], 2022).

Amerika Çevre Koruma Ajansı (U.S. EPA) hava kirliliğini, “kirletici maddelerin havada insan sağlığına zarar verebilecek veya çevresel etkilere yol açabilecek düzeyde bulunması” şeklinde açıklamaktadır (Zencirci ve Işıklı, 2017).

Hava kirliliği, “hava bileşimini değiştiren katı, sıvı ve gaz halde bulunabilen kirleticilerin, insan sağlığına, canlı hayatına ve ekolojik dengeye zarar verecek ya da yaşamdan maddi nesnelerden yararlanılmasını engelleyecek miktar veya sürede atmosferde bulunmasıdır” olarak da ifade edilmiştir (Polat, 2016).

Müezzinoğlu (2000) ise hava kirliliğini “bir veya daha fazla kirletici türünün açık havada (bina dışında) insan, bitki ve hayvan yaşamına, ticari veya kişisel eşyalara, ekolojik dengeye ve çevre kalitesine zarar veren miktarda belli bir derişim ve sürenin üzerinde bulunması” olarak tanımlamaktadır (akt. Partigöç ve Çubukçu, 2017). Havada bulunan kirleticiler iç mekan hava kirleticileri ve dış mekan hava kirleticileri olarak iki gruba ayrılır (Çilingir, 2016). Tablo 2.1’de iç ve dış ortam hava kirleticileri özetlenmiştir. En yaygın hava kirleticileri (kriter kirleticiler); partikül madde (PM), yer seviyesinde ozon (O₃), kükürt dioksit (SO₂), karbon monoksit (CO), azot dioksit (NO₂) ve kurşun (Pb) dur (U.S. Environmental Protection Agency [U.S. EPA], 2015a).

Havada farklı birçok kirletici bulunmakta fakat kentsel hava kirliliği çoğunlukla PM ve SO₂ konsantrasyonlarının belirlenmesiyle tespit edilmektedir (Orman korumu, 2008).

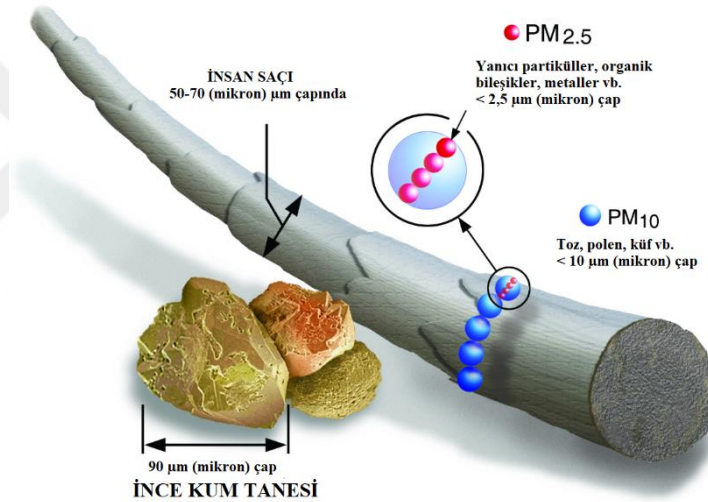
Tablo 2.1.Hava kirleticileri (US EPA, 2014a; US EPA, 2014b; US EPA, 2017; NSW Health, 2022; Çilingir 2016 dan oluşturulmuştur)

Dış ortam hava kirleticileri	
Partikül madde	<i>PM₁₀</i> : Atmosferdeki en büyük partikül maddeleri kategorisidir ve çapları 10 mikrometre veya daha küçük olan parçacıklar <i>PM_{2,5}</i> : Çapı 2,5 mikrometre veya daha küçük olan partikül maddeler <i>PM₁</i> : Çapı 1 mikrometre veya daha küçük olan partiküller <i>Ultra İnce Partiküller (UFP)</i> : Çapı 0,1 mikrometre veya daha küçük olan partiküller İnorganik ve organik maddelerden oluşan partiküller
Azot dioksit (NO ₂)	Fosil yakıtların yanması ve endüstriyel faaliyetler sonucu ortaya çıkar. Özellikle trafik kaynaklıdır. Solunum yoluyla akciğerlere nüfuz ederek solunum sistemi sorunlarına neden olabilir.
Ozon (O ₃)	Güneş ışığı, azot dioksitler ve uçucu organik bileşiklerin etkileşimi sonucu oluşan bir gazdır. Ozon, solunum yoluyla alındığında solunum sorunlarına neden olabilir.
Kükürt Dioksit (SO ₂)	Fosil yakıtların yanması, endüstriyel prosesler ve yanardağ faaliyetleri gibi kaynaklardan salınır. Solunum yoluyla alındığında solunum yolu rahatsızlıklarına ve asit yağmurlarına yol açabilir.
Karbonmonoksit (CO)	Tam yanmamış fosil yakıtlar sonucunda oluşan bir gazdır. Zehirleyici bir gaz olup, özellikle kapalı alanlarda doğal gaz, odun ve kömür gibi yakıtların yanması sonucu ortaya çıkar.
Kurşun (Pb)	Endüstriyel faaliyetler, fosil yakıt kullanımı, ev ve bahçelerde boya kullanımı, madencilik faaliyetleri sonucu oluşabilmektedir. Çocuklarda beyin ve sinir sistemi sorunlarına neden olabilir. Hamile kadınlarda da bebek gelişimini etkiler.
Bazı toksik maddeler	Asetaldehid, arsenik bileşikleri, benzen, kadmiyum bileşikleri, karbon tetraklorür, kloroform, kömür ocağı gazları, etilen oksit, formaldehid, heksaklorobenzen, mangan bileşikleri, civa bileşikleri, nikel bileşikleri ve vinil klorür.
İç ortam hava kirleticileri	
Uçucu Organik Bileşenler (UOB)	Birçok ev eşyası, mobilya, temizlik malzemeleri, boya, tutkal ve diğer kimyasal ürünlerde bulunan maddelerdir. Kokulu kimyasalların ve çeşitli gazların buharlaşmasıyla ortaya çıkar. Bunlar formaldehit ve benzen gibi maddeleri içerebilir ve uzun süreli maruz kalma ciddi sağlık sorunlarına yol açabilir.
Sigara Dumanı	Sigara içme veya pasif olarak sigara dumanına maruz kalma, iç ortamdaki en yaygın ve zararlı hava kirleticilerinden biridir. Sigara dumanı, binlerce kimyasal madde içerir ve kanserojen maddeler, nikotin ve zararlı partiküller gibi sağlık açısından riskli bileşenler içerir.
Evcil Hayvan Tüyleri ve Döküntüleri	Evcil hayvanların tüyleri ve döküntüleri iç ortamda alerjik reaksiyonlara neden olabilir.
Ev Tozu Akarı ve Polenler	Ev tozu akarları, evlerdeki halılar, yatak, yastık ve mobilyalar gibi yerlerde yaşayan mikroskopik organizmalardır. Bu akarlar ve polenler, alerjik reaksiyonlara neden olabilen solunum yolu irritasyonuna yol açabilir.
Küf ve Mantarlar	Nemli veya küf bulunan ortamlarda büyüeyen küf ve mantarlar, iç ortamlarda hava kalitesini olumsuz etkileyebilir. Küf ve mantarların sporları solunum yoluyla alınabilir ve alerjik reaksiyonlara veya solunum yolu rahatsızlıklarına yol açabilir.
Kimyasal Temizlik Ürünleri	Evde kullanılan çeşitli temizlik malzemeleri, kimyasal ürünler, iç ortamda havaya zararlı maddeler salabilir. Bu kimyasal maddeler, havadaki parçacıklar, uçucu organik bileşikler ve diğer zararlı bileşenlerdir.

2.1. PM Kaynakları

U.S. EPA'ya göre; "PM, partikül madde anlamına gelmekte (partikül kirliliği olarak da adlandırılır) ve havada bulunan katı partiküller ve sıvı damlacıkların karışımın" ifade etmektedir (U.S. Environmental Protection Agency [U.S. EPA], 2016).

Partiküler madde, dış hava kirliliğinin önemli bir bileşeni olarak kabul edilir ve Dünya Sağlık Örgütü'ne (DSÖ) bağlı Uluslararası Kanser Araştırma Ajansı (IARC) tarafından insanlar için kanserojen olarak sınıflandırılmıştır (International Agency for Research on Cancer [IARC-WHO], 2013). Partiküler madde, çaplarına göre farklı sınıflandırmalara tabi tutulur. Şekil 2.1, PM'lerin sınıflandırılmasını göstermektedir (U.S. Environmental Protection Agency [U.S. EPA], 2016). Bu partiküller, aerodinamik çaplarına bağlı olarak tanımlanır; genellikle 10 mikrometre (μm) ve daha küçük çaplı olanlar solunabilir parçacıklar PM_{10} olarak adlandırılırken, genellikle 2,5 mikrometre (μm) ve daha küçük çaplı olan ise ince solunabilir parçacıklar $\text{PM}_{2.5}$ olarak adlandırılmaktadır (U.S. Environmental Protection Agency [U.S. EPA], 2016). Aynı zamanda, partiküler maddelerin boyutları yani büyüklükleri havada kalma sürelerini belirlemektedir (Mayruk, 2005).

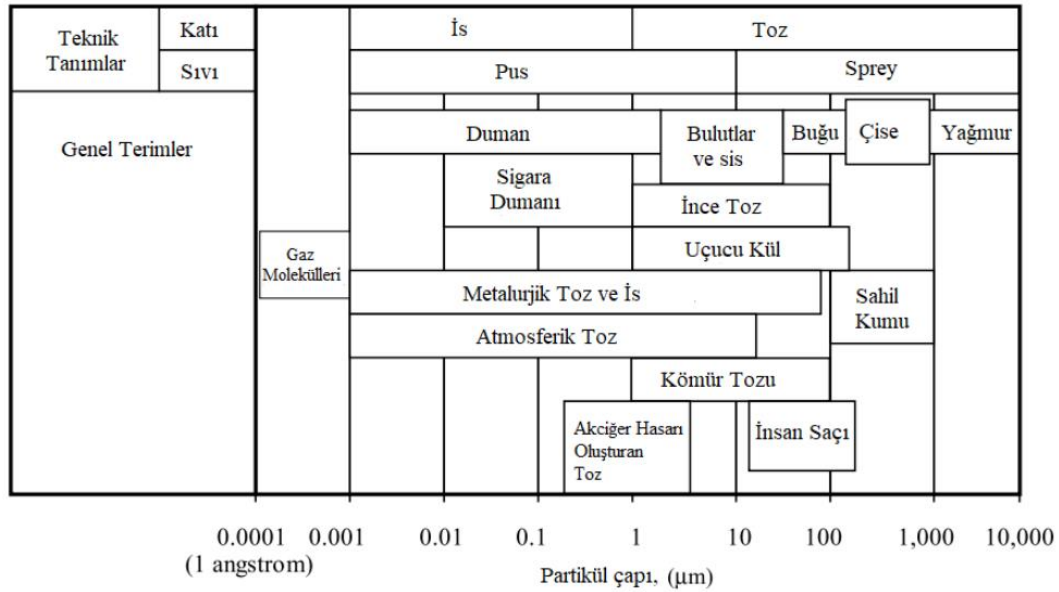


Şekil 2.1. Partikül maddelerin sınıflandırılması (U.S. Environmental Protection Agency [U.S. EPA], 2016).

Hava kirleticileri, kaynak çıkışına göre iki ana gruba ayrılır: birincil (primer) ve ikincil (sekonder) kirleticiler. Birincil kirleticiler, doğrudan kaynaklardan yayılan hava kirleticisi özelliklere sahip bileşenlerdir. Bunlar, havada doğrudan var olan kirleticileri ifade ederler. Öte yandan, atmosferdeki çeşitli reaksiyonlar sonucu birincil kirleticilerin etkileşimiyle oluşan bileşenler sekonder kirleticiler olarak adlandırılır ("Hava kirliliği," t.y.) Partikül maddeler birincil kirleticilere örnek olarak verilebilir ve bunların kaynaklarına bağlı olarak havaya direkt olarak yayıldıkları görülür (örneğin SiO_2 , toz, mist, tütsü, is, duman). Ayrıca, bu partikül maddeler atmosferdeki

diğer gazlarla tepkimeye girerek ikincil kirleticilere (örneğin Londra sisi veya endüstriyel duman) dönüşebilirler (“Hava kirliliği,” t.y.). Partikül maddelerin, aerodinamik çaplarına göre havada kalma süreleri ve taşınma mesafeleri değişiklik gösterir. Örneğin, PM_{2.5}'in aerodinamik çapı daha küçük olduğu için havada uzun süre (saatler, haftalar) kalabilir ve uzak mesafelere taşınabilirken, PM₁₀ daha büyük çaplı olduğu için havada kısa süreli (dakikalar, saatler) kalır ve daha kısa mesafelere taşınabilir (Özbeyaz ve ark., 2016).

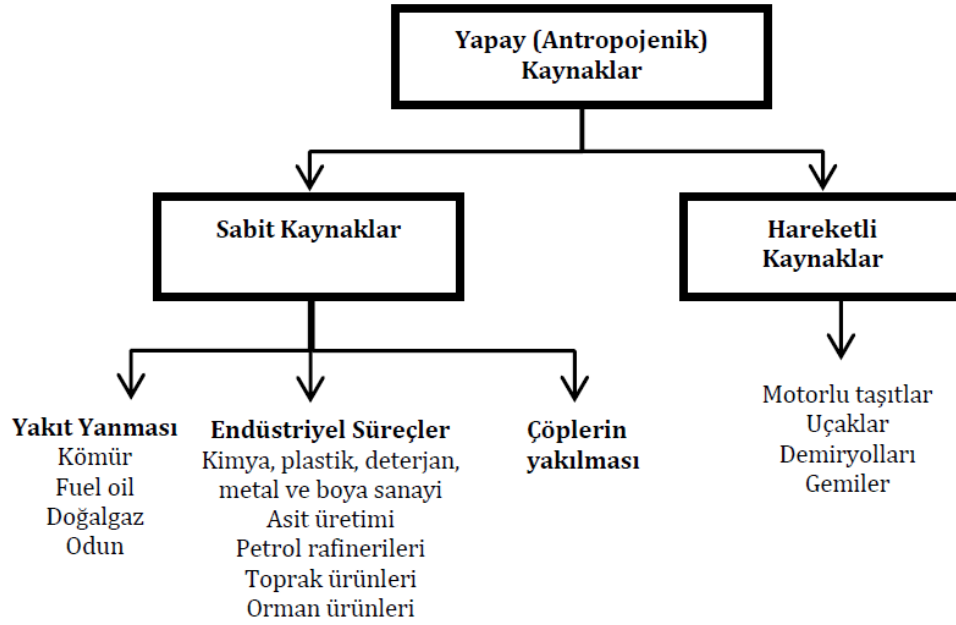
Hava içinde bulunan partiküller, filtrasyon mekanizmaları aracılığıyla incelenebilir. Bu mekanizma yanında, partikül boyutlarına göre sınıflandırmanın yanı sıra partiküllerin çeşitli şekillere sahip olması nedeniyle saçtığı ışığın çapları da dikkate alınmalıdır. Kısacası, partikül maddeler, özgün yapı ve şekilleri nedeniyle sınıflandırmayı karmaşık hale getirebilirler (Ayturan 2019). PM kirleticileri, boyutları çok küçük olduğundan milimetre cinsinden ifade edilen µm (mikrometre) birimi ile ölçülür. Havada bulunan partikül maddelerin terimsel dağılımı Şekil 2.2’de gösterilmiştir.



Şekil 2.2. Partikül madde dağılımı ve terimleri (Ayturan, 2019).

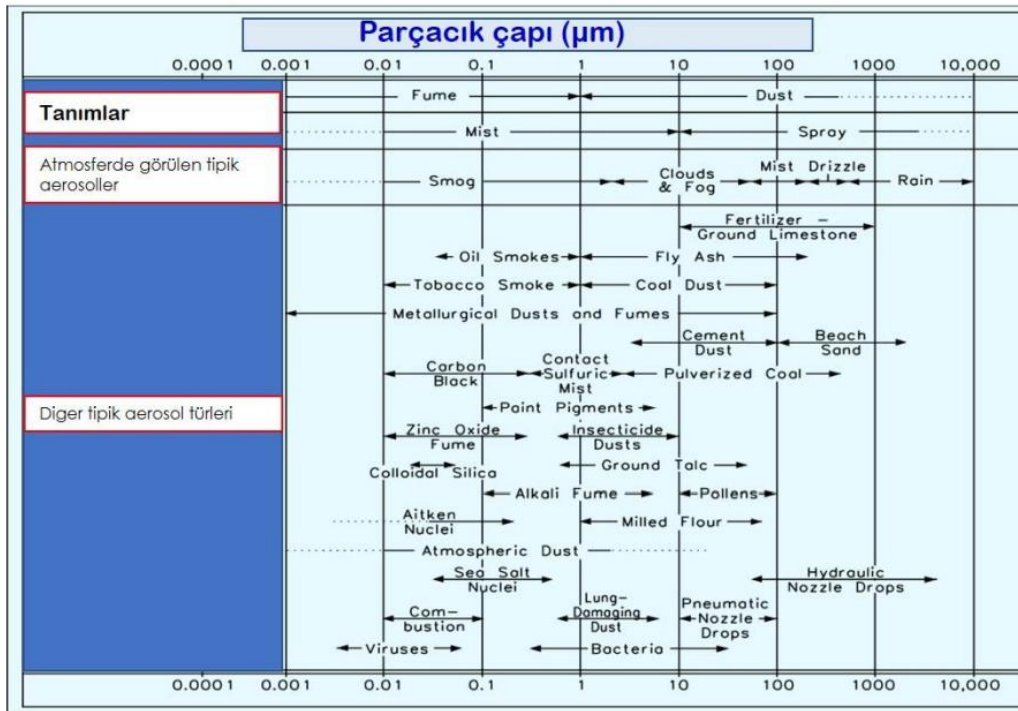
Hava kirliliği kaynakları, doğal ve yapay olarak iki ana grupta sınıflandırılabilir. İnsan etkinlikleri sonucu meydana gelen hava kirliliği kaynakları yapay olarak adlandırılır ve antropojenik kaynaklar olarak ifade edilir (Partigöç ve Çubukçu, 2017). Şekil 2.3’te antropojenik kaynaklar sınıflandırılmıştır.

Dünyada doğal 5000 milyon ton/yıl ve antropojenik 500 milyon ton/yıl PM oluşmaktadır (Can, 2016). Dünya nüfusunun artmasıyla birlikte tüketim talepleri de artmakta ve bu da üretim tesislerinin sayısının ve kapasitelerinin artmasına yol açmaktadır. Bu bağlamda, antropojenik kaynaklı salınımın insan sağlığı ve çevre üzerindeki etkileri gelecekte daha da önem kazanacaktır.



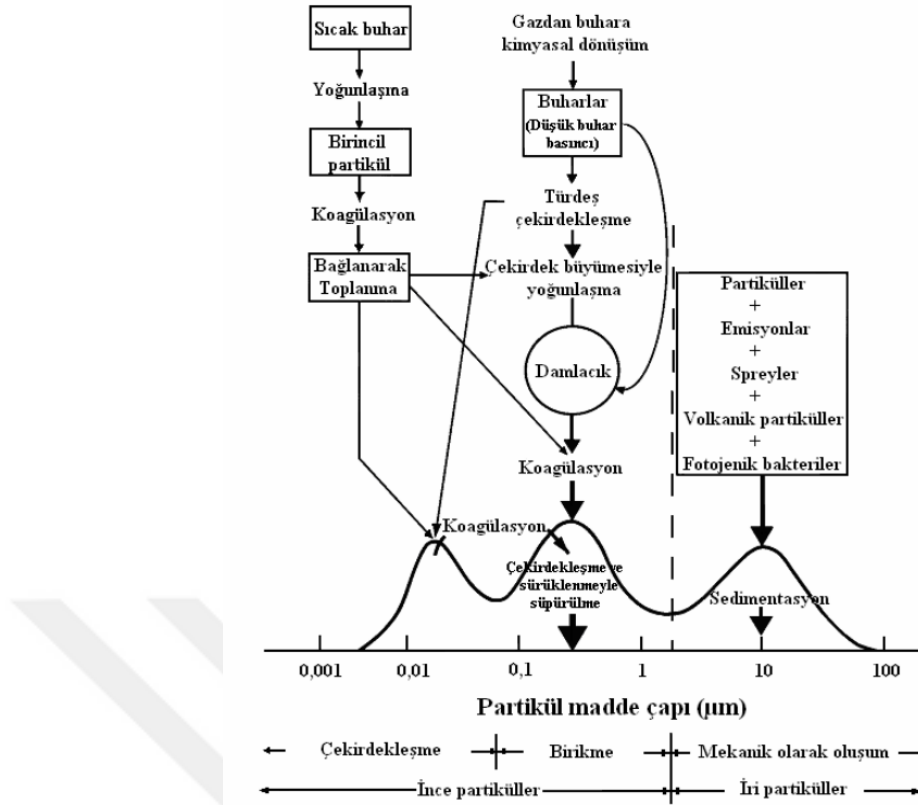
Şekil 2.3. Antropojenik kaynakların sınıflandırılması (Partigöç ve Çubukçu 2017).

Kömürlü termik santraller, endüstriyel işlemlerden kaynaklanan emisyonlar ve taşıtlardan kaynaklanan emisyonlar, öne çıkan antropojenik partiküler madde kaynaklarına örnek olarak verilebilir (Arı ve ark., 2008). Şekil 2.4’te atmosferde yaygın olarak bulunan parçacıkların boyut aralıkları ve kaynakları sunulmuştur (Tuncel, 2016).



Şekil 2.4. Parçacık boyutları ve kaynakları (Tuncel, 2016).

PM₁₀ ve PM_{2.5} genellikle farklı emisyon kaynaklarından salınır ve bu nedenle farklı kimyasal bileşenlere sahiptir. PM₁₀'ün büyük bir kısmı ve PM_{2.5} kirliliğinin önemli bir bölümü benzin, yağ, dizel yakıt veya ahşap malzemelerin yanması sonucu meydana gelen emisyonlardan kaynaklanır. PM₁₀'ün diğer kaynakları arasında şantiyeler, çöp ve atık yakma alanları, tarım sahaları, orman yangınları, çeşitli endüstriyel etkinlikler, açık alan tozları, polen ve bakteri parçaları yer almaktadır (California air resources board, t.y.). Şekil 2.5'te PM oluşum ve uzaklaştırma süreçleri gösterilmektedir. Toz oluşumunun %75'i doğal kaynaklı iken %25'i antropojenik faaliyetler sonucu meydana gelmektedir (Tuncel 2016). EPA verilerine göre, PM₁₀'ün büyük bir bölümü doğal kaynaklı toz salınımından kaynaklanmaktadır (Tuncel, 2016).



Şekil 2.5. Partikül madde oluşum aşamaları (Can, 2016).

Atmosferde bulunan doğal parçacık türleri; rüzgarın taşıdığı toz parçacıkları, deniz tuzu parçacıkları, volkanik patlamaların sonucunda oluşan parçacıklar ve doğal orman yangınlarından kaynaklanan parçacıklardır (Tuncel, 2016). Atmosferde en yüksek yoğunluğa sahip olan parçacıklardan biri mineral tuzlarıdır; toprak erozyonu, kum fırtınaları, toprak parçacıklarının yoldan ve topraktan havalandığı durumlarla meydana gelir. Diğerleri ise deniz tuzu parçacıklarıdır ve okyanus ve denizlerden salınmaktadır (Tuncel, 2016).

Doğal kaynaklardan ortaya çıkan diğer bir parçacık türü olan volkanik küller ise troposfere ve stratosfere kadar yükselme potansiyeline sahiptir; aynı zamanda rüzgar gibi meteorolojik etkilerle atmosferdeki diğer parçacıklara karışabilir (Tuncel, 2016). Doğal orman yangınları sonucunda atmosfere gazlar ve PM kompleksleri yayılabilir, Hinds (1999) yılda 20 Tg parçacığın atmosfere yayıldığını belirtmektedir (Tuncel, 2016).

Antropojenik kaynaklardan yani insan faaliyetleri sonucu oluşan PM oluşumunda, özellikle yanma işlemleri, oldukça önemlidir (Tuncel 2016). Antropojenik PM

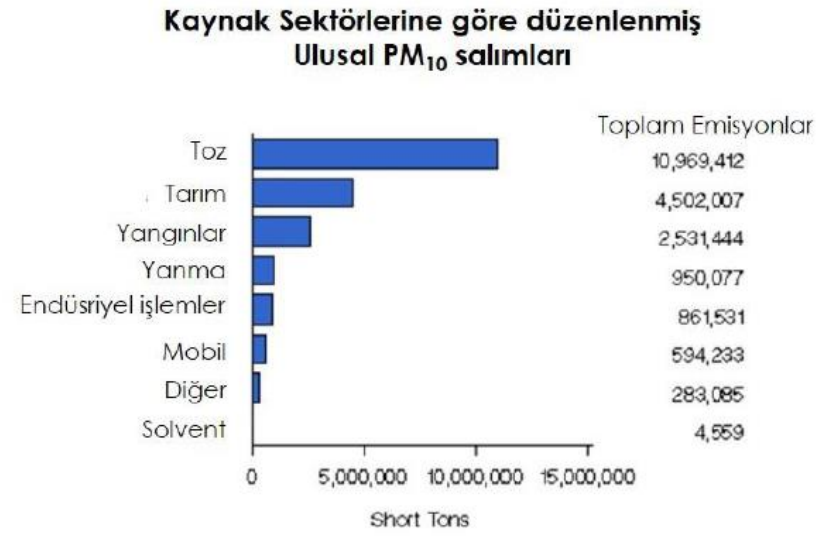
kaynaklar; trafik, endüstriyel faaliyetler, kömürlü termik santraller, atık bertarafı, rafineriler, ısınma amacıyla odun ve kömür yakılması gibi etkinlikler bu kategoriye örnek olarak verilebilir (Can, 2016).

Yüksek miktarda karbon ve kükürt içeren kömürün tam olarak yanmadığı durumlarda yüksek seviyelerde partikül madde oluşabilir (Tuncel, 2016). İnce partikül madde kaynağı olarak öne çıkan güç santralleri, özellikle kömürlü güç santralleridir ve bu santrallerden SiO_2 , Al_2O_3 , Fe_2O_3 emisyonları yayılır. Çevresel İşbirliği Komisyonu (CEC, 2005) raporuna göre, ABD’de yüksek PM katkısına sahip güç santrallerinin %96,4’ü kömür, %2,8’i fuel-oil ve %2’si doğalgaz kullanmaktadır (Tuncel, 2016).

Endüstri kaynaklı PM salınımına örnek olarak demir-çelik sanayisi, çimento fabrikaları, petrol rafinerileri, inşaat faaliyetleri vb. verilebilir. Endüstri kaynaklı PM’ler, yüksek eser toksik maddeler (As, Cd, Cr, Cu, Pb, Zn, Hg, Ni, Se ve V vb) içerdiğinden dolayı sağlığa olumsuz etkileri yüksektir (Tuncel, 2016). Şekil 2.6 EPA’nın ABD’deki PM kaynakları ve salınım envanterine ilişkin bilgiler sunmaktadır (Tuncel 2016).

Önceden ince partikül madde olan tanımladığımız $\text{PM}_{2.5}$, fabrikalar, enerji santralleri, motorlu taşıtlar ve uçakların yakıt artıkları, odun ve kömür kullanımı, orman ve tarımsal yanmalar, volkanik patlamalar ve toz fırtınaları gibi çeşitli kaynaklardan kaynaklanabilir (Uslu, 2020).

Ayrıca, dizel yakıt dumanında bulunan $\text{PM}_{2.5}$ ’un, okul servislerinde, trafiğin rüzgarın ters istikametinde kalan bölgelerinde ve kamyon bekleme alanlarında yoğunlaştığı belirtilmektedir (Çapraz, 2013).



Şekil 2.6. ABD’de PM kaynaklarına göre salınım miktarları (Tuncel, 2016).

2.2. Hava Kirliliğinin İnsan Sağlığına Etkisi

Güzel ve Özer (2022)’nin yaptığı bir çalışmada bir insanın yaşamı boyunca yaklaşık 250 milyon litre hava tükettiğini belirtmektedir (Güzel ve Özer, 2022). Bu noktada kaliteli havanın önemi ortaya çıkmaktadır.

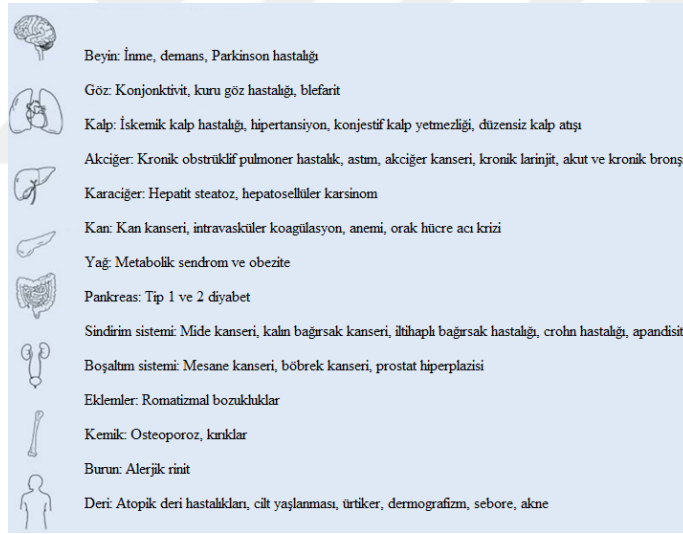
Aynı araştırmada, Türkiye’de PM₁₀ ve PM_{2.5} kirliliğinin durumu incelenmiştir. 2009-2019 yılları arasında PM₁₀ kirliliğinin ortalama 294,56 bin ton (en yüksek 2012 yılında 344,52 bin ton) ve konsantrasyonunun azalmakta olduğu, ancak PM_{2.5} kirliliğinin 195,08 bin ton (en yüksek 2019 yılında 202,21 bin ton) ve konsantrasyonunun artmaya doğru bir eğilim gösterdiği belirtilmektedir (Güzel ve Özer 2022).

Avrupa Çevre Ajansı (AÇA), Avrupa’da hava kirliliğinin insan sağlığı ve çevre üzerinde olumsuz etkiler yarattığını vurgularken, AB ve DSÖ’nün belirlediği yasal

sınırların hala aşıldığını ve bu durumun sonucunda Avrupa'da yılda yaklaşık 400 bin erken ölümün yaşandığını ifade etmektedir (Avrupa Çevre Ajansı [AÇA], 2021a). AÇA ayrıca, partikül madde (PM), azot dioksit (NO₂) ve yer seviyesinde ozonun (O₃), Avrupa'da insan sağlığı ve çevre açısından en büyük zararı veren kirleticiler olduğunu belirtmektedir (Avrupa Çevre Ajansı [AÇA], 2021a).

