



İstatistiksel ve Makine Öğrenmesi Modelleri Kullanılarak Yer Seviyesi Ozon Konsantrasyonunun Tahmin Edilmesi: İzmir, Türkiye

Hakan ÇELİKİTEN^{1*}

¹Kafkas Üniversitesi, Mühendislik Mimarlık Fakültesi, İnşaat Mühendisliği Bölümü, Kars, Türkiye

Makale Bilgisi

Research article
Received: 15/05/2026
Revision: 07/06/2026
Accepted: 08/06/2026

Anahtar Kelimeler

Ozon Tahmini
İzmir
Çoklu Doğrusal Regresyon
Makine Öğrenmesi

Article Info

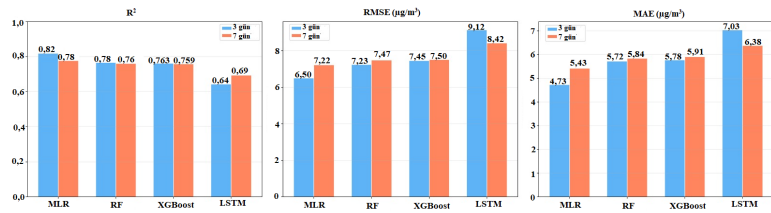
Araştırma makalesi
Başvuru: 15/05/2026
Düzeltilme: 07/06/2026
Kabul: 08/06/2026

Keywords

Ozone prediction
İzmir
Multiple Linear Regression
Machine Learning

Grafik Özet (Graphical/Tabular Abstract)

İzmir'e ait günlük meteorolojik ve NO_x verileri kullanılarak ozon tahmin modelleri geliştirilmiştir. Çoklu doğrusal regresyon ve makine öğrenmesi yöntemlerinin ozon tahmin performansları karşılaştırılmıştır. / Ozone prediction models were developed using daily meteorological and NO_x data from İzmir. The ozone prediction performance of multiple linear regression and machine learning methods was compared.



Şekil A: Modellerin ozon tahmin performansları karşılaştırması / Figure A: Performance comparison of models in ozone prediction.

Önemli Noktalar (Highlights)

- İzmir ilinde iki yıllık günlük meteoroloji ve NO_x verileri (2024–2025) ile Ozon tahmin edildi. / Two years of daily meteorological and NO_x data (2024–2025) from İzmir province were used to predict ozone levels.
- Ozon tahmini için Çoklu Doğrusal Regresyon, Rastgele Orman, XGBoost ve Uzun Kısa Süreli Bellek modelleri değerlendirildi. / Multiple Linear Regression (MLR), Random Forest (RF), XGBoost, and Long Short-Term Memory (LSTM) models were evaluated for ozone prediction.
- Regresyon ve makine öğrenmesi modelleri, yapılandırılmış gecikmeli girdiler kullanarak tahmin doğruluğunu daha da iyileştirdi. / Regression and machine learning models further improved prediction accuracy by using structured lagged inputs.

Amaç (Aim): Bu çalışmanın amacı, İzmir şehrine ait günlük meteorolojik parametreler ve NO_x verilerini kullanarak ozon (O₃) konsantrasyonlarını yüksek doğrulukla tahmin etmektir. Ayrıca girdi değişkenlerinin modellerin ozon konsantrasyon tahmin performansında göreceli katkısını belirlemektir. / The aim of this study is to predict ozone concentrations with high accuracy using daily meteorological parameters and NO_x data from city of İzmir. Another objective is to determine the relative contribution of the input variables to the ozone concentration prediction performance of the models.

Özgünlük (Originality): Çalışmanın özgün yönü, Çoklu Doğrusal Regresyon, Random Forest, XGBoost ve LSTM modellerinin aynı veri seti üzerinde karşılaştırılması ve yapılandırılmış gecikmeli girdilerin tahmin performansına etkisinin değerlendirilmesidir. / The novelty of this study lies in the comparison of Multiple Linear Regression, Random Forest, XGBoost, and LSTM models on the same dataset, as well as the evaluation of the effect of structured lagged inputs on prediction performance.

Bulgular (Results): Ozon tahmini için; MLR, RF, XGBoost ve LSTM modellerinde 3 ve 7 gün geriye bakış → 1 gün ileriye bakış olmak üzere iki aşamalı tasarım uygulanmıştır. 3 günlük geçmiş O₃ hareketli ortalamaları ile MLR, XGBoost, RF ve LSTM modelleri ile R² değerleri sırasıyla 0,82; 0,76; 0,78 ve 0,70 olarak elde edilmiştir. / For ozone forecasting, a two-stage design was applied to the MLR, RF, XGBoost, and LSTM models, with a 3- and 7-day retrospective look and a 1-day forward look. With 3-day historical ozone moving averages, the R² values for the MLR, XGBoost, RF, and LSTM models were obtained as 0.82, 0.76, 0.78, and 0.70, respectively.

Sonuç (Conclusion): Gecikmeli özniteliklerin (3 ve 7 günlük gecikmeler) ve hareketli ortalamaların (3 ve 7 günlük) modellere dahil edilmesinin, modellerin O₃ tahmin performansını önemli ölçüde iyileştirdiği görülmüştür. / The inclusion of lag features (3-day and 7-day lags) and moving average features (3-day and 7-day moving averages) significantly enhanced the O₃ prediction performance of the models.



İstatistiksel ve Makine Öğrenmesi Modelleri Kullanılarak Yer Seviyesi Ozon Konsantrasyonunun Tahmin Edilmesi: İzmir, Türkiye

Hakan ÇELİKİTEN^{1*} ¹Kafkas Üniversitesi, Mühendislik Mimarlık Fakültesi, İnşaat Mühendisliği Bölümü, Kars, Türkiye

Makale Bilgisi

Araştırma makalesi
Başvuru: 15/05/2026
Düzeltilme: 07/06/2026
Kabul: 08/06/2026

Anahtar Kelimeler

Ozon Tahmini
İzmir
Çoklu Doğrusal
Regresyon
Makine Öğrenmesi

Öz

Ozon (O₃), özellikle büyük şehirlerde halk sağlığı açısından önemli risklere ve küresel ısınmaya neden olan sera gazıdır. Bu çalışma, 2024 ile 2025 yılları arasında Türkiye'nin İzmir ilinde ölçülen günlük konsantrasyon verilerinden ozon konsantrasyon değerlerini tahmin etmeyi amaçlamaktadır. Analiz, meteorolojik parametrelere (sıcaklık, rüzgar hızı, bağıl nem, yağış, bulut örtüsü, güneşlenme süresi ve güneş radyasyonu) ve azot oksit konsantrasyonlarına dayalı olarak bir gün sonraki ozon konsantrasyonlarını tahmin etmeye dayanmaktadır. Tahmin için; Çoklu Doğrusal Regresyon (MLR), Rastgele Orman (RF), Aşırı Gradyan Artırma (XGBoost) ve Uzun Kısa Süreli Bellek (LSTM) modelleri kullanılmıştır. Tahmin performansını artırmak için, gecikmeli modeller geliştirilmiş ve değerlendirilmiştir. Model performansı, determinasyon katsayısı (R²), kök ortalama kare hatası (RMSE) ve ortalama mutlak hata (MAE) metrikleri kullanılarak değerlendirilmiştir. Çalışma; sadece meteorolojik parametreler ve NO_x konsantrasyonlarının girdi değişkeni olduğu durumlarda modellerin tahmin performanslarının düşük olduğunu ortaya koymuştur. Bu amaçla modellerde 3 ve 7 gün geriye bakış → 1 gün ileriye bakış olmak üzere iki aşamalı tasarım uygulanmıştır. Ozon tahmininde en yüksek doğruluk R²=0,82, RMSE=6,50 ve MAE=4,73 değerleriyle MLR modeliyle elde edilmiştir. Modellerden elde edilen özellik önem analizi, geçmiş 3 ve 7 günlük O₃ verilerinin en etkili tahmin edici olarak belirlemiştir. Makine ve derin öğrenme algoritmalarının daha karmaşık ve veri setinin çok olduğu durumlarda kullanması önerilmektedir. Ayrıca O₃ tahmin modellerinde trafik yoğunluğu ve uçucu organik bileşik konsantrasyonlarının modellere dahil edilmesinin daha iyi performans metrikleri sunması öngörülmektedir.

