Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Nigde Omer Halisdemir University Journal of Engineering Sciences

ISSN: 2564-6605

Derleme / Review

HAVA KALİTESİ PARAMETRELERİNİN TAHMİNİ VE MEKANSAL DAĞILIMI İÇİN MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN KULLANILMASI

Yeşim DOKUZ[1, *], Aslı BOZDAĞ[2], Öznur Begüm GÖKÇEK[3]

[1] *Niğde Ömer Halisdemir Üniversitesi, Bilgisayar Mühendisliği Bölümü, 51240, Niğde Türkiye*
[2] *Niğde Ömer Halisdemir Üniversitesi, Harita Mühendisliği Bölümü, 51240, Niğde Türkiye*
[3] *Niğde Ömer Halisdemir Üniversitesi, Çevre Mühendisliği Bölümü, 51240, Niğde Türkiye*

ÖZET

Hava kirliliği, nüfus ve endüstrileşmenin artması ile birlikte günümüzde küresel boyutta yaşanan sorunlardan biri haline gelmiştir. Bu nedenle hava kirletici parametreler düzenli aralıklarla ölçülmeli ve ölçüm sonuçları değerlendirilerek gerekli tedbirler alınmalıdır. Hava kirliliğinin önlenmesi amacıyla kirletici parametreler bir model kapsamında ele alınması ve tahmin sonuçlarının mekânsal olarak değerlendirilmesi gerekmektedir. Son zamanlarda hava kirliliğine yönelik objektif ve daha hassas sonuçların elde edilmesi için yapay zeka teknolojilerine ait makine öğrenmesi algoritmalarından yararlanılarak hava kalitesi tahmini ve mekânsal dağılımına ilişkin çalışmalar yapılmaktadır.

Bu çalışmada, öncelikli olarak hava kirletici parametrelerin özellikleri, çevreye olan etkileri ve bu parametrelerin tahmin edilmesi ve izlenmesinin gerekliliği açıklanmıştır. Ardından bu parametrelerin değerlendirilmesinde uygulanan makine öğrenmesi yöntemlerinin neler olduğu; hangi parametrelerin kullanıldığı, kullanım amaçları, kısıtlılıkları, elde edilen doğruluk düzeyleri ve arazi kullanım ile ilişkisi açısından incelenerek kullanılan yöntemlere ve çalışma prensiplerine ilişkin detaylı bilgi verilmiştir. Bu çalışma, hava kalitesinin iyileştirilerek sürdürülebilir bir çevrenin elde edilmesinde hangi parametreler hangi yöntem kullanılarak nasıl bir analiz ile incelenmeli sorusuna ilişkin seçim karmaşasının çözümlenmesine yönelik gelecek çalışmalara bir fikir sunmaktadır.

Anahtar kelimeler: Hava kirletici parametre, Makine öğrenmesi, Mekânsal dağılım, Yapay zeka, Tahmin modellemesi.

USE OF MACHINE LEARNING METHODS FOR ESTIMATION AND SPATIAL DISTRIBUTION OF AIR QUALITY PARAMETERS

ABSTRACT

Air pollution, has an increasing problem in the global scale along with increasing population and industrialization. Therefore, air pollutant parameters should be measured at regular intervals and necessary measures should be taken by evaluating the measurement results. In order to prevent air pollution, pollutant parameters should be considered within the scope of a model and the estimation results should be evaluated spatially. Recently, in order to obtain objective and more accurate results for air pollution, machine learning algorithms belonging to artificial intelligence technologies are being used and studies on air quality estimation and spatial distribution are performed.

In this study, the properties of air pollutant parameters, their effects on the environment and the necessity of estimation and monitoring of these parameters are explained. Then, what are the machine learning methods used in the evaluation of these parameters; which parameters are used, the purposes of use, their limitations and the obtained accuracy levels and relationship with land use are examined and detailed information is given about the methods and working principles. This study provides an insight into future studies to resolve the confusion of the choice of which parameters should be analyzed by using which methods and methods to achieve a sustainable environment by improving air quality.

Keywords: Air pollution parameter, Machine learning, Spatial distribution, Artificial intelligence, Predictive modelling.