AÇA (2023), hava kirliliğinin uzun ve kısa süreli maruziyetlerinin inme, kronik obstrüktif akciğer hastalığı, solunum yolu ve akciğer kanserleri, ağırlaştırılmış astım, alt solunum yolu enfeksiyonları gibi çeşitli hastalıklara yol açabileceğini belirtmektedir (Avrupa Çevre Ajansı [AÇA], 2023).

Schraufnagel ve ark., (2019) tarafından gerçekleştirilen küresel bir çalışma, uzun süreli hava kirliliğinin vücuttaki organlara etki edebileceğini ve mevcut sağlık sorunlarını daha karmaşık hale getirebileceğini göstermiştir (Schraufnagel ve ark., 2019) Şekil 2.7'de ise hava kirliliğinin etkilediği organlar ve neden olduğu hastalıklar gösterilmektedir (Schraufnagel ve ark., 2019).



Şekil 2.7. Hava kirliliğinden etkilenen organlar ve hastalıklar (Schraufnagel D.E. ve ark., 2019 makalesinden uyarlanmıştır).

PM, insan solunum ve kardiyovasküler sistemlerinde birikerek ciddi sağlık sorunlarına neden olabilir. PM konsantrasyonunun yüksek olduğu ortamlar, solunum sistemi ve kalp damar sistemi üzerinde olumsuz etkiler yaratır ve PM kirliliği ile hastalıklar arasında önemli bir ilişki bulunmaktadır (Özbeyaz, 2016). PM₁₀ (çapı <10 µm) parçacıklar ciğerlere ve bazıları kan dolaşımına kadar ilerleyebilirken, ince PM_{2.5} (çapı <2,5 µm) parçacıklar sağlık üzerinde en büyük riski oluşturur (U.S. EPA, 2016).

PM'nin insan sağığına olan olumsuz etkisi özellikle solunum sistemi üzerindedir (Hetland ve ark, 2004; akt. Demir ve ark., 2010). AÇA (Avrupa Çevre Ajansı) 2019 yılında, 27 AB üye devletinde PM_{2.5} kirliliğıyle ilişkilendirilen yaklaşık 307.000 erken ölümün olduğı belirtilmektedir (Avrupa Çevre Ajansı [AÇA], 2021b).

Yapılan birçok çalışma, PM'nin insan sağığı üzerindeki olumsuz etkilerini incelemiştir. Yılmaz ve ark. (2022) Muş İli'nde gerçekleştirdikleri hava kalitesi değerlendirmesi çalışmasında PM₁₀ değerlerinin çoğunlukla Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliğı'nde belirlenen yıllık sınır değerleri üzerinde olduğunu tespit etmiştir (Yılmaz ve ark., 2022). Ayrıca, yapılan diğer çalışmalar, PM_{2.5} maruziyeti ile olumsuz sağık etkileri arasındaki ilişkileri doğrulamaktadır (Pun ve ark. 2017; Guan ve ark., 2021; Miller ve Xu, 2018; Yu ve ark., 2023).

Xue ve ark. (2022) yaptığı bir çalışma, hamile kadınların düşük ve orta gelirli ülkelerde PM_{2.5} maruziyetinin önlenmesi (sınır değerlere getirilmesi) ile anne sağığıının iyileştirilebileceğini ve PM_{2.5} maruziyetinin 137 ülkede ölü doğumların %39,7'sine katkıda bulunduğunu göstermektedir (Xue ve ark., 2022).

Tarih boyunca dünya genelinde ve ülkemizde, PM kaynaklı çevre felaketleri ve kazaları birçok trajik olaya neden olmuştur. Bu olaylar, hava kirliliğinin insan sağığına ve çevreye olan etkilerini gözler önüne sermektedir:

- 1930 yılında Belçika'da meydana gelen Meuse Vadisi sisi, 63 kişinin hayatını kaybetmesine neden olmuştur (Akdur, 2009).
- 1952 yılında İngiltere'nin Londra şehrinde yaşanan öldüren sis (the killer fog), 4000'den fazla kişinin yaşamını yitirmesine yol açmıştır (Wikipedia, 2023),
- 1948 yılında Amerika'nın Donora şehrinde gerçekleşen Donora sisi, 20 kişinin hayatını kaybetmesine sebep olmuştur (Gürsoy, 2022).
- 1971 yılında başlayan ve hala devam eden Türkmenistan doğalgaz çukuru yangını, uzun yıllardır çevre ve insan sağığıını etkilemektedir (Çevre online, 2023).
- 2006 yılında Endonezya'nın Java adasında yaşanan volkan patlaması sonucu oluşan sıcak kükürt çamuru, ciddi sağık sorunlarına neden olmuştur (Çevre online, 2023).

- 1962 yılında ABD'nin Pensilvanya eyaletinde bulunan dev kömür madeni yangını, çevre ve insan sağlığını olumsuz etkilemeye devam etmektedir (Çevre online, 2023).

Türkiye’de yakın geçmişte 28 temmuz 2021’de Antalya-Manavgat yangını (Akın, 2021), 3 temmuz 2020 Sakarya havai fişek fabrikası patlaması (Keskin, 2020), 1 kasım 2022 Bursa tekstil fabrikası yangını (Ataş, 2022), 10 haziran 2023 İstanbul Başakşehir İkitelli Organize Sanayi Bölgesinde çıkan fabrika yangını (Anadolu Ajansı, 2023) gibi olaylar, çevre ve insan sağlığına ciddi zararlar vermiştir.

2.3. Yasal Mevzuat

Hava kalitesinin iyileştirilmesi ve hava kirleticilerin atmosferdeki konsantrasyonlarının sınırlandırılması amacıyla ülkeler ve sağlık organizasyonları tarafından çeşitli sınır değerler ve yasal mevzuatlar belirlenmiştir. Bu önlemler, insan sağlığına ve çevreye yönelik olumsuz etkilerin azaltılmasını hedeflemektedir. Aşağıda, bu konuda oluşturulan bazı ulusal/uluslararası yasal düzenlemelere örnekler verilmiştir:

- Dünya Sağlık Örgütü (DSÖ), hava kalitesinin iyileştirilmesi amacıyla partikül madde, ozon, azot dioksit ve kükürt dioksit gibi kirleticiler için hava kalitesi klavuzları yayınlamıştır (World Health Organization [WHO], 2005).
- Türkiye, CLRTAP Sınırlar Ötesi Uzun Menzilli Asidifikasyon, Ötrifikasyon ve Yer Seviyesi Ozon Kirliliğinin Önlenmesine Yönelik Hava Kirliliği Sözleşmesi’ ne 18 Nisan 1983 tarihinde taraf olmuştur (AB Çevre Müktesebat, t.y.).
- EMEP [Avrupa İzleme ve Değerlendirme Programı] Finansman Protokolü’ne 1985 yılında taraf olmuştur (AB Çevre Müktesebat, t.y.).
- Ülkemiz’de, hava kalitesi yönetimi için gerekli düzenlemeler Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği (HKDYY) ile sağlanmıştır. Bu yönetmelik, 06/06/2008 tarihli 26898 sayılı Resmi Gazete’de yayınlanarak yürürlüğe girmiştir. HKDYY, Avrupa Birliği (AB) çevre mevzuatına (AB’nin 96/62/EC, 99/30/EC, 2000/69/EC, 2002/3/EC ve 2004/107/EC sayılı direktiflerine) paralel olarak hazırlanmıştır.

Tablo 2.2’de hava kalitesinin iyileştirilmesi için oluşturulan bazı ulusal/uluslararası yasal düzenlemeler bir araya getirilmiştir.

Tablo 2.2. Hava kirliliği yönetimi kapsamında bazı ulusal/uluslararası mevzuatlar.

Uluslararası mevzuat	
96/62/EC Hava kalitesi çerçeve direktifi 16/2001/EC Çevresel havada ozon direktifi 1999/30/EC+2001/744/EC (düz.) Havadaki sülfür dioksit, azot dioksit ve azot oksitler, partiküller ve kurşuna ilişkin direktif 2000/69/EC Havada bulunan benzen ve karbonmonoksit limit değerleri direktifi 2001/81/EC Atmosferdeki kirleticilerin miktarları için tavan değerler direktifi 1999/94/EC Yolcu otomobillerinden çıkan CO₂ emisyonları ve yeni araçlarda yakıt tasarrufu direktifi 94/63/EC Petrol depolama ve dağıtımını kaynaklı uçucu organik karbonlara ilişkin direktif 98/70/EC+2000/71/EC (düz.) Petrol ve dizel yakıt kalitesi direktifi 1999/32/EC Likit yakıtlardaki sülfür içeriğinin azaltılmasına ilişkin direktif 97/68/EC+2001/63/EC (düz.) Hareketli makinalara takılan motorlardan çıkan kirlilikle ilgili direktif	(Şenoğlu, 2006 akt. Orhun, 2013)
2004/107/EC Dış ortam havası içerisindeki arsenik, kadmiyum, cıva, nikel ve polisiklik aromatik hidrokarbonlara ilişkin direktif 2008/50/EC CAFE Avrupa için daha temiz hava ve dış ortam hava kalitesine ilişkin Avrupa Parlamentosu ve konsey direktifi 2000/69/EC Dış ortam havasındaki benzen ve karbon monoksit için sınır değerlere ilişkin direktif. 2002/3/ED Dış ortam havası içerisindeki ozona ilişkin direktif	(AB Çevre Müktesebat, t.y.)
Ulusal mevzuat	
07/02/2009 tarihli ve 27134 sayılı Isınmadan Kaynaklanan Hava Kirliliğinin Kontrolü Yönetmeliği 06/06/2008 tarihli ve 26898 sayılı Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği 03/07/2009 tarihli ve 27277 sayılı Sanayi Kaynaklı Hava Kirliliğinin Kontrolü Yönetmeliği 06/10/2009 tarihli ve 27368 sayılı Bazı Akaryakıt Türlerindeki Kükürt Oranının Azaltılmasına İlişkin Yönetmelik 19/07/2013 tarihli ve 28712 sayılı Koku Oluşturan Emisyonların Kontrolü Hakkında Yönetmelik 11/03/2017 tarihli ve 30004 sayılı Egzoz Gazı Emisyonu Kontrol Yönetmeliği	T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Çevre Yönetimi Genel Müdürlüğü (2023)

HKDYY yönetmeliğinin amacı “hava kirliliğinin çevre ve insan sağlığı üzerindeki zararlı etkilerini önlemek veya azaltmak için hava kalitesi hedeflerini tanımlamak ve oluşturmak, tanımlanmış metotları ve kriterleri esas alarak hava kalitesini değerlendirmek, hava kalitesinin iyi olduğu yerlerde mevcut durumu korumak ve diğer durumlarda iyileştirmek, hava kalitesi ile ilgili yeterli bilgi toplamak ve uyarı eşikleri aracılığı ile halkın bilgilendirilmesini sağlamaktır” (HKDYY, 2008).

Bu amaç doğrultusunda sülfürdioksit (SO₂), azot dioksit (NO₂), azot oksit (NO_x), PM₁₀, kurşun (Pb), benzen, karbonmonoksit (CO), arsenik (As), kadminyum (Cd), nikel (Ni), civa (Hg), benzo(a) piren, ozon (O₃), polisiklik aromatik hidrokarbonlar (PAH) gibi kirleticiler için belirli limit değerler ve uyarı eşikleri belirlenmiştir. Bu önlemler çevre ve insan sağlığı üzerindeki zararlı etkileri önlemek, azaltmak veya engellemek amacıyla alınmıştır. Ülkemizde PM_{2.5} için henüz herhangi bir sınır değeri belirlenmemişken, Dünya Sağlık Örgütü (DSÖ) ve Avrupa Birliği (AB) tarafından sınır değerleri belirlenmiştir. PM₁₀ için ulusal ve uluslararası mevzuatta belirlenen sınır değerleri Tablo 2.3'te, PM_{2.5} için ise uluslararası mevzuatta belirlenen sınır değerleri Tablo 2.4'te bir araya getirilmiştir.

Tablo 2.3. PM₁₀ için sınır değerleri.

	PM ₁₀ için limit değerler (µm/m ³)		Ref.
	Günlük	Yıllık (ort.)	
Avrupa Birliği	50*	40	European commission (t.y)
Dünya Sağlık Örgütü (DSÖ)	45	15	WHO (2021).
HKDYY	50	40	HKDYY, (2008).
Çevre Koruma Ajansı (EPA)	150	-	EPA Federal Register (2020)

*1 takvim yılında 35 defadan fazla geçilmeyecektir.

Tablo 2.4. PM_{2.5} için sınır değerleri.

	PM _{2.5} için limit değerler (µm/m ³)		Ref.
	Günlük	Yıllık (ort.)	
Avrupa Birliği	-	20*	European commission (t.y)
Dünya Sağlık Örgütü (DSÖ)	15	5	WHO (2021).
HKDYY	-	-	HKDYY (2008).
Çevre Koruma Ajansı (EPA)	35	12/15	EPA Federal Register (2020)

* 2020 yılı için ikinci aşama limit değeri

3. LİTERATÜR ARAŞTIRMASI

Havada bulunan kirleticilerin tespiti, insan sağlığına etkilerinin belirlenmesi ve bu etkilerin azaltılması konusu, çeşitli disiplinler arasında iş birliği gerektiren önemli bir konudur. Mühendislik, tıp, fen bilimleri ve diğer alanlar, hava kirliliği sorununun anlaşılması ve çözümü için bir araya gelmektedir. Günümüzde artan hava kirliliği düzeyleri ve buna bağlı sağlık sorunlarının artışı, bilim insanlarının bu konuya daha fazla odaklanmasını gerektirmektedir. Hava kirleticilerinin ölçümü ve tespiti karmaşık bir alan iken, yüksek doğrulukla çalışan tahmin modelleri, yerel yönetimlere ve karar vericilere önemli bir bilgi kaynağı sağlayabilir.

Türkiye'de çevre mühendisliği alanında yapay zeka tabanlı hava kirliliği tahmin modeli konulu çalışmalar oldukça sınırlıdır. Bu konuda yapılan literatür araştırması, geleneksel zaman serisi modelleme yöntemlerinin yanı sıra yapay zeka tabanlı yaklaşımların (makine öğrenmesi ve derin öğrenme) hava kirliliği parametrelerinin tahmini için ulusal ve uluslararası düzeyde önemli çalışmalar içerdiğini göstermiştir. Konu ile ilgili incelenen çalışmalar ilerleyen kısımlarda özetlenmiştir.

3.1. Geleneksel Zaman Serisi Yöntemleri

Akgül I. (1994)' tarafından yapılan tanıma göre, zaman serisi “zaman içinde meydana gelen gözlemlerin bir araya toplanması” anlamına gelir (Akgül, 1994). Bir başka şekilde ifade edilecek olursa zaman serisi “her biri belirli bir zamanda (t) kaydedilen bir dizi x_t gözlemidir” (Karakoyun, 2018).

Zaman serileri, gözlemlerin belirli bir zaman aralığında toplandığı kesikli zaman serisi veya sürekli olarak devam eden sürekli zaman serisi olarak adlandırılabilir (Akgül, 1994). Hava kalitesi verileri sürekli ve çok değişkenli zaman serileridir; her ölçüm belirli bir zaman noktasını temsil eder, bu ölçümler bir şekilde önceki ölçümlerle ilişkilidir ve dolayısıyla birbirine bağımlıdır (Freeman ve ark., 2018). Zaman serisi analizi, gelecekteki değişiklikleri tahmin etmeyi amaçlar.

Zaman serisi analizi, tahmin mekanizmalarına ihtiyaç duyulan birçok alanda önemlidir. Tıp, sosyal bilimler, ekonomi gibi alanlarda sıkça rastlanan zaman serisi verilerinin incelenmesi için zaman serisi analizlerine başvurulur (Karakoyun, 2018).

Zaman serisi analizleri genel olarak basit ortalama yöntemi, hareketli ortalama yöntemi, eksponansiyel (üssel) düzgünleştirme yöntemi, trend analizi yöntemi ve mevsimsel dalgalanmaları ve trende oranlama yöntemini içerir (Yavuz, t.y.). Zaman serisi analiz yöntemleri genellikle serinin durağan (stationary) olduğu durumlarda kullanılır, yani ortalamada ve varyansta sistemli bir değişim olmamalı ve periyodik değişimler olmamalıdır (Akgül, 1994).

Şekil 3.1’de zaman serisi tahmin yöntemlerinin bazıları bir araya getirilmiştir (Akgül, 1994). Zaman serisi analizleri, regresyon analizinden farklı olarak birbirini takip eden gözlemlere dayanır. Geçmiş gözlemlerden yola çıkarak geleceği tahmin etmek, daha doğru tahminler yapmamıza yardımcı olur (Akgül, 1994).



Şekil 3.1. Zaman serileri analiz yöntemleri (Akgül, 1994'den uyarlanmıştır).

Uygun ve doğruluğu yüksek tahmin modelinin seçilmesi birçok faktöre dayanır. Bu faktörler arasında geçmiş verilerin elde edilebilirliği, tahminin içeriği, amacı ve kullanım yeri, tahmin dönemi, mevsimsel ve trendsel değişkenlerin varlığı, analistin yetkinliği, kullanılan bilgisayar programlarının yetenekleri, harcanacak zaman ve tahmin maliyeti gibi etkenler bulunmaktadır (Akgül, 1994).

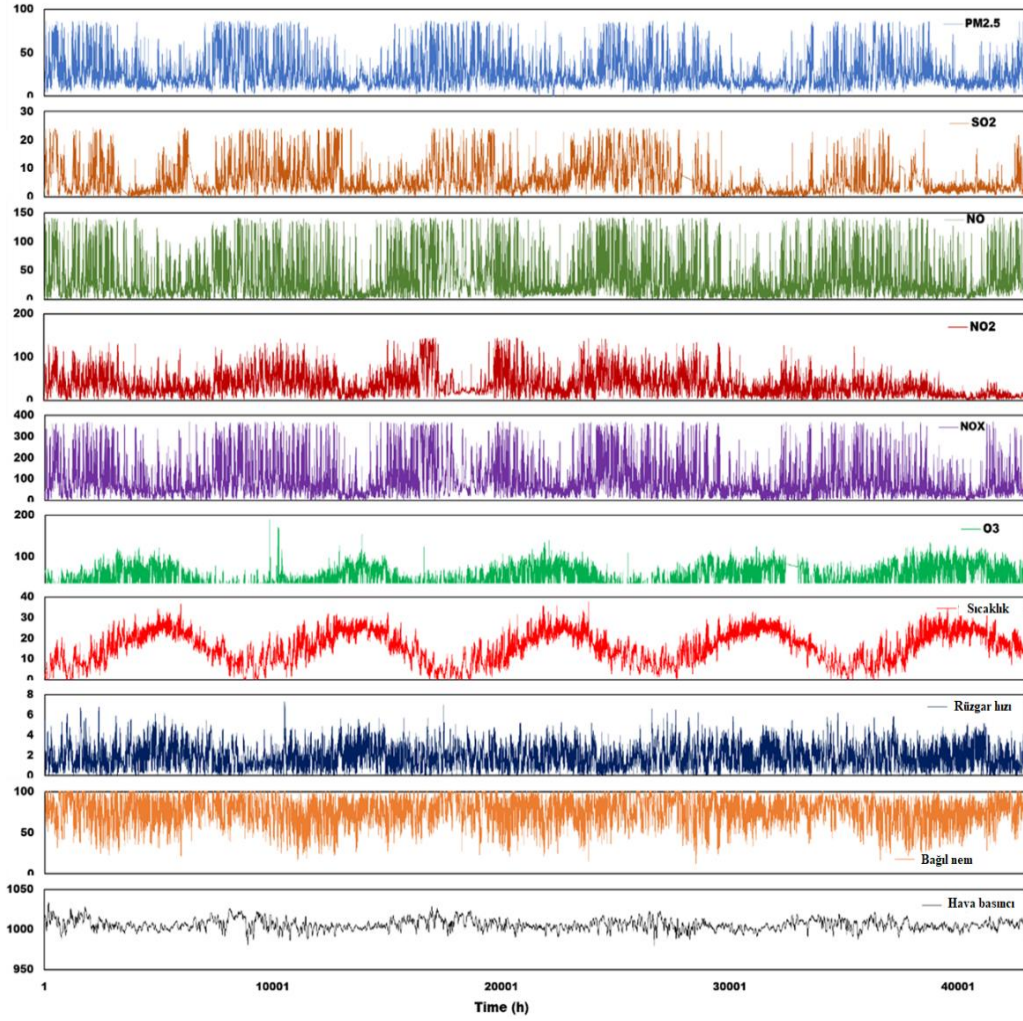
Aksoy ve ark. (2020) tarafından yapılan bir çalışmada, Türkiye'nin gelecekteki sıcaklık tahmini için 2840 aylık ortalama sıcaklık değerini içeren bir zaman serisi veri seti kullanılmıştır. Çalışmada veri setinin durağan özellikte olduğu tespit edilmiş ve bu nedenle ARIMA modeli tercih edilmiştir. Çalışmanın performans değerlendirmesinde, statik ve dinamik tahmin modellerinde sırasıyla %96,68 ve %97,19 oranında determinasyon katsayısı (R^2) elde edilmiştir (Aksoy ve Salman, 2020).

Liu ve ark. (2021) tarafından gerçekleştirilen bir çalışmada, minyatür hava kalitesi detektörünün izleme verilerini düzeltmek için Temel Bileşen Regresyonu (PCR), Destek Vektör Regresyonu makinesi (SVR) ve Otoregresif Hareketli Ortalama (ARMA) modelini birleştiren bir hava kalitesi tahmin modeli oluşturulmuştur. Bu çalışma, önerilen PCR-SVR-ARMA kombinasyon tahmin modelinin diğer yöntemlere göre daha iyi bir tahmin etkisine sahip olduğunu göstermiştir (Liu ve ark., 2021).

Turgut ve Temiz (2015) tarafından yapılan bir çalışmada, Ankara'nın Sıhhiye bölgesindeki PM_{10} ölçüm değerleri üzerinden Box-Jenkins yöntemi kullanılarak zaman serisi analizi yapılmış ve gelecekteki PM_{10} değerlerinin tahmini gerçekleştirilmiştir. Bu çalışmanın sonucunda, bölgenin kötü hava kalitesine sahip olduğu ve gelecekteki ortalama PM_{10} değerinin 83,21 olduğu belirlenmiştir (Turgut ve Temiz, 2015).

Tez çalışmasında çalışılan veri setlerinden biri olan ve 2. araştırma konusunda (AK-2) kullanılan veri setine ait grafikler Şekil 3.2'de verilmiştir. Bu veri seti, gerçek hava kirleticisi parametrelerini ve meteorolojik parametreleri içeren tipik birçok değişkenli zaman serisini içermektedir. Grafik incelendiğinde, hava kalitesi ve meteorolojik verilerde düzensizliklerin olduğu görülmektedir. Aynı zamanda sıcaklık verileri sadece mevsimsel ve yıllık değişim modellerini göstermektedir. Meteorolojik durumlar ile hava kirliliği konsantrasyonları arasında etkileşim bulunmaktadır. Yüksek

rüzgar hızı $PM_{2.5}$ konsantrasyonlarını düşürür, yüksek nem hava kirliliğini artırır ve yüksek atmosferik basınç ise hava kalitesini olumlu etkiler (Du ve ark. 2021).



Şekil 3.2. 01/01/2019-30/11/2019 tarihleri arasında Kağıthane ilçesi için saatlik hava kalitesi ve meteorolojik veriler zaman serisi grafiği (Eren ve ark. 2023).

3.2. Makine Öğrenmesi Yöntemleri

Mevcut hava kirliliği tahminleri için genellikle sayısal ve istatistiksel yöntemler yaygın olarak kullanılmaktadır. Sayısal modeller, büyük ölçekte hava kalitesini modellemek için meteorolojik prensipleri ve matematiksel yöntemleri kullanırken, atmosferik fizik ve kimyasını da dikkate alır (Yan ve ark., 2021).

Geleneksel regresyon yöntemleri, bilgisayar bilimi ve teknolojisindeki ilerlemelerle birlikte istatistiksel tahmin yöntemlere evrilmiştir. Bu gelişmelerle birlikte, yüksek boyutlu uzayda doğrusal olmayan hava kirliliği modellemesi için yapay zeka (YZ) algoritmaları kullanılmıştır (Eren ve ark. 2023).

Son yıllarda, hava kirleticilerini tahmin etmek için yapay zeka yöntemlerinin kullanımına dair birçok çalışma yürütülmektedir. Örneğin, Polonya yerleşim yerlerinde $PM_{2.5}$ ve PM_{10} tahmin sistemi (Czernecki ve ark., 2021) Santiago de Chile’de saatlik $PM_{2.5}$ tahmini (Perez ve Menares, 2018), Bişkek’te günlük $PM_{2.5}$ tahmini (Isaev ve ark., 2022), Tahran’daki $PM_{2.5}$ tahmin sistemi (Karimian ve ark., 2019) sayılabilir.

Ayrıca, hava kirliliği tahmini için en yaygın kullanılan yapay zeka tabanlı yöntemler arasında yapay sinir ağı (YSA) (Cabaneros ve ark., 2019), destek vektör makinesi (SVM) (Leong ve ark., 2020), rastgele orman (RF) (Rubal ve Kumar, 2018) ve derin öğrenme (DL) (Bui ve ark., 2018) yer almaktadır.

Bu yöntemler arasında DL’in hava kirliliği tahmini için yaygınlaşması, daha fazla katman ve geniş veri setlerinin kullanılabilmesi, tüm katmanların eşzamanlı olarak işlenmesiyle daha doğru sonuçların elde edilmesi gibi avantajlara sahip olmasından kaynaklanmaktadır (Bekkar ve ark., 2021).

3.3. Derin Öğrenme Yöntemleri

Derin öğrenme algoritmaları (LSTM, CNN, RNN, vb.) özellikle çok parametrelili, karmaşık ve doğrusal olmayan süreçlerin modellenmesi için uygundur. Örneğin, LSTM, doğrusal olmayan gerçek dünya durumlarını doğru bir şekilde temsil ettiği, uzun geçmiş verilerinin etkisini dikkate aldığı ve çoklu girdileri veya çok değişkenli serileri çözdüğü için hava kirliliği modellemesinde en yaygın kullanılan DL algoritmasıdır (Bekkar ve ark., 2021; Lu ve ark. 2021).

Karimian ve ark. (2019) çalışmasında, farklı zaman aralıklarında $PM_{2.5}$ konsantrasyonlarını tahmin etmek için çoklu toplama regresyon ağaçlarına (MART), derin ileri beslemeli sinir ağına (DFNN) ve LSTM’ye dayalı modellerin tahmin performanslarını karşılaştırmıştır. Çalışma, LSTM modelinin diğer iki modele kıyasla zaman serisi verilerindeki zamansal bağımlılıkları yakaladığını ve $R^2=0,80$ ile $PM_{2.5}$

konsantrasyonları için en iyi tahmini verdiğini ortaya koymuştur (Karimian ve ark., 2019).

Pak ve ark. (2020) çalışmasında, çalışmasında, ertesi günün ortalama günlük $PM_{2.5}$ konsantrasyonlarını tahmin etmek için 384 istasyondan elde edilen 3 yıllık hava kalitesi ve meteorolojik verilere dayalı olarak hibrit bir evrişimli sinir ağı ve uzun kısa süreli bellek (CNN-LSTM) modelini kullanmışlardır. Bu modelin, çok katmanlı algılayıcı (MLP) ve LSTM modellerine kıyasla daha iyi kararlılık ve tahmin performansına sahip olduğu sonucuna varmışlardır (Pak, ve ark., 2020).

Menares ve ark. (2021) çalışmasında, maksimum $PM_{2.5}$ konsantrasyonlarını tahmin etmek için bir LSTM, DFFNN ve MLP modeli önermiştir. Uygun kirletici ve meteorolojik değişkenler seçildiğinde LSTM modellerinin $PM_{2.5}$ tahminine daha iyi yanıt verdiği ($R=0,86$) sonucuna varmışlardır. LSTM modelleri, bellek birimleri aracılığıyla zaman içinde $PM_{2.5}$ tahmini için yararlı olan önemli sinoptik kalıpları (synoptic patterns) hatırlayabilir. LSTM algoritması, geçmiş hava kalitesi ve meteorolojik verileri kullanarak $PM_{2.5}$ 'i tahmin etmede güvenilir ve doğru olduğu görülmüştür (Menares ve ark., 2021).

Bekkar ve ark. (2021) çalışmasında, Çin' in Pekin şehrine ait geçmiş kirletici verileri, $PM_{2.5}$ ve meteorolojik verileri içine alan veri setleri ile CNN-LSTM hibrit modelinin tahmin performansını değerlendirmiştir. Çalışma, LSTM, Bi-LSTM, GRU, Bi-GRU, CNN ve hibrit bir CNN-LSTM modeli tahmin performansını incelemiş ve analiz sonuçlarına göre CNN-LSTM hibrit modelin karşılaştırılan tüm modellerden yüksek tahmin performansına ($R^2=0,989$) sahip olduğu ortaya konulmuştur (Bekkar ve ark., 2021).

Aydın ve ark. (2021) çalışmasında, hava kalitesinin bir önceki koşullar ile bağlantılı olduğu düşüncesinden yola çıkarak LSTM modelini PM_{10} değeri tahmini için önermiştir. Model performansını değerlendirme çalışması yapmış ve klasik DL yöntemi ile karşılaştırmıştır. Klasik DL yönteminin LSTM yöntemine göre düşük performansa sahip olduğu ve ayrıca, LSTM modelinin veri kaybı söz konusu olduğunda daha az etkilendiği sonucuna varmıştır (Aydın ve ark., 2021).

Tao ve ark. (2019) çalışmasında, $PM_{2.5}$ konsantrasyonu tahmini için evrişim tabanlı çift yönlü kapılı tekrarlayan birim (CBGRU) yöntemi ile kısa vadeli bir tahmin modeli

geliştirmiştir. UCI Machine Learning Repository'deki Pekin $PM_{2.5}$ veri seti kullanılmıştır. CBGRU modeli RMSE:14,53; MAE:10,48; SMAPE:0,20 değerlerine sahiptir. Ayrıca geleneksel makine öğrenimi modelleri (SVR, DTR, GBR) ve klasik derin öğrenme modelleri (RNN, GRU, LSTM) ile karşılaştırma çalışması yapılmıştır. Tahmin sonuçları incelendiğinde diğer tahmin modellerine göre, CBGRU modelinin hatasının daha düşük ve tahmin performansının daha iyi olduğu ortaya konmuştur. Sonuçlar, CBGRU yönteminin bir zaman serisi olan $PM_{2.5}$ tahmini için uygun ve diğer yöntemler ile rekabet edebilir olduğunu göstermektedir (Tao ve ark, 2019).