Prediction Ground-Level Ozone Concentrations Using Statistical and Machine Learning Models: Izmir, Türkiye

Article Info

Research article
Received: 15/05/2026
Revision: 07/06/2026
Accepted: 08/06/2026

Keywords

Ozone prediction
Izmir
Multiple Linear
Regression
Machine Learning

Abstract

Ozone (O₃) is a greenhouse gas that poses significant risks to public health, particularly in large cities, and contributes to global warming. This study aims to identify ozone trends and forecast concentration values based on daily concentration data measured in Izmir, in western Turkey, between 2024 and 2025. The analysis is based on predicting the next day's ozone concentrations using meteorological parameters (temperature, wind speed, relative humidity, precipitation, cloud cover, sunshine duration, and solar radiation) and nitrogen oxide concentrations. For the prediction, Multiple Linear Regression (MLR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) models were used. To improve forecasting performance, lagged models were developed and evaluated. Model performance was assessed using the coefficient of determination (R²), root mean square error (RMSE), and mean absolute error (MAE) metrics. The study revealed that model prediction performance was low when only meteorological parameters and NO_x concentrations were used as input variables. To address this, a two-stage design was applied to the models, consisting of a 3-day and 7-day lookback followed by a 1-day lookahead. The highest accuracy in ozone forecasting was achieved with the MLR model, yielding R² = 0.82, RMSE = 6.50, and MAE = 4.73. The feature importance analysis derived from the models identified both the 3-day lagged ozone concentration (O_{3_lag3}) and the 7-day lagged ozone concentration (O_{3_lag7}) as the most effective predictors. The use of machine learning and deep learning algorithms is recommended for more complex scenarios and situations with large datasets. Additionally, it is anticipated that incorporating traffic density and volatile organic compound concentrations into O₃ prediction models will yield better performance metrics.

1. GİRİŞ (INTRODUCTION)

Kentsel hava kirliliği, özellikle nüfus yoğunluğunun yüksek olduğu yerlerde hızla artmakta ve ciddi çevresel sorunlara neden olmaktadır [1]. Önemli hava kirleticilerinden biri olan ozon (O_3), azot oksitlerle (NO_x) ve uçucu organik bileşikler (VOC'ler) gibi öncül kirleticilerin katıldığı karmaşık fotokimyasal reaksiyonlar sonucunda oluşan reaktif bir gazdır [2]. Ozon azot oksitlerle fotokimyasal reaksiyona girerek azot dioksit (NO_2) ve oksijen (O_2) oluşturur [3]. Bu reaksiyon troposferde düşük ozon konsantrasyon seviyelerine yol açar. Ancak, insan kaynaklı VOC'ler hidroksil radikalleri ile reaksiyona girerek troposferde ozon moleküllerinin oluşumuna ve birikmesine yol açar [4].

Son yıllarda, hızlı sanayileşme, kentleşme, artan araç emisyonları ve küresel ısınma, özellikle gelişmekte olan ve sanayileşen bölgelerde yüzey seviyesindeki O_3 konsantrasyonlarında sürekli bir artışa katkıda bulunmuştur [5]. Sıcaklık yükseldikçe ve güneş radyasyonunun şiddeti arttıkça, O_3 oluşumu da kademeli olarak artmaktadır [6]. Ozon konsantrasyonunun aşırı yükselmesi ekolojik çevre ve insan sağlığı üzerinde de olumsuz etkilere yol açmaktadır [7]. Stratosferik O_3 güneşten gelen zararlı ultraviyole (UV) radyasyonunu emerek yaşamı korurken, troposferik veya yer seviyesindeki O_3 canlı sistemler için son derece zararlıdır [8].

Ozon önemli bir hava kirleticisi ve sera gazıdır. Küresel ısınma nedeniyle, dünyanın dört bir yanında aşırı sıcaklık olaylarında önemli ve sürekli bir artış meydana gelmekte olup [6], bu olaylar sosyoekonomik kalkınma, insan sağlığı, nüfus göçü, gıda güvenliği, karasal ve sucul ekosistemler üzerinde önemli etkilere sahiptir [9]. Daha yüksek sıcaklıklar, solunum sistemini tahriş eden ve astım ve kronik bronşit gibi durumları kötüleştiren daha fazla yer seviyesinde ozon oluşumuna yol açabilir [10]. İnsanların ozona kısa vadeli maruziyetinin, özellikle solunum ve kardiyovasküler sistemler üzerinde ve uzun vadeli maruziyetin de solunum yolu ölümleri, çocuklarda yeni başlayan astım ve astımlılarda artan solunum semptom etkileri ile ilgili ilişkiler bildirilmiştir [11]. Ayrıca bazı çalışmalar, ozonun epilepsi üzerindeki etkisini vurgulayarak, çevresel faktörlerin nöbetlerin patofizyolojisine katkıda bulunduğu karmaşık mekanizmaları ortaya koymaktadır [12]. Troposferik ozonun; bitkiler, ormanlar, tarım

ürünleri ve mikroflora üzerinde de önemli etkileri vardır. Bitkilerde nekrotik lezyonlar, kloroz, bodur bitki büyümesi, kızarma, bronzlaşma, beneklenme vb. gibi hasarın kesin belirtileri de dahil olmak üzere çeşitli hasarlara sebep olur [8].

Ozonun ekosistem ve insan sağlığı üzerindeki olumsuz etkilerinden dolayı, hava kalitesinin etkin yönetimi ve erken uyarı sistemlerinin geliştirilmesi için ozon konsantrasyonlarının doğru şekilde tahmin edilmesi büyük önem taşır. Günümüzde ozon üzerine yapılan çalışmalar iki ana başlıkta toplanmaktadır. İlk yaklaşım, ozon konsantrasyonundaki değişimlerin insan sağlığı, ekolojik çevre ve tarım üzerindeki etkilerini incelemektedir. İkinci yaklaşım ise, ozon konsantrasyonunun tahmini ve erken uyarı sistemlerinin geliştirilmesine odaklanmaktadır [7]. Bu amaçla literatürde gerçek zamanlı hava kirliliği konsantrasyonlarını tahmin etmek için deterministik modeller, istatistiksel yöntemler ve makine öğrenmesi ve yapay zeka tabanlı modeller yaygın olarak kullanılmaktadır. Geleneksel istatistiksel modellerde Doğrusal Regresyon (LR), Çoklu Doğrusal Regresyon (MLR), Doğrusal Olmayan Genişletme (nonlinear extension), geleneksel otoregresif entegre hareketli ortalama (ARIMA) gibi teknikler kullanılmaktadır [2], [13]. Buna karşılık son yıllar da Yapay Sinir Ağları (ANN), Random Forest (Karar Ağaçları), Destek Vektör Makineleri (SVM), Uzun Kısa Süreli Bellek (LSTM), Derin evrimsel sinir ağı (CNN) gibi makine öğrenmesi ve yapay zeka algoritmaları hava kalitesi ve kirletici konsantrasyonlarının tahmininde başarılı şekilde uygulanmıştır [2], [14], [15]. Ayrıca modellerin tahmin sınırlamalarını aşmak ve ozon konsantrasyonunun doğru ve etkin tahmini için hibrit modeller de geliştirilmiştir [7], [16], [17].