* Corresponding author / Sorumlu yazar. Tel.:0 388 2254018 ; e-mail / e-posta: ytorun@ohu.edu.tr,
Geliş / Recieved: 02.12.2019 Kabul / Accepted: 20.01.2020 doi: 10.28948/ngumuh.654092

1. GİRİŞ

Günümüzde insanların çoğu iş ve eğitim gibi birçok nedenden dolayı nüfusu yüksek olan şehirlerde yaşamaktadırlar. Şehirlerde nüfus artıkça sanayi ve trafiğinde etkisiyle birçok çevresel sorun ortaya çıkmaktadır. Hava kirliliği de bu sorunların başında gelmektedir. Hava kirliliği, canlıların yaşamını ve sağlığını olumsuz etkileyen veya havada bulunan yabancı maddelerin, olması gerekenin üzerinde miktar ve yoğunluğa ulaşması ile oluşmaktadır. Hava kirliliği, artan nüfus, kentselleşme ve endüstrinin gelişmesiyle birlikte giderek artan bir önem arz etmektedir. Hava kirliliği insan kaynaklı olarak trafik, ısınma ve sanayiden kaynaklanabileceği gibi çöl tozları, polenler, rüzgar etkisiyle oluşan tozlar, yanardağ faaliyetleri ve orman yangınları gibi doğal kaynaklı da olabilmektedir. Hava kirleticileri parametrelerinden en önemlileri, Partikül Madde (PM), Kükürt dioksit (SO_2), Karbon monoksit (CO), Karbondioksit (CO_2), Ozon (O_3), Azotoksitler (NO_x) ve Hidrokarbonlar (HC)'dır. Bu parametrelerin çok düşük konsantrasyonlarının solunması bile canlıların yaşamını olumsuz etkilemektedir. Bu parametrelerin neden olduğu hava kirliliğinin önlenebilmesi için kirliliği oluşturan hava kalitesi parametrelerinin o bölgeyi temsil edecek noktalardan belirli zamanlarda ölçülmesi ve bu ölçülen sonuçlara göre gereken tedbirlerin alınması gerekmektedir [1]. Ölçülecek olan parametrelerin değerleri ulusal ve uluslararası değerlendirmeler sonucunda belirlenmiştir. Bu ölçümlerden istenen sonuçların yeni teknolojilerle belirli bir düzeyin altında tutulması ve geleceğe yönelik sağlıklı tahminlerde bulunulması gelmektedir. Bu yönde günümüzde yapay zeka teknolojilerinden makine öğrenmesi algoritmalarından yararlanılarak hava kirliliğine yönelik objektif ve daha hassas sonuçlar elde edilen çalışmalar yapılmaktadır.

Makine öğrenmesi, verilen bir problemi probleme ait ortamdan edinilen veriye göre modelleyen bilgisayar mühendisliği yöntemlerinin genel adıdır. Var olan veri setleri ve kullanılan makine öğrenmesi yöntemleri ile oluşturulan model, en yüksek performansı elde etmek üzerine kurulmaktadır. Makine öğrenmesi için çeşitli görevler tanımlanmıştır. Bu görevlerin başında sınıflandırma, regresyon ve kümeleme gelmektedir. Sınıflandırma görevinde eldeki eğitim verileri kullanılarak girdi parametrelerinden bir model kurulmakta ve kurulan bu model, yeni gelen verileri ön tanımlı sınıflardan birine atamaktadır. Algoritmalar tarafından geliştirilen bu modelin başarısı test verileri kullanılarak değerlendirilmektedir. Regresyon görevinde, sınıflandırma algoritmaları kullanılmakta, ancak ön tanımlı sınıflar yerine sürekli bir değer aralığı bulunmaktadır. Kümeleme görevinde, girdi parametrelerinden oluşan model ön tanımlı sınıflar olmadan girdi verilerinin aldığı değerlere göre kümeler oluşturulmaktadır. Literatürde en yaygın kullanılan makine öğrenmesi algoritmaları; Destek Vektör Makineleri, Lojistik Regresyon, Naive Bayes, K-En Yakın Komşu, Rastgele Orman ve Karar Ağacı algoritmalarıdır.