Li ve ark. (2020), CNN-LSTM derin öğrenme algoritmalarını hibrit bir şekilde kullanarak Pekin bölgesindeki $PM_{2.5}$ verilerini tahmin etti. Çalışmada, zaman serileri analizinde CNN yönteminin ve LSTM yönteminin avantajları birleştirilerek RMSE değeri 10s/epoch'da 1 gün için 18,99 olan başarılı bir model geliştirilmiştir (Li ve ark, 2020).

Li ve Shen (2019), Çin'de hava kirliliği problemi ciddi boyutlarda olduğu ve $PM_{2.5}$ 'un önde gelen kirlilik parametresi olması nedeniyle tahmin doğruluğunu artırma için tamamlayıcı toplu ampirik mod ayrıştırma (CEEMD) ve LSTM algoritmalarından oluşan hibrit bir model üstünde çalışmıştır. Çalışma koşulları; mini batch boyutu 32, Adam optimizasyonu, kayıp fonksiyonu MSE ve 200 gizli katman nöron sayısıdır. CEEMD-LSTM (MAE:10,79; RMSE:14,64; MAPE:19,071) daha sonra ELM (MAE:25,37; RMSE:33,76; MAPE:50,82), SVR (MAE:21,61; RMSE:29,36; MAPE:45,27) ve LSTM (MAE:16,65; RMSE:23,31; MAPE:33,67) modelleri ile karşılaştırılmış ve daha yüksek tahmin yeteneğine sahip olduğu belirtilmiştir. Ayrıca LSTM li modellerin hafıza kapasitesi nedeniyle seçildiği belirtilmektedir (Li ve Shen, 2019).

Yapılan literatür çalışması neticesinde tez çalışmasında 2 ayrı araştırma konusu belirlenmiş ve bu iki özel araştırma hedefi ve ilgili adımlar aşağıda sunulmuştur.

AK-1 Özel Hedefleri: VÖİ-ÖS'li/VÖİ-ÖS'siz Veri Setlerinin PM_{10} Tahminine Etkisinin Belirlenmesi

1. PM_{10} Tahmini için LSTM Modeli Geliştirme:

- a. PM_{10} konsantrasyonlarını tahmin etmek için bir LSTM modeli tasarlanması ve eğitilmesi.
 - b. Eğitilen modelin performansının RMSE, MAE ve R^2 gibi performans metrikleri kullanılarak değerlendirilmesi.
2. VÖİ-ÖS'li ve VÖİ-ÖS'siz Veri Setlerinin Farklı LSTM Modelleri Performans Üzerindeki Etkisi:
- a. VÖİ-ÖS'li ve VÖİ-ÖS'siz veri setleri kullanılarak farklı LSTM modellerinin (Yalın, Çift Yönlü ve Yığılmış LSTM) eğitilmesi.
 - b. Eğitilen LSTM modellerinin test veri seti ile RMSE, MAE ve R^2 gibi performans metrikleri ile karşılaştırılması.

AK-2 Özel Hedefleri: Farklı Derin Öğrenme Modellerinin Saatlik $PM_{2.5}$ Konsantrasyon Tahmini için Karşılaştırmalı Analizi

1. Saatlik $PM_{2.5}$ Konsantrasyon Tahmini İçin Derin Öğrenme Modeli Geliştirme:
 - a. Veri setinde önceki 24 saatin verilerini girdi olarak kullanarak 25. saatlik $PM_{2.5}$ konsantrasyonunu tahmin edecek şekilde zaman pencereleme yapılması.
 - b. Saatlik $PM_{2.5}$ konsantrasyonlarını tahmini için LSTM, RNN ve GRU farklı derin öğrenme modellerinin geliştirilmesi.
 - c. Geliştirilen derin öğrenme modelleri performansının, MSE, RMSE, MAE, MAPE ve R^2 gibi performans metrikleri kullanılarak değerlendirilmesi.
2. Model Genelleştirme Yeteneği ve Uygulanabilirlik Değerlendirmesi:
 - a. Potansiyel olarak benzer özelliklere sahip farklı bölgelerden alınan veriler kullanılarak, geliştirilen modelin $PM_{2.5}$ konsantrasyonu tahmini için genelleştirme yeteneğinin test edilmesi.

Bu iki araştırma konusu ile çalışma, hava kirliliği tahmini için derin öğrenme yöntemlerinin etkili kullanımını ve performansını değerlendirmeyi amaçlamaktadır.

3.4. Araştırma Hedefi ve Bilimsel Katkı

Günümüzde özellikle sanayi bölgelerine yakın şehirlerde hava kirliliği hem çevresel hem de halk sağlığı açısından önemli bir sorun olarak öne çıkmaktadır. Önümüzdeki dönemlerde, hava kirliliği kaynaklı ani olaylar gerçekleşirse, bu sadece insan sağlığını etkilemekle kalmayacak, aynı zamanda sanayi tesislerinin işleyişini yavaşlatma gibi yaşamsal konularda da zorluklara neden olabilecektir. Bu bağlamda, hava kirliliği tahmini, çevresel etkilerin belirlenmesi ve olası felaketlerin önceden önlenmesi için alınacak tedbirlerin hızla hayata geçirilmesi gereken ilk adımdır. Doğru ve güvenilir tahminler, gelecekteki değişiklikleri öngörmek ve olumsuz etkileri azaltmak için gereken önlemleri almak açısından kritik bir rol oynamaktadır.

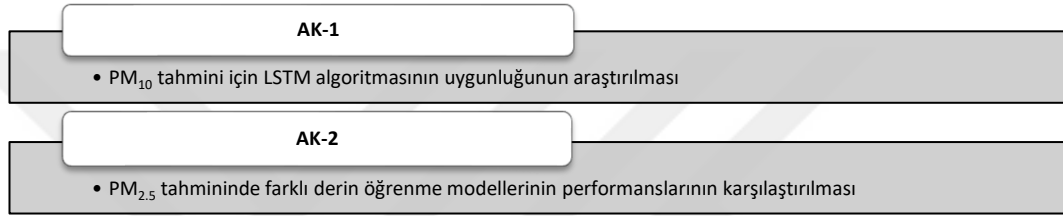
Yapay zeka tabanlı teknolojilerin geliştirilmesi, doğal ve sosyal bilimlerde aktif olarak kullanılmaktadır. Bu bağlamda, bu tez çalışması, partikül madde türlerinin derin öğrenme algoritmaları ile tahmin edilmesini ve en uygun modelin belirlenmesini amaçlamaktadır. Çalışma, Türkiye’de nüfus yoğunluğunun ve sanayi faaliyetlerinin en yüksek olduğu İstanbul ilini ele alarak, hava kirliliği tahminleri için YZ tekniklerinden olan LSTM (Yalın, Çift yönlü ve Yığılmış), GRU ve RNN modellerini kullanarak hava kirliliği tahminleri yapmıştır. Bu çalışma, özellikle insan sağlığı üzerinde olumsuz etkisi olan PM_{10} ve tehlikeli $PM_{2.5}$ kirleticileri için uygun modellerin tespit edilmesine yönelik olarak Çevre Mühendisliği alanında YZ uygulamalarının yaygınlaşmasına katkıda bulunmayı amaçlamaktadır.

Bu çalışmanın özellikle, hızla artan sanayileşme ve şehir nüfusu gibi faktörlerle birlikte, yerel yönetimlerin hava kirliliğine karşı stratejiler geliştirmeleri ve akıllı şehir kavramını desteklemeleri gereken bir dönemde, literatüre ve denetleyici kuruluşlara önemli bir katkı sağlayabileceği düşünülmektedir.



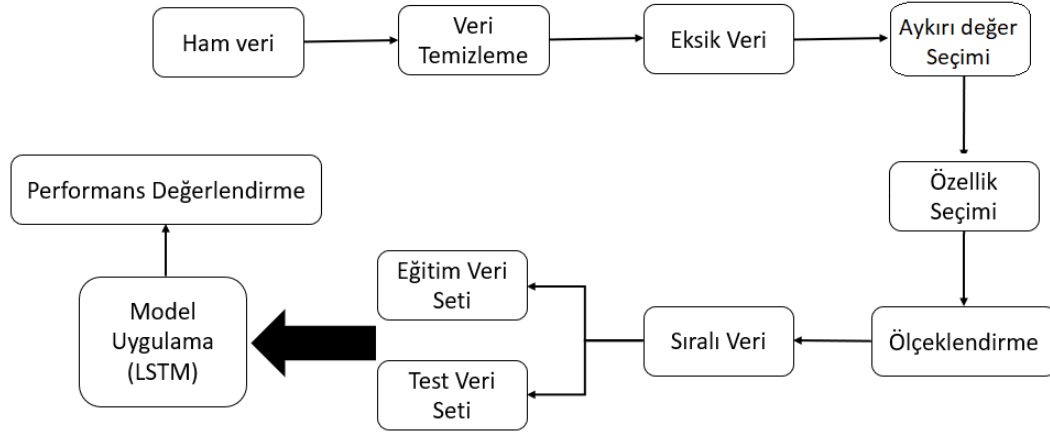
4. MATERYAL VE METOT

Bu tez çalışması, partikül madde kaynaklı hava kirliliğini (PM_{10} ve $PM_{2.5}$) tahmin etmek amacıyla tahmin performansı yüksek ve güvenilir modeller oluşturarak, bu modellerin karar vericilere bilinçli katkı sağlama amacını taşımaktadır. Bu amacı gerçekleştirmek için, Şekil 4.1'de gösterildiği şekilde, belirlenen hedefleri yerine getirmek üzere 2 ayrı araştırma konusu (AK) üzerinde çalışılmıştır.



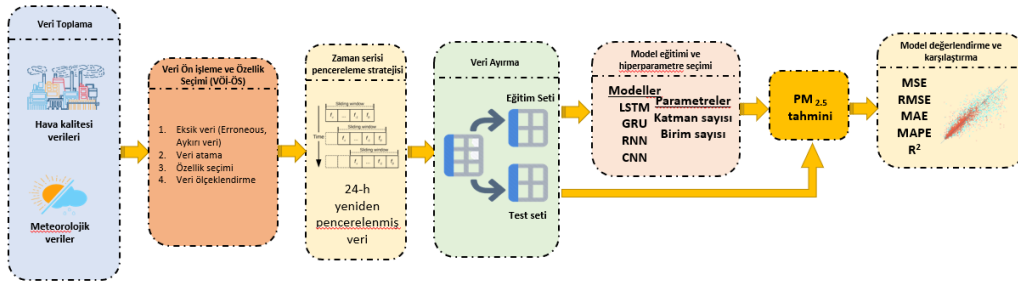
Şekil 4.1. Tez araştırma konuları

Araştırma Konusu 1 (AK-1) kapsamında, PM_{10} hava kirliliği parametresinin tahmini için LSTM algoritmasının etkinliği incelenmektedir. Bu amaç doğrultusunda, meteorolojik ve hava kalitesi verileri bir araya getirilerek PM_{10} tahmini için bir veri seti oluşturulmuştur. Ardından, eksik veri tespiti ve doldurulması, aykırı veri belirleme ve düzeltme, özellik seçimi, verinin ölçeklendirilmesi gibi Veri Ön İşleme ve Özellik Seçimi (VÖİ-ÖS) adımı uygulanmıştır. Bu aşamada amaç, ham verinin Veri Ön İşleme aşamalarından geçirilerek model performansının artırılmasıdır. VÖİ-ÖS uygulanan ve uygulanmayan verilerle ayrı ayrı modelleme çalışmaları gerçekleştirilmiştir. Son olarak, oluşturulan modellerin performansı değerlendirilerek tahmin doğruluk oranı incelenmiştir. Çalışmada izlenen yol Şekil 4.2.'de verilmiştir.



Şekil 4.2. Araştırma konusu 1 akış şeması.

Araştırma Konusu 2 (AK-2) çerçevesinde, PM_{2.5} hava kirliliği parametresinin tahmini için farklı derin öğrenme modellerinin performansları karşılaştırılarak en uygun model araştırılmıştır. Bu amaçla, meteorolojik ve hava kalitesi verileri bir araya getirilerek PM_{2.5} tahmini veri seti oluşturulmuştur. Bu veri setine Veri Ön İşleme ve Özellik Seçimi (VÖİ-ÖS) adımı uygulanmıştır. Daha sonra modelleme aşamasına geçilmiş ve farklı derin öğrenme algoritmaları kullanılarak PM_{2.5} tahmini çalışmaları gerçekleştirilmiştir. Modellerin performansı değerlendirilerek en iyi model belirlenmiştir. Son olarak, farklı veri setleri kullanılarak seçilen en iyi modelin genelleme kabiliyeti incelenmiştir. Çalışmada izlenen yol Şekil 4.3'te verilmiştir.



Şekil 4.3. Araştırma konusu 2 akış şeması.

4.1. Çalışma Alanı

Türkiye'nin nüfusu, Türkiye İstatistik Kurumu (TÜİK) Adrese Dayalı Nüfus Sistemi 2021 verilerine göre 84 milyon 680 bin 273 kişi olarak kaydedilmiştir (TÜİK, 2022). Ülkemizde yaşayan nüfusunun yaklaşık %70'i PM₁₀ limit değerlerinin aştığı şehirlerde yaşamaktadır (TMMOB, 2018). Türkiye'nin en kalabalık şehri olan İstanbul, 2021 yılı

verilerine göre yaklaşık 15,84 milyon insanın yaşadığı bir yerdir. Bu şehirde kilometrekare başına düşen nüfus yoğunluğu 3049 kişi (Türkiye: 110 kişi/km²) olarak ölçülmüştür (TÜİK, 2022). Nüfus artışı, beraberinde motorlu taşıt ve konut sayısındaki artışı getirmekte ve bu da araç emisyonlarını ve kentsel ısınmayı artırmaktadır. İstanbul, ticaret, endüstri, lojistik, turizm, kültür, eğitim ve sanat gibi birçok faaliyetin yoğun olarak bir arada bulunduğu bir şehirdir. Ancak İstanbul'un çeşitli bölgeleri Çevre Mühendisleri Odası'nın 2018 raporuna göre PM₁₀ kirliliği açısından yasal düzenlemelerin ötesinde değerlere sahiptir (TMMOB, 2018). İstanbul'un farklı topografik özellikleri nedeniyle değişen bölgesel iklim koşulları ve buna bağlı olarak hava kirleticileri bulunmaktadır. Ayrıca İstanbul, Avrupa'dan esen batı rüzgarları ve genellikle sonbahar ve ilkbaharda Sahra Çölü'nden (Afrika'dan) gelen tozlu rüzgarlar nedeniyle önemli ölçüde partikül madde kirliliğine maruz kalmaktadır (Ağaç, 2016). Bu faktörler, İstanbul'un farklı bölgelerinde düşük hava kalitesi değerlerinin gözlemlenmesine neden olmaktadır. İlgili tez çalışması için her iki araştırma konusu için de çalışma alanı olarak İstanbul İli ve bu bölgedeki hava kalitesi istasyonları seçilmiştir.

4.2. Hava Kirliliği Verileri

Halihazırda, İstanbul'un çeşitli bölgelerinde 38 hava kalitesi izleme istasyonu (HKİİ) mevcuttur (İstanbul Büyükşehir Belediyesi [İBB], t.y.). Bu istasyonlardan seçilen toplam 11 istasyonun verileri kullanılarak modelleme çalışmaları gerçekleştirilmiştir. Tüm çalışmada veri kaynakları olarak belirlenen hava kalitesi izleme istasyonları Tablo 4.1'de ayrıntılı bir şekilde listelenmiştir. İstasyon verilerinin kalite güvencesi ve kontrolü; eksik veriler, yinelenen ölçümler, cihaz hataları gibi önemli sorunlardan kaçınmak ve istasyon ölçümlerinin doğru bir şekilde temsil edilmesini sağlamak amacıyla büyük bir öneme sahiptir. Yerel istasyonlardan elde edilen veriler (büyük veri), internet aracılığıyla Gölbaşı-Ankara'da bulunan T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Veri İşletim Merkezi'ne (Çevresel Etki Değerlendirmesi İzin ve Denetim Genel Müdürlüğü bünyesindeki Laboratuvar, Ölçüm ve İzleme Dairesi Başkanlığı) iletilmektedir. Bu veriler, cihazların kalibrasyon ve alarm bilgileri göz önünde bulundurularak doğrulama çalışmalarıyla değerlendirilmekte ve ilgili merkez tarafından geçerliliği sağlanmaktadır (SİM, 2022a).

Tablo 4.1. Veri kaynakları olarak seçilen istasyonlar.

İstasyon Adı			Tipi	Araştırma Konusu			
				AK -1		AK-2	
				Model Geliştirme	Genelleme	Model Geliştirme	Genelleme
Başakşehir İzleme İstasyonu (BHKİİ)	Hava Kalitesi	Sanayi-Kentsel		✓			
Kağıthane İzleme İstasyonu (KHKİİ)	Hava Kalitesi	Kentsel			✓	✓	
Aksaray İzleme İstasyonu (AHKİİ)	Hava Kalitesi	Trafik					✓
Avcılar İzleme İstasyonu (AvHKİİ)	Hava Kalitesi	Şehir-Arkaplan					✓
Beşiktaş İzleme İstasyonu (BeHKİİ)	Hava Kalitesi	Trafik			✓		✓
Kadıköy İzleme İstasyonu (KaHKİİ)	Hava Kalitesi	Kentsel					✓
Kartal İzleme İstasyonu (KarHKİİ)	Hava Kalitesi	Kentsel					✓
Maslak İzleme İstasyonu (MHKİİ)	Hava Kalitesi	Kentsel					✓
Sultangazi İzleme İstasyonu (SHKİİ)	Hava Kalitesi	Kentsel-Trafik					✓
Tuzla İzleme İstasyonu (THKİİ)	Hava Kalitesi	Kentsel					✓
Ümraniye İzleme İstasyonu (ÜHKİİ)	Hava Kalitesi	Şehir-Arkaplan					✓

- Tez çalışması AK-1 PM₁₀ tahmini için kullanılan hava kirliliği veri seti:

Bu bölümde model geliştirme aşamasında, Başakşehir Hava Kalitesi İzleme İstasyonu'ndan (BHKİİ) 01/03/2013 - 30/11/2019 tarihleri arasında alınan veriler üzerinde çalışılmıştır (SİM, 2019). Veri seti saatlik olarak ölçülen PM₁₀ (µg/m³), SO₂

($\mu\text{g}/\text{m}^3$), NO ($\mu\text{g}/\text{m}^3$), NO₂ ($\mu\text{g}/\text{m}^3$), NO_x ($\mu\text{g}/\text{m}^3$), O₃ ($\mu\text{g}/\text{m}^3$) ve CO ($\mu\text{g}/\text{m}^3$) olmak üzere toplam 7 farklı kirletici parametreyi içermektedir. İlgili istasyon, deniz seviyesinden 142 metre yükseklikte olup karayolundan 34 metre uzaklıktadır (Enlem: 41.0955, Boylam: 28.7898). İstasyonun ölçüm yüksekliği 4 metredir (BHKİİ, 2021).

- Tez çalışması AK-2 PM_{2.5} tahmininde derin öğrenme modelleri performans karşılaştırılması için kullanılan hava kirliliği veri seti:

Bu bölümünde model geliştirme sürecinde, Kağıthane Hava Kalitesi İzleme İstasyonu'ndan (KHKİİ) 01/01/2015- 30/11/2019 tarihleri arasında toplam 43057 saatlik veri kullanılarak çalışma yapılmıştır (SİM, 2022b). Veri seti saatlik PM_{2.5} ($\mu\text{g}/\text{m}^3$), SO₂ ($\mu\text{g}/\text{m}^3$), NO ($\mu\text{g}/\text{m}^3$), NO₂ ($\mu\text{g}/\text{m}^3$), NO_x ($\mu\text{g}/\text{m}^3$) ve O₃ ($\mu\text{g}/\text{m}^3$) olmak üzere toplam 6 farklı kirletici parametreyi içermektedir. İlgili istasyon, deniz seviyesinden 91 metre yükseklikte olup karayolundan 6 metre uzaklıktadır (Enlem: 41.0923, Boylam: 28.9747). İstasyonun ölçüm yüksekliği 4 metre'dir (KHKİİ, 2022).

Ayrıca, çalışma kapsamında belirlenen modelin PM_{2.5} tahmininde genelleme yeteneğini değerlendirmek için Tablo 4.1'de belirtilen diğer istasyon verileri kullanılmıştır. Model genelleme yeteneğini değerlendirmek için 01/01/2019 00:00 ile 30/11/2019 (her istasyon için toplam 7960 saat) tarihlerini kapsayan diğer dokuz farklı izleme istasyonu tarafından sağlanan veri setleri kullanılmıştır (SİM, 2022b). Hava kalitesi veri setleri, PM_{2.5} ($\mu\text{g}/\text{m}^3$), NO ($\mu\text{g}/\text{m}^3$), NO₂ ($\mu\text{g}/\text{m}^3$), NO_x ($\mu\text{g}/\text{m}^3$), O₃ ($\mu\text{g}/\text{m}^3$) ve SO₂ ($\mu\text{g}/\text{m}^3$) kirleticilerinden oluşmaktadır.

4.3. Meteorolojik Veriler

Tez çalışmasının AK-1 bölümünde PM₁₀ tahmini için kullanılan meteorolojik parametreler veri seti aşağıdaki gibidir:

Bu bölümde, çalışma kapsamında atmosferik sıcaklık (°C), rüzgar yönü, rüzgar hızı (m/s), bağıl nem (%), ve atmosferik basınç (mbar) olmak üzere 5 farklı meteorolojik parametreye ait veriler TC Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Meteoroloji Genel Müdürlüğü'nden (MGM) temin edilmiştir (MGM, 2020).

Tez çalışmasının AK-2 bölümünde PM_{2.5} tahmininde derin öğrenme modellerinin performans karşılaştırılması için kullanılan meteorolojik parametreler veri seti aşağıdaki gibidir:

Bu bölümde, çalışma kapsamında atmosferik sıcaklık (°C), rüzgar hızı (m/s), bağıl nem (%), ve atmosferik basınç (mbar) olmak üzere 4 farklı meteorolojik parametreye ait veriler TC Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Meteoroloji Genel Müdürlüğü'nden (MGM) temin edilmiştir (MGM, 2022). Çalışmanın genelleme kabiliyeti bölümünde ise aynı meteorolojik parametreler tercih edilmiştir.

4.4. Veri Setine İstatistiksel Bakış

Veri setlerinin araştırmacının bir sonraki adımı planlaması için istatistiksel açıdan yorumlanması gerekmektedir. AK-1 ve AK-2 çalışmalarında kullanılan veri setlerinin istatistiksel araçlar kullanılarak elde edilen bilgiler Tablo 4.2 ve Tablo 4.3'te verilmiştir. Bu tablolarda, hava kirliliği ve meteorolojik parametrelerin ortalama, standart sapma, maksimum ve minimum değerleri hesaplanmıştır. Sunulan parametreler, veri ön işleme ve modelleme süreçlerinin girdi parametrelerini içermekte olup, aynı zamanda PM₁₀ ve PM_{2.5} gibi hedef parametreleri içermektedir.

Tablo 4.2. BHKİİ veri seti istatistiksel dağılımı.

Özellikler	Hava kirliliği parametreleri							Meteorolojik parametreler				
	PM ₁₀ (µg/m ³)	SO ₂ (µg/m ³)	O ₃ (µg/m ³)	CO (µg/m ³)	NO (µg/m ³)	NO ₂ (µg/m ³)	NO _x (µg/m ³)	Bağıl nem (%)	Rüzgar yönü	Rüzgar hızı (m/s)	Atmosferik sıcaklık (°C)	Atmosferik basınç (mbar)
Mevcut veri sayısı	56178	55783	55555	56175	52307	56045	50103	47275	56076	56942	56896	56955
Eksik veri sayısı	2983	3378	3606	2986	6854	3116	9058	11886	3085	2219	2265	2206
Eksik veri yüzdesi (%)	6,03	6,13	6,13	5,07	11,7	5,27	31,2	20,5	5,69	4,23	3,83	4,2
Ort.	55,5	5,53	56,9	504	13,5	32	46	82	118	2,62	15,5	994
Std.	43,4	8,23	32,1	368	29,2	25,4	67,3	17,3	96,6	1,55	7,54	70
Min	0	-59,7	-2,86	-215	-5,15	0,16	-4,98	0	0	0	-6,2	0
25%	29,7	1,9	33,9	300	3,2	14,4	9,1	71,7	34,4	1,4	9,5	995
50%	44,3	3,37	59,3	400	3,84	24	26,6	87	70,2	2,3	15,7	998
75%	66,4	6,51	79	570	9,98	41,4	51,9	94,5	198	3,55	21,8	1003
Maks.	865	230	1310	20710	551	260	1102	100	360	12,9	36,2	1022

Tablo 4.3. KHKİİ veri seti istatistiksel dağılımı.

Özellikler	Hava kirliliği parametreleri						Meteorolojik parametreler			
	PM _{2.5} (µg/m ³)	SO ₂ (µg/m ³)	NO (µg/m ³)	NO ₂ (µg/m ³)	NO _x (µg/m ³)	O ₃ (µg/m ³)	Atmosferik sıcaklık °C	Rüzgar hızı m/s	Bağıl nem %	Atmosferik basınç mbar
Eksik veri ve yüzdesi	1570 (3,6%)	1444 (3,4%)	1593 (3,7%)	1516 (3,5%)	1545 (3,6%)	1473 (3,4%)	-	884 (2,1%)	966 (2,2%)	851 (2,0%)
Ort.	26,20	5,91	33,92	36,53	93,18	44,81	15,67	1,87	77,48	1006,11
Std.	17,02	5,13	33,27	26,45	80,48	28,97	7,63	1,17	16,75	6,51
Min.	0,01	0,03	0,01	0,01	0,04	0,02	0,01	0,02	12,52	979,76
25%	14,51	2,34	11,17	17,22	37,81	19,10	9,36	0,92	66,73	1001,70
50%	20,75	3,92	20,53	29,42	64,32	43,07	15,85	1,68	80,08	1005,44
75%	32,31	7,83	44,78	49,18	121,14	68,17	22,09	2,69	91,14	1010,18
Maks.	86,69	24,30	142,29	144,89	370,65	189,79	37,68	7,32	100,00	1033,30

4.5. Modelleme Süreci

Modelleme süreci uzun ve dikkatli bir çalışma gerektirmektedir. Modelleme süreci genel olarak şu adımlardan oluşmaktadır: Sorunu tanımlama, ilgili verilerin toplanması, keşifsel veri analizi, veri ön işleme, model seçimi, modelin eğitilmesi ve performans değerlendirmesi. Modelleme süreci, Şekil 4.4 de verilmiştir.



Şekil 4.4. Modelleme süreci (Yalçın, 2021)'den uyarlanmıştır.

4.5.1. Veri ön işleme ve özellik seçimi (VÖİ-ÖS)

Zaman serilerindeki eksik veya hatalı veriler, insan hataları, cihaz arızaları veya kesinti süreleri gibi nedenlerle oluşabilir. Bu tür eksik ve hatalı verilerin doğru bir şekilde ele alınmaması, tahmin modelleri tarafından güvenilir olmayan sonuçlara yol açabilir (Mir ve ark., 2022). Bu nedenle, modelleme süreci öncesinde veri setlerinin incelenerek veri kalitesinin artırılması ve model performansının iyileştirilmesi gereklidir.

Veri ön işleme, verinin başka bir işlem için hazırlanmasını içerir ve klasik veri madenciliği görevlerinin ötesine geçer. Bu aşama, modellerin geliştirilmesi için bir araç olarak önemli bir rol oynamaktadır (García ve ark., 2016). Veri ön işleme, örnekleme, dönüşüm, gürültü giderme, eksik veri atama, normalleştirme ve özellik çıkarma gibi temel yöntemleri içermektedir (Lawton, t.y.).

Bu çalışmada, tahmin modelinin geliştirilmesi için kullanılacak ham veri setinin hazırlanması için Veri Ön İşleme ve Özellik Seçimi (VÖİ-ÖS) adımı gerçekleştirilmiştir. Bu kapsamda AK-1 çalışmasında BHKİİ verilerine eksik veri analizi, aykırı veri tespiti, veri ölçeklendirme ve özellik seçimi yapılmıştır. VÖİ-ÖS uygulanmış veri setleri ve uygulanmamış veri setleri ile tahmin çalışmaları gerçekleştirilmiş ve model performansları karşılaştırılmıştır. Bu model performans sonuçları, “Bölüm 5: Araştırma Bulguları” bölümünde ayrıntılı bir şekilde sunulmuştur.

4.5.1.1. Eksik veri analizi

Hava kalitesi izleme istasyonlarında, makine arızaları, rutin bakım, insan hataları ve diğer faktörler nedeniyle sık sık sorunlar yaşanabilir. Bu tür durumlarda, istasyonlardaki veriler her zaman eksiksiz bir şekilde kaydedilemeyebilir. Özellikle hava kalitesi ve meteorolojik parametreler gibi yüksek oranda eksik veri içeren özellikler, model tahminlerinin doğruluğunu olumsuz etkileyebilir. Bu nedenle, bu istasyonlardan elde edilen verilerde eksik ve anormal değer analizlerinin yapılması gerekmektedir. Benzer şekilde, bir özellikteki negatif değerlerin sayısı da veri analizlerini olumsuz etkileyebilir. Bu bağlamda, AK-1 çalışmasında öncelikle veri setinden yüksek oranda eksik veri içeren özelliklerin çıkarılması işlemi uygulanmıştır.

Aynı şekilde, analizimize bir özelliğin gözlenen değerlerini dahil etmek amacıyla eşik değerleri belirlemiş bulunmaktayız. Bu eşik değerler aşağıdaki şekildedir:

- (1) Eğer bir özelliğin ölçülen değerlerinin toplam sayısının %30'u eksik ise, bu özelliği veri kümesinden çıkarılması.
- (2) Her gözlemin dört ölçümü olması gerektiğini göz önünde bulundurarak, dörtten az ölçümü olan herhangi bir gözlemin çıkarılması (Aksangür ve ark., 2022).