Bu çalışmada, Türkiye'nin Ege bölgesinde bulunan nüfus ve sıcaklığın yüksek olmasından dolayı İzmir ili seçilmiştir. Şehirdeki sabit hava kalitesi izleme istasyonundan elde edilen 2024-2025 yılları arasındaki NO_x konsantrasyonu ve sıcaklık, rüzgar hızı, nispi nem, yağış, güneşlenme süresi, güneş radyasyonu ve bulutluluk verileri kullanılarak günlük ozon konsantrasyonu tahmin edilmiştir. Çalışmada Çoklu Doğrusal Regresyon (MLR), Rastgele Orman (RF) ve Uzun Kısa Süreli Bellek (LSTM) modelleri kullanılmış ve model performansları değerlendirilmiştir. Bu araştırmanın temel amacı, girdi değişkenlerinin modellerin ozon konsantrasyon tahmin performansında göreceli

Bu çalışmada, İzmir'deki O₃ konsantrasyonlarının modellenmesinde birincil girdiler arasında meteorolojik değişkenler dikkate alınmıştır. Buna göre; günlük olarak ortalama rüzgar hızı (m/s), hava sıcaklığı (°C), toplam yağış (mm), nispi nem (%), bulut örtüsü (sekiz okta) güneş radyasyonu (kwsaat/m²) ve günlük güneşlenme süresi (saat) verileri kullanılmıştır. Analizde kullanılan meteorolojik veriler, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı bünyesindeki Türkiye Devlet Meteoroloji Servisi tarafından işletilen Meteorolojik Veri Bilgi Satış ve Sunum Sistemi (MEVBIS) aracılığıyla Güzelyalı da bulunan İzmir Meteoroloji Bölge Müdürlüğü'nün istasyonundan günlük olarak elde edilmiştir.

2.2. Ön Veri Analizi (Preliminary Data Analysis)

Bu çalışmada kullanılan veri setinin yapısını ve bütünlüğünü incelemek için ön istatistiksel analiz

yapılmıştır. O₃, NO_x (NO+NO₂), günlük ortalama rüzgâr hızı (m/s), hava sıcaklığı (°C), toplam yağış (mm), bağıl nem (%), bulut örtüsü (sekiz okta) ve günlük güneşlenme süresi (saat) olmak üzere temel parametrelerin özet istatistikleri verilmiştir (Tablo 1). Her değişken için geçerli gözlem sayısı, ortalama, standart sapma ve minimum/maksimum değerler bildirilmiştir. Her değişken için geçerli gözlem sayısı, ortalama, standart sapma ve minimum/maksimum değerler rapor edilmiştir. Sadece altı meteorolojik parametre ile ve bu meteorolojik parametrelere ek olarak NO_x O₃ değerlerini çıktı olarak tahmin etmeyi amaçlayan modellere girdi değişkenleri olarak sağlanmıştır. Eksik değerler doğrusal interpolasyon ile doldurulmuştur. Tablo 1'de bildirilen 25. yüzdellik ve 75. yüzdellik değerleri, çeyrekler arası aralık yöntemi kullanılarak aykırı değerleri belirlemek için kullanılmıştır.

Table 1. Çalışmada kullanılan kirletici konsantrasyonları ve meteorolojik parametrelere ilişkin tanımlayıcı istatistiklerin özeti (Descriptive statistical summary of the pollutant concentrations and meteorological variables employed in this study)

	O ₃ (µg/m ³)	NO _x (µg/m ³)	Güneşlenme süresi (saat)	(8 okta)	Nispi nem (%)	Rüzgâr hızı (m/s)	Yağış (mm)	Sıcaklık (°C)	Güneş Radyasyonu (kW saat/m ²)
Sayı	732	732	732	732	732	732	732	732	732
Ortalama	22,62	55,55	8,77	2,25	54,32	2,69	1,31	18,25	4,90
Standart sapma	13,04	34,23	4,10	2,12	14,68	0,82	4,98	7,72	2,38
Min.	2,27	15,48	0	0	21,1	0,9	0	0,4	0
25%	12,79	30,26	6,25	0,2	43,25	2,1	0	12	2,8
75%	28,91	73,17	12,40	4,00	64,75	3,20	0	25,50	7,20
Maks.	60,04	212,6	14,3	8	93,3	5,6	45,7	34	9

Tablo 1'de gösterildiği gibi, veri seti 732 gözlem içermektedir. Günlük ortalama rüzgâr hızı 2,69 m/s olarak kaydedilirken, ortalama nispi nem yaklaşık %54,32 ve ortalama güneşlenme süresi 8,77 saat olarak belirlenmiştir. Günlük ortalama sıcaklık değerleri 0,4°C ile 34°C arasında değişmekte olup, bölgede havanın sıcak olduğunu göstermektedir. Bulutluluk ise 0 ile 8 okta arasında değişmektedir. Günlük toplam global güneş radyasyonu da 0 ile 9 kW saat/m² arasında değişmiş olup ortalama 4,90'dır. Bu tanımlayıcı istatistikler, model

geliştirme öncesinde veri seti hakkında temel bir anlayış sağlamaktadır.

2.3. Model Tasarımı ve Tahmin Çerçevesi (Model Design and Prediction Framework)

Çalışma kapsamında O₃ tahmin performansını değerlendirmek için dört tahmin modeli; Çoklu Doğrusal Regresyon (MLR), Rastgele Orman (RF), Aşırı Gradyan Artırma (XGBoost) ve Uzun Kısa Süreli Bellek (LSTM) uygulanmıştır. MLR, bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkiyi, eğitim verilerine dayalı

parametreleri tahmin ederek modellemek için kullanılan temel bir istatistiksel yöntemdir. RF ve XGBoost, sınıflandırma ve regresyon dahil olmak üzere çeşitli makine öğrenimi görevlerini çözmek için uygun, ölçeklenebilir, ağaç tabanlı algoritmalarıdır. LSTM ise uzun veri dizileri üzerinde bilgi öğrenmek ve hatırlamak için tasarlanmış özel bir tekrarlayan sinir ağı türüdür. Model çalışma prensipleri ve temel algoritmaları detaylı olarak Çelikten, H. (2026)'da açıklanmıştır [15], [22].

Tahmin modelleri, farklı senaryolarda çalıştırılmış ve model performansları ortalama mutlak hata (MAE), kök ortalama kare hatası (RMSE) ve determinasyon katsayısı (R^2) gibi ölçütler kullanılarak değerlendirilmiştir (Denklem 1, 2, 3).