Hava kalitesi parametrelerinin tahmini için literatürdeki çalışmalar incelendiği zaman, hava kirliliği üzerine farklı ülkelerde farklı şehirlere ait veri setleri ve çeşitli makine öğrenmesi yöntemleri kullanılarak çeşitli çalışmalar yapılmıştır. Bu yöntemler temel olarak, geleneksel sınıflandırma algoritmaları ve derin öğrenme yöntemleri olmak üzere iki gruba ayrılabilir. Geleneksel sınıflandırma algoritmaları, veri madenciliği görevlerini yerine getirmek için geliştirilmiş olan sınıflandırma algoritmalarını ifade etmektedir. Bu algoritmalara örnek olarak LASSO Regresyon [2-4], Destek Vektör Makinesi (Support Vector Machines – SVM) [5-8], Rastgele Orman (Random Forest) [9-12], k En Yakın Komşu algoritmaları (k Nearest Neighbor – kNN) [13-14] sayılabilir. Derin öğrenme yöntemleri olarak kullanılan yöntemler ise Derin Sinir Ağları (Deep Neural Networks – DNN) [15-18], Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN) [19-22] ve Yinelemeli Sinir Ağları (Recurrent Neural Networks – RNN) [23, 24] olarak sayılabilir.

Bu çalışmada, ilk olarak hava kirletici parametrelerin genel özellikleri, çevresel etkileri ve bu parametrelerin tahmin edilmesi ve izlenmesinin gerekliliği açıklanmıştır. Ardından hava kalitesinin tahmin edilmesinde kullanılan yöntemler ve çalışma prensiplerine ilişkin detaylı bilgi verilmiştir. Son olarak literatür kapsamında uygulanan yöntemler; kullanılan veri seti, kullanım amaçları, kısıtlılıkları ve elde edilen doğruluk düzeyleri açısından incelenmiştir. Bu çalışma, hava kalitesinin iyileştirilerek sürdürülebilir bir çevrenin elde edilmesinde hangi parametreler hangi yöntem kullanılarak nasıl bir analiz ile incelenmeli sorusuna ilişkin seçim karmaşasının çözümlenmesine yönelik gelecek çalışmalara bir fikir sunmaktadır.

2. HAVA KALİTESİ PARAMETRELERİ, ÖZELLİKLERİ ve ETKİLERİ

Hava içerisinde farklı oran ve çeşitlikte birçok kimyasalı bir arada bulundururken bu kimyasalların belli oranların üzerine çıkması halinde havanın yapısı bozulmaya başlamakta bu da hava kirliliği olarak adlandırılmaktadır. Hava kirletici parametrelerin havanın yapısını bozması nedeniyle izlenmesi ve takip edilmesi gerekmektedir. Hava kirleticiler, toz şeklindeki kirletici maddeler ve gaz (SO_2, NOx, HC, CO, CO_2, O) olmak üzere genel olarak iki alt grupta toplanmaktadır.

Toz şeklindeki hava kirletici parametreler arasında yer alan partikül maddeler aerodinamik çaplarına göre adlandırılmaktadır. Hava kirliliği izleme parametrelerinden biri olan PM_{10}, aerodinamik çapları 10 μm'den küçük olan, $PM_{2,5}$ ise aerodinamik çapları 2,5 μm'den küçük olan partiküllerdir. Kaynağı genellikle fabrika, enerji tesisi, yakma tesisi, inşaat kaynaklı antropojenik faaliyetler gibi yangın, toz taşınımı gibi doğal kaynaklardır. Hayatımızın her alanında karşılaştığımız partikül maddeler sis, duman, toz gibi partikül çeşitlerinin yanı sıra bakteri, mantar, yosun gibi biyolojik çeşitleri de mevcuttur [25, 26]. Partikül maddenin oluşumuna sebep olan kirletici kaynaklar; evsel ısınma, endüstriyel faaliyetler ve trafiktir [27].