Bu aşamada, eksik veriler 0 olarak kabul edildi. Analiz sırasında, eksik verilerin modelin sinir ağına iletilmesini engellemek amacıyla bir veri maskeleye işlemi kullanıldı. Bu veri maskeleye, hassas verilerin gerçekçi olmayan ancak gerçek verileri gizleyen bir yöntemdir (Ravikumar ve ark., 2011). Çalışmada eksik veriler 0 değerleriyle dolduruldu ve LSTM modellemesinde 0 değerleri filtrelemek için bir maskeleye katmanı eklenmiştir.

AK-2 veri setinde ise VÖİ-ÖS aşamasında, başlangıçta hatalı veriler ve aykırı değerler eksik veri olarak kabul edilmiş, daha sonra doğrusal interpolasyon yöntemiyle doldurulmuştur (Zhang ve Thorburn, 2022). Doğrusal interpolasyon yöntemi, eksik ve eksik olmayan değerler arasında doğrusal bir ilişki olduğunu varsayan kullanışlı bir veri doldurma yöntemidir. Doğrusal interpolasyon fonksiyonu denklem 4.1’de görüldüğü gibi ifade edilebilir (Huang, 2021).

$$f_1(x) = b_0 + b_1(x - x_0) \quad (4.1)$$

x: bağımsız değişken

x_0 : bağımsız değişkenin bilinen bir değeri

$f_1(x)$: bağımsız değişkenin bir x değeri için bağımlı değişken değeri

Denklem 1’de b_0 ve b_1 ’i tanımlamak için aşağıdaki denklem 4.2 ve 4.3 kullanılır

$$b_0 = f(x_0) \quad (4.2)$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \quad (4.3)$$

4.5.1.2. Aykırı veri analizi

Aykırı değerler, veri kümesinin genel yapısından anormal şekilde uzak olan veri noktalarıdır. Bu tür aykırı değerler, hatalı veri girişi, ölçüm hataları veya verinin özelliği gibi çeşitli nedenlerle ortaya çıkabilir. Bununla birlikte, hatalı sensör okumaları da bu veri kümesindeki aykırı değerlerin muhtemel kaynaklarından biridir.

Aykırı değer tespiti yöntemleri, değişken sayısına bağlı olarak tek değişkenli ve çok değişkenli olarak iki kategoriye ayrılabilir. Z puanı, bir veri noktasının veri kümesinin ortalama değerine göre kaç standart sapma uzaklıkta olduğunu gösteren bir ölçüdür. Bu çalışmada kullanılan Z puanı, denklem 4.4'e göre hesaplanmıştır (Dayanıklı, 2021).

$$Z = \frac{x - \bar{x}}{\sigma} \quad (4.4)$$

x: gözlem değeri

$\bar{x}$: ortalama değer

σ : standart sapma

4.5.1.3. Veri ölçeklendirme

Veri ölçeklendirme adımı, verileri daha işlevsel bir yapıya dönüştürmeyi amaçlar. Model eğitimi için kullanılan veri setlerinin öncesinde özel bir aralığa ölçeklendirilmesi genellikle tercih edilir. Bu yaklaşımda, en küçük değer 0, en büyük değer ise 1 olarak ayarlanarak, diğer tüm veri noktaları da bu [0-1] aralığına dağıtılır (Aksangür ve ark. 2022). Bu çalışmada, tüm girdi ve çıktı verileri (özellikler) [0-1] aralığına getirmek amacıyla denklem 4.5 kullanılmıştır. Bu ölçekleme işlemi, scikit-learn kütüphanesinde MinMaxScaler fonksiyonu olarak adlandırılır ve denklem 4.5'e göre hesaplanır (Guo ve ark., 2020):

$$Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (4.5)$$

Z_i = i zamanda 0 ile 1 arasında normalize edilmiş değer

x_i = x değişkeni için gözlemlenen n. değer

$\min(x)$ = veri setindeki minimum değer

$\max(x)$ = veri setindeki maksimum değer

4.5.1.4. Özellik seçimi

Bir modelin doğru ve güvenilir tahminler yapabilmesi için, temsili özelliklerin (parametrelerin) güvenilir ve tutarlı veri setlerine sahip olması gerekmektedir. Özelliklerin ve verilerin, modelin amaçlarına uygun şekilde hazırlanması gereklidir. Veri setinde modelleme kriterlerini karşılamayan özelliklerin tespit edilmesi ve çıkarılması önemlidir. Veri setlerinin düzenlenmesi model performansını artırır (Tao ve ark., 2019).

Doğru ve verimli bir modelleme için, tahmin edilen parametre üzerinde az etkisi olan bağımsız ve özdeş girdilerin çıkarılması, boyutun azaltılması açısından önemlidir. Bağımsız özellikler arasındaki ilişki, genellikle korelasyon analizi ile incelenir. Bu nedenle kullanılan özellikler arasında korelasyon analizi yapılmıştır. Korelasyon analizi, değişkenler arasındaki ilişkinin yönünü ve gücünü belirler. Yüksek korelasyon, iki veya daha fazla değişkenin güçlü bir ilişkiye sahip olduğunu, zayıf korelasyon ise değişkenlerin neredeyse hiç ilişkili olmadığını gösterir. Çoklu regresyon analizinde, bağımsız değişkenler arasında yüksek korelasyon istenmemektedir (Graham, 2003).

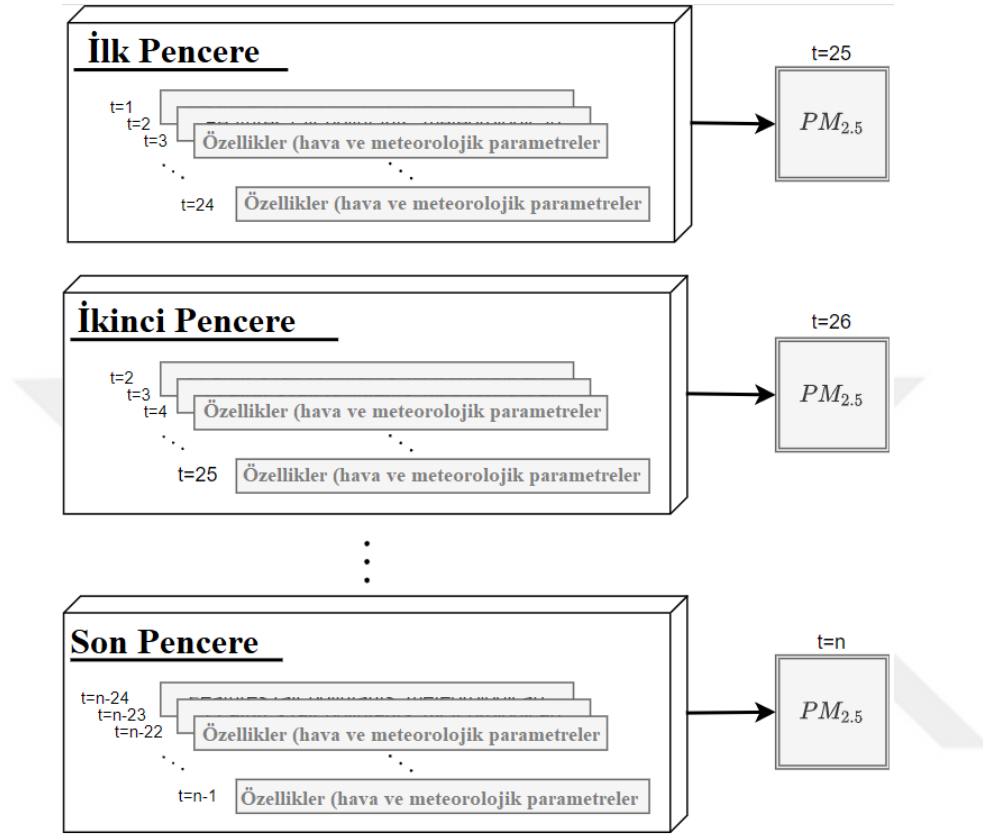
Bazı bağımsız özellikler arasında yüksek doğrusal korelasyon olduğunda, bağımlı değişken üzerindeki etkisi azalabilir. Pearson korelasyon katsayısı, özellikler arasındaki ilişkinin yönünü ve gücünü değerlendirmek için kullanılır. Bu çalışmada, özelliklerin Pearson korelasyonları hesaplanarak özellik çıkartma işlemi gerçekleştirilmiştir. Seçim kriterleri şunlardır:

- Özellikler arası korelasyon pozitif ve negatif yönde $<0,60$ olmalıdır.
- Hedef özellik ile diğer özelliklerin arasındaki korelasyon pozitif ve negatif yönde $>0,1$ olarak belirlenmiştir (Aksangür ve ark., 2022).

4.5.2. Zaman serisi çerçevesi oluşturma stratejisi

Zaman serisi tahmini, gelecekteki bilgileri öngörmek için geçmiş ve mevcut örnekleri kullanma işlemidir. Bu çalışma, denetimli öğrenme problemi olan zaman serisi çerçevesi oluşturma stratejisi üzerine odaklanmaktadır, aynı zamanda “sonraki saat tahmini” olarak da bilinir. Bu stratejinin temel amacı, tarihsel veri kümelerini ($t = 1, t = 2, \dots, t = 24$) ve sonraki saat veri kümesini ($t = 25$) oluşturmak için orijinal veri kümesini bölerek mevcut verilerden tahmin veri kümeleri oluşturmaktır. Ardından

model, gelecekteki değerleri tahmin etmek için hedef veri seti ile birleştirilebilen bir giriş zaman serisi veri seti oluşturmak üzere tüm veri kümelerini birleştirir ($t = 25, t = 26, \dots, t = n$). Bu strateji, Şekil 4.5'te özetlendiği gibi uygulanmaktadır (Eren ve ark. 2023).



Şekil 4.5. Zaman serisi çerçevesi oluşturma stratejisi (Eren ve ark. 2023)

4.5.3. Derin öğrenme uygulamaları

Geleneksel makine öğrenimi yöntemleri, ham doğal verileri işleme yetenekleri nedeniyle yetersiz kalmaktadır (LeCun ve ark., 2015). Deng (2014) makalesindeki bir tanım (tanım-2) derin öğrenmeyi, “veriler arasındaki karmaşık ilişkileri modellemek için çoklu temsil düzeylerini öğrenmeye yönelik algoritmalarla ilgili bir alt alan” olarak açıklamaktadır. Bu alt alan, daha yüksek seviyeli özelliklerin ve kavramların daha düşük seviyedeki özellikler açısından tanımlanabileceği bir özellik hiyerarşisi oluşturan derin mimariyi kullanır. Bu tür modellerin çoğunun denetimsiz temsil öğrenimine dayandığı belirtilir (Wikipedia, 2012; akt. Deng ve Yu, 2014).

Tahmin modeli geliştirme çalışmalarının temel hedefi, seçilen veri seti için güvenilir ve kabul edilebilir performans gösterecek uygun algoritmayı belirlemektir. Son

dönemde, tahmin amaçlı derin öğrenme modellerinin yükselişi, bir dizi yeni algoritmanın ortaya çıkmasına yol açmıştır. Bu nedenle, çeşitli derin öğrenme algoritmalarını birleştirmenin model tahmin performansı üzerindeki etkisinin incelenmesi önemli bir boyut olarak kabul edilebilir.

- Model Genelleme kabiliyeti

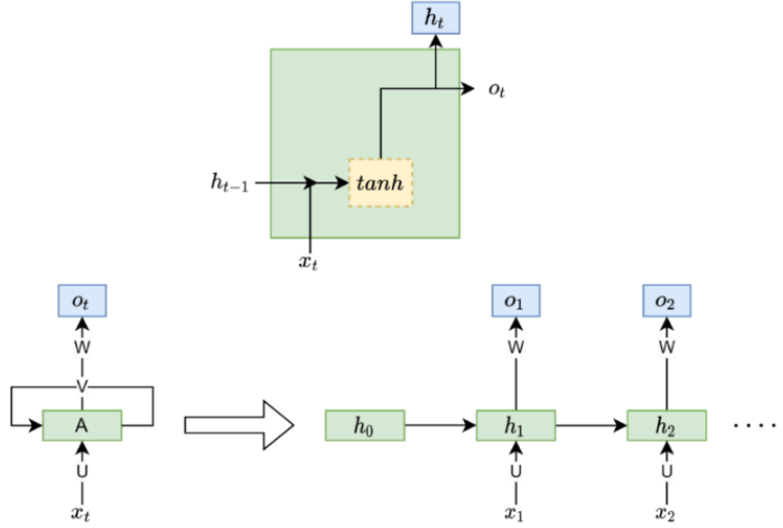
Geliştirilen modelin genellemeye açık olması, yani daha önce görmediği yeni örneklerin tahmininde ne kadar başarılı olduğu, modelin benzer özelliklere sahip farklı bölgelere ve veri setlerine uygulanabilirliği ve yaygınlaştırılabilirliği açısından kritik bir öneme sahiptir. Bu çalışmada, AK-1 ve AK-2 çalışmalarında belirlenen modellerin farklı istasyon verileriyle test edilerek genelleme yetenekleri ortaya konmuştur.

Çalışmada uygulanan RNN, LSTM ve GRU yöntemlerine ait bilgiler alt başlıklar halinde oluşturulmuştur.

4.5.3.1. Yinelemeli sinir ağlar (RNN)

Yinelemeli Sinir Ağları (RNN), insan beyni sinir bağlantılarını taklit eden ve bilgi depolamak için bir belleğe sahip olmayı amaçlayan yapay sinir ağlarının bir çeşididir. RNN'ler ses tanıma (Graves ve ark., 2013), kelime tahmini ve çevirisi (Kim ve Lee, 2016; Park ve ark., 2018; Zoph ve ark., 2016), zaman serileri (Canizo ve ark., 2019; Tokgöz ve Ünal, 2018; Walid ve Alamsyah, 2017) gibi veri noktalarının birbirine bağımlı olduğu sıralı verileri öğrenmek ve tahmin yapmak amacıyla geliştirilmiştir.

RNN'ler, çıkış değerini üretmek için önceki durumlardan gelen bilgiler de kullanılır. RNN'ler, klasik bir yapay sinir ağı gibi girdi, gizli ve çıktı katmanlarına sahiptir (Assaad ve ark., 2008). Şekil 4.6'da gösterildiği gibi, RNN'ler birbirini etkileyen ve birbirine bağlı bir zincir olarak temsil edilebilir. Her düğüm, sinir ağı hücresini tek bir zaman adımında gösterir. X , giriş özellik değerini ve X_t ise özelliğin t zamanındaki durumunu gösterir. O_t , t zamanındaki çıkış değerini, h_t ise gizli durumu gösterir.



Şekil 4.6. Temel RNN yapısı (Eren ve ark. 2023)

RNN içerisinde girdi katmanından gizli katmana, gizli katmandan ise çıktı katmanına kadar bir bağlantı vardır. Bu bağlantılar gizli durum ve çıkış ağırlık matrisleri U, W ve V olarak gösterilmektedir. Büyük/küçük değerleri uygun bir aralığa getirmek için toplam bir aktivasyon fonksiyonu (sigmoid, tanh, relu vb.) ile çarpılır. Denklem 4.6 ve 4.7’de verilmiş olan denklemler ile hesaplamalar yapılmaktadır (Pascanu ve ark., 2014).

$$h_t = \tanh(W h_{t-1} + U x_t) \quad (4.6)$$

$$y_t = \tanh(V h_t) \quad (4.7)$$

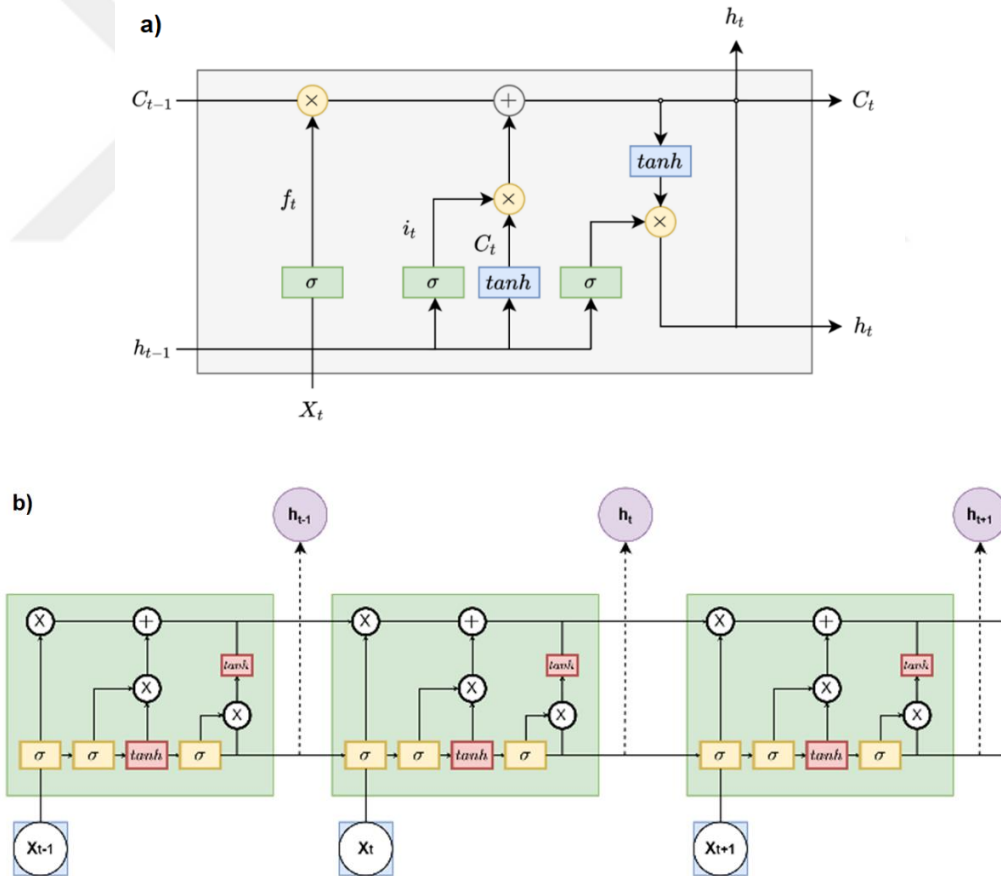
4.5.3.2. Uzun kısa süreli bellek (LSTM)

RNN uygulamalarında ağırlıkları ayarlamamızı sağlayan değer gradyan olarak ifade edilir. Uzun ve birbirine bağlı ağlarda hatanın etkisi azalır ve gradyan kaybolmaya başlayabilir. Bu probleme kaybolan gradyan problemi denir. Uzun kısa süreli bellek (LSTM) yöntemi bu sorunu çözmek için geliştirilmiş yöntemlerden biridir (Ergüder, 2019). LSTM ilk olarak 1997 yılında Sepp Hochreiter ve Jürgen Schmidhuber (Hochreiter ve Schmidhuber, 1997) tarafından ortaya atılmış ve daha sonra Gers ve ark (1999)’ tarafından geliştirilmiştir (Gers ve ark., 1999). LSTM, sıralı verilerde, yani uzun zaman adımlı zaman serilerinde kanıtlanmış başarılı bir uygulamaya sahiptir

(Cao ve ark., 2019; Chimmula ve Zhang, 2020). LSTM yapısında kaybolan gradyan problemine çözüm olarak hafıza birimine gösterilir (Greff ve ark., 2016).

LSTM, modelde uzun vadeli bağımlılıkların hatırlanmasına izin verdiği için zaman serileri analizi için kullanılır. Doğrusal olmayan bir ilişkiye sahip olan zaman serilerinde LSTM'nin performansı her zaman yüksektir (Freeman ve ark., 2018). LSTM'ler, RNN'ler gibi bellek hücreleri taşırlar (Abidogun, 2005).

LSTM'lerde RNN'lere ek olarak; giriş (input gate), unut (forget gate) ve çıkış (output gate) kapıları olmak üzere üç kapı hücresi ile bellek iyileştirilmiştir. LSTM, giriş verileri üzerinde kendi kendini eğiten bir dizi kapı ve geri besleme döngüsü kullanır. Derin öğrenme uygulamalarında σ olarak ifade edilen aktivasyon fonksiyonu için genellikle sigmoid, relu, tanh kullanılır (Freeman ve ark, 2018). Bir LSTM hücresinin yapısı Şekil 4.7a ve b'de verilmiştir.



Şekil 4.7. LSTM yapısı a)birim b) döngü (Eren ve ark. 2023).

LSTM’lerde her kapıda ayrı bir çarpım işlemi gerçekleştirilir (4.9). Öncelikle f_t denklemi ile önceki durum unutulmuş değişken ile çarpılır. Daha sonra giriş kapısında değişkenin güncellenip güncellenmeyeceği ve depolanacağı belirlenir (4.10). Son olarak, durumlar çıkış kapısına gider ve çıkış değişkeni belirlenir (4.11). Hesaplama kullanılan denklemler aşağıda verilmiştir (Gers ve ark. 1999; Colah, 2015; Eren ve ark., 2023).

$$\sigma(t) = \frac{1}{1+e^{-t}}, \tanh(t) = \left(\frac{e^t - e^{-t}}{e^t + e^{-t}} \right) \quad (4.8)$$

$$f_t = \sigma(W_f(h_{t-1}, X_t) + B_f) \quad (4.9)$$

$$i_t = \sigma(W_i(h_{t-1}, X_t) + B_i) \quad (4.10)$$

$$O_t = \sigma(W_o(h_{t-1}, X_t) + B_o) \quad (4.11)$$

AK-1 çalışmasında, Yalın, Çift yönlü ve Yığılmış olmak üzere üç LSTM varyantı kullanarak saatlik PM₁₀ konsantrasyonlarını tahmin etmek için VÖİ-ÖS’siz ve VÖİ-ÖS’li modeller geliştirdik. Üstünde çalışılan üç farklı LSTM varyantının özellikleri aşağıda kısaca özetlenmiştir.

- Yalın (vanilla) LSTM: bir tahmin yapmak için bir çıktı katmanına ve tek bir gizli LSTM birimi katmanına sahip bir LSTM modelidir (Brownlee, 2018).
- Çift yönlü (bi-directional) LSTM: LSTM modelinin giriş dizisini hem ileri hem de geri öğrenmesine izin veren ve de her iki yorumu birleştirerek sıralı tahmini sorunları için yardımcı olabilecek bir LSTM varyantıdır (Zvornicanin, 2022). Çift yönlü LSTM lere geribesleme katmanı bulunmaktadır. Graves and Schmidhuber (2005) tarafından BiRNN ve LSTM’nin birleştirilmesi ile geliştirilmiştir. Çift yönlü LSTM’ ye ait denklem (4.12-4.18)’ de verilmiştir (Yu ve ark., 2019).

$$\tilde{f}_t^L = \sigma(W_{\tilde{f}h}^L h_{t+1}^L + W_{\tilde{f}x}^L h_t^{L-1} + b_{\tilde{f}}^L) \quad (4.12)$$

$$\tilde{i}_t^L = \sigma(W_{\tilde{i}h}^L h_{t+1}^L + W_{\tilde{i}x}^L h_t^{L-1} + b_{\tilde{i}}^L) \quad (4.13)$$

$$\tilde{c}_t^L = \tanh(W_{\tilde{c}h}^L h_{t+1}^L + W_{\tilde{c}x}^L h_t^{L-1} + b_{\tilde{c}}^L) \quad (4.14)$$

$$\tilde{c}_t^L = \tilde{f}_t^L \cdot \tilde{c}_{t+1}^L + i_t^L \cdot \tilde{c}_t^L \quad (4.15)$$

$$\tilde{o}_t^L = \sigma(W_{\tilde{o}h}^L h_{t+1}^L + W_{\tilde{o}x}^L h_t^{L-1} + b_{\tilde{o}}^L) \quad (4.16)$$

$$\tilde{h}_t^L = \tilde{o}_t^L \cdot \tanh(\tilde{c}_t^L) \quad (4.17)$$

$$y_t = W_{\tilde{h}y} \vec{\tilde{h}}_t + W_{\tilde{h}y} \tilde{h}_t + b_y \quad (4.18)$$

-Yığılmış (stacked) LSTM: Derinliği ve kapasitesi az olan modellere kıyasla yığılmış LSTM modelde derinlik sağlar. Yığılmış LSTM'ler çok katmanlı ve bağlantılı bir yapıya sahiptir (Yu ve ark. 2019). Bir yığılmış LSTM mimarisi, birbirinin üzerine yerleştirilmiş birden çok gizli LSTM katmanına sahiptir. Yukarıdaki bir LSTM katmanını, aşağıdaki LSTM katmanına tek bir değer çıktısı yerine bir sıra çıktısı sağlar. Spesifik olarak, tüm girdi zamanı adımları için bir çıktı zamanı adımı yerine, girdi zamanı adımı başına bir çıktı sağlar (Brownlee, 2017). Yığılmış LSTM' ye ait denklem (4.19-4.24)' de verilmiştir (Yu ve ark., 2019).

$$f_t^L = \sigma(W_{fh}^L h_{t-1}^L + W_{fx}^L h_t^{L-1} + b_f^L) \quad (4.19)$$

$$i_t^L = \sigma(W_{ih}^L h_{t-1}^L + W_{ix}^L h_t^{L-1} + b_i^L) \quad (4.20)$$

$$\tilde{c}_t^L = \tanh(W_{\tilde{c}h}^L h_{t-1}^L + W_{\tilde{c}x}^L h_t^{L-1} + b_{\tilde{c}}^L) \quad (4.21)$$

$$c_t^L = f_t^L \cdot c_{t-1}^L + i_t^L \cdot \tilde{c}_t^L \quad (4.22)$$

$$o_t^L = \sigma(W_{oh}^L h_{t-1}^L + W_{ox}^L h_t^{L-1} + b_o^L) \quad (4.23)$$

$$h_t^L = o_t^L \cdot \tanh(c_t^L) \quad (4.24)$$

Derin öğrenme yöntemleri, modellerin hiperparametre setini seçme sorununu çözmek için birden fazla deneme çalışması yapan bir doğru parametre ayarlama işlemine ihtiyaç duyar. Derin sinir ağlarında gizli katman sayısı (number of hidden layers) ve katmandaki nöronların sayısı (neuron units) önemli hiperparametrelerdir. Ayrıca, LSTM modelleri için dikkate alınması gereken birkaç hiperparametre daha vardır. Bunlar her yineleme için işlenen eğitim seti örneklerinin miktarı olan yığın boyutu (batch size) ve optimize edici (optimizer) dir. Bu çalışmada, bir deneme-yanılma

yaklaşımı kullanarak ilk olarak 72'lik bir toplu yığın boyutu seçilmiştir. Küçük yığın boyutu, yerel minimuma yaklaşmazken, büyük yığın boyutu ise yineleme başına çok uzun süre alır. Çalışmada, daha az bellek kullanımı ve sparse parametrelerde en iyi şekilde çalıştığı için optimize edici olarak ADAM optimize edici seçilmiştir (Kingma ve Ba, 2014). Üçüncü parametre olarak sönümleme (dropout) adı verilen; fazla uydurmayı azaltmak için bir düzenleme yöntemi kullanılır. Bu yöntemde, bazı birimler afin dönüşümler kullanılarak ağdan rastgele düşürülür (Srivastava ve ark., 2014).

Çalışmada geliştirilen LSTM modelinin karara bağlı parametre seçimleri yapılmıştır. Seçilen parametre ve değerler Bölüm 5.1.2. LSTM Tahmin Sonuçları kısmında Tablo 5.2' de sunulmuştur.

4.5.3.3. Geçişli tekrarlayan birim (GRU)

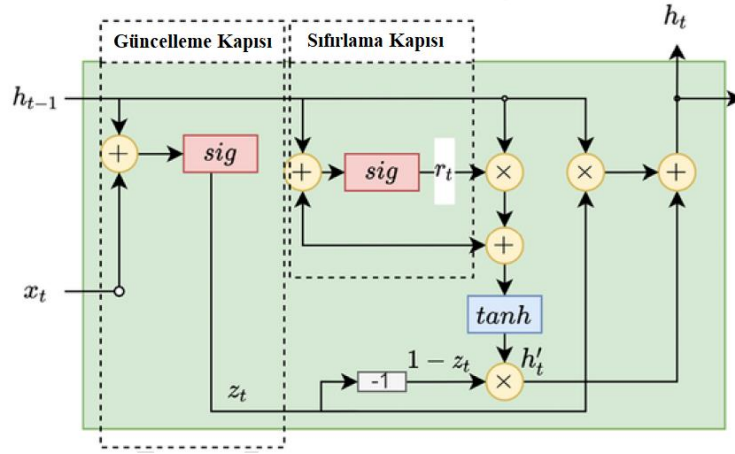
GRU birimi, LSTM'ler gibi, RNN'nin bir çeşidi olarak geliştirilmiştir. GRU'lar LSTM'lerden daha az karmaşık bir yapıya sahiptir. LSTM'lerdeki giriş, unut ve çıkış hücreleri, GRU'larda Şekil 4.8'de gösterildiği gibi çıkış (output gate), güncelleme (updategate-z) ve sıfırlama (reset gate-r) kapıları olarak güncellenmiştir. Bu kapılar sayesinde mevcut durumun geçmişteki durumdan etkilenip etkilenmeyeceği belirlenir. Sıfırlama kapısında hafızadaki giriş ile yeni girişin nasıl birleştirileceğine karar verilir (4.25). Bilginin bellekte ne kadar süre kalacağı ise güncelleme kapısı tarafından belirlenir (4.26). GRU'nun üç kapısında yapılan hesaplamalarda kullanılan denklemler aşağıda verilmiştir (Cho ve ark., 2014; Eren ve ark., 2023).

$$r = \sigma (W_r (h_{t-1}, X_t) + U_r \cdot X_t) \quad (4.25)$$

$$z = \sigma (W_z (h_{t-1}, X_t) + U_z \cdot X_t) \quad (4.26)$$

$$c = \tanh (W_c (h_{t-1} * r) + U_c \cdot X_t) \quad (4.27)$$

$$h_t = (z * c) + ((1 - z) * h_{t-1}) \quad (4.28)$$



Şekil 4.8. GRU yapısı (Eren ve ark. 2023)

4.6. Performans Değerlendirme İndeksleri

Bu tez çalışmasında, modellerin tahmin performansını değerlendirmek ve önerilen yöntemin etkinliğini belirlemek için kullanılan değerlendirme ölçütleri şunlardır: Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hatası (RMSE), Ortalama Kare Hata (MSE), Ortalama Mutlak Yüzde Hatası (MAPE) ve Belirleme Katsayısı (R^2). Seçilen ölçütlerin formülasyonu denklem 4.29-4.33 arasında sunulmuştur (Adhikari ve Agrawal, 2013; Eren ve ark., 2019).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.29)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4.30)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.31)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4.32)$$

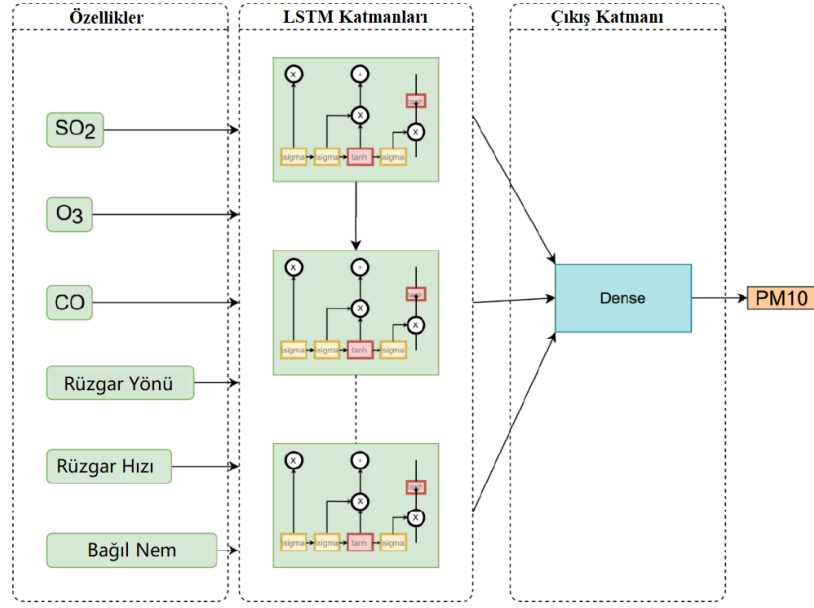
$$R^2 = \left[\frac{n \sum_{i=1}^n y_i \hat{y}_i - (\sum_{i=1}^n y_i)(\sum_{i=1}^n \hat{y}_i)}{\sqrt{[n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2][n \sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2]}} \right]^2 \quad (4.33)$$

Burada n , gözlem sayısıdır, γ_i tahmin değer ve $\hat{y}_i$ gözlem değeridir. Değerler ne kadar küçükse, modelin tahmin performansı o kadar iyi demektir. Ayrıca, R^2 , tahmin sonuçlarının gerçek verilere uygunluğunu değerlendirmek için kullanılır. Değer 1'e ne kadar yakınsa, veri uyum derecesi o kadar yüksek ve tahmin etkisi o kadar iyi olur.