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (3)$$

Yukarıdaki denklemlerde, y_i , i-inci veri noktasının gerçek değerini temsil ederken, \hat{y}_i karşılık gelen tahmin edilen değeri ifade eder. \bar{y}_i terimi, zaman serisinin anlamını ve n değişken sayısını temsil eder. Tahmin performanslarının karşılaştırılması için klasik ve geliştirilmiş (Lag-3 gün, Lag-7 gün) modeller çalıştırılmıştır. Modeller, t anındaki 3 ve 7 günlük geçmiş zaman penceresi kullanılarak, bir

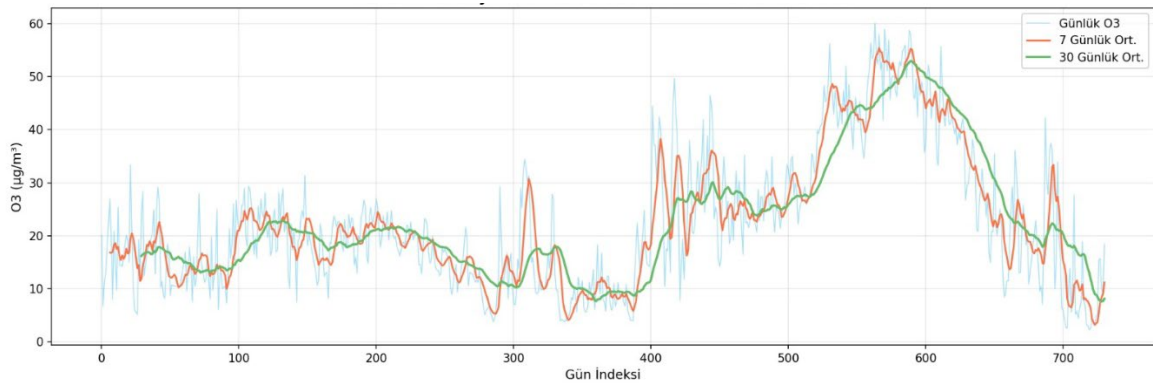
gün sonraki Ozon değerlerini tahmin etmek üzere tasarlanmıştır. Veriler kronolojik bölme yöntemleri kullanılarak bölümlere ayrıldı; gözlemlerin ilk %80'i (daha önceki dönem) eğitim için, kalan %20'si (en yeni dönem) ise test için kullanılmıştır.

Senaryo 1: Modellere girdi değişkeni olarak sadece meteorolojik parametreler (sıcaklık, nispi nem, yağış, bulutluluk, güneş radyasyonu, güneşlenme süresi ve rüzgar hızı) eklenmiştir.

Senaryo 2: Modellere girdi değişkeni olarak meteorolojik parametrelerin yanı sıra NO_x konsantrasyonları ve hedef değişkenin (O_3) geçmiş değerleri (gecikmeli değişkenler) ilave edilmiştir.

3. BULGULAR (RESULTS)

Bu çalışmada, İzmir şehrindeki hava kalitesi izleme istasyonundan günlük olarak tutarlı ve kesintisiz Ozon verilerinin mevcut olduğu 2024-2025 dönemi incelenmiştir. İstatistiksel analiz, çalışma döneminde ozonun günlük ortalama konsantrasyonunun $22,62 \mu\text{g}/\text{m}^3$ olduğunu ve minimum $2,27 \mu\text{g}/\text{m}^3$ ile maksimum $60,04 \mu\text{g}/\text{m}^3$ arasında değiştiğini ortaya koymuştur (Şekil 2). $13,04 \mu\text{g}/\text{m}^3$ 'lük standart sapma, ölçüm serisi boyunca değişkenlik olduğunu göstermektedir. Ozon için ölçüm periyodu sırasında Ulusal Hava Kalitesi Sınır Değeri (maksimum günlük 8 saatlik ortalama: $120 \mu\text{g}/\text{m}^3$) aşılmamıştır. Ortalama NO_x konsantrasyonu ise $55,55 \mu\text{g}/\text{m}^3$ olarak kaydedilmiş olup, gözlemlenen değerler $15,48 \mu\text{g}/\text{m}^3$ ile $212,6 \mu\text{g}/\text{m}^3$ arasında değişmektedir.



Şekil 2. İzmir - Eğitim istasyonu O_3 konsantrasyon dağılımı (Distribution of O_3 concentrations at the Izmir training station)

İzmir şehrindeki ozon konsantrasyonlarının analizi, belirgin mevsimsel değişkenlik göstermektedir; kış

aylarında daha düşük değerler, yaz aylarında ise daha yüksek konsantrasyonlar gözlemlenmiştir. Bu

mevsimsel deęişiklik, meteorolojik koşulların ozon seviyelerini etkilediğini düşündürmektedir. Bu dinamikleri yakalamak için, modelleme sürecinde kapsamlı bir girdi deęişkenleri kümesi kullanılmıştır; bunlar arasında meteorolojik parametreler (sıcaklık, nispi nem, rüzgar hızı, yağış, güneşlenme süresi, güneş radyasyonu ve bulut örtüsü) ile birlikte eş kirletici konsantrasyonu (NO_x) yer almaktadır.

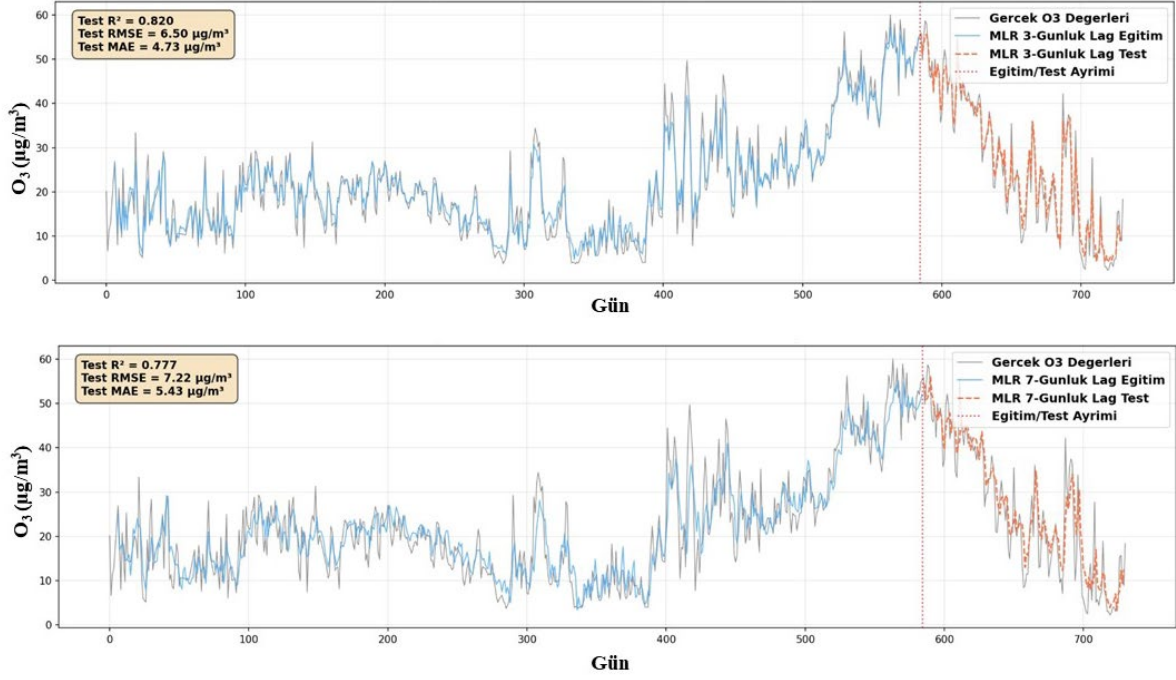
Ozon tahmin performansını deęerlendirmek için dört tahmin modeli Çoklu Doğrusal Regresyon (MLR), Rastgele Orman (RF), Aşırı Gradyan Artırma (XGBoost) ve Uzun Kısa Süreli Bellek (LSTM) uygulanmıştır. Tablo 2 de, klasik modeller (girdi deęişkeni temel meteorolojik parametreler ve temel meteorolojik parametreler + NO_x) için O_3 tahmin performansları gösterilmiştir.