Gaz şeklindeki hava kirletici parametrelerinden biri olan Kükürt dioksit (SO_2), oda sıcaklığında renksiz, yanıcı olmayan ve keskin kokuya sahip bir gazdır [28]. Yüksek kükürt oranına sahip fosil yakıtlardan özellikle kömür, linyit ve yağların yanması ile ortaya çıkan kükürt oksitlerin büyük bir kısmı antropojenik faaliyetler sebebiyle havaya salınmaktadır. SO_2, atmosferde birkaç gün boyunca kalabilmekte ve meteorolojik faktörlere bağlı olarak yüzlerce kilometre uzağa taşınabilmektedir [29]. Bu parametrenin etkisinin azaltılması için yenilenebilir enerjinin hayatın her alanında kullanılmaya başlaması ve daha az zararlı yakıtların kullanılması gerekmektedir [25].

Hava kirleticilerden bir diğeri olan NOx'ler NO ve NO_2 olarak karşımıza çıkmaktadır. Bu gazlar, fosil yakıt yakma tesislerinden kaynaklanabildikleri gibi, nitrik asit üreten çeşitli kimya sanayisinden de atmosfere bırakılmaktadır. NOx'ler, atmosferde nitrik asit oluşturarak asit yağmurlarına neden oldukları gibi troposferde de ozon oluşumuna neden oldukları için dikkat edilmesi gereken bir parametredir [30]. Ayrıca, NO_2 gazının yüksek konsantrasyonlarda olması ciddi düzeyde solunum yolu tahribatına neden olabilmektedir. Dolayısıyla azot oksitlerin canlı sağlığına zarar vermeyecek şekilde dönüşümleri yapılarak giderilmesi gerekmektedir.

Karbonmonoksit (CO), renksiz, kokusuz, tatsız ve havanın ortalama mol ağırlığına yakın mol ağırlığında bir gaz olup varlığı fark edilemeyen bir gazdır. Karbon içeren yakıtların tam yanmaması sonucu ortaya çıkar. İnsan sağlığı açısından da CO önemlidir. Bilinen gaz zehirlenmeleri tam yanmamış artık gazların solunmasından kaynaklanmaktadır. Ayrıca karbon monoksitin iki önemli kaynağı kapalı mekânlardaki sigara dumanı ve otomobillerin egzoz gazıdır. Bu durum dolaşım, kalp ve kan hastalığı bulunan kişileri olumsuz etkilemektedir [31].

Ozon, atmosferde en yoğun (%90) olarak troposfer ve stratosfer tabakalarında bulunmaktadır. Ozon belirli kimyasal tepkime neticesinde aktif oksijen atomlarını oluşturur. Bu aktif atomlar atmosferdeki çeşitli organik ve anorganik kimyasal kirleticileri oksitleyerek tüketir. Bu reaksiyonlar üst atmosferde görülür. Yer seviyesi UV bulunmasa bile görünür ışınlar bir miktar ozon oluşturabilmektedir. Azot oksitler ve uçucu organik kirleticilerinin temel kaynakları olan trafik, solvent kullanımı ve sanayi tesisleri dolaylı olarak yer seviyesi ozon kirliliği kaynağıdır. Özellikle NO_x ve hidrokarbonların güneş ışığı ile tepkimeye girmesi sonucu ortaya çıkmakta olup insan sağlığını olumsuz yönde etkilendiğinden kötü ozon olarak nitelendirilmektedir[25, 26]. Yer seviyesindeki ozona maruz kalanlar arasında herkes farklı risk taşımakta olup, çocuklar daha çok risk altındadır.