4.7. Kullanılan Programlama Dili Kütüphaneleri

- Tez çalışması AK-1 için modelleme süreci aşamaları:

Modelleme için Pandas Matplotlib, Numpy, Scikit-learn ve Keras kütüphaneleri kullanılarak Jupyter Notebook kod geliştirme ortamında Python programlama dilinde veriler işlenmiştir. Matplotlib veri görselleştirme için geliştirildi. Çalışmada, tahmin ve gerçek değer arasındaki farkları görmek için kullanılmıştır. Numpy, bilimsel bir hesaplama kitaplığıdır. Sayısal olarak ilk sürüm 1995'te Jim Hugunin tarafından oluşturulmuş ve ardından 2006 yılında Travis Oliphant tarafından oluşturularak NumPy olarak yayınlanmıştır (Van Der Walt ve ark., 2011). Genellikle 2D veriler için kullanılır, görselleştirme için kullanılmıştır. Scikit-learn ise makine öğrenimi algoritmalarını içerir. Uygulamalar ve sonuçların değerlendirilmesi için kullanılmıştır. İlk olarak David Cournapeau tarafından 2007'de bir Google yaz kodu projesi olarak geliştirilmiştir (Pedregosa ve ark., 2011). Farklı kitaplıklarla uyumludur (örneğin, Numpy, SciPy, vb.). Eğitim ve test setlerini ayırmak ve verileri ölçeklendirmek için kullanılmıştır. Keras, baş yazar François Chollet (Chollet, 2015) tarafından ONEIROS (Açık Uçlu Nöro-Elektronik Akıllı Robot İşletim Sistemi) projesinin bir parçası olarak 2015 yılında geliştirilmiştir. Yapay sinir ağları ve derin yapay sinir ağları içerir. Bu çalışmada zaman serileri analizi için LSTM birimleri üzerinde öğrenme gerçekleştirilmiştir. Tez çalışması AK-1 de kullanılan LSTM tahmin modeli mimarisi, Şekil 4.9'da gösterilmektedir.



Şekil 4.9. AK-1 LSTM mimarisi (Aksangür ve ark. 2022)

- Tez çalışması AK-2 için modelleme süreci aşamaları:

Bu çalışmada derin öğrenme ve temel model uygulamaları için Keras (sürüm 2.6.0), Tensorflow arka ucu ve Scikit-learn (Abadi ve ark., 2016; Chollet 2015; Pedregosa ve ark., 2011) içeren Python paketleri kullanılmıştır. Sonuçların görselleştirilmesi için seaborn ve matplotlib (Hunter, 2007; Waskom, 2021), deneysel tasarımlar için ise papermill kullanılmıştır. Ağır iş yükleri için NVIDIA'nın Tesla T4 GPU'su ile donatılmış Google Colab'da çalıştırılmıştır. Modelleme sürecinin devamı ise, Intel(R) Core (TM) i7-5600U CPU @ 2.60GHz 2.59 GHz 12,0 GB RAM'e sahip bir bilgisayarda yürütüldü.

Model geliştirme süreci aşağıda açıklanan iki aşamada gerçekleştirilmiştir.

- Aşama 1: Gizli katman sayısı ve katmandaki nöron sayısı seçimi.

Bu aşamada $PM_{2.5}$ tahmininde kullanılan LSTM modelini oluşturmak için gizli katman sayısı ve nöron sayısı optimizasyonları yapılmıştır. Geliştirilen modelde kullanılan diğer parametreler ise öğrenme oranı, aktivasyon fonksiyonu, optimize edici (optimizer), döngü sayısı (epoch), yığın boyutu (batch size) ve kayıp fonksiyonunu içermektedir.

- Aşama 2: Farklı derin öğrenme algoritmalarının tahmin performans değerlendirmesi.

Bu aşamada GRU + GRU, LSTM+LSTM, RNN + RNN, GRU + LSTM, GRU + RNN, LSTM+RNN, LSTM+GRU, RNN + GRU, RNN + LSTM algoritmaları PM_{2.5} tahmin performansları değerlendirilmiştir. Modeller, güçlü değerlendirme sonuçları elde etmek için on kez çalıştırıldı.



5. ARAŞTIRMA BULGULARI

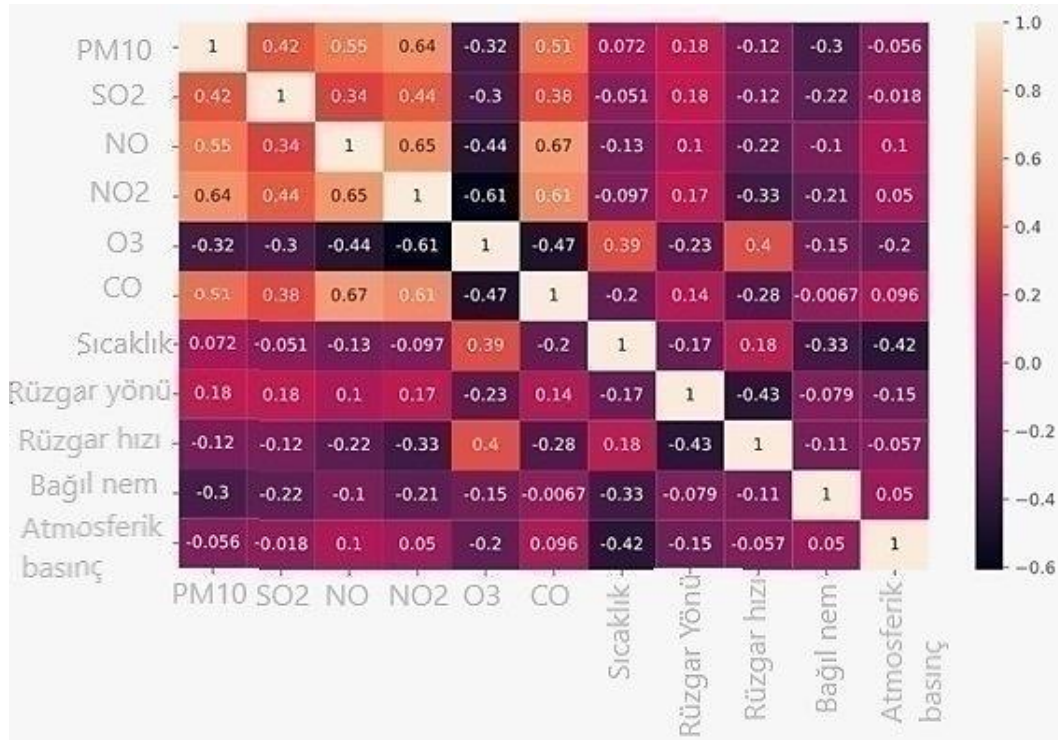
5.1. PM₁₀ Tahmininde LSTM'nin Uygunluğu

Bu bölümde, tez çalışmasının VÖİ-ÖS adımı sonuçları ve LSTM tahmin sonuçları olmak üzere iki aşamada sonuçlar elde edilmiştir.

5.1.1. Veri ön işleme ve özellik seçimi adımı sonuçları

AK-1 kapsamında yürütülen tez çalışmasının ilk aşamasında, VÖİ-ÖS'nin veri setine etkilerini içeren sonuçlar elde edilmiştir. BHKİİ veri seti, 59161 gözlem değeri ve 12 özellik içermektedir (Tablo 5.1). Ancak, veri seti eksik, hatalı ve aykırı değerler içermektedir. Bu nedenle, VÖİ-ÖS adımı kullanılarak modelleme öncesi veriler iyileştirilmiş ve hazır hale getirilmiştir. Bu aşamada uygulanan detaylı prosedürler Bölüm 4.3.1'de veri ön işleme aşamasında ayrıntılı olarak açıklanmıştır. Ayrıca, veri setinde bulunan özellikler incelenmiş ve özellik seçimi (ÖS) işlemi ile PM₁₀ konsantrasyonu tahmini için en uygun özellikler belirlenmiştir.

Bu çalışmada, özellikler arasındaki korelasyonun pozitif ve negatif yönlerde 0,60'tan küçük olması ve aynı zamanda diğer özellikler ile hedef özellik arasındaki korelasyonun pozitif ve negatif yönlerde 0,10'dan büyük olması şartları göz önünde bulundurularak özellik indirgeme işlemi gerçekleştirilmiştir. Bu aşama sayesinde, veri setindeki özellikler arasındaki çoklu bağlantı sorunu çözülmüş ve ayrıca hedef özellik ile düşük korelasyona sahip olan özellikler elenmiştir. AK-1 veri kümesinde bulunan özellikler ve aralarındaki korelasyon ilişkileri, Pearson korelasyon katsayıları kullanılarak oluşturulan bir renk skalasında Şekil 5.1'de görselleştirilmiştir.



Şekil 5.1. Özellikler arası korelasyon ısı haritası.

Şekil 5.1 incelendiğinde, birbiriyle yüksek korelasyona sahip özellikler ve hedef özellik (PM₁₀) üzerinde düşük etkiye sahip korelasyonlar veri setinden çıkarılmıştır. Örneğin, Şekil 5.1, CO, NO₂ ve NO özelliklerinin birbirleriyle yüksek korelasyona sahip olduğunu göstermektedir. Bu durumda, CO diğer iki özelliği (NO, NO₂) temsil edebilir. NO ve NO₂ özellikleri arasındaki yüksek korelasyon, çoklu bağlantı sorununa neden olmaktadır, bu nedenle bu özellikler veri setinden çıkarılmıştır. CO ve NO_x kirleticileri arasındaki yüksek korelasyon katsayısı (r) literatürde de belirtilmiştir, çünkü bu kirleticiler genellikle benzer kaynaklardan salınmaktadır (Erbaşlar ve Taşdemir 2007). Olcese ve ark. (2001), tarafından yapılan bir çalışmada, CO ve NO_x arasında $r = 0,751-0,866$ ile doğrusal bir ilişki olduğu ortaya konulmuştur (Olcese ve ark. 2001). Aynı şekilde, iki farklı noktada yapılan farklı bir çalışmada ise CO ve NO arasındaki ilişki en yüksek $r = 0,86$ ve $0,91$ iken, CO ve NO₂ arasındaki ilişki en yüksek $r = 0,76$ ve $0,79$ olarak belirlenmiştir (Erbaşlar ve Taşdemir, 2007).

Buna ek olarak, literatürde atmosferik sıcaklık ve atmosferik basınç özellikleri ile PM değişkenleri arasındaki ilişkilerin düşük korelasyon katsayılarına sahip olduğu belirtilmektedir (Tao ve ark., 2019). Bu nedenle, atmosferik sıcaklık ve atmosferik basınç özellikleri veri setinden çıkarılmıştır.

Tablo 5.1, BHKİİ verilerine ait VÖİ-ÖS'li ve VÖİ-ÖS'siz veri setleriyle ilgili temel bilgileri özetlemektedir. Tablonun ilk bölümü, ön işleme adımı öncesi veri setini göstermektedir. Bu aşamada eksik veri noktaları "mevcut olmayan veri" olarak kaydedilmiştir. Sıfır ve negatif değerler, hava kalitesi izleme istasyonlarında meydana gelen makine arızaları veya rutin bakımlar nedeniyle oluşan kayıp verileri temsil etmektedir. Negatif ve sıfır okumalar, olası yanlış veya hatalı sensör okumaları olarak kabul edilmiştir. Bu şekilde, "mevcut olmayan veri" ile sıfır ve negatif değerler toplamda eksik verileri oluşturmaktadır. Tablonun ikinci bölümü ise VÖİ-ÖS sonrası veri seti hakkında bilgi içermektedir. Bu aşamada dörtten az ölçüm içeren gözlemler de kaldırılmıştır.

BHKİİ veri seti aykırı değerler içerebilir ve bunların tespiti için Z skoru yöntemi kullanılmıştır. Z puanı, bir veri noktasının ortalama değerden ne kadar uzakta olduğunu gösteren bir ölçüdür. Bu yöntem, veri kümesi için normal dağılım varsayımını gerektirir. Bu çalışmada, veri setinin normal dağılımı sağladığı kabul edilerek PM_{10} değişkeninin p-değeri=1 <0,05 olan Shapiro-Wilk testi uygulanmıştır. Bu test sonucunda elde edilen aykırı değerler, Z skoru +/- 3 dağılım aralığında olan veriler olarak belirlenmiştir. Ardından bu aykırı değerler veri kümesinden çıkarılmıştır.

Tablo 5.1. VÖİ-ÖS öncesi ve sonrası her bir hava kirliliği ve meteorolojik parametrenin veri setlerine ait tanımlayıcı istatistikler.

Özellikler		Hava kirliliği parametreleri							Meteorolojik parametreler				
		PM ₁₀ (µg/m ³)	SO ₂ (µg/m ³)	O ₃ (µg/m ³)	CO (µg/m ³)	NO (µg/m ³)	NO ₂ (µg/m ³)	NO _x (µg/m ³)	Bağıl nem (%)	Rüzgar yönü	Rüzgar hızı (m/s)	Atmosferik sıcaklık (°C)	Atmosferik basınç (mbar)
Veri ön işleme ve öz nitelik seçimi öncesi (VÖİ-ÖS' siz)	#Mevcut veri	56178	55783	55555	56175	52307	56045	50103	47275	56076	56942	56896	56955
	# Mevcut olmayan veri	2983	3378	3606	2986	6854	3116	9058	11886	3085	2219	2265	2206
	# O ve negatif veri	585	248	23	16	79	0	9390	278	283	281	0	278
	% Eksik veri*	6,03	6,13	6,13	5,07	11,7	5,27	31,2	20,5	5,69	4,23	3,83	4,20
	Ort.	55,5	5,53	56,9	504	13,5	32	46	82	118	2,62	15,5	994
	Std	43,4	8,23	32,1	368	29,2	25,4	67,3	17,3	96,6	1,55	7,54	70
	Min	0	-59,7	-2,86	-215	-5,15	0,16	-4,98	0	0	0	-6,2	0
	25%	29,7	1,9	33,9	300	3,2	14,4	9,1	71,7	34,4	1,4	9,5	995
	50%	44,3	3,37	59,3	400	3,84	24	26,6	87	70,2	2,3	15,7	998
	75%	66,4	6,51	79	570	9,98	41,4	51,9	94,5	198	3,55	21,8	1003
	Maks.	865	230	1310	20710	551	260	1102	100	360	12,9	36,2	1022
Veri ön işleme ve öz nitelik seçimi sonrası (VÖL-ÖSli)	# Veri	54454	54597	55373	55123	Çıkarılan özellikler			46671	55792	56170	Çıkarılan özellikler	
	Ort.	52,3	5,07	56,6	473				82,7	118	2,58		
	Std	32,6	4,93	30,1	257				15,6	96,5	1,46		
	Min	0,69	0,01	0,01	0,01				33,9	0,1	0,08		
	25%	29,9	1,9	33,9	300				72,4	34,7	1,41		
	50%	44	3,3	59,2	400				87,4	71,5	2,3		
	75%	64,7	6,3	79	550				94,7	198	3,5		
	Maks.	186	28,6	153	1606				100	360	7,26		

* Kayıp veri yüzdesi, gözlem süresi dikkate alınarak her bir parametre için 59161 saat gözlem değeri kullanılarak hesaplanmıştır.

5.1.2. LSTM tahmin sonuçları

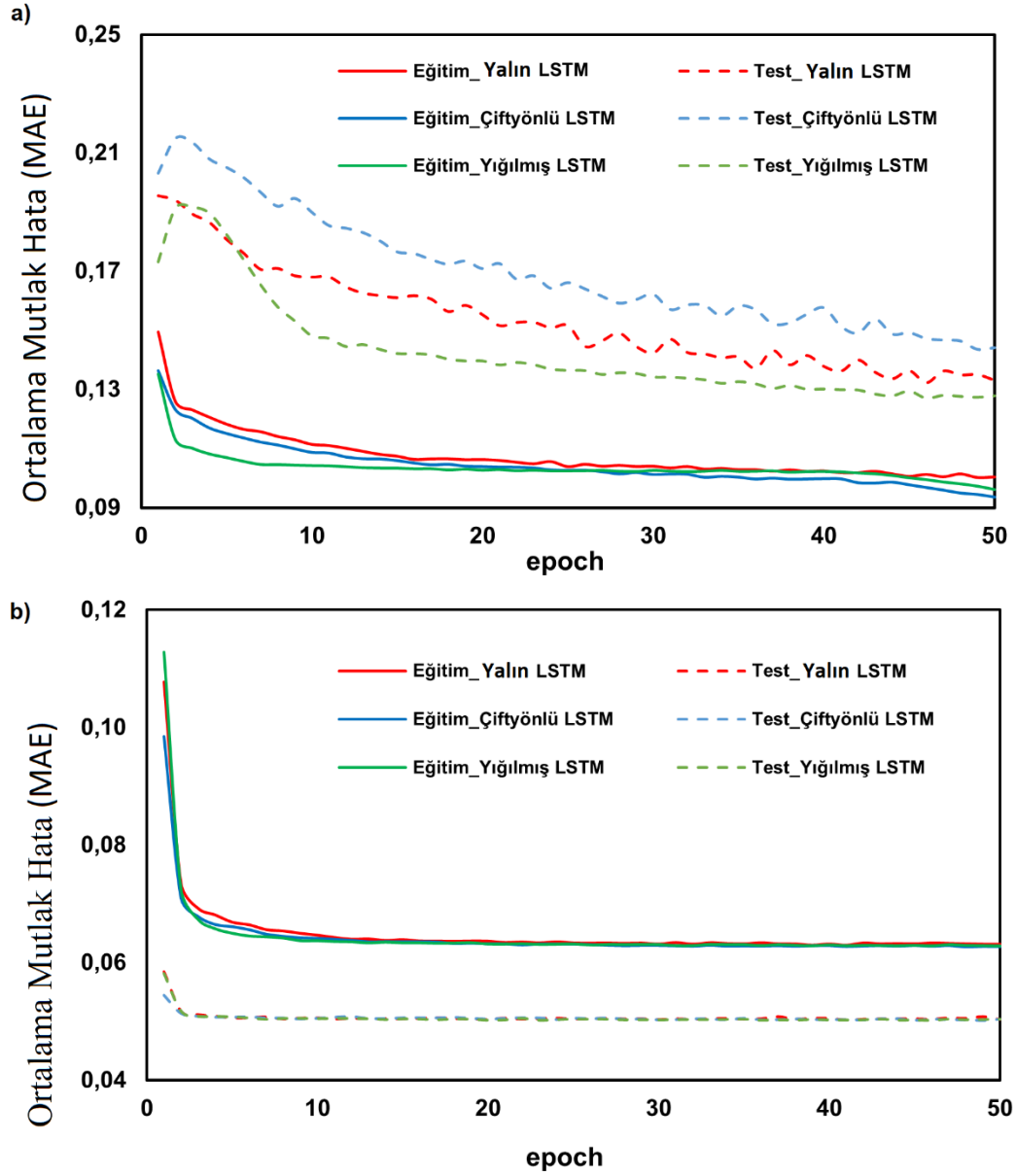
AK-1 kapsamında yapılan tez çalışmasının ikinci aşamasında, VÖİ-ÖS'li ve VÖİ-ÖS'siz veri setlerinin derin öğrenme algoritmasına etkileri incelenmiştir. Bu çalışmada analiz için BHKİİ veri seti kullanılmıştır. Analiz safhasında, PM_{10} değerlerinin önceki döngülerinden elde edilen veriler, özellik grubuna dahil edilmiştir. Bu şekilde, önceki veriler LSTM modeline girdi olarak sunulmuştur. BHKİİ veri seti analiz için hazır hale getirildikten sonra, eğitim ve test setleri olarak ayrılmıştır. Bu bağlamda, ilk üç yılın verileri eğitim için kullanılırken, geri kalan veri seti modelin test edilmesinde kullanılmıştır. LSTM modeli, toplamda 50 birimli tek LSTM katmanı ve bir çıktı katmanından oluşmaktadır. LSTM modelinin eğitiminde, kayıp fonksiyonunu en aza indirmek için Adaptif Moment Tahmini (Adam) optimize edici yöntemi kullanılmıştır. Tüm bu ayrıntılar Tablo 5.2'de yer alan LSTM modeli parametreleri olarak sunulmuştur.

Tablo 5.2. Geliştirilen LSTM modelleri karar bağı parametreleri.

Modeller	Parameterler
Yalın LSMT	50 birimli tek LSTM katmanı, aktivasyon fonksiyonu = tanh, dropout oranı = 0,2, epoch = 50, yığın boyutu = 72, kayıp fonksiyonu = MAE, optimize edici= ADAM.
Çift yönlü LSTM	50 birimli tekli BiLSTM katmanı, aktivasyon fonksiyonu = tanh, dropout oranı = 0,2, epoch = 50, yığın boyutu = 72, kayıp fonksiyonu = MAE, optimize edici = ADAM.
Yığılmış LSTM	Her katmanı 50 birim olan iki gizli katman, aktivasyon fonksiyonu = tanh, dropout oranı=0,2, epoch = 50, yığın boyutu = 72, kayıp fonksiyon = MAE, optimize edici = ADAM.

Şekil 5.2'de, VÖİ-ÖS'siz ve VÖİ-ÖS'li LSTM modellerinin üç farklı varyantının tahmin performansları eğitim ve test süreçleri için hata oranları (MAE) açısından karşılaştırılmıştır. VÖİ-ÖS prosedürü uygulanmadığında, tüm LSTM modellerinin eğitim seti için MAE değerleri 50 epoch sonrasında yaklaşık 0,1 seviyesindeyken (Şekil 5.2a), VÖİ-ÖS uygulandığında bu değerlerin 0,06 seviyesine düştüğü

gözlemlenmiştir (Şekil 5.2b). Sonuç olarak, modellerin tahmin doğruluğu, VÖİ-ÖS yöntemi kullanılarak önemli ölçüde artırılmıştır.

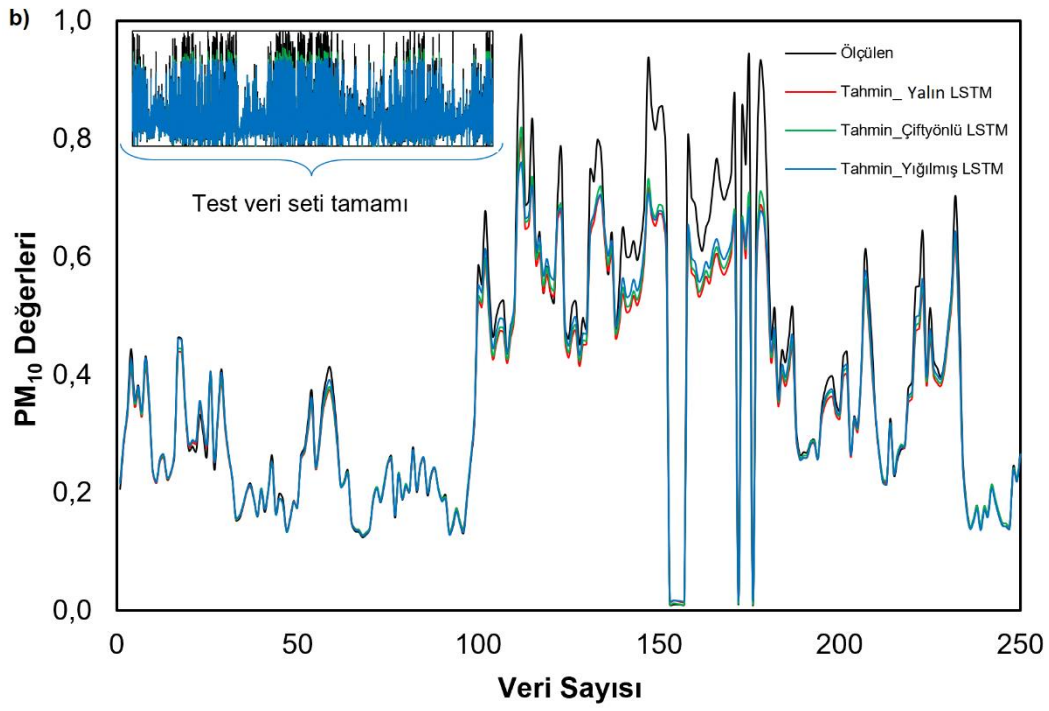
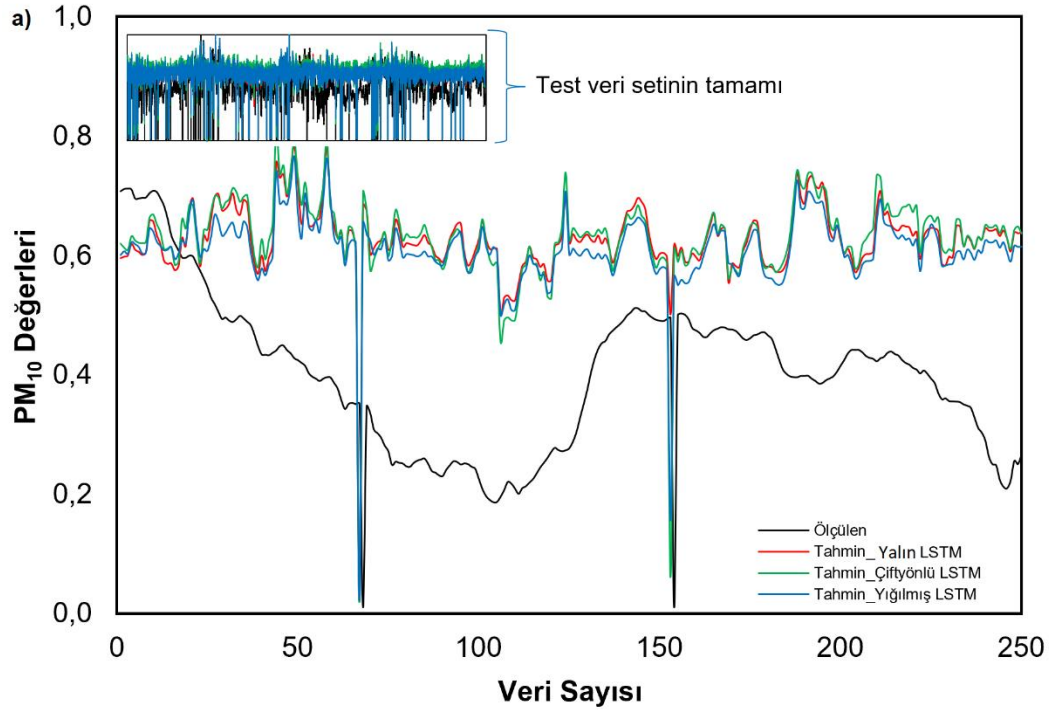


Şekil 5.2. Model tahmin performansı a) VÖİ-ÖS'siz LSTM b) VÖİ-ÖS'li LSTM.