Tablo 2. Meteorolojik parametrelere dayalı O_3 tahmin modellerinin performans metrikleri (Performance metrics of O_3 prediction models based on meteorological parameters)

Model	RMSE	MAE	R^2
(Temel Meteorolojik Parametreler)			
MLR	13,08	10,42	0,28
RF	14,15	11,36	0,16
LSTM	13,01	10,36	0,29
Model			
(Temel Meteorolojik Parametreler + NO_x)			
MLR	12,84	10,25	0,31
RF	14,65	11,81	0,10
LSTM	12,79	10,16	0,31

Temel meteorolojik parametreler girdi deęişkeni olduğunda LSTM diğer modellere kıyasla düşük hata deęerlerine sahip (RMSE: 13,01; MAE: 10,36) ve en yüksek R^2 deęerini (0,29) üretmiştir. MLR sonuçları da LSTM'ye oldukça yakın çıkmıştır. RF modeli ise daha yüksek hata ve daha düşük R^2 (0,16) ile en zayıf performansı göstermiş. NO_x deęişkeni eklendiğinde MLR ve LSTM modellerinde küçük ama tutarlı bir iyileşme görülmüştür. RF modelinin düşük performansı, veri yapısının karar ağacı tabanlı yaklaşım için yeterince uygun olmadığını göstermektedir. Ayrıca tüm modellerde R^2 deęerlerinin düşük kalması (~ 0.3) ozon oluşumunun oldukça karmaşık bir süreç olduğunu ve ozon seviyelerinin yalnızca kullanılan deęişkenlerle tam açıklanamadığını ortaya koymuştur.

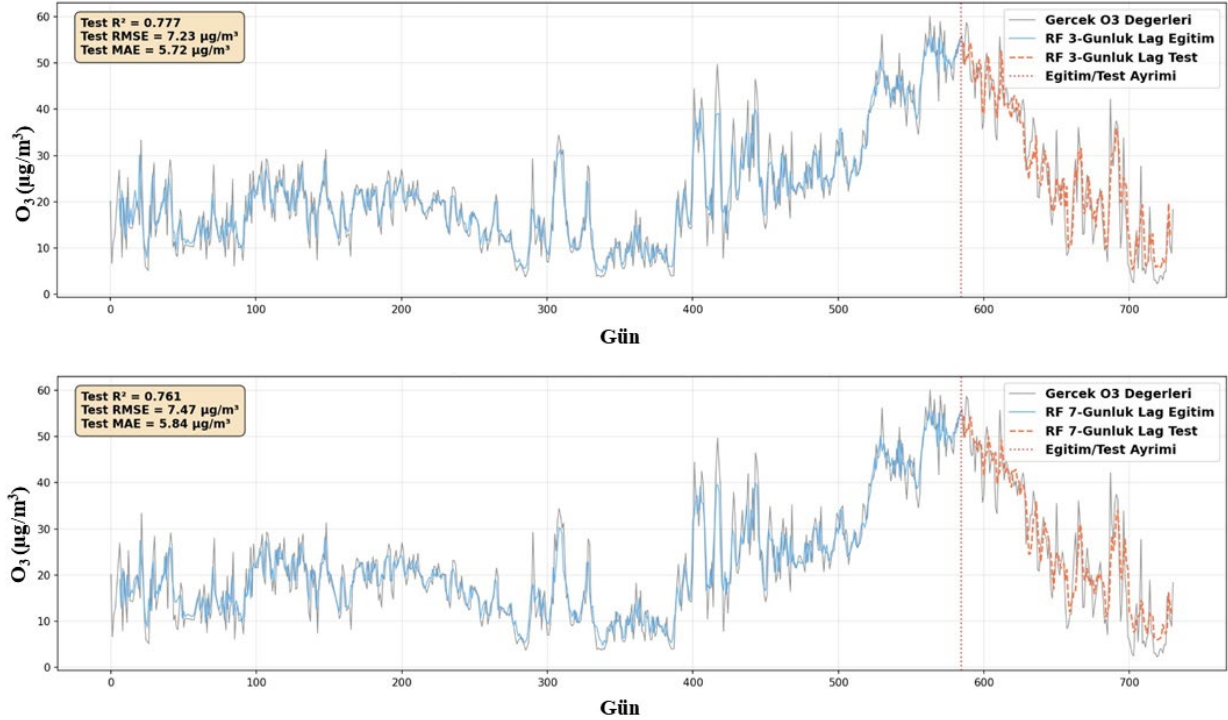
Model performanslarının düşük olması nedeniyle geliştirilmiş modellerde girdi deęişkeni olarak meteorolojik parametreler, NO_x konsantrasyonu ve O_3 konsantrasyonlarının otokorelasyonunu (geçmişe bağımlılığını) yansıtan 3 ve 7 günlük gecikmeli deęişkenler (Lag 3 – Lag 7) modele dahil edilmiştir. 3 ve 7 günlük lag deęişkenlerinin modellere dahil edilmesi ile zamansal bağımlılıkların temsil edilmesi sağlanmıştır. Şekil 3'te, MLR modeli kullanılarak gözlemlenen ve tahmin edilen O_3 konsantrasyonlarını karşılaştıran bir dağılım grafięi sunmaktadır; X ekseninde 2024-2025 yıllarındaki günler (1 Ocak 2024 – 31 Aralık 2025), Y ekseninde O_3 konsantrasyon deęerleri yer almaktadır.



Şekil 3. O₃ tahmini için MLR model sonuçları (Lag 3 ve Lag 7 gün) (MLR model results for O₃ prediction using 3-day and 7-day lag features)

Şekil 3'te gösterildiği gibi, MLR (Lag - 3 ve Lag - 7 gün) modelinin klasik konfigürasyonlara kıyasla yeniden eğitildiğinde tahmin doğruluğunda önemli bir iyileşme olduğunu göstermektedir. Lag 3 günlük projeksiyon için O₃, minimum hata (RMSE ≈ 6,5

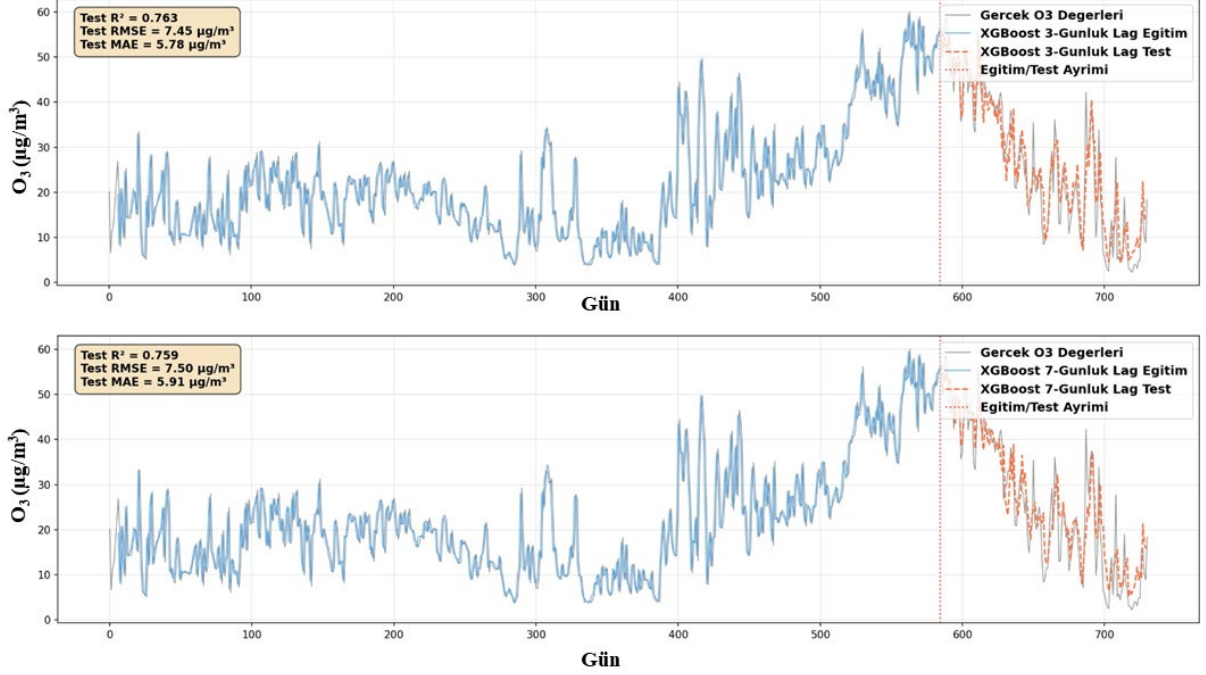
µg/m³) ile iyi bir doğruluk (R² ≈ 0,82) elde ederken, Lag 7 güne göre daha iyi bir performans sergilemektedir (R² ≈ 0,77, RMSE ≈ 7,22). Benzer şekilde RF modeli için elde edilen sonuçlar Şekil 4'te gösterilmiştir.