Hava kirletici parametrelerin insan sağlığı başta olmak üzere ekosistem, iklim değişikliği, ekonomi, kültürel miras ve yapılar üzerine çeşitli olumsuz etkileri bulunmaktadır. İnsan sağlığına etkileri; solunum sistemi ve özellikle akciğer olmak üzere, bronşit, eklem romatizması, raşitizm ve çeşitli kalp hastalıkları göz yanmaları, görme bulanıklığı, nefes darlığı, iştahsızlık, kan zehirlenmesi vb.' dir. Sağlık açısından havadaki tozların tane irilik dağılımları önemli olmaktadır. Özellikle 1-5 mikron, 0,1-1 mikron ve <0,1 mikron çaplı tozların havadaki konsantrasyonlarının dağılımlarının bilinmesi önem taşır. 5 mikrondan iri taneciklerin insan vücuduna girme şansı yoktur. 1 mikrondan ince olan tanecikler ise gaz gibi davranarak üst ve alt solunum yollarını aşıp akciğer bölmelerine girebilmektedir [32]. Ekosistem üzerine etkileri; kirletici parametreler bitki örtüsü üzerine direkt, su ve toprak üzerine ise dolaylı bir şekilde etki ederek ekosistemin doğal dengesini bozmaktadır. Asit yağmurları zamanla nehir, göl, bataklık ve okyanus gibi yüzey sularının asitlik derecesinin artmasına neden olur. Bu durum bu alanlarda yaşayan balık ve kurbağa gibi canlıların popülasyonlarının azalmasına yol açar. İklim değişikliği üzerine etkileri; kömür, petrol ve petrol türevleri, doğal gaz gibi fosil yakıtların kullanımı, termik santrallerden, büyükbaş hayvan çiftliklerine kadar havaya karbondioksit salınmasına yol açan birçok faaliyetler hava kirliliğine ve küresel ısınmaya neden olmaktadır. Bu ısınma buzulları eriterek ve sera etkisi yaratarak olağandışı iklim olaylarına neden olmaktadır [33]. Kültürel miras ve yapılar üzerine etkileri; hava kirletici parametreler yapıların üzerinde korozyon, biyodegredasyon ve topraklaşma gibi zararlar oluşturmaktadır. Mimari yapının hava kirliliği etkisi altında strüktür bozulmasına uğraması özellikle asit yağmuru etkisiyle ile binaların çatı ve duvar gibi dış kısımlarının kararmasına ve aşınmasına neden olmaktadır.

3. HAVA KALİTESİ TAHMİNİNDE KULLANILAN YÖNTEMLER

Bu bölümde, hava kalitesi tahmini için kullanılan makine öğrenmesi yöntemleri alt başlıklar halinde anlatılmıştır. Öncelikle geleneksel sınıflandırma algoritmaları olan istatistiksel tabanlı LASSO regresyon algoritması, Destek Vektör Makinesi (SVM), Rastgele Orman ve k En Yakın Komşu (kNN) algoritmaları anlatılmıştır. Ardından derin öğrenme yöntemleri olan Derin Sinir Ağları (Deep Neural Networks - DNN), Yinelemeli Sinir Ağları (Recurrent Neural Networks - RNN) ve Evrişimli Sinir Ağları (Convolutional Neural Networks - CNN) sunulmuştur.

3.1. Geleneksel Sınıflandırma Algoritmaları

Bu bölümde, makine öğrenmesi algoritmalarından geleneksel sınıflandırma algoritmaları ve bu algoritmaların temel çalışma prensipleri tanıtılmıştır.

3.1.1. LASSO regresyon algoritması

LASSO (Least Absolute Shrinkage and Selection Operator) regresyon algoritması, lineer regresyonun model karmaşıklığının düşürmek ve modelin veriye bağlı aşırı öğrenmesinin önüne geçmek için önerilmiş bir istatistiksel regresyon algoritmasıdır [34]. LASSO regresyon algoritması, girdi parametrelerinin önemlerini artırıp azaltarak, regresyon modelinin daha iyi sonuç üretmesini sağlamaktadır. Bu şekilde hem modelin aşırı öğrenmesini önlemekte hem de parametre seçimini kendi içerisinde yapmaktadır. LASSO regresyon algoritmasında L1 düzenleştirme yaklaşımı kullanarak katsayılarının mutlak değerinin bir oranını optimizasyon işlemine dahil etmektedir. Bu sayede, parametrelerin sonuç üzerindeki etkisi düzenleştirilmiş olmaktadır.

3.1.2. Destek vektör makinesi (SVM) algoritması

SVM algoritması iki veya daha fazla boyuttaki noktaları birbirlerinden ayıran bir çizgi, düzlem veya hiperdüzlem üreterek sınıflandırma yapmaya çalışan ayrımcı bir sınıflandırma algoritmasıdır [35, 36]. SVM algoritması, farklı sınıflara ait noktalar arasındaki uzaklığı maksimize edecek en uygun doğruyu bulmaya çalışarak sınıflar arasındaki ayrımı mümkün olduğu kadar iyi belirlemeye çalışmaktadır. SVM algoritması, sınıflandırma yapmak için fonksiyon tipi, C ve gamma adında üç parametre kullanmaktadır. Fonksiyon tipi, girdi verisinin karakteristiğine göre lineer, non-lineer, polinomiyal veya radyal tabanlı fonksiyon olarak seçilebilmektedir. C ve gamma parametreleri, SVM algoritmasının aşırı veya zayıf öğrenmesini engellemek için kullanılan parametrelerdir.