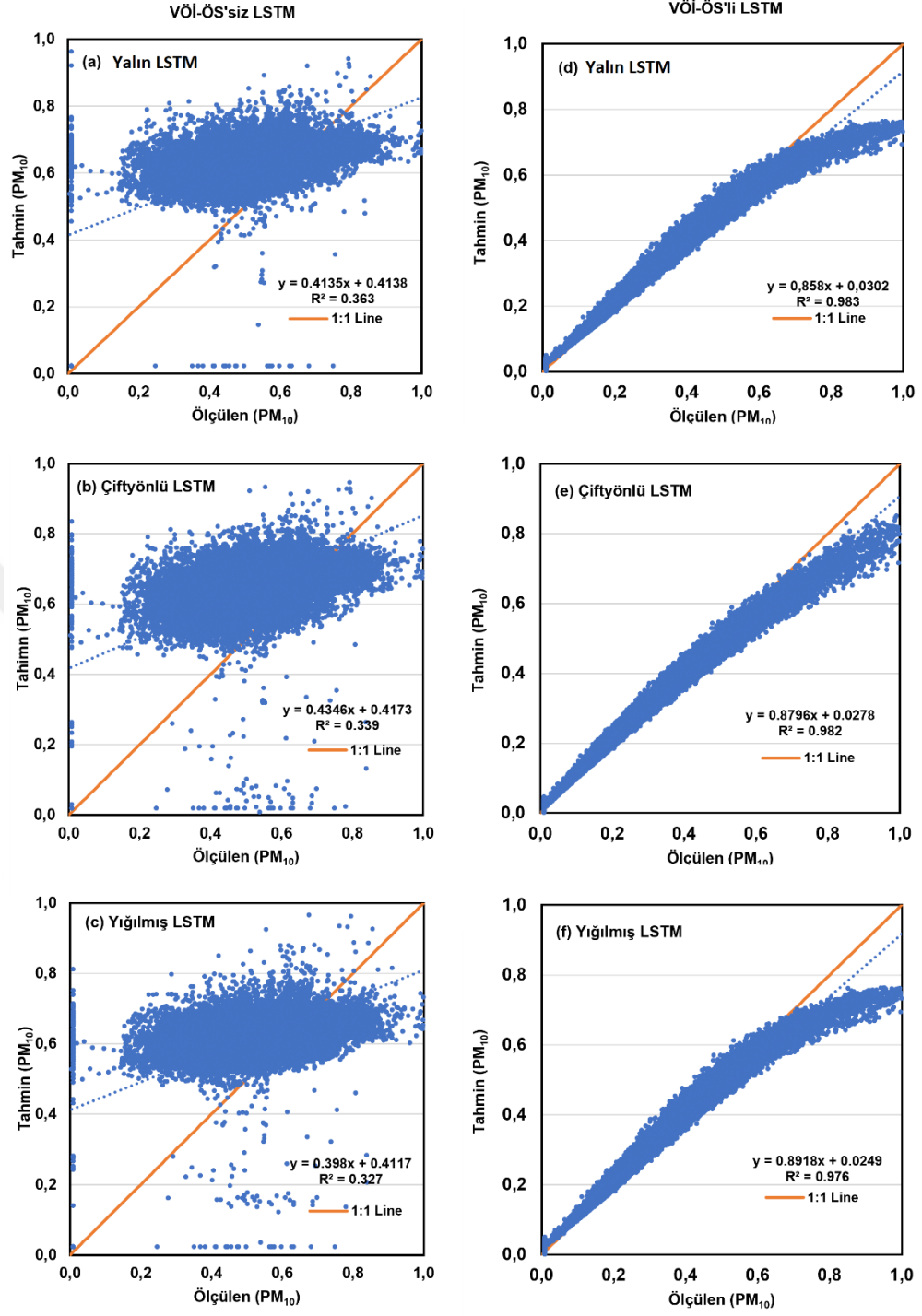
Şekil 5.3'te, PM_{10} model tahminleri VÖİ-ÖS'siz ve VÖİ-ÖS'li üç farklı LSTM varyantı için ölçülen değerlerle karşılaştırılmıştır. Grafik üzerinde büyük miktarda veri olduğundan, okunabilirlik açısından sadece ilk 250 sonuç görüntülenmiştir. Şekil 5.3a, VÖİ-ÖS'siz üç LSTM modelinin tahmin değerlerini ölçülen değerlerle karşılaştırmaktadır. Öte yandan, Şekil 5.3b VÖİ-ÖS'li üç LSTM modelinin tahmin değerlerinin ölçülen değerlerle iyi bir uyum içinde olduğunu göstermektedir. PM_{10} 'un

model tahminlerini ve ölçülen değerlerin dağılımı Şekil 5.4 (Şekil 5.4a-c: VÖİ-ÖS'siz, Şekil 5.4d-f: VÖİ-ÖS'li) olarak sunulmaktadır. Tüm modellerin tahmin edilen ve ölçülen değerlerinin birbirine yakın olması, başarılı bir tahmin performansına sahip olduklarını göstermektedir. VÖİ-ÖS adımının, modellerin belirleme katsayısı (R^2) değerlerini yükselttiği gözlemlenmiştir. Bu durum, VÖİ-ÖS adımının modellerin tahmin performansını artırdığını göstermektedir. Aynı yönde, VÖİ-ÖS'li ve VÖİ-ÖS'siz üç LSTM modelinin RMSE, MAE ve R^2 değerlerini Şekil 5.5'te sunmaktayız. Bu grafik, tüm VÖİ-ÖS'li LSTM modellerinin RMSE ve MAE hata ölçütleri açısından önemli bir iyileşme sağladığını vurgulamaktadır. Ayrıca, tahmin edilen ve gözlemlenen PM_{10} değerleri arasındaki artık hatalar en yüksek PM_{10} değerlerinde meydana geldiği Şekil 5.6'da gösterilmiştir.

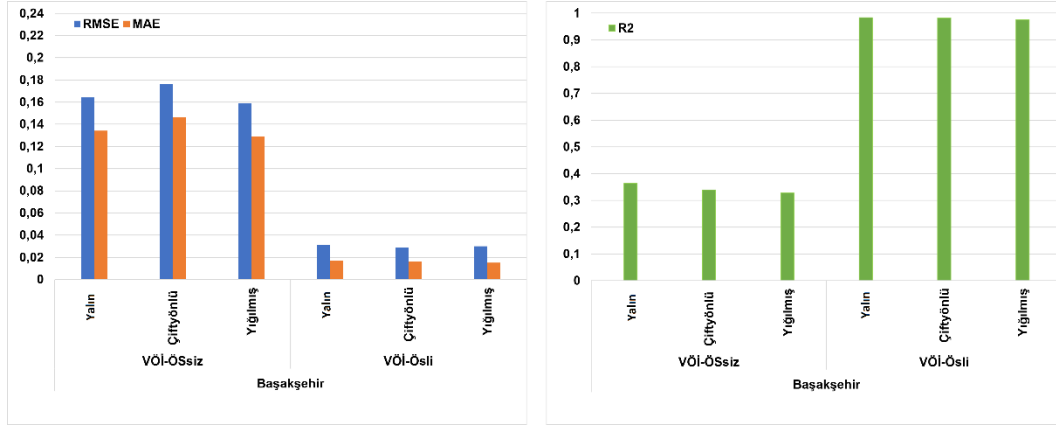




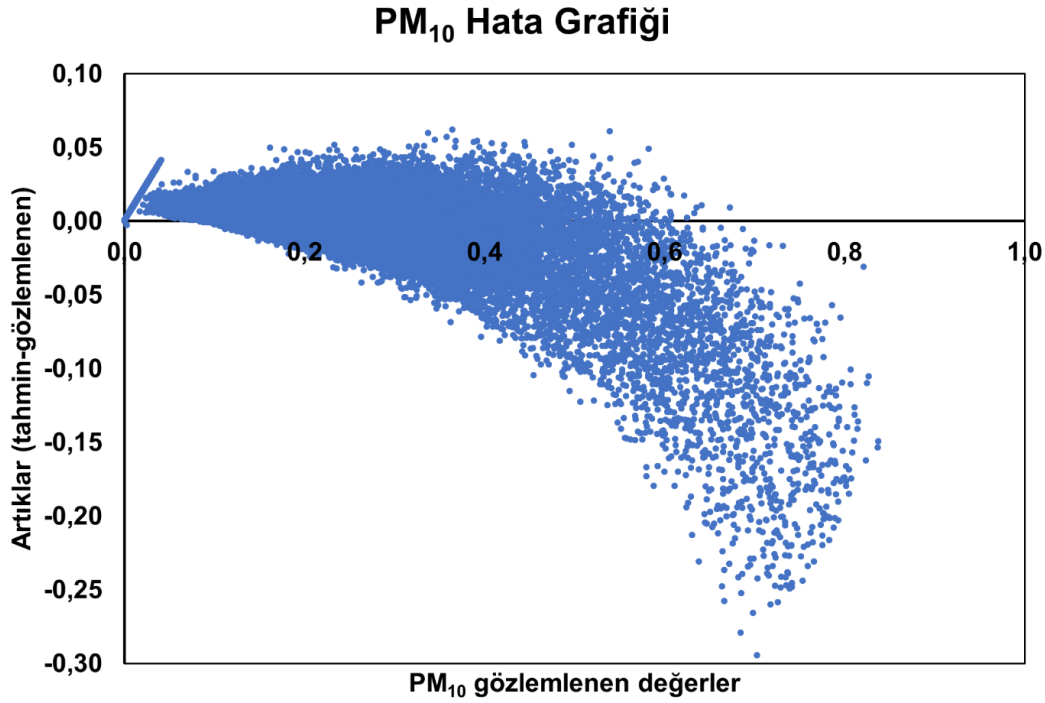
Şekil 5.3. PM₁₀ için ölçülen ve tahmin edilen değerlerin karşılaştırılması a) VÖİ-ÖS'siz LSTM modelleri b) VÖİ-ÖS'li LSTM modelleri.



Şekil 5.4. Model tahminleri ve ölçülen değerler arasındaki dağılım grafikleri a-c)VÖİ ÖS'siz LSTM modelleri d-f)VÖİ-ÖS'li LSTM modelleri.



Şekil 5.5.VÖİ-ÖS'siz ve VÖİ-ÖS'li LSTM modellerinin RMSE, MAE ve R^2 değerlerinin çubuk grafiği.



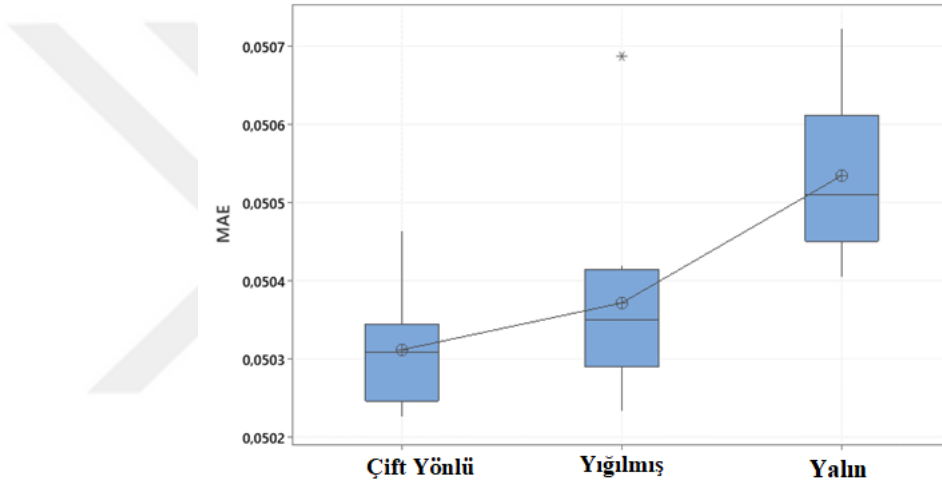
Şekil 5.6. En iyi LSTM modelinin ilgili hatalarının gözlemlenen PM₁₀ değerlerine karşı çizilmiş dağılım grafiği.

- İstatistiksel Değerlendirme

Bu kısım, önerilen LSTM modellerinin performansını değerlendirmek amacıyla yapılan istatistiksel karşılaştırmaları içermektedir. VÖİ-ÖS'li LSTM modellerinin tahmin performanslarının, VÖİ-ÖS'siz LSTM modellerinden daha üstün olduğu sonucuna ulaşılmıştır. Bu nedenle, önerilen LSTM modelleri ile modeller arasındaki

MAE değerlerinin daha ayrıntılı bir analizini gerçekleştirmek adına 30 farklı rasgele deney yapılmıştır.

Şekil 5.7, 30 farklı çalıştırmada elde edilen test verileri için MAE değerlerinin kutu grafiğini sunmaktadır. Kutu grafikleri, veri kümesinin dağılımı ve aykırı değerler hakkında bilgi sağlar. Şekil 5.7'den görüleceği üzere, Yalın LSTM modelinin MAE değerlerinin medyan çizgisi daha yüksektir, bu da en düşük performansı sergilediğini gösterir. Yığılmış ve çift yönlü LSTM modelleri ise daha benzer performanslar elde etmiştir. Elde edilen sonuçlara göre, medyan değerler Çift Yönlü LSTM modelinde 0,05031, Yığılmış LSTM modelinde 0,05037 ve Yalın LSTM modelinde 0,05053 olarak gözlenmiştir.



Şekil 5.7. Çoklu bağımsız çalışmalarda önerilen LSTM modelleri için MAE performans değerlerinin kutu grafiği.

Önerilen LSTM modellerinin performans farkını değerlendirmek amacıyla tek yönlü bir ANOVA testi uygulanmıştır. ANOVA testinin sonuçları, Tablo 5.3'te sunulmaktadır. Bu sonuçlar, $\alpha = 0,05$ anlamlılık düzeyinde sıfır hipotezini ($p = 0,001$) reddetmek için yeterli kanıt sağlamaktadır. Bu da, en az bir modelin, diğer modellerden önemli ölçüde farklı bir performans sergilediği anlamına gelir. Modellerin performanslarının istatistiksel olarak anlamlı olup olmadığını belirlemek için %95 güven düzeyinde Tukey ve Fisher ikili karşılaştırmaları gerçekleştirilmiştir. Genel sonuçlar Tablo 5.4'te sunulmaktadır. Elde edilen sonuçlara göre, Yalın LSTM modeli A grubunda yer alırken diğer modeller B grubunda yer almaktadır. Bu, Yalın LSTM modelinin diğer iki modele istatistiksel olarak farklı olduğunu göstermektedir.

Tablo 5.3. Geliştirilmiş LSTM modelleri tek yönlü ANOVA sonuçları.

Metot

Null hipotez

Hepsi eşit

Alternative hipotez

Hepsi eşit değil

Anlamlılık düzeyi

$\alpha = 0,05$

Analiz için eşit varyanslar varsayılmıştır.

Faktör Bilgisi

Faktor

Düzeyler

Değerler

Model

3

Çift yönlü ; Yığılmış; Yalın

Varyans Analizi

Kaynak

DF (Serbestlik derecesi)

Adj SS

Adj MS

F-Değer

P-Değer

Model

2

0,000001

0,000000

52,35

0,000

Hata

87

0,000001

0,000000

Toplam

89

0,000001

Model Özeti

S

R-sq

R-sq(adj)

R-sq(pred)

0,0000859

54,62%

53,57%

51,43%

Ortalamalar

Model

N

Ort.

StDev

95% CI

Çift Yönlü

30

0,050309

0,000054

(0,050278; 0,050340)

Yığılmış

30

0,050325

0,000095

(0,050293; 0,050356)

Yalın

30

0,050513

0,000101

(0,050482; 0,050544)

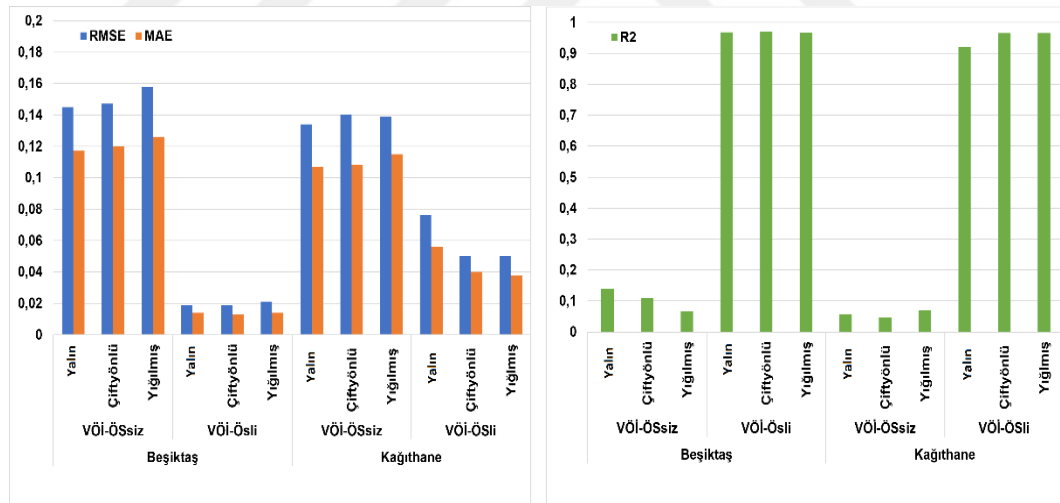
Pooled StDev = 0.00392617

Tablo 5.4. Tukey ve Fisher test gruplama bilgileri.

Tukey İkili Karşılaştırma				Fisher İkili Karşılaştırma			
Model	N	Ort.	Gruplama	Model	N	Ort.	Gruplama
Yalın	30	0,050513	A	Yalın	30	0,050513	A
Yığılmış	30	0,050325	B	Yığılmış	30	0,050325	B
Çift Yönlü	30	0,050309	B	Çift Yönlü	30	0,050309	B

- Önerilen modelin farklı bölgeler için uygulanması:

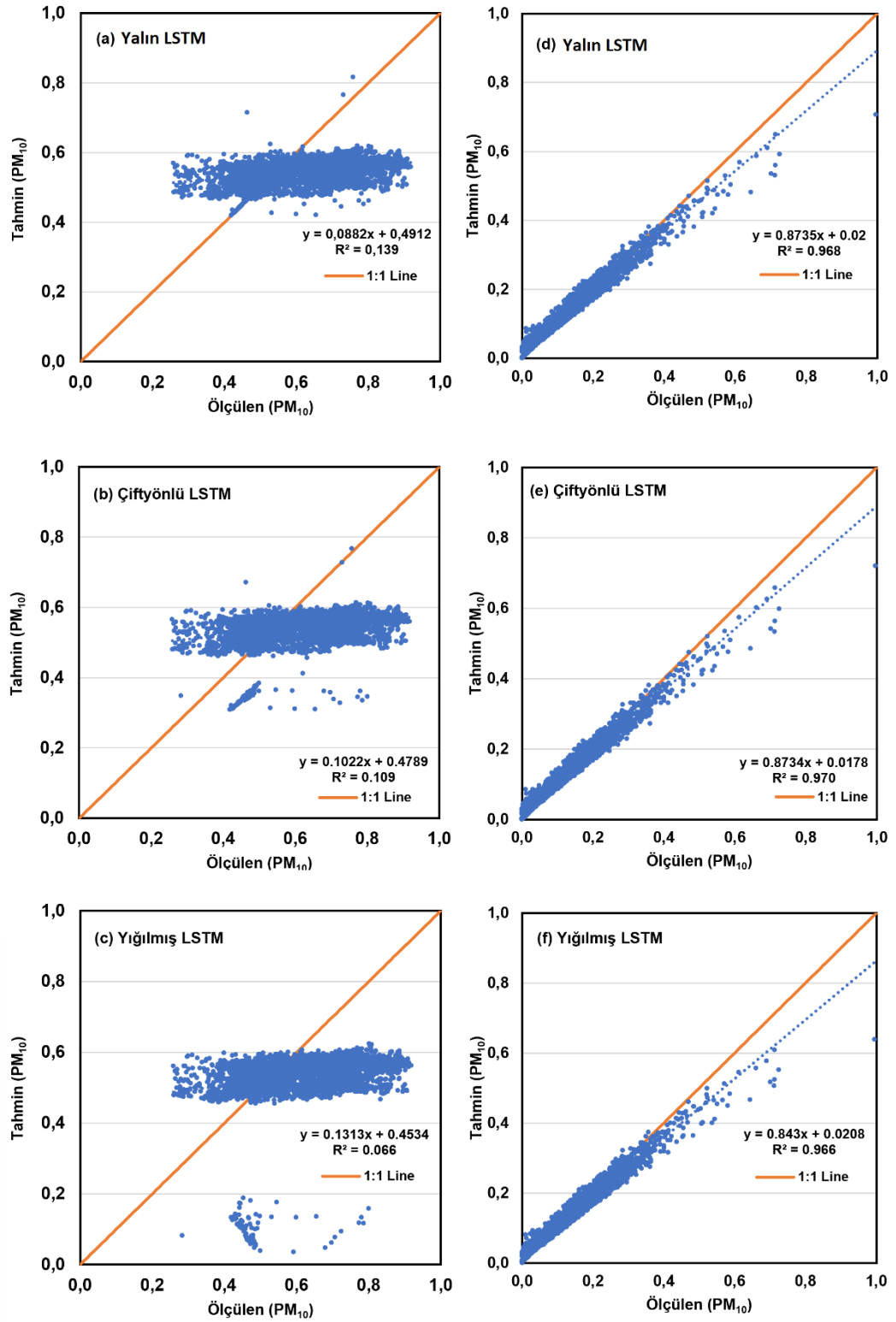
Bu bölümde, önerilen modelin genelleme yeteneği, iki farklı bölgenin veri setlerini kullanarak değerlendirilmiştir. Beşiktaş ve Kağıthane Hava Kalitesi İzleme İstasyonları'ndan (SİM, 2019) hava kalitesi veri setleri ve Devlet Meteoroloji Genel Müdürlüğü'nden (MGM, 2020) temin edilen meteorolojik veri setleri bu amaçla kullanılmıştır. Elde edilen LSTM model performans sonuçları Şekil 5.8'de sunulmuştur. Şekil 5.8'den görüleceği üzere, LSTM modellerinde VÖİ-ÖS kullanımının tahmin performansını önemli ölçüde artırdığı gözlenmektedir.



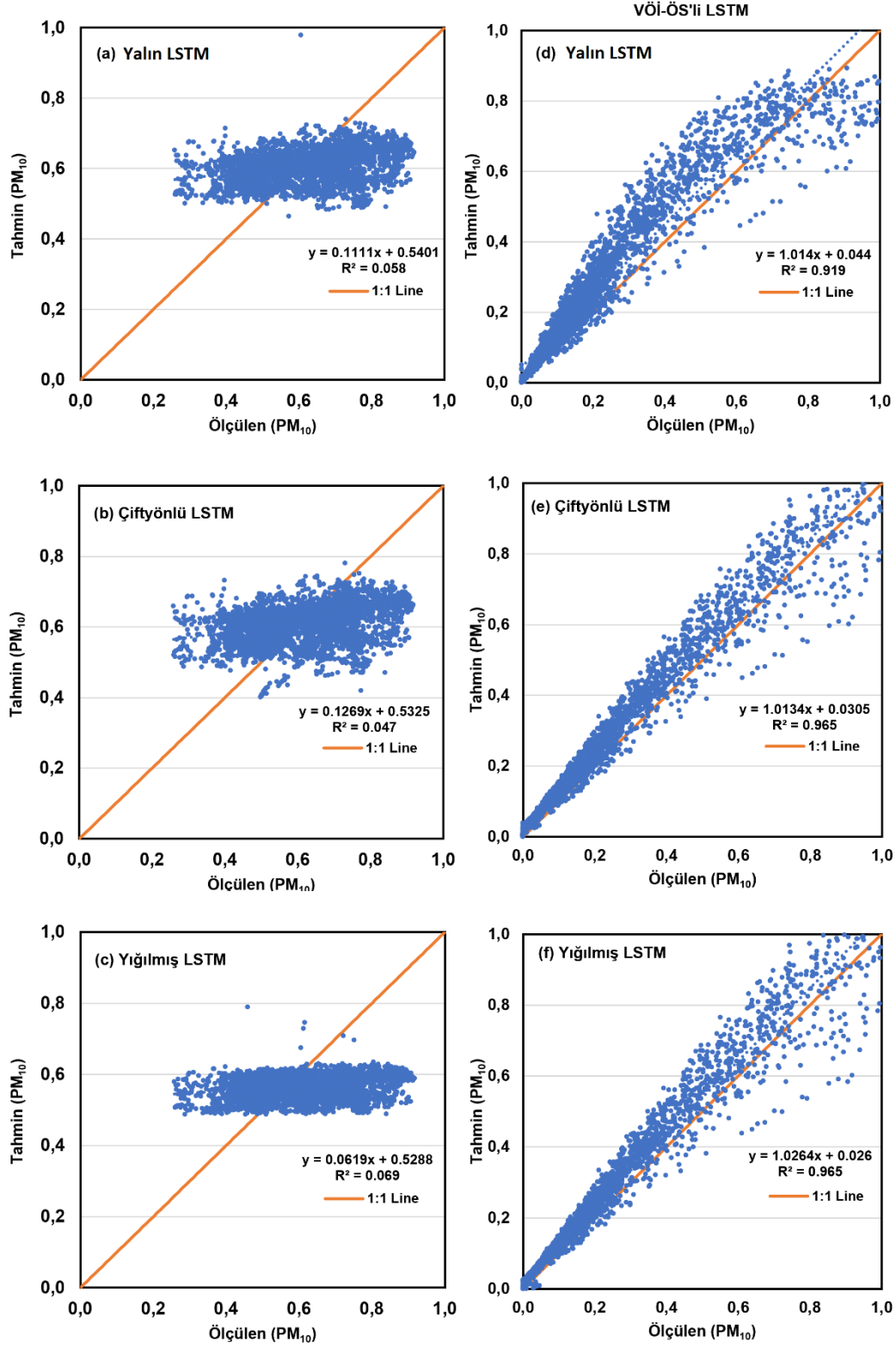
Şekil 5.8. Beşiktaş ve Kağıthane bölgeleri için VÖİ-ÖS'siz ve VÖİ-ÖS'li LSTM modellerinin RMSE, MAE ve R^2 değerlerinin çubuk grafikleri.

Şekil 5.9 ve 5.10'da, Beşiktaş ve Kağıthane bölgeleri için sırasıyla tahmin edilen ve ölçülen değerlerin dağılımını gösteren grafikler sunulmaktadır. Bu grafiklerden görüleceği üzere, LSTM modellerinde VÖİ-ÖS kullanılmadığında tahmin edilen ve ölçülen değerler arasındaki uyum zayıfken, VÖİ-ÖS kullanıldığında ise mükemmel bir uyum gözlenmektedir. Sonuç olarak, VÖİ-ÖS ile önerilen LSTM modelleri farklı bölgeler için olağanüstü tahmin performansına sahiptir.





Şekil 5.9. Beşiktaş bölgesi için model tahmini ve ölçülen değerler arasındaki dağılım a-c) VÖİ-ÖS'siz LSTM modelleri d-f) VÖİ-ÖS'li LSTM modelleri.



Şekil 5.10. Kağıthane bölgesi için model tahminleri ile ölçülen değerler arasındaki dağılım grafikleri a-c)VÖİ-ÖS'siz LSTM modelleri d-f)VÖİ-ÖS'li LSTM modelleri.

5.2. PM_{2.5} Tahmininde DÖ Modellerinin Performanslarının Karşılaştırılması

Tez çalışması, AK-2 kapsamında derin öğrenme algoritmalarını kullanarak saatlik PM_{2.5} konsantrasyonunu tahmin etmek amacıyla iki aşamalı bir modelleme yaklaşımıyla ilerlemiştir. Bu bölümde, LSTM modeli için yapılan hiperparametre belirleme sonuçları (gizli katman sayısı ve katmandaki nöron sayısı seçimi) ile farklı derin öğrenme algoritmalarının tahmin performansını değerlendirme sonuçları olmak üzere iki ayrı aşamada sonuçlar elde edilmiştir. Geliştirilen tüm modellerin performans sonuçları Tablo 5.5'te sunulmuştur.

5.2.1. Gizli katman sayısı ve katmandaki nöron sayısı seçimi

Tez çalışmasının AK-2 kapsamında 1. aşamada, gizli katman sayısı ve katmandaki nöron sayısının seçimi ile ilgili çalışmaların sonuçları elde edilmiştir. Bu çalışmada, model farklı gizli katman sayıları (1/2/4) ve katmandaki nöron sayıları (5/10/20/30/40) ile test edilmiştir. Modelin diğer kullanılan parametreleri şunlardır: öğrenme oranı: 0,01, aktivasyon fonksiyonu: ReLU, optimize edici: ADAM, epoch: 50, yığın boyutu: 32 ve kayıp fonksiyonu: MAE olarak belirlenmiştir (Eren ve ark., 2023). Aktivasyon fonksiyonu olarak çabuk işlem hızı ve daha çok tercih edildiği için ReLU (Doğrultulmuş Doğrusal Birim) fonksiyonu tercih edilmiştir (Szandała, 2021). Güçlü sonuçlar elde etmek için tüm LSTM modelleri on kez çalıştırılarak ortalama performans değerleri hesaplanmış ve Tablo 5.5'in ilk bölümünde sunulmuştur.

LSTM model performanslarının değerlendirilmesi sonucunda, iki gizli katman ve her bir katmanda 30 nöron bulunan Model 9'un en iyi performansı gösterdiği belirlenmiştir. Model 9, bu performans değerlerine göre saatlik PM_{2.5} konsantrasyonlarını doğru bir şekilde tahmin etmiştir. Bu bağlamda, MSE=24,95 ve 25,20; RMSE=4,771 ve 4,818; MAE=3,099 ve 3,129; MAPE=0,13 ve 0,132 ve R²=0,976 ve 0,966 değerleri ile en yüksek performansı sergilemiştir.

Tablo 5.5. PM_{2.5} tahmininde model performans sonuçları.

Modeller	Gizli Katman	Nöron sayısı	Algoritma	Eğitim					Test					
				MSE	RMSE	MAE	MAPE	R ²	MSE	RMSE	MAE	MAPE	R ²	
Aşama 1	Model#1	1	5	LSTM	27,45	5,002	3,228	0,138	0,956	28,19	5,137	3,315	0,142	0,930
	Model#2	1	10	LSTM	27,04	4,966	3,221	0,138	0,956	27,89	5,122	3,323	0,142	0,926
	Model#3	1	20	LSTM	30,79	5,262	3,405	0,154	0,935	31,56	5,394	3,490	0,158	0,912
	Model#4	1	30	LSTM	45,93	8,920	4,656	0,201	0,904	33,16	5,160	3,581	0,201	0,880
	Model#5	1	40	LSTM	30,13	5,196	3,395	0,147	0,944	31,19	5,379	3,514	0,153	0,911
	Model#6	2	5	LSTM	46,94	9,008	4,693	0,202	0,896	48,15	9,240	4,815	0,207	0,873
	Model#7	2	10	LSTM	26,93	4,941	3,219	0,135	0,963	27,27	5,003	3,260	0,136	0,951
	Model#8	2	20	LSTM	25,92	4,902	3,136	0,133	0,973	26,11	4,937	3,159	0,134	0,965
	Model#9	2	30	LSTM	24,95	4,771	3,099	0,130	0,976	25,20	4,818	3,129	0,132	0,966
	Model#10	2	40	LSTM	66,44	13,02	6,156	0,270	0,838	67,40	13,21	6,245	0,274	0,826
	Model#11	4	5	LSTM	168,4	33,49	13,77	0,617	0,515	169,1	33,64	13,84	0,620	0,512
	Model#12	4	10	LSTM	127,0	25,25	10,70	0,476	0,647	127,0	25,24	10,70	0,476	0,647
	Model#13	4	20	LSTM	187,3	37,45	15,21	0,682	0,461	184,8	36,94	15,00	0,673	0,468
	Model#14	4	30	LSTM	167,4	33,40	13,72	0,614	0,522	166,9	33,29	13,67	0,612	0,523
	Model#15	4	40	LSTM	206,0	41,14	16,59	0,745	0,397	204,2	40,78	16,45	0,739	0,401
Aşama 2	Model#16	2	30	GRU+ GRU	25,81	4,853	3,132	0,133	0,975	26,05	4,900	3,162	0,134	0,965
	Model#17	2	30	LSTM+LSTM	24,60	4,739	3,056	0,129	0,980	24,84	4,785	3,086	0,130	0,970
	Model#18	2	30	RNN+RNN	31,99	5,281	3,407	0,154	0,934	31,54	5,371	3,465	0,157	0,917
	Model#19	2	30	GRU+LSTM	67,86	13,15	6,239	0,278	0,835	68,35	13,241	6,283	0,280	0,828
	Model#20	2	30	GRU+RNN	26,76	4,940	3,214	0,134	0,964	27,31	5,042	3,281	0,137	0,944
	Model#21	2	30	LSTM+RNN	46,47	8,967	4,686	0,202	0,897	27,53	5,093	3,291	0,139	0,936
	Model#22	2	30	LSTM+GRU	29,19	5,099	3,305	0,141	0,949	30,10	5,258	3,409	0,145	0,920
	Model#23	2	30	RNN+GRU	32,79	5,453	3,521	0,194	0,911	33,93	5,641	3,642	0,201	0,880
	Model#24	2	30	RNN+LSTM	48,92	9,184	4,796	0,205	0,879	31,15	5,352	3,522	0,153	0,911

5.2.2. Farklı derin öğrenme algoritmalarının tahmin performans değerlendirilmesi

Tez çalışmasının AK-2 kapsamında, 2. aşamada farklı derin öğrenme algoritmalarının tahmin performanslarının değerlendirildiği çalışmaların sonuçları elde edilmiştir. Bu aşamada, KHKİİ verileri kullanılmıştır. Çalışmada GRU+GRU, LSTM+LSTM, RNN+RNN, GRU+LSTM, GRU+RNN, LSTM+RNN, LSTM+GRU, RNN+GRU, RNN+LSTM olmak üzere 9 farklı hibrit model üzerinde çalışılmış ve sonuçlar elde edilmiştir. Farklı eğitim algoritmaları kullanılan tüm modeller 10 kez çalıştırılmış ve bu 10 çalıştırmanın ortalamaları Tablo 5.5'in 2. aşama bölümünde sunulmuştur. Modeller, aşama 1'de optimize edilen sonuçlara dayanarak iki gizli katman ve her bir katmanda 30 nöron içermektedir.