Şekil 4. O₃ tahmini için RF model sonuçları (Lag 3 ve Lag 7 gün) (RF model results for O₃ prediction using 3-day and 7-day lag features)

Şekil 4'te görüldüğü üzere RF modeli için Lag 3 günlük eğitim sonunda R^2 , RMSE ve MAE değerleri sırasıyla 0,77; 7,23 ve 5,72 olarak elde edilmiştir. Buna karşılık geçmiş 7 günlük O_3 verilerine dayalı olarak bir sonraki gün tahmin sonuçları incelendiğinde model performansı azda olsa düşmüştür ($R^2_{Lag7} = 0,76$, $RMSE_{Lag7} = 7,47$ ve $MAE_{Lag7} = 5,84$). Klasik modellerle kıyaslandığında ($R^2 = 0,16$) 3

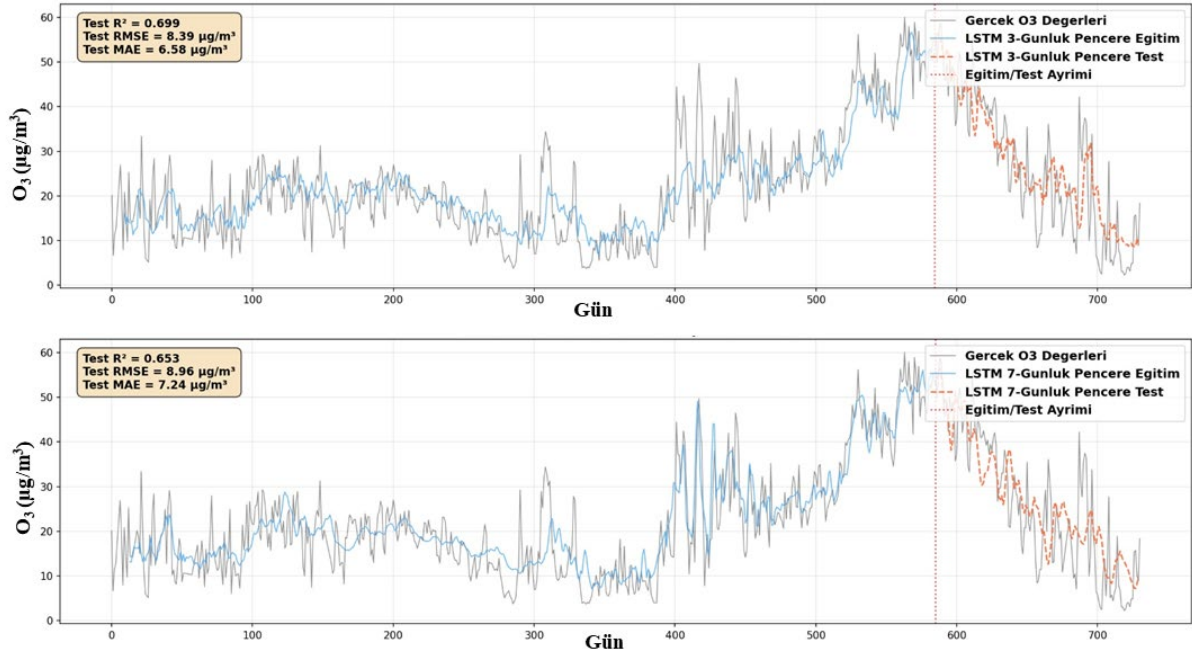
günlük Lag eğitiminin RF model performansını önemli derecede iyileştirildiği görülmüştür. XGBoost modeli için 3 ve 7 günlük model eğitiminin sonuçları Şekil 5'te gösterilmiştir.



Şekil 5. O_3 tahmini için XGBoost model sonuçları (Lag 3 ve Lag 7 gün) (XGBoost model results for O_3 prediction using 3-day and 7-day lag features)

XGBoost modeli her iki gecikme yapısında da benzer performans göstermiştir. Lag-3 modeli için $R^2=0,76$, $RMSE=7,45$ ve $MAE=5,78$ elde edilirken, Lag-7 modeli için $R^2=0,759$, $RMSE=7,50$ ve $MAE=5,91$ bulunmuştur. Sonuçlar, Lag-3

modelinin hata metrikleri açısından çok az daha iyi performans gösterdiğini, ancak iki model arasında genel olarak önemli bir fark olmadığını ortaya koymaktadır. LSTM modeli ile elde edilen O_3 tahmin sonuçları da Şekil 6.'da gösterilmiştir.



Şekil 6. O₃ tahmini için LSTM model sonuçları (Lag 3 ve Lag 7 gün) (LSTM model results for O₃ prediction using 3-day and 7-day lag features)

LSTM Lag-3 modeli için $R^2=0.699$, $RMSE=8,39$ ve $MAE=6,58$ elde edilirken, Lag-7 modeli için $R^2=0,65$, $RMSE=8,96$ ve $MAE=7,24$ bulunmuştur. Sonuçlar, Lag-3 modelinin hata metrikleri açısından MLR, RF ve XGBoost ile benzer şekilde çok az daha iyi performans gösterdiğini, ancak iki model arasında genel olarak önemli bir fark olmadığını ortaya koymaktadır.

Literatürde yapılan çalışmalar incelendiğinde İsveç'te Massimo Stafoggia ve ark. 2005-2016 yılları arasında günlük ozon konsantrasyonlarını tahmin etmek için RF modeli kullanmış ve hava sıcaklığı, bulut örtüsü, atmosfer basıncı ve kar albedosu gibi çeşitli meteorolojik değişkenleri dikkate almıştır. Çalışmada R^2 değerleri 0,52- 0,65, RMSE değerleri ise 10,5-13,8 arasında değişmiştir [23]. Yılmaz, A. (2021) makine öğrenme algoritmaları ile ozon seviyesi tahmini ile ilgili yaptığı çalışmada RF, XGBoost modelleri için RMSE değerlerini sırasıyla 5,07 ile 2,65 olarak bildirmiş, hibrit bir makine öğrenmesi algoritmasını ($RMSE=1,21$) önermiştir [16]. Liu ve ark. 2013-2017 yılları arasındaki günlük maksimum 8 saatlik ortalama ozon gözlemlerini, eş zamanlı ozon geri kazanımlarını, aerosol yeniden analizini, meteorolojik parametreleri ve arazi kullanım verilerini birleştirerek, eXtreme Gradient Boosting algoritmasına dayalı ulusal bir MDA8 tahmin modeli oluşturmuşlardır. Modelde, günlük düzeyde yıllık, saha bazlı ve örnek bazlı çapraz doğrulama şemaları için R^2 değerleri sırasıyla 0,61, 0,64 ve 0,78 olarak bildirilmiştir [24]. Andrew C. Comrie