3.1.3. Rastgele orman algoritması

Rastgele orman algoritması, karar ağaçlarını temel alarak sınıflandırma veya regresyon görevlerini yerine getiren topluluk tabanlı bir algoritmadır [37]. Rastgele orman algoritmasının temelinde karar ağaçları bulunmaktadır. Rastgele orman algoritması girdi parametreler seti içerisinden rastgele parametreler seçerek bu parametreler ile çok sayıda karar ağacı oluşturmaktadır. Oluşturulan bu karar ağaçlarının sonuçları kullanılarak algoritmanın sınıflandırma çıktısı hesaplanmaktadır. Bu sayede, birbirinden bağımsız veya ilgisiz parametreler farklı karar ağaçlarıyla değerlendirildiği için rastgele orman algoritması başarılı sonuçlar üretmektedir.

3.1.4. k En yakın komşu (kNN) algoritması

k En Yakın Komşu algoritması, literatürde yaygın bir şekilde kullanılan ve tembel öğrenmeye dayalı olarak sınıflandırma ve regresyon görevlerini yerine getiren bir algoritmadır [38]. kNN algoritması, eğitim aşamasında tespit edilmiş olan k sınıf merkezlerini dikkate alarak, test değerlerinin bu sınıf merkezlerine olan uzaklıklarına göre sınıflandırma işlemini yapmaktadır. Sınıf merkezlerine yakınlık ölçütü olarak Öklid, Minkowski, Manhattan uzaklıkları gibi çeşitli uzaklık ölçekleri kullanılmaktadır. kNN algoritması rastgele *k* adet sınıf merkezi tanımlayarak algoritmaya başlamakta eğitim verilerini bu sınıf merkezlerine olan yakınlıklarına göre sınıflandırmaktadır. Daha sonra iteratif olarak sınıf merkezlerini eğitim verilerinin ortasına kaydırmakta ve yeniden sınıflandırma yapmaktadır. Yeterli başarım sağlandığı zaman kNN algoritması sınıflandırma modelini üretmiş olmaktadır.

3.2. Derin Öğrenme Yöntemleri

Bu bölümde, en yaygın olarak kullanılan ve başarılı olan derin öğrenme yöntemleri ve bu yöntemlerin temel çalışma prensipleri tanıtılmıştır.

3.2.1. Derin sinir ağları (Deep Neural Networks - DNN)

Derin sinir ağları (DNN), sınıflandırma görevlerini yerine getirmek için girdi parametreleri ile çıktı değeri arasında bağlantı kurmak için sinir ağı yapısı kuran ve bu sinir ağıyla parametrelere ait ağırlıkları değiştirerek sinir ağını eğiten derin öğrenme yöntemlerinden biridir [39]. DNN'de her bir katmanda çok sayıda nöron bulunmakta ve bu nöronlar arasındaki bağlantılar için büyük boyutlu ağırlık matrisleri üretilmektedir. Büyük boyutlu bu matrislerin eğitimi için de derin öğrenme yöntemleri kullanılarak ağırlıkların optimizasyonu sağlanmakta ve bu sayede sınıflandırma görevleri başarılı bir şekilde yerine getirilmektedir. DNN'de, eğitim aşamasında ileri yayılımda eğitim verileri sinir ağı üzerinden geçirilmekte ve buna göre bir sonuç değeri üretilmektedir. Üretilen sonucun gerçek sonuçla olan farkı ise tekrar sinir ağına geriye yayılım algoritmalarıyla yayılmakta ve ağırlıklar hata oranına göre güncellenmektedir. Bu aşamalar eğitim işlemi sonuna kadar devam etmekte ve derin öğrenme modelini oluşturmaktadır.