Tüm modellerin performans sonuçları karşılaştırıldığında, model#16 (GRU + GRU), model#17 (LSTM+LSTM) ve model#20'nin (GRU+RNN) birbirine oldukça yakın performans değerlerine sahip olduğu tespit edilmiştir. Bununla birlikte, model#17'nin sonuçları, eğitim ve test için sırasıyla MSE = 24,60 ve 24,84; RMSE = 4,739 ve 4,785; MAE = 3,056 ve 3,086; MAPE = 0,129 ve 0,13 ve $R^2 = 0,98$ ve 0,97 ($p < 0,05$) değerleri ile en iyi performansa sahip olduğunu göstermektedir.

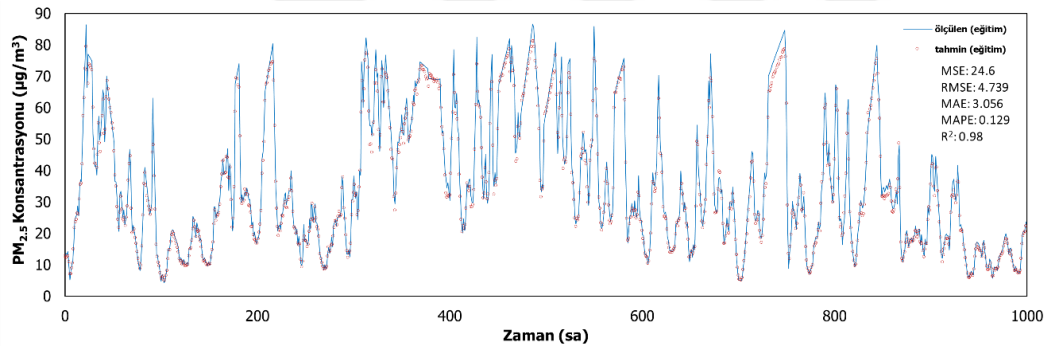
Sonuç olarak, model#17 (LSTM+LSTM), tahmin doğruluğu açısından diğer geliştirilen modellere göre daha yüksek performans göstermiştir. Öte yandan, model #19 (GRU + LSTM), bu çalışmada en düşük tahmin performansı sonuçlarına sahip olduğu görülmektedir. Model#17 için seçilen hiperparametreler Tablo 5.6'da özetlenmiştir.

Şekil 5.11'de, LSTM+LSTM modelinden elde edilen PM_{2.5} tahmini değerleri ile ölçülen değerler karşılaştırılmıştır. Şekil, grafiğin daha iyi okunabilmesi için ilk 1000 saatlik verilerin karşılaştırmasını sunmaktadır.

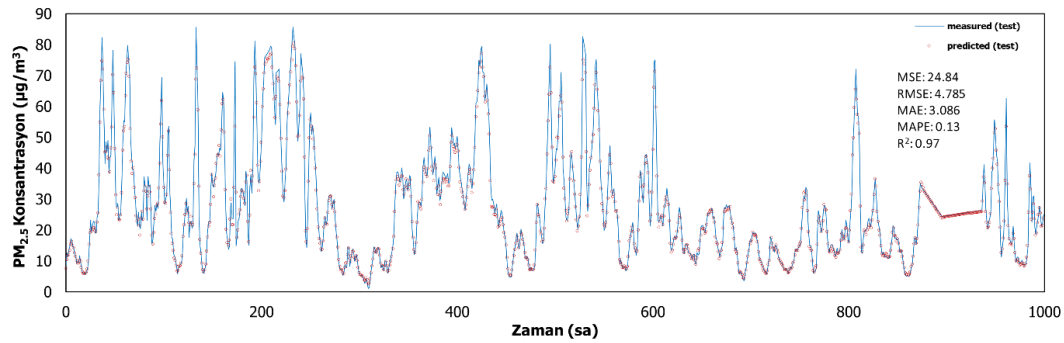
Tablo 5.6. Model#17 ait hiperparametre deęerleri.

Hiperparametre	Deęer
Yığın boyutu	32
Optimize edici	Adam
Kayıp fonksiyonu	MAE
Gizli katman sayısı	2
Her katmandaki nöron sayısı	30
Öęrenme oranı	0,01
Katmanlarda kullanılan modeller	LSTM+LSTM
Dropout oranı	0
Epoch	50
Aktivasyon fonksiyonu	relu
Eęitim oranı	0,80
n saat	24

a)

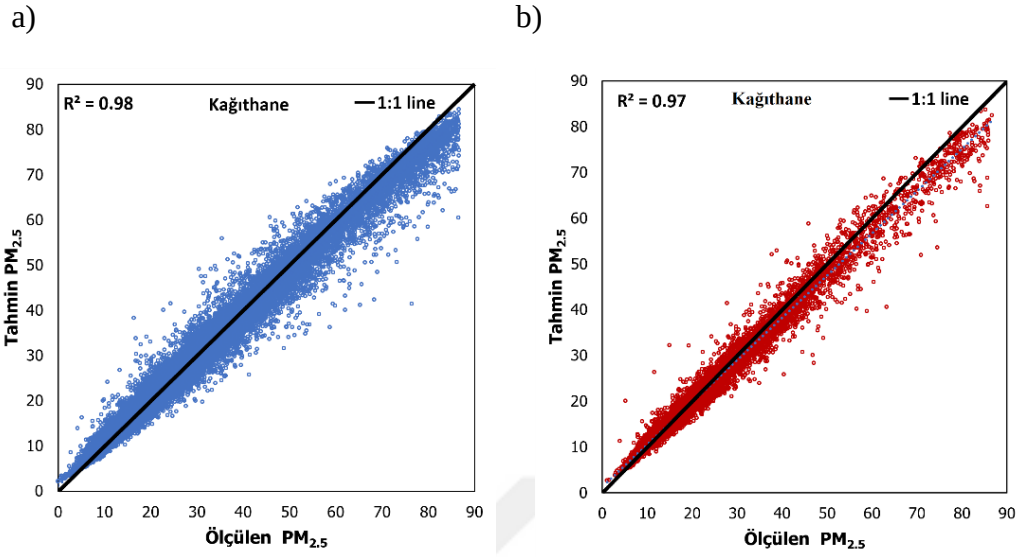


b)



Şekil 5.11. LSTM+LSTM modeli kullanılarak tahmin edilen ve ölçülen PM_{2.5} deęerlerinin karşılaştırılması a)KHKİİ eęitim veri seti ve b)KHKİİ test veri seti (p<0,05).

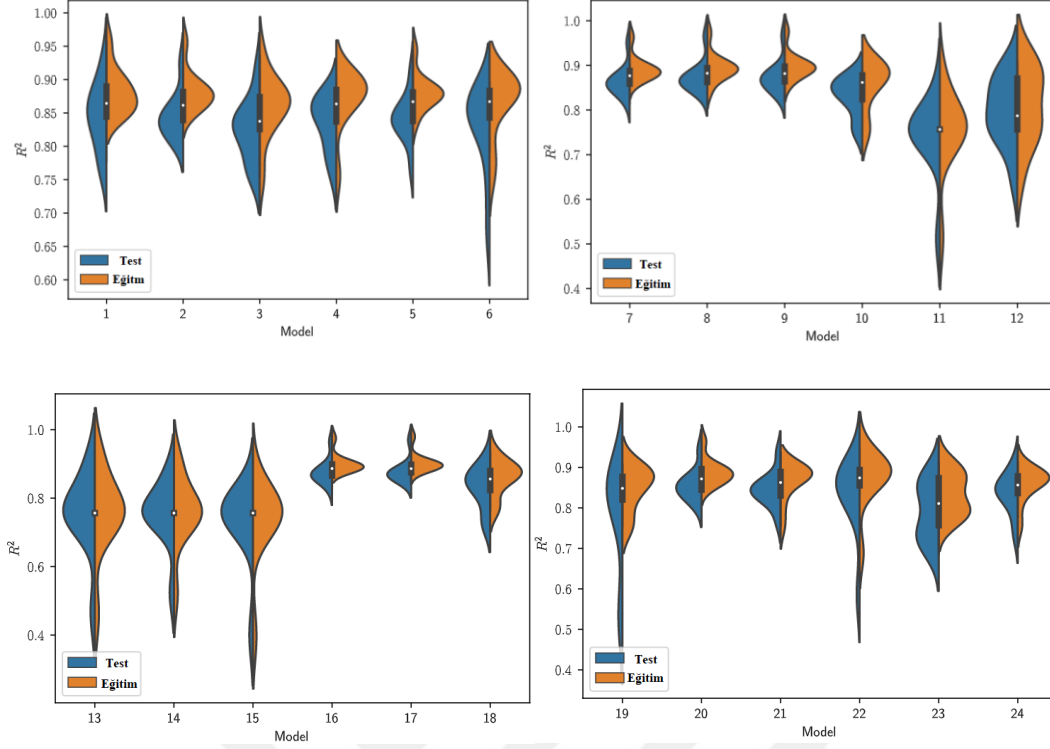
Benzer şekilde, tahmin edilen ve ölçülen değerler için dağılım grafikleri Şekil 5.12'de verilmiştir.



Şekil 5.12. Model#17 tahmin edilen ve ölçülen PM_{2.5} değerlerinin dağılım grafikleri
a) KHKİİ eğitim veri seti ve b) KHKİİ test veri seti ($p < 0,05$).

Model performanslarını özetlersek, Şekil 5.13'te tüm modellerin R² değerlerine yönelik keman grafikleri oluşturulmuştur. Keman grafiği, deney sonuçlarının dağılımını ve ortalamasını yansıtmaktadır. Grafikte beyaz bölgeler medyan değerlerini temsil ederken, mavi ve turuncu bölgeler model performansının dağılımını göstermektedir. Bu analiz sonucunda, model#9'un ilk aşamada en iyi sonuca sahip olduğu, ikinci aşamada ise model#17'nin (LSTM+LSTM) yüksek performans sergilediği anlaşılmaktadır.

Genel olarak, Tablo 5.5 ve Şekil 5.11 a ve b'de gösterildiği gibi, LSTM+LSTM modeli diğer modellere kıyasla en üstün performansa sahiptir. Tahmin doğruluğu açısından değerlendirildiğinde, model#9 ve model#17 (LSTM+LSTM) en düşük MSE, RMSE, MAE ve MAPE değerlerine sahipken, en yüksek R² değerlerine sahip modeller olarak öne çıkmaktadır.



Şekil 5.13. Tüm modellerin keman grafikleri a)Model#1-Model#6 b)Model#7-Model#12 c)Model#13-Model#18 d) Model#19-Model#24.

- Farklı çalışma sonuçları ile karşılaştırma

Tez çalışması AK-2 kapsamında yapılan çalışmalar sonucunda en iyi performans sonuçlarına sahip olan hibrit model, model#17 olarak belirlenmiştir. Model#17 sonuçları, literatürde model tahmin performanslarına dayalı olarak yayınlanan derin öğrenme ile $PM_{2.5}$ konsantrasyonlarını tahmin etmek için geliştirilen model sonuçları ile karşılaştırılmış ve sonuçlar Tablo 5.7’de bir araya getirilmiştir. Tablo 5.7’de karşılaştırma için seçilen ilk çalışma Huang ve Kuo (2018)’ nin Çin’in başkenti olan Pekin’de yapay zeka modelleriyle $PM_{2.5}$ konsantrasyon tahmini çalışmasıdır. Çalışmada APNet olarak adlandırılmış CNN+LSTM algoritması kullanılmıştır. APNet’in performansını destek vektör makineleri (Support Vector Machine-SVM), rastgele orman (Random Forest-RF), karar ağacı (Decision Tree-DT), çok katmanlı algılayıcılar (Multi-layer Perceptron-MLP), Evrişimli sinir ağları (Convolutional Neural Network-CNN) ve uzun kısa süreli bellek (Long Short-Term Memory-LSTM) modelleri ile karşılaştırmışlardır. Çalışma ile CNN-LSTM (APNet) modelinin en yüksek tahmin doğruluğunu sahip olduğu belirlenmiştir (Huang ve Kuo, 2018).

Bir diğer çalışmada ise Li ve ark. (2020), bir sonraki gün $PM_{2.5}$ konsantrasyonunu tahmini için uygun model belirleme çalışmaları yapmıştır. Çalışmada, Çin’in Pekin

şehrindeki PM_{2.5} konsantrasyonunun saatlik değerlerini ve meteorolojik verileri kullanılmıştır. Bu çalışma, hibrit model (CNN-LSTM) performansının tekli LSTM' den daha yüksek olduğunu bulmuştur. Ayrıca çalışmada, çok değişkenli modellerin performanslarının tek değişkenli modellere göre daha yüksek olduğunu belirtilmektedir (Li ve ark., 2020).

Son olarak, Ma ve ark. (2019), PM_{2.5} tahmininde yapay sinir ağlarının (çift yönlü Uzun Kısa Süreli Bellek-BLSTM ve transfer öğrenme tekniği) tahmin doğruluğunu araştırmıştır. Çalışma ile önerilen transfer öğrenme tabanlı çift yönlü uzun kısa süreli bellek (transfer learning based-bidirectional long short-term memory-TL-BLSTM) modelinin daha büyük zamansal çözümlerde daha iyi performans gösterdiği sonucuna varılmıştır (Ma ve ark., 2019).

Tüm bu çalışmalar performans indeksine göre değerlendirildiğinde, PM_{2.5} tahmin performansları için, derin öğrenme modellerinin hava kirleticileri gelecek tahmini için uygun olduğu anlaşılmaktadır. Tablo 5.7'de görüldüğü gibi önerilen derin öğrenme modelinin sonuçlarının diğer literatür çalışmaları ile uyumlu olduğu görülmüştür.

Tablo 5.7. Mevcut literatür ile karşılaştırma.

Referans	Çalışma alanı	Method	Performans İndeksi		
			MAE	RMSE	R ²
(Huang ve Kuo, 2018).	PM _{2.5} ve meteorolojik veri (Pekin, Çin).	ApNet (CNN - LSTM)	14,634	24,228	-
(Li ve ark., 2020).	PM _{2.5} ve meteorolojik veri (Pekin, Çin).	CNN-LSTM (Multivariate)	13,969	17,930	-
(Ma ve ark., 2019).	PM _{2.5} (Guangdong, Çin)	TL- BLSTM	6,184	8,653	-
Model#17	PM _{2.5} ve meteorolojik veri (Kağıthane, İstanbul, Türkiye)	LSTM-LSTM	3,086	4,785	0,97

5.2.3. Önerilen model genelleme kabiliyetinin değerlendirilmesi

Tez çalışmasının bu bölümünde, başarılı bulunan model#17 (LSTM+LSTM), İstanbul iline bağlı Aksaray, Avcılar, Beşiktaş, Kadıköy, Kartal, Maslak, Sultangazi, Tuzla ve Ümraniye ilçelerinde bulunan dokuz farklı hava kalitesi izleme istasyonunun veri setlerini kullanarak genelleme kabiliyeti değerlendirilmiştir. Modelin bu istasyonlar için genelleme performans sonuçları Tablo 5.8'de sunulmuştur.

Tablo 5.8. Dokuz farklı istasyon model#17 genelleme kabiliyeti performansı (p<0,05).

Station	MSE	RMSE	MAE	MAPE	R ²	SD	SD/RMSE
Aksaray	29,41	5,531	3,570	0,152	0,932	16,220	2,932
Avcılar	28,52	5,364	3,462	0,147	0,961	12,512	2,262
Beşiktaş	30,02	5,646	3,644	0,155	0,913	13,023	2,355
Kadıköy	28,46	5,353	3,455	0,147	0,963	14,118	2,552
Kartal	30,09	5,659	3,652	0,155	0,911	21,257	3,843
Maslak	29,70	5,585	3,605	0,153	0,923	13,209	2,388
Sultangazi	29,99	5,640	3,640	0,155	0,914	12,762	2,307
Tuzla	30,32	5,702	3,681	0,156	0,904	14,476	2,617
Ümraniye	30,06	5,652	3,648	0,155	0,912	14,025	2,536

Önerilen model#17 (LSTM+LSTM), tüm istasyonlar için mükemmel performans değerlerine sahip olduğu gibi, genelleme performansları incelendiğinde en üstün performansın Kadıköy ve Avcılar istasyonlarında elde edildiği tespit edilmiştir.

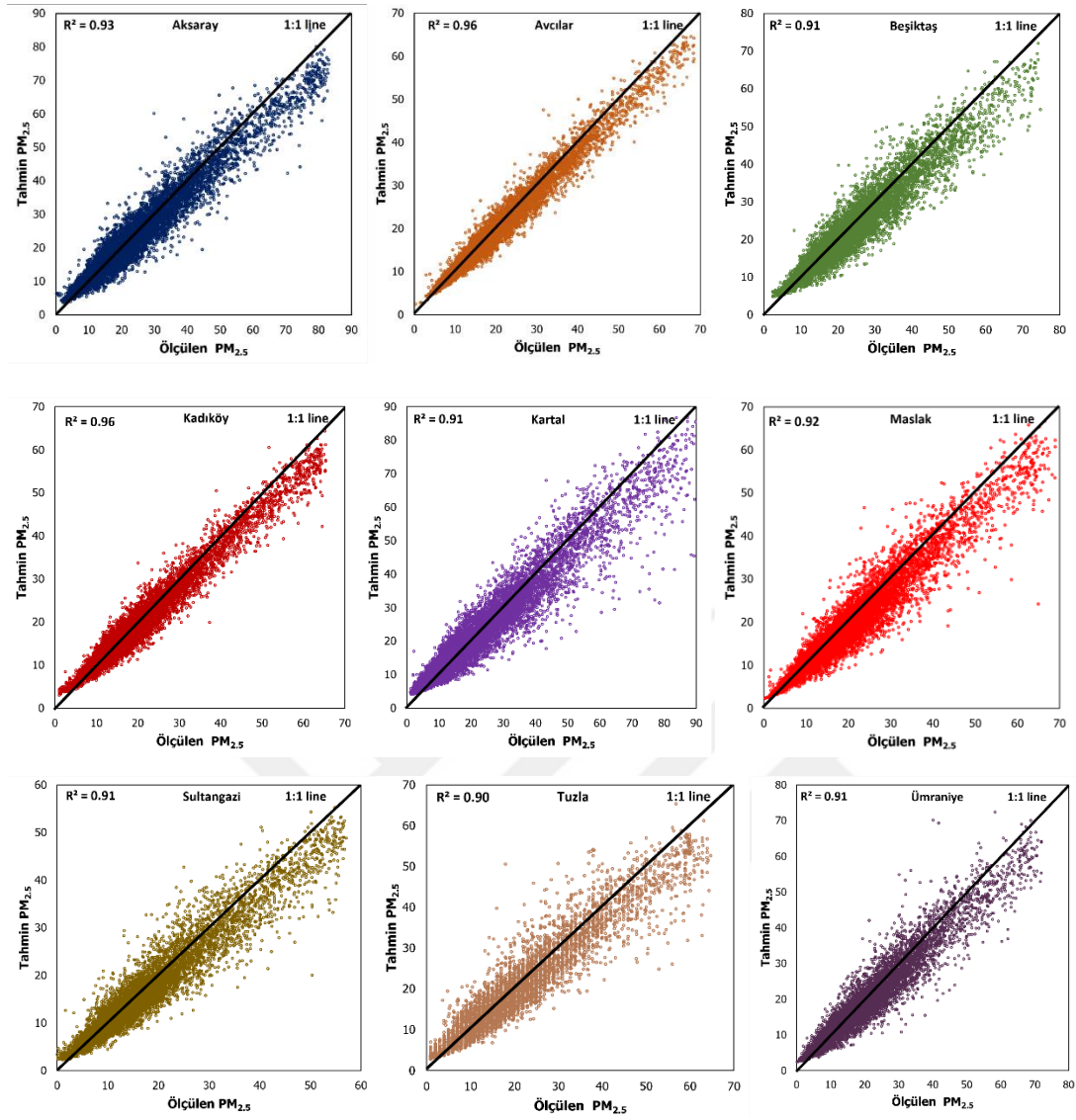
Ayrıca, RMSE'nin geçerliliğini test etmek amacıyla, Ritter ve Muñoz-Carpena (2013) tarafından önerilen bir yaklaşım kullanılmıştır (Ritter ve Muñoz-Carpena, 2013). RMSE performansını değerlendirmenin yanı sıra, SD (Standart Sapma) değerlerinin de dikkate alınmasını önerir. RMSE'nin optimal değeri sıfır olarak kabul edilse de SD değerleri yüksek olan veri setlerinde RMSE'nin yüksek çıkması daha olasıdır. Bu bağlamda, algoritmanın performansını daha kapsamlı bir şekilde değerlendirmek için Tablo 5.8'de SD/RMSE oranı verilmiştir.

Ritter ve Muñoz-Carpena (2013), SD/RMSE oranına dayalı olarak dört performans derecelendirmesi önermektedir. Bu öneri Tablo 5.9'da sunulmuştur.

Tablo 5.9. SD/RMSE performans dereceleri (Ritter ve Muñoz-Carpena (2013)'den uyarlanmıştır)

SD/RMSE	Performans derecesi
$\geq 3,2$	çok iyi
2,2-3,2	iyi
1,2-2,2	kabul edilebilir
$<1,7$	yetersiz

Bu çalışmada, Kartal ilçesinin performansının "çok iyi" performans derecesine sahip olduğu, diğer ilçelerin ise "iyi" performans derecesine sahip olduğu gözlemlenmektedir. Şekil 5.14'te, dokuz bölge için tahmin edilen ve ölçülen $PM_{2.5}$ değerleri arasındaki dağılımı gösteren grafikler sunulmuştur. Bu grafiklerden görüleceği üzere, ölçülen ve tahmin edilen $PM_{2.5}$ konsantrasyonları arasında mükemmel bir uyum gözlemlenmektedir. Sonuç olarak, önerilen model#17 (LSTM+LSTM), İstanbul metropoliten kentinde saatlik $PM_{2.5}$ tahmini yapmak için üstün genelleme yeteneğine sahiptir.



Şekil 5.14. Model#17 genelleme kabiliyeti dağılım grafikleri (Aksaray, Avcılar, Beşiktaş, Kadıköy, Kartal, Maslak, Sultangazi, Tuzla, Ümraniye)

6. TARTIŞMA VE SONUÇLAR

Bu çalışma, partikül madde kaynaklı hava kirliliğinin neden olduğu sağlık risklerini, çevresel etkileri ve ekonomik kayıpları azaltmak için en etkili yöntemlerden biri olan gelecekteki PM konsantrasyonlarını tahmin edebilen modellerin oluşturulması üzerine odaklanmaktadır. Gerçek dünya verileri kullanılarak yapılan araştırmada, saatlik PM₁₀ tahmini için Yalın, Çift yönlü ve Yığılmış LSTM olmak üzere üç farklı LSTM varyantının uygunluğu incelenmiş; ayrıca saatlik PM_{2.5} konsantrasyonlarını tahmin etmek için LSTM, RNN ve GRU gibi derin öğrenme algoritmaları kullanılarak bir derin öğrenme modeli geliştirilmiştir. Çalışmanın içeriği, VÖİ-ÖS adının modelleme performanslarına etkisi ve modellerin genelleme yeteneklerinin belirlenmesi amacıyla yapılan ek çalışmalarla zenginleştirilmiştir. Bu şekilde, çalışma konuları daha geniş bir perspektiften ele alınarak araştırma yanı güçlendirilmiştir.

Bu çalışmanın ana sonuçlarını özetleyecek olursak;

- Veri ön işleme (VÖİ) ve özellik seçimi (ÖS) süreçleri, orijinal veri setlerindeki hatalı veya eksik verileri düzelter ve gereksiz özellikleri ortadan kaldıran önemli adımlardır. VÖİ-ÖS'li geliştirilmiş LSTM modelleri, PM₁₀ ve PM_{2.5} hava kirlleticilerinin konsantrasyonlarını doğru bir şekilde tahmin edebilme yeteneğine sahiptir.
- Önerilen modeller, belirli bir bölgedeki hava kalitesini izlemek ve iyileştirmek için önemli bir araç olarak kullanılabilir. Saatlik hava kirleticilerinin (örneğin PM₁₀) konsantrasyonlarını tahmin etmek, şehir yöneticilerinin hızlı ve etkili kararlar almasına yardımcı olabilir.

- Farklı derin öğrenme algoritmaları arasında yapılan karşılaştırmalarda, geliştirilen LSTM+LSTM (model#17) modelinin en iyi tahmin performansına sahip olduğu belirlenmiştir. Bu model saatlik PM_{2.5} konsantrasyonlarını yüksek doğrulukla tahmin edebilmiştir. Ayrıca, model#17 farklı bölgelerden alınan verilere uygulandığında da üstün genelleme yeteneği göstermiştir.

- Önerilen modelin performans sonuçları, literatürde yayınlanan diğer çalışmalarla karşılaştırılmış ve önerilen modelin daha iyi performans sergilediği görülmüştür. Bu, modelin hava kirliliği tahmininde etkili bir araç olduğunu vurgulamaktadır.

Sonuç olarak, önerilen model sayesinde PM kirliliği tahmin edilerek olası olumsuz etkilerin belirlenmesi mümkün hale gelir ve çeşitli hava kirliliği azaltma senaryoları ve uyarıları hazırlanarak halk sağlığının korunması için gerekli önlemler alınabilir. Ancak, çalışmanın bazı sınırlamaları da göz önünde bulundurulmalıdır. Bu sınırlamalar şunlarla ilgilidir:

Diğer Hava Kirleticileri: Çalışmada PM_{2.5} ve PM₁₀ konsantrasyonlarına odaklanılmış olsa da, diğer hava kirleticileri (SO₂, NO_x, O₃, CO gibi) için model etkinliği araştırılmamıştır. Bu nedenle bu hava kirleticilerinin tahmininde modelin performansının nasıl olduğu belirlenmelidir.

Yeni Algoritmalar ve Teknikler: Derin öğrenme alanında hızla gelişen yeni algoritmalar ve teknikler bulunmaktadır. Çalışma bu algoritmaların model performansına etkisini değerlendirmemiş ve gelecekteki gelişmeleri içermemiştir.

Genelleme Yeteneği: Çalışma İstanbul ili üzerine odaklanmış ve model genelleme yeteneği bu bölgede test edilmiş olsa da, farklı coğrafi bölgelerdeki genelleme yeteneği olarak test edilmemiştir. Bu nedenle modelin farklı bölgelerdeki performansı daha detaylı olarak araştırılmalıdır.

Bu çalışma ile Türkiye'de Çevre Mühendisliği alanında hava kirliliği tahmini konusuna yeni bir bakış açısı getirilmiştir. İstanbul gibi büyük ve hızla büyüyen metropollerdeki hava kirliliği sorununa karşı etkili bir çözüm sunulmuştur. LSTM (Yalın-Çift yönlü ve Yığılmış), GRU ve RNN gibi derin öğrenme modelleri, akıllı şehir yaklaşımıyla hava kirliliğini yönetmeye yardımcı olabilir. Bu çalışma aynı zamanda literatüre ve denetleyici kurumlara da katkı sağlamayı amaçlamaktadır. Yerel yönetimlerin hava kirliliği kaynaklı yönetim stratejilerini belirlemesine ve akıllı şehir anlayışını desteklemesine yardımcı olabilir.