yapay sinir ağlarının kullanım potansiyelini geleneksel regresyon modelleri ile karşılaştırmalı olarak incelediği çalışmada gecikmeli veri içeren çoklu regresyon modellerinin performansının, sinir ağlarına oldukça yakın çıktığını belirtmiştir [25]. Hashim NIM. ve ark. meteorolojik veriler (bağıl nem, sıcaklık, UVB, rüzgâr hızı ve rüzgâr yönü) ve kirletici konsantrasyonlarının dahil edildiği verilerle Çoklu Doğrusal Regresyon (MLR), İleri Beslemeli Yapay Sinir Ağı, Radyal Tabanlı Fonksiyon Ağı ve bunların geliştirilmiş versiyonlarını kullanarak günlük O₃ konsantrasyonu tahmin etmişlerdir. Sonuçlara göre yüzey ozon seviyesini en iyi açıklayan model, en düşük hata değerlerine ($MAE = 6,06$; $RMSE = 7,77$) ve en yüksek R^2 değerlerine (0,91) sahip olan MLR modeli olduğunu bildirmişlerdir [26]. Bu çalışmada da benzer şekilde 3 günlük ve 7 günlük geçmiş ozon verilerinin dahil edildiği geliştirilmiş modellerde MLR modelinin diğer modellere kıyasla daha iyi performans ortaya koyduğu belirlenmiştir. Bu durum bugünkü ozon seviyesinin, önceki günlerin ozonundan önemli ölçüde etkilendiğini ve bu ilişkinin büyük ölçüde doğrusal olmasından olabilir. LSTM'nin çok büyük üstünlük sağlayamaması ise; veri miktarının sınırlı olması (saatlik değil günlük) ve lag değişkenlerinin zaten temel bilgiyi taşıması ile açıklanabilir.

Ülkemizde yapılan çalışmalar incelendiğinde, Erzincan ilinde 4 saate kadar gecikmeli değerler kullanılarak O₃ değerlerinin en doğru şekilde LSTM algoritması ($R^2 = 0,93$ ve $RMSE = 6,02$) elde

edildiği bildirilmiştir [14]. Çelik, G., ve ark. Marmara Bölgesinde (İstanbul, Sakarya ve Edirne) beş farklı regresyon algoritması ile girdi değişkenleri meteorolojik parametreler (rüzgar hızı, rüzgar yönü, nem, sıcaklık, basınç, yağış) ve kirletici konsantrasyonlarını (NO_x , $PM_{2.5}$, PM_{10} , SO_2 , NO , NO_2 ve CO) kullanmışlardır. Çalışma kapsamında en başarılı sonuçların Random Forest ($R^2 > 0,75$) ile elde edildiğini rapor etmişlerdir [27]. Ay ve İkinci; Sakarya, Kocaeli ve Çanakkale ilinin O_3 konsantrasyonlarının tahminini girdi olarak kirliliğe sebep olan parametreler (PM_{10} , SO_2 , NO , NO_2 ve O_3) kullanarak XGBoost, Yapay Sinir Ağları ve LSTM modelleri ile yapmışlardır. Eğitim kümesi (21500) ve test kümesi (4079) sayısının yüksek olduğu bu çalışmada LSTM modelini önermişlerdir [28]. Bu çalışmada MLR modelinin makine öğrenmesi algoritmalarına göre daha yüksek performans göstermesi veri sayısının az olmasından kaynaklanabilir. Bu sonuç daha basit, daha yorumlanabilir ve daha düşük hesaplama maliyetli istatistiksel modellerin hava kirliliği çalışmalarında veri setinin az ve karmaşık olmadığı kirletici konsantrasyonlarının tahmin edilmesinde hâlâ iyi alternatifler olduğunu desteklemektedir.

4. SONUÇLAR (CONCLUSIONS)

Bu çalışma, Türkiye'nin İzmir bölgesinde günlük O_3 tahminlerinin iki yıllık analizini sunmaktadır. Bu amaçla MLR, RF, XGBoost ve LSTM model performansları karşılaştırılmıştır. Kullanılan tahmin modellerinde sıcaklık, rüzgar hızı, yağış, bulutluluk, nispi nem, güneşlenme süresi, güneş radyasyonu ve değişkenleri ile modellerin performansının düşük olduğu ($R^2 \sim 0.3$) belirlenmiştir. 3 ve 7 günlük geriye dönük bir bakış açısı altında çoklu doğrusal regresyon ve makine öğrenimi algoritmaları karşılaştırılmıştır. Meteorolojik parametrelere ve NO_x girdi değişkenlerinin yanı sıra gecikmeli özelliklerin (Lag 3 ve Lag 7), hareketli ortalamaların (3 günlük ve 7 günlük) dahil edilmesinin, modellerin O_3 tahmin performansını önemli ölçüde iyileştirdiği görülmüştür. 3 günlük O_3 hareketli ortalamaları ile MLR modeli R^2 , RMSE ve MAE sırasıyla 0,82; 6,5 ve 4,73 değerleriyle en iyi performans elde etmiştir. Bu durum önceki gün ozon konsantrasyon değerleri ile bugünkü ozon seviyelerinin arasında doğrusal ilişki olduğunu ortaya koymaktadır. LSTM, XGBoost ve RF modelleri ise sırasıyla 0,70; 0,76 ve 0,78 R^2 değerleriyle birbirlerine yakın tahmin üretmiştir. Girdi veri seti sayısının az olması makine ve derin öğrenme model tahmin performanslarını etkilediği gözlenmiştir. Bu sonuç, veri setinin az ve karmaşık olmadığı durumlarda basit ve düşük hesaplama maliyetli istatistiksel modellerin kirletici

konsantrasyonlarının tahmin edilmesinde iyi alternatifler olduğunu göstermektedir. Bundan sonraki çalışmalarda ozon tahmin modellerine uçucu organik bileşik konsantrasyonları, sınır tabaka yüksekliği ve trafik yoğunluğu gibi ek parametrelerin dahil edilmesi ve saatlik veri setleriyle model tahmin performanslarının araştırılması önerilmektedir.

TEŞEKKÜR (ACKNOWLEDGMENTS)

Yazar, bu çalışmada kullanılan verilerin sağlanmasındaki desteklerinden dolayı Çevre, Şehircilik ve İklim Değişikliği Bakanlığı'na ve Devlet Meteoroloji Hizmetleri Genel Müdürlüğü'ne içtenlikle teşekkür eder. / The author would like to express sincere gratitude to the Ministry of Environment, Urbanization and Climate Change and the Turkish State Meteorological Service for their valuable support in providing the data used in this study.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Hakan ÇELİKTEN: Verilerin elde edilmesi, modelleme ve makalenin yazım süreci yazar tarafından gerçekleştirilmiştir.