3.2.2. Yinelemeli sinir ağları (Recurrent Neural Networks - RNN)

Yinelemeli sinir ağları (RNN), veri kümesindeki verilerin bir sırayla geldiği ve t zamanındaki verinin önceki zamanlardaki verilerden etkilendiği durumlar için özelleştirilmiş bir sinir ağı yapısıdır [40]. RNN, sinir ağı üzerinde dolaşırken önceki zamanlardaki değerleri belirli bir oranda ilgili zamandaki değeri üretmek için kullanmaktadır. Bu yapısından dolayı, RNN, sıralı bilgileri işlemede ve zamana bağlı tahminler üretmekte diğer sinir ağı yapılarına göre daha başarılı olmaktadır. Bununla birlikte, zamana göre bağımlılıklar kurmasından dolayı RNN'in eğitimi klasik sinir ağlarının eğitiminden daha zordur ve özel eğitim algoritmaları ve yeni sinir ağı modelleri geliştirilmiştir. RNN sinir ağını kullanan en yaygın sinir ağı modeli Uzun Kısa Dönem Hafıza (Long-Short Term Memory – LSTM) sinir ağı modeli geliştirilmiştir.

LSTM sinir ağı, RNN sinir ağı modeli üzerine geliştirilmiş, RNN'deki zamanda bağımlılıkları da içeren ancak RNN'in eğitiminde yetersiz kalan kısımların üstesinden gelmek için önerilmiştir [41]. LSTM sinir ağında nöronların yapısı değiştirilmiş ve nöronlar içerisine giriş, unutma ve çıkış kapıları eklenmiştir. Bu sayede, tüm veriler üzerinde zamanda bağımlılık engellenmiş ve belirli adım sayısından sonra ağın eski bilgileri unutması ve daha yeni bilgilerle sonuç üretmesi sağlanmıştır.

3.2.3. Evrişimli sinir ağları (Convolutional Neural Networks – CNN)

Evrişimli Sinir Ağları (CNN), özellikle görüntü veri kümeleri ve diğer iki boyutlu veri kümeleri üzerinde etkin bir şekilde kullanılan bir sinir ağı yapısıdır [42]. CNN çok az sayıda önişlem aşaması gerektirmekte ve tüm işlemleri sinir ağı içerisinde gerçekleştirmektedir. CNN'de farklı amaçlara yönelik katmanlar bulunmakta ve bu katmanlar uygulama gereksinimlerine göre kullanılmaktadır. Bu katmanlardan en önemlisi, evrişim (convolution) katmanıdır ve bu katmanda veri kümesindeki özellikler tespit edilmektedir. Ayrıca, havuz (pooling), düzleştirme (flattening), ve standart sinir ağı katmanları da kullanılmaktadır. Evrişim katmanında ilgili örneğin ayırt edici özellikleri tespit edilmekte, havuz katmanında ağırlık sayılarının azaltılması gerçekleştirilmekte, düzleştirme katmanında önceki katmanlardaki verileri sinir ağı katmanı için hazırlamakta ve sinir ağı katmanında sınıflandırma işlemi yapmaktadır. CNN'deki katmanlar uygulamanın gereksinimlerine göre istenilen sayıda ve sırada kullanılarak güçlü sınıflandırma sistemleri geliştirilmektedir.

4. HAVA KALİTESİ TAHMİNİ İÇİN YAPILAN ÇALIŞMALAR

Hava kalitesi değişimin doğru tahmin edilmesi, çevresel izleme, ekolojik sistemlerin sürdürülebilirliği, insan sağlığı ve sürdürülebilir kentsel planlamada kritik ve belirleyici bir rol oynamaktadır. Bu nedenle hava kirliliği arttıkça hava kalitesinin değişimini izleme, tahmin etme, yönetme ve önlemeye yönelik yaklaşımlar da giderek artmaktadır. Son yıllarda araştırmacılar tarafından analitik yöntemlerin dışında deterministtik yöntemler, istatistik yöntemler, makine öğrenmesi ve derin öğrenme yöntemleri gibi bir takım yöntemler hava kalitesini tahmin etmek amaçlı önerilmiştir. Ancak bu yöntemlerin kendi içinde veri hacmi, incelenen bölge, meteorolojik değişkenler ve süre gibi kısıtları bulunmaktadır.

Veri hacmi açısından hava kalitesini etkileyen parametrelerin (PM_{10}, $PM_{2.5}$, NO, NO_X, CO, O_3, SO_2) sayısı ve bu parametrelerin incelenen bölgeye etkilerinin belirlenmesinde tahmin zorluğu gelmektedir. Örneğin bu parametrelerden sadece $PM_{2.5}$'un çevre bileşenlerinin (Al, Cu, Fe, K, Ni, Pb, S, Ti, V ve Zn) ele alınması söz konusu olduğunda incelenecek veri hacmi oldukça artacaktır.

İncelenen bölge ve bu bölgedeki verinin dağılımının ortaya konulması diğer bir kısıt olarak karşımıza çıkmaktadır. Örneğin sanayi kenti, turizm bölgesi ya da trafik yoğunluğu bulunan bir bölgede bu bölgeye yönelik parametre seçimi ve veri tasarımı kullanılan yöntemler açısından oldukça önemlidir.

Meteorolojik değişkenler hava kalitesini oldukça değiştiren parametrelerdir. Rüzgar hızı, rüzgar yönü, güneş radyasyonu, bağıl nem, basınç ve sıcaklık hava kalitesini iyileştiren ya da düşüren etmenlerdir. Küresel ısınma ve iklimsel değişimlerin etkisiyle farklılaşan bu etmenler hava kalitesi değişiminin izlenmesi, tahmin edilmesi ve yönetilmesi için depolanması ve analiz edilmesi gereken önemli değişkenlerdir. Hava kalitesi değişiminin tahmin edilmesinde sonuçların doğruluğu artırmak için bu meteorolojik değişkenlerin tanıtılması ve değişime olan etkilerinin belirlenmesi oldukça önemlidir.

Süre açısından her bir hava kalitesi parametresinin saatlik, günlük, haftalık, aylık ve yıllık olmak üzere farklı zaman serilerinde değerlendirilmesi çok büyük bir veri hacmi ile uğraşmayı gerektirmektedir. Uzun zaman serilerinde büyük veri hacminin depolanarak meteorolojik parametreler ile yüksek doğrulukta tahminler elde edilmesi oldukça karmaşık bir değerlendirme sürecini gerektirmektedir. Bu nedenle birçok yöntemin hava kalitesi tahmininde doğruluk açısından yetersiz kalmasına neden olmaktadır.

Literatürde kullanılan yöntemler incelendiğinde deterministtik yöntemler, yüksek maliyetli hesaplamalar ve parametre tanımlaması için özel bilgiler gerektirirken, istatistiksel yöntemler ise tahmin performansı, doğrusal varsayım ve çoklu doğrusallık sorunu nedeniyle sınırlılıklar taşımaktadırlar [43]. Son yıllarda sıklıkla hava kalitesi tahmini için kullanılan makine

öğrenimi yöntemi ise uzun vadeli zaman serilerinde yer alan büyük hacimli verilerden hava kirleticilerinin bağımlılıklarını öğrenememektedir. Bu nedenle günlük, haftalık, aylık ve yıllık zaman serilerinde yer alan büyük hacimli verileri kullanılarak uzun kısa süreli bellek (LSTM) modelini transfer eden ve hava kalitesini yüksek doğrulukla tahmin edebilecek derin öğrenmeye dayalı yöntem birçok çalışma tarafından önerilmektedir.

Hava kalitesi tahmininde önemli olan diğer bir husus veri hacmi, incelenen bölge, meteorolojik değişkenler ve süre kısıtlarına uygun yöntem seçildikten sonra elde edilen tahmin değerlerinin bölgesel analizler şeklinde haritalandırılarak görsel olarak sunumudur. Sonuçların haritalandırılarak değişimin sunulması merkezi ve yerel yönetimlerin hava kalitesine yönelik durumu değerlendirebilmesi ve gerekli önlemleri almasında oldukça önemlidir. Özellikle bir kent yönetiminin bu kentin hava kalitesinin değişimini yansıtan haritasına bağlı olarak mekânsal ve stratejik