KAYNAKLAR

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X., 2016. Tensorflow: a system for large-scale machine learning. *In: 12th USENIX Symposium on Operating Systems Design and Implementation*. pp. 265–283.
- Abidogun, O. A. (2005). *Data mining, fraud detection and mobile telecommunications: Call pattern analysis with unsupervised neural networks* [Doktora Tezi]. University of the Western Cape.
- Adhikari, R. ve Agrawal, R. K., 2013, An Introductory Study on Time Series Modeling and Forecasting, Lambert Academic Publishing.
- Ağaç, K. (2016). İstanbul kağıthane bölgesinde yüksek PM₁₀ konsantrasyonlarının meteorolojik olarak incelenmesi [Yüksek Lisans Tezi], İstanbul Teknik Üniversitesi.
- Akdur, R., (2009). Hava Kirliliği ve Çözümler. Halk Sağlığı İle İlgili Güncel Sorunlar ve Yaklaşımlar, s:151-155 *Ankara Tabip Odası yayını* Ankara
- Akgül, I. (1994). Zaman Serisi Analizi ve Öngörü Modelleri. *Öneri Dergisi*, 1(1), 52-69. DOI: 10.14783/maruoneri.698511
- Akın, İ. (2021). Manavgat'ta büyük orman yangını <https://www.sozcu.com.tr/2021/gundem/manavgatta-buyuk-orman-yangini-mahalleler-bosaltilmaya-baslandi-6561403/>
- Aksangür, İ., Eren, B., Erden, C. (2022). Evaluation of data preprocessing and feature selection process for prediction of hourly PM10 concentration using long short-term memory models. *Environmental Pollution* - Vol.311 - ISSN : 0269-7491 - DOI : 10.1016/j.envpol.2022.119973 - English - Article - 2022 - WOS:000861288900001
- Aksoy, B., Khaled O., ve Salman, M. (2020). ARIMA modeli kullanılarak Türkiye'deki iklim sıcaklıklarının geleceğe yönelik tahminlenmesi. *Uluslararası Teknolojik Bilimler Dergisi*, 12(2), 69-76.
- Anadolu Ajansı (2023). İstanbul'daki fabrika yangını 70 saat sonra söndürüldü <https://www.trthaber.com/haber/turkiye/istanbuldaki-fabrika-yangini-70-saat-sonra-sonduruldu-774593.html>
- Arı, A., Argante, J., Meliefste, K., Gaga, E. O., Yay, O. D., Örnektekin, S., Döğeroğlu, T., & Doorn, W.V., (2008, Ekim, 22-25). İskenderun ve Payas'ta Atmosferik PM₁₀ ve PM_{2.5} Derişimlerinin İncelenmesi. *Hava Kirliliği ve Kontrolü Ulusal Sempozyumu*. Hatay, Türkiye.

- Assaad, M., Boné, R., & Cardot, H. (2008). A new boosting algorithm for improved time-series forecasting with recurrent neural networks. *Information Fusion*, 9(1), 41-55.
- Ataş, H. (2022). Bursa’da devam eden fabrika yangınıyla ilgili hava kirliliği uyarısı <https://www.sozcu.com.tr/2022/gundem/bursada-devam-eden-fabrika-yanginiyla-ilgili-hava-kirliligi-uyarisi-7456575/>
- Avrupa Birliği Çevre Müktesebatına Uyum Çalışmaları (t.y.). <https://www.ab.gov.tr/files/SEPBB/cevrefaslidokumanlar/abcevremuktesebatin-uyumcalismalari.pdf>
- Avrupa Çevre Ajansı (2021a, 4 Şubat). Hava kalitesinin iyileştirilmesi insanların sağlığını iyileştirir, üretkenliğini artırır. <https://www.eea.europa.eu/tr/isaretler/aca-isaretler-2020/articles/hava-kalitesinin-iyilestirilmesi-insanlarin-sagligini> adresinden 31 Nisan 2023 tarihinde alınmıştır.
- Avrupa Çevre Ajansı (2021b, 15 Kasım). Air quality in Europe 2021-Health impacts of air pollution in Europe, 2021. <https://www.eea.europa.eu/publications/air-quality-in-europe-2021/health-impacts-of-air-pollution>
- Avrupa Çevre Ajansı (2023, 3 Nisan). Air pollution: how it affects our health. <https://www.eea.europa.eu/themes/air/health-impacts-of-air-pollution>
- Aydin, S., Taşyürek, M., ve Öztürk, C. (2021). Derin Öğrenme Yöntemi ile İç Anadolu Bölgesi ve Çevresi Hava Kirliliği Tahmini. *European Journal of Science and Technology*. 29 168-173 DOI: 10.31590/ejosat.1019618
- Ayturan, Y. A. (2019). *Derin öğrenme ile havadaki partikül madde konsantrasyonunun tahmini* [Yüksek Lisans Tezi] KTO Karatay Üniversitesi
- Bekkar, A., Hssina, B., Douzi, S., & Douzi, K. (2021). Air-Pollution Prediction in Smart City, Deep Learning Approach. *Journal of Big Data*, 8(1), 161. <https://doi.org/10.1186/s40537-021-00548-1>
- BHKİİ, 2021. Hava kalitesi izleme projesi - İstanbul büyükşehir belediyesi. <https://havakalitesi.ibb.istanbul/Pages/AirQualityDetails/6d169752-eabc-4f7e-87ea-6cf10b452486>
- Brownlee, J. (2017, Ağustos 17). Stacked Long Short-Term Memory Networks. *Machine Learning Mastery*. <https://machinelearningmastery.com/stacked-long-short-term-memory-networks/>
- Brownlee, J. (2018, Kasım 14). How to Develop LSTM Models for Time Series Forecasting. *Machine Learning Mastery*. <https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/>
- Bui, T.-C., Le, V.-D., & Cha, S.-K. (2018). A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM. arXiv:1804.07891 [cs, stat]. <http://arxiv.org/abs/1804.07891>
- Cabaneros, S. M., Calautit, J. K., & Hughes, B. (2019). A Review Of Artificial Neural Network Models For Ambient Air Pollution Prediction. *Environmental Modelling and Software*, 119, 285-304. DOI:10.1016/j.envsoft.2019.06.014

- California air resources board, (t.y.). Inhalable Particulate Matter and Health (PM_{2.5} and PM₁₀). <https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health>.
- Can, E. (2016). *Kütahya ili genelinde partikül madde kirletici kaynaklarının belirlenmesi* [Yüksek Lisans Tezi] Anadolu Üniversitesi.
- Canizo, M., Triguero, I., Conde, A., & Onieva, E. (2019). Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing*, 363, 246-260. DOI: 10.1016/j.neucom.2019.07.034
- Cao, J., Li, Z., & Li, J. (2019). Financial Time Series Forecasting Model Based On CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, 519, 127-139. DOI: 10.1016/j.physa.2018.11.061
- Chimmula, V. K. R., & Zhang, L. (2020). Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864. DOI: 10.1016/j.chaos.2020.109864
- Cho, K., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. arXiv:1406.1078v3 [cs.CL] 3 Sep 2014
- Chollet, F. (2015). Keras. <https://github.com/fchollet/keras>
- Colah, C. (2015 Agust 27) Understanding LSTM Networks. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Czernecki, B., Marosz, M., & Jędruszkiewicz, J. (2021). Assessment of Machine Learning Algorithms in Short-term Forecasting of PM₁₀ and PM_{2.5} Concentrations in Selected Polish Agglomerations. *Aerosol and Air Quality Research*, 21(7), 200586. DOI:10.4209/aaqr.200586
- Çapraz, Ö. (2013). *İstanbul'da 2007-2012 yılları arasındahavakirliliğininölümler üzerindeki etkilerinin modellenmesi* [Yüksek Lisans Tezi] İstanbul Teknik Üniversitesi.
- Çevre İklim ve Sağlık İçin İş Birliği Projesi (2022). Hava Kirliliğinin Sağlık Etkileri, https://www.env-health.org/wp-content/uploads/2022/03/Hava_Kirliligi_Bilgi_Notu.pdf adresinden 01.05.2023 tarihinde alınmıştır.
- Çevre Online (2023, 14 Mart). Çevre Felaketleri. <https://cevreonline.com/cevre-felaketleri/>
- Çilingir, B. M. (2016). Hava Kirliliği ve Akciğer Air pollution and lung. *J CONTEMP MED*, 2016;6 (Case Reports): 131-137. DOI: 10.16899/ctd. 80586
- Dayanıklı, A. S. (2021, Şubat 11). Aykırı Değer (Outlier) Analizi Nedir? Uç Değerler Nasıl Tespit Edilir? *Ravenfo*. <https://ravenfo.com/2021/02/11/aykiri-deger-analizi/>
- Demir, G., Özdemir, H., Borucu, G., Yiğit, S., ve Ak, N. (2010). İstanbul'daki Çocuk Oyun Parkları'nda Partikül Madde (PM_{2.5} ve PM₁₀) Kirliliğinin İncelenmesi. *Ekoloji*, Ekoloji 20, 77, 72-79. doi: 10.5053/ekoloji.2010.7711
- Deng, L. & Yu, D., (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, 7(3-4), 197-387. DOI: 10.1561/20000000039

- Du, S., Li, T., Yong, Y., & Horng S-J. (2021). Deep Air Quality Forecasting Using Hybrid Deep Learning Framework. *IEEE Transactions on Knowledge and Data Engineering*. 33(6) 2412-2424. DOI:10.1109/TKDE.2019.2954510
- Enironmental Protection Agency [EPA] (2020). Review of the National Ambient Air Quality Standards for Particulate Matter. Federal Register /82684 Vol. 85, No. 244 / Friday, December 18, 2020 / Rules and Regulations
- Erbaşlar, T., Taşdemir, Y. (2007) Bursa Atmosferinde Ölçülen Klasik Hava Kirleticilerin Birbirleri ile Olan İlişkileri. *Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, Cilt 12, Sayı 2, 2007
- Eren, B., Yaqup, M., Eyüpoğlu, V. (2019). A comparative study of artificial neural network models for the prediction of Cd removal efficiency of polymer inclusion membranes. *Desalination and Water Treatment*. 143 (2019) 48–58 March DOI: 10.5004/dwt.2019.23531.
- Eren, B., Aksangür, İ., Erden, C. (2023) Predicting Next Hour Fine Particulate Matter (PM_{2.5}) in The Istanbul Metropolitan City Using Deep Learning Algorithms With Time Windowing Strategy, *Urban Climate*, 48 (2023) 101418 DOI:10.1016/j.uclim.2023.101418
- Ergüder, H. (2019). Recurrent Neural Network Nedir? *Medium*. <https://medium.com/@hamzaerguder/recurrent-neural-network-nedir-bdd3d0839120>
- European commission (t.y) EU air quality standards https://environment.ec.europa.eu/topics/air/air-quality/eu-air-quality-standards_en
- Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting Air Quality Time Series Using Deep Learning. *Journal of the Air & Waste Management Association*, 68(8), 866-886. <https://doi.org/10.1080/10962247.2018.1459956>
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big Data Preprocessing: Methods and Prospects. *Big Data Analytics*, 1(1), 9. DOI: 10.1186/s41044-016-0014-0
- Gers, F.A., Schmidhuber, J., & Cummins, F., 1999. Learning to forget: continual prediction with LSTM, in: 1999 ninth international conference on artificial neural networks ICANN 99. Conf. Publ. No. 470. *Presented at the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99* 2, 850–855. DOI:10.1049/cp:19991218. . Conf. Publ. No. 470
- Graham, M.H., 2003. Confronting 541 Multicollinearity In Ecological Multiple Regression. *Ecology* 84, 542 2809–2815. <https://doi.org/10.1890/02-3114>
- Graves, A., Mohamed, A., Hinton, G., (2013). Speech recognition with deep recurrent neural networks. 2013 IEEE international conference on acoustics, speech and signal processing. 6645–6649.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. DOI: 10.1109/TNNLS.2016.2582924

- Guan, Y., Xiao, Y., Rong, B., Zhang, N., & Chu, C. (2021). Long-term health impacts attributable to PM2.5 and ozone pollution in China's most polluted region during 2015–2020. *Journal of Cleaner Production*, 321, 128970. <https://doi.org/10.1016/j.jclepro.2021.128970>
- Guo, C., Liu, G., & Chen, C.-H. (2020). Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network. *Wireless Communications and Mobile Computing*, 2020, 1-13. DOI: 10.1155/2020/8854649
- Gürsoy, U. (2022, 5 Aralık). İnsan Kaynaklı Çevre Yıkımlarında Bugün-Büyük Londra Dumanlı Sisi. <https://yesilgazete.org/insan-kaynakli-cevre-yikimlarinda-bugun-buyuk-londra-dumanli-sisi/>
- Güzel, Ş., ve Özer, P. (2022). Türkiye’de Hava Kirliliği ve Sağlık Harcamaları. *Sağlık ve Sosyal Refah Araştırmaları Dergisi*. 4(2). 186-202. E-ISSN: 2667-8217 <https://doi.org/10.55050/sarad.1138629>
- Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği [HKDYY], (2008, 6 Haziran). Resmi Gazete (Sayı: 26898). <https://www.resmigazete.gov.tr/eskiler/2008/06/20080606-6.htm>
- Hava Kirliliği ve Atmosfer (t.y.). https://acikders.ankara.edu.tr/pluginfile.php/81428/mod_resource/content/0/3
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, C.-J., & Kuo, P.-H. (2018). A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors*, 18(7), Article 7. DOI: 10.3390/s18072220
- Huang, G., 2021. Missing data filling method based on linear interpolation and lightgbm. *J. Phys. Conf. Ser.* 1754, 012187. <https://doi.org/10.1088/1742-6596/1754/1/012187>.
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9, 90-95. DOI: 10.1109/MCSE.2007.55
- International Agency for Research on Cancer (2013, 17 Ekim). IARC: Outdoor Air Pollution A Leading Environmental Cause Of Cancer Deaths. https://www.iarc.who.int/wp-content/uploads/2018/07/pr221_E.pdf
- Isaev, E., Ajikeev, B., Shamyranov, U., Kalnur, K., Maisalbek, K., & Sidle, R. C. (2022). Impact of Climate Change and Air Pollution Forecasting Using Machine Learning Techniques in Bishkek. *Aerosol and Air Quality Research*, 22(3), 210336. DOI: 10.4209/aaqr.210336
- İklimBU, Boğaziçi Üniversitesi. (2019, Eylül 10). Hava. <http://climatechange.boun.edu.tr/hava/> adresinden 28 Nisan 2023 tarihinde alınmıştır.
- İstanbul Büyükşehir Belediyesi [İBB], (t.y.). Hava kalitesi izleme projesi. <https://havakalitesi.ibb.gov.tr/Pages/AirQuality>
- Karakoyun, E. Ş. (2018). *Derin öğrenme ile zaman serilerinin gerçek zamanlı tahmini* [Yüksek Lisans Tezi], Necmettin Erbakan Üniversitesi.

- Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X., & Sachdeva, S. (2019). Evaluation of Different Machine Learning Approaches to Forecasting PM2.5 Mass Concentrations. *Aerosol and Air Quality Research*, 19(6), 1400-1410. DOI:10.4209/aaqr.2018.12.0450
- Keskin A. (2020). BTÜ Kimya Mühendisliği Öğretim Üyesi Doç. Dr. Hülya Koyuncu Sakarya'daki Patlamayı Değerlendirdi. <https://btu.edu.tr/tr/haber/detay/4898/bt%C3%BC-kimya-m%C3%BChendisli%C4%9Fi-%C3%B6%C4%9Fretim-%C3%BCyesi-do%C3%A7-dr-h%C3%BClya-koyuncu-sakaryadaki-patlamay%C4%B1-de%C4%9Ferlendirdi>
- KHKİİ, 2022. Kağıthane 2 (Kentsel) | Hava Kalitesi İzleme Projesi—İstanbul Büyükşehir Belediyesi. <https://havakalitesi.ibb.gov.tr/Pages/AirQualityDetails/d9f15eae-128a-4a71-ad30-141b4728a520>
- Kim, H., & Lee, J.-H. (2016). A recurrent neural networks approach for estimating the quality of machine translation output. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 494-498.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. DOI: 10.48550/arXiv.1412.6980
- Lawton, G. (t.y.) Data Preprocessing: Definition, Key Steps and Concepts. *Data Management*. <https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing>
- LeCun, Y., Bengio, Y., & Hinton, G., Deep Learning (2015). *Nature*. 521(7553) 436-444 DOI: 10.1038/nature14539
- Leong, W. C., Kelani, R. O., & Ahmad, Z. (2020). Prediction of Air Pollution Index (API) Using Support Vector Machine (SVM). *Journal of Environmental Chemical Engineering*, 8(3), 103208. DOI: 10.1016/j.jece.2019.103208
- Li, J., & Shen, J. (2019). Prediction of PM2.5 Concentration Based on CEEMD-LSTM Model. *Chinese Control Conference (CCC)*, 8439-8444. DOI: 10.23919/ChiCC.2019.8866134
- Li, T., Hua, M., & Wu, X. (2020). A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5). *IEEE Access*, 8, 26933-26940. DOI: 10.1109/ACCESS.2020.2971348
- Liu, B., Jin, Y., & Li, C. (2021). Analysis and prediction of air quality in Nanjing from autumn 2018 to summer 2019 using PCR-SVR-ARMA combined model. *Nature, Scientific Reports*, 11(1), <https://doi.org/10.1038/s41598-020-79462-0>
- Lu, X., Sha, Y.H., Li, Z., et al., 2021. Development and application of a hybrid long-short term memory – three dimensional variational technique for the improvement of PM2.5 forecasting. *Sci. Total Environ.* 770, 144221. <https://doi.org/10.1016/j.scitotenv.2020.144221>.

- Ma, J., Cheng, J. C. P., Lin, C., Tan, Y., & Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214, 116885. DOI: 10.1016/j.atmosenv.2019.116885
- Mayruk, A. (2005). *Yüreğir ve seyhan (adana) ilçelerinde ana arterlerdeki toz ve gürültü dağılım haritalarının hazırlanması* [Yüksek Lisans Tezi] Çukurova Üniversitesi.
- Menares, C., Perez, P., Parraguez, S., & Fleming, Z. L. (2021). Forecasting PM2.5 Levels in Santiago de Chile Using Deep Learning Neural Networks. *Urban Climate*, 38 (100906) DOI: 10.1016/j.uclim.2021.100906
- Miller, L., & Xu, X. (2018). Ambient PM_{2.5} Human Health Effects-Findings in China and Research Directions. *Atmosphere*, 9(11), 424. <https://doi.org/10.3390/atmos9110424>.
- Mir, A. A., Kearfott, K. J., Çelebi, F. V., & Rafique, M. (2022). Imputation By Feature Importance (IBFI): A Methodology to Envelop Machine Learning Method For Imputing Missing Patterns In Time Series Data. *Plos One*, 17(1), e0262131. DOI: 10.1371/journal.pone.0262131
- NSW Health (2022) Outdoor air pollution <https://www.health.nsw.gov.au/environment/air/Pages/outdoor-air-pollution.aspx>
- Olcese, L. E., Gustavo, G. P., & Beatriz, M. T. (2001) An inexpensive method to estimate CO and NOx emissions from mobile sources. *Atmospheric Environment*, 35, 6213–6218
- Orhun, Z. (2013). *Türkiye coğrafi bölgelerinde örnek şehirlerde ölçülen PM10 değerlerinin değişiminin incelenmesi* [Yüksek Lisans Tezi] İstanbul Teknik Üniversitesi.
- Orman Koruma (2008, 29 Şubat). Hava kirliliği. https://www.ktu.edu.tr/dosyalar/ormankoruma_d0053.pdf adresinden 28 Nisan 2023 de alınmıştır.
- Özbeyaz, A., Tufaner, F., ve Demirci, Y. (2016, Ekim, 26-28). Partikül Madde İle İlişkili Görüntüleri Kullanarak Hava Kirliliği Tahminine Ait Bir Model Tasarımı. *1st International Mediterranean Science and Engineering Congress, Çukurova University, Congress Center*, 969-975, Paper ID:302 Adana / Türkiye
- Pak, U., Ma, J., Ryu, U., Ryom, K., Juhok, U., Pak, K., and Pak, C. (2020). Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Science of The Total Environment*, 699, 133561. DOI: 10.1016/j.scitotenv.2019.07.367
- Park, H., Cho, S., & Park, J. (2018). Word RNN as a baseline for sentence completion. *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 183-187.
- Partigöç, N. S., ve Çubukçu, K.M. (2017). Hava Kirliliği ve Kent ilişkisine Ampirik Bakış: Ekolojik Sürdürülebilirlik Ekseninde Bir Değerlendirme. *Akademia Disiplinlerarası Bilimsel Araştırmalar Dergisi*, 3 (2), 28-45. ISSN: 2548-0987

- Pascanu, R., Gulçehre, C., Cho, K. and Bengio Y. (2014). How to Construct Deep Recurrent Neural Networks. arXiv:1312.6026v5 [cs.NE] 24 Apr 2014
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 2825-2830
- Perez, P., & Menares, C. (2018). Forecasting of Hourly PM2.5 in South-West Zone in Santiago de Chile. *Aerosol and Air Quality Research*, 18(10), 2666-2679 DOI: 10.4209/aaqr.2018.01.0029
- Polat, E. E. (2016). *Gaziantep'in trafik kaynaklı hava kirliliğinin belirlenmesi üzerine bir çalışma* [Yüksek Lisans Tezi]. Kahramanmaraş Sütçü İmam Üniversitesi.
- Pun, V. C., Kazemiparkouhi, F., Manjourides, J., Suh, H. H. (2017). Long-Term PM2.5 Exposure and Respiratory, Cancer, and Cardiovascular Mortality in Older US Adults. *Am J Epidemiol.* 2017 Oct 15;186(8):961-969. doi: 10.1093/aje/kwx166. PMID: 28541385
- Ravikumar, G. K., Manjunath, T. N., Ravindra, S. Hegadi, & Umesh I. M., (2011). A Survey on Recent Trends, Process and Development in Data Masking for Testing. *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814
- Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33-45. DOI: 10.1016/j.jhydrol.2012.12.004
- Rubal, & Kumar, D. (2018). Evolving Differential Evolution Method With Random Forest For Prediction of Air Pollution. *Procedia Computer Science*, 132, 824 833. DOI: 10.1016/j.procs.2018.05.094
- Schraufnagel, D.E., Balmes, J.R., Cowl, C. T., De Matteis, S., Jung, S. H., Mortimer, K., Perez-Padilla, R., Rice M. B., Riojas-Rodriguez, H., Sood, A., Thurston, G. D., To, T., Vanker, A., & Wuebbles D. J. (2019). Air Pollution and Noncommunicable Diseases. *CHEST* 2019, 155(2): 417-426. DOI: 10.1016/j.chest.2018.10.041
- SİM (2019). Sürekli İzleme Merkezi. T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı <https://www.havaizleme.gov.tr/Services/AirQuality/Default.ltr.aspx>
- SİM (2022a). Sürekli İzleme Merkezi. T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. Ulusal Hava Kalitesi İzleme Ağı (UHKİA) <https://sim.csb.gov.tr/Intro/Uhkia>
- SİM (2022b). Sürekli İzleme Merkezi. T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı <https://www.havaizleme.gov.tr/Services/AirQuality/Default.ltr.aspx>.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.

- Szandała, T., (2021) Review and comparison of commonly used activation functions for deep neural networks. In: Bhoi, A.K., Mallick, P.K., Liu, C.-M., Balas, V.E. (Eds.), Bio-Inspired Neurocomputing. Springer, Singapore, pp. 203–224. DOI: 10.1007/978-981-15-5495-7_11
- Tao, Q., Liu, F., Li, Y., & Sidorov, D. (2019). Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU. *IEEE Access*, 7, 76690-76698. DOI: 10.1109/ACCESS.2019.2921578
- Tokgöz, A., ve Ünal, G. (2018). A RNN based time series approach for forecasting turkish electricity load. 2018 26th Signal Processing and Communications Applications Conference (SIU), 1-4.
- Tuncel, S. G., (2016). Atmosferden PM₁₀, PM_{2.5} ve PM₁ parçacıkları eş zamanlı olarak toplayabilecek, ekonomik bir örnekleme sisteminin geliştirilmesi (Program Kodu: 1005 Proje No: 114Y160).
- Turgut, D., ve Temiz, İ. (2015). Time Series Analysis And Forecasting For Air Pollution In Ankara: A Box-Jenkins Approach. *Alphanumeric Journal*, 3(2). <https://doi.org/10.17093/aj.2015.3.2.5000148347>
- TÜİK. (2022). Address Based Population Registration System Results for 2021. <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuclari-2021-45500>
- Türkiye Cumhuriyeti Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. Çevre Yönetimi Genel Müdürlüğü (2023) <https://cygm.csb.gov.tr/yonetmelikler-i-440>
- Türkiye Cumhuriyeti Çevre, Şehircilik ve İklim Bakanlığı, Meteoroloji Genel Müdürlüğü [MGM], (2020). <https://mgm.gov.tr/>
- Türkiye Cumhuriyeti Çevre, Şehircilik ve İklim Bakanlığı, Meteoroloji Genel Müdürlüğü [MGM] (2022). <https://mgm.gov.tr/>
- Türkiye Mühendis ve Mimar Odaları Birliği [TMMOB], (2018). Hava Kirliliği Raporu 2018, Çevre Mühendisleri Odası [ÇMO]. https://www.cmo.org.tr/genel/bizden_detay.php?kod=100058&tipi=91&sube=0 23 Nisan 2021 tarihinde alınmıştır.
- U.S. Environmental Protection Agency (2014a). What is particle pollution <https://www.epa.gov/pmcourse/what-particle-pollution>
- U.S. Environmental Protection Agency (2014b). Criteria Air Pollutants <https://www.epa.gov/criteria-air-pollutants>
- U.S. Environmental Protection Agency (2015a). Managing Air Quality - Air Pollutant Types. <https://www.epa.gov/air-quality-management-process/managing-air-quality-air-pollutant-types>
- U.S. Environmental Protection Agency, (2016, Nisan 19). Particulate Matter (PM) Basics [Overviews and Factsheets]. <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>
- U.S. Environmental Protection Agency (2017). Indoor Pollutants and Sources <https://www.epa.gov/indoor-air-quality-iaq/indoor-pollutants-and-sources>
- Uslu, E. E. (2020, 17 Aralık). PM_{2.5} Nedir? Neden Tehlikelidir?. <http://www.eeu.com.tr/pm2-5-nedir-neden-tehlikelidir/> adresinden 28 Nisan 2023’de alınmıştır.

- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22-30. DOI: 10.1109/MCSE.2011.37
- Walid, A., & Alamsyah, I. U. (2017). Recurrent neural network for forecasting time series with long memory pattern. *J. Phys.: Conf. Ser.*, 824(1), 012038.
- Waskom, M. L. (2021). Seaborn: Statistical Data Visualization. *Journal of Open Source Software*, 6(60), 3021. DOI: 10.21105/joss.03021
- Wikipedia (2023, 7 Nisan). Londra Öldüren Sis'i. https://tr.wikipedia.org/wiki/Londra_%C3%96ld%C3%BCren_Sis%27i#cite_note-EHP_112_1-2
- World Health Organization (2005). WHO Air Quality Guidelines For Particulate Matter, Ozone, Nitrogen Dioxide And Sulfur Dioxide. https://apps.who.int/iris/bitstream/handle/10665/69477/WHO_SDE_PHE_OEH_06.02_eng.pdf?sequence=1&isAllowed=y
- World Health Organization (2021). WHO Air Quality Guidelines For Particulate Matter, Ozone, Nitrogen Dioxide And Sulfur Dioxide. <https://apps.who.int/iris/bitstream/handle/10665/345329/9789240034228-eng.pdf?sequence=1&isAllowed=y>
- World Health Organization (2022, 17 Aralık). Air pollution. <https://www.who.int/health-topics/air-pollution>
- Xue, T., Tong, M., Li, J., Wang, R., Guan, T., Li, J., Li, P., Liu, H., Lu, H., Li, Y., & Zhu, T. (2022). Estimation of Stillbirths Attributable to Ambient Fine Particles in 137 Countries. *Nature Communications*, 13(1) 6950. <https://doi.org/10.1038/s41467-022-34250-4>
- Yalçın, B. (2021, Ocak 23). Makine Öğreniminde Veri Ön İşlemedeki Temel İşlemler ve Uygulanması. *Mert Mekatronik*. <https://mertmekatronik.com/makine-ogreniminde-kullanilan-kutuphaneler-ve-veri-on-isleme-adimlari>
- Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, 169, 114513. DOI: 10.1016/j.eswa.2020.114513
- Yavuz, A. (t.y.) Zaman serisi analizi, Hatay Mustafa Kemal Üniversitesi. <http://www.mku.edu.tr/files/339-8868446a-e4c0-4070-8530-b8beccc1b0ec.pdf>
- Yılmaz, Z., Karagözoğlu, M. B., Köker, B., ve Selçuk, S. F. (2022). Limit Values for the Density of Pollutant Parameters in the Atmosphere: Sample Study Air Pollution of Mus Province. *Kırklareli Üniversitesi Mühendislik ve Fen Bilimleri Dergisi*, 8-1 (2022) 69-87. <https://doi.org/10.34186/klujes.1109956>
- Yu, Y., Si, X., Hu, C. and Zhang, J. (2019). A Review of Recurrent Neural Networks:LSTM Cells and Network Architectures. *Neural Computation* 31, 1235–1270 DO:10.1162/neco_a_01199
- Yu, Y., Zou, W. W., Jerrett, M., & Meng, Y.-Y. (2023). Acute Health Impact of Convectional and Wildfire-related PM2.5: A narrative review. *Environmental Advances*, 12(2023) 100179. <https://doi.org/10.1016/j.envadv.2022.100179>

- Zencirci, S. A. ve Işıklı, B. (2017). Hava kirliliği. *Türk Dünyası Uygulama ve Araştırma Merkezi Halk Sağlığı Dergisi*. 2 (2) 24-36 ISSN: 2564-6311
- Zhang, Y., Thorburn, P.J., 2022. Handling missing data in near real-time environmental monitoring: a system and a review of selected methods. *Futur. Gener. Comput.Syst.* 128, 63–72. <https://doi.org/10.1016/j.future.2021.09.033>.
- Zoph, B., Vaswani, A., May, J., & Knight, K. (2016). Simple, fast noise-contrastive estimation for large rnn vocabularies. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1217-1222.
- Zvornicanin, E. (2022, Ocak 25). Differences Between Bidirectional and Unidirectional LSTM | Baeldung on Computer Science. <https://www.baeldung.com/cs/bidirectional-vs-unidirectional-lstm>





ÖZGEÇMİŞ

Ad-Soyad : İpek AKSANGÜR

ÖĞRENİM DURUMU:

- **Lisans** : 2009, Sakarya Üniversitesi, Mühendislik Fakülte, Çevre Mühendisliği
- **Yükseklisans** : 2015, Uludağ Üniversitesi, Çevre Mühendisli Anabilim Dalı, Çevre Mühendisliği Programı

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2010-2019 yılları arasında Çevre Danışmanlık firmalarında Çevre Mühendisi olarak çalıştı.

TEZDEN TÜRETİLEN ESERLER:

- **Eren, B., Aksangur, İ., Erden C.** 2023. Predicting Next Hour Fine Particulate Matter (PM2.5) in The Istanbul Metropolitan City Using Deep Learning Algorithms With Time Windowing Strategy, *Urban Climate*, 48 (2023) 101418 [DOI:10.1016/j.uclim.2023.101418](https://doi.org/10.1016/j.uclim.2023.101418)
- **Aksangür, İ., Eren, B., Erden, C.,** 2022. Evaluation of data preprocessing and feature selection process for prediction of hourly PM10 concentration using long short-term memory models. *Environmental Pollution* - Vol.311 - ISSN : 0269-7491 - [DOI : 10.1016/j.envpol.2022.119973](https://doi.org/10.1016/j.envpol.2022.119973)