The data acquisition, modeling, and manuscript writing were carried out by the author.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

KAYNAKLAR (REFERENCES)

- [1] Ding, J. Ren, C. Wang, J. Feng, Z. and Cao, S.-J. Spatial and temporal urban air pollution patterns based on limited data of monitoring stations, Journal of Cleaner Production, 434, p. 140359, 2024; doi: 10.1016/j.jclepro.2023.140359.
- [2] Zhou, Z. Qiu, C. and Zhang, Y. A comparative analysis of linear regression, neural networks and random forest regression for predicting air

- ozone employing soft sensor models. *Scientific Reports*, vol. 13(1), 1–23, 2023; doi: 10.1038/s41598-023-49899-0.
- [3] Latif, S. D., Lai, V., Hahzaman, F. H., Ahmed, A. N., Huang, Y. F., Birima, A. H., & El-Shafie, A. Ozone concentration forecasting utilizing leveraging of regression machine learnings: A case study at Klang Valley, Malaysia. *Results in Engineering*, vol. 21, p. 101872, 2024; doi: 10.1016/j.rineng.2024.101872.
- [4] Demir S. and Saral, A. Identification and Apportionment of Sources of Ozone-forming Potential for Proper Reduction Strategies CLEAN – Soil, Air, Water. 41(2), 107–112, 2013; doi: 10.1002/clen.201100704.
- [5] Faizan, M. and Çelikten, H. Mitigation of Ozone (O₃) Effects and Enhanced Air-Purifying Capacity of Plants by Carbon-Based Nanomaterials. *MAS Journal of Applied Sciences*. 11(1), 168–180, 2026; doi: 10.5281/zenodo.19127754.
- [6] He, Z. Wu, Z. Herzog, O. Hei, J. Li, L. and Li, X. Compound health effects and risk assessment of extreme heat and ozone air pollution under climate change: A case study of 731 urban areas in China. *Sustainable Cities and Society*, vol. 119, p. 106084, 2025; doi: 10.1016/j.scs.2024.106084.
- [7] Tang, H. Bhatti, U. A. Li, J. Marjan, S. Baryalai, M. Assam, M., ... & Mohamed, H. G. A New Hybrid Forecasting Model Based on Dual Series Decomposition with Long-Term Short-Term Memory. *International Journal of Intelligent Systems*, 2023(1), 9407104, 2023; doi: 10.1155/2023/9407104.
- [8] Karmakar, S. Baul Das, A. Gurung, C. and Ghosh, C. Effects of Ozone on Plant Health and Environment: A Mini Review, *Res. Jr. Agril. Sci*, 13, 612-619. 2022.
- [9] He W. and Chen, H. More extreme-heat occurrences related to humidity in China. *Atmospheric and Oceanic Science Letters*, 16(5), 100391, 2023; doi: 10.1016/j.aosl.2023.100391.
- [10] Ajayi, F. Adewale, A. and Osho, M. Integrating Engineering Innovations to Enhance Environmental Resilience: Evaluating the Impact of Greenhouse Gas Emissions, Ozone Depletion, And Aquatic Ecosystem Degradation on Public Health, *International Journal of Computer Applications Technology and Research*, 13, 16–36, 2024; doi: 10.7753/IJCATR1310.1002.
- [11] Nuvolone, D. Petri, D. and Voller, F. The effects of ozone on human health. *Environ Science and Pollution Research*, 25(9), 8074–8088, 2018; doi: 10.1007/s11356-017-9239-3.
- [12] Aljafen, B. N. Shaikh, N. AlKhalifah, J. M. & Meo, S. A. Correction: Effect of environmental pollutants particulate matter (PM_{2.5}, PM₁₀), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO) and ground level Ozone (O₃) on epilepsy.,” *BMC Neurology*, 25(1), 1–1, 2025; doi: 10.1186/s12883-025-04142-3.
- [13] Eslami, E. Choi, Y. Lops, Y. and Sayeed, A. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing & Applications*, 32(13), 8783–8797, 2020; doi: 10.1007/s00521-019-04282-x.
- [14] Katipoğlu, O. M. Elshaboury, N. Kartal, V. Ertugay, N. Kilinc, H. C. Şenocak, S. & Pande, C. B. Application of Time Series Methodologies for Robust Forecasting of Atmospheric Pollutant Concentrations: PM₁₀, SO₂, NO₂, NO_x, and O₃ in an Urban Environment. *Environmental Quality Management*, 35(2), 70206, 2025; doi: 10.1002/tqem.70206.
- [15] Çelikten, H. A Multi-Output Deep Learning Framework for Simultaneous Forecasting of PM₁₀ and Air Quality Index in High-Altitude Basins: A Case Study of Iğdir, Türkiye. *Sustainability*, 18(8), 2026; doi: 10.3390/su18083883.
- [16] Yılmaz, A. Ozone Level Prediction with Machine Learning Algorithms. *Journal of Aeronautics & Space Technologies / Havacılık ve Uzay Teknolojileri Dergisi*, 14(2), 177–183, 2021.
- [17] Carbo-Bustinza, N. Iftikhar, H. Belmonte, M. Cabello-Torres, R. J. De La Cruz, A. R. H. and López-Gonzales, J. L. Short-Term Forecasting of Ozone Concentration in Metropolitan Lima Using Hybrid Combinations of Time Series Models. *Applied Sciences*, 13(18), 2023; doi: 10.3390/app131810514.
- [18] Birim, N. G. Turhan, C. Atalay, A. S. and Gokcen Akkurt G. The Influence of Meteorological Parameters on PM₁₀: A Statistical Analysis of an Urban and Rural Environment in Izmir/Türkiye. *Atmosphere*, 14(3), 421, 2023; doi: 10.3390/atmos14030421.
- [19] Toz, A. C. Buber, M. Koseoglu, B. and Sakar, C. An estimation of shipping emissions to analysing air pollution density in the Izmir Bay. *Air Quality, Atmosphere & Health*, 14(1), 69–81, 2021; doi: 10.1007/s11869-020-00914-7.
- [20] Gümüş, M. G. Investigation of Seasonal Relationships Between Air Pollution and Plant Health With Satellite Imagery: Turkey-Izmir

- Province Case (2020-2024). vol. 12, pp. 7–23, 2025.
- [21] Elbir, T. (2002). Application of an ISCST3 model for predicting urban air pollution in the Izmir metropolitan area,” *International Journal of Environment and Pollution*, 18(5), 498-507.
- [22] Çelikten, H. Evaluating machine learning (RF, XGBoost) and statistical model (MLR) for PM10 and air quality prediction: A case from Kars, Türkiye. *Atmospheric Pollution Research*, 17(6), 102975, 2026; doi: 10.1016/j.apr.2026.102975.
- [23] Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., ... & Bellander, T. A Random Forest Approach to Estimate Daily Particulate Matter, Nitrogen Dioxide, and Ozone at Fine Spatial Resolution in Sweden. *Atmosphere*, 11(3), 239, 2020; doi: 10.3390/atmos11030239.
- [24] Liu, R. Ma, Z. Liu, Y. Shao, Y. Zhao, W. and Bi, J. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environment International*, vol. 142, p. 105823, 2020; doi: 10.1016/j.envint.2020.105823.
- [25] Comrie, A. C. Comparing Neural Networks and Regression Models for Ozone Forecasting. *Journal of the Air & Waste Management Association*, 47(6), 653–663, 1997; doi: 10.1080/10473289.1997.10463925.
- [26] Hashim, N. M. Noor, N. M. Ul-Saufie, A. Z. Sandu, A. V. Vizureanu, P., Deák, G., & Kheimi, M. Forecasting Daytime Ground-Level Ozone Concentration in Urbanized Areas of Malaysia Using Predictive Models. *Sustainability*, 14(13), 7936, 2022; doi: 10.3390/su14137936.
- [27] Çelik, G. Sülün, Ş. Keçeci, S. and Ergenekon, P. Ozon Konsantrasyonunun Marmara Bölgesi İstasyonlarındaki Hava Kirliliği ve Meteorolojik Ölçümler Kullanılarak Makine Öğrenmesi Modelleriyle Tahmini, *Sura Akademi*. 9, 9–23, 2025.
- [28] Ay, Ş. and Ekinci, E. Comparison of Machine Learning and Deep Learning Methods for Modeling Ozone Concentrations. *jista*, 5(2), 106–118, Sep. 2022; doi: 10.38016/jista.1054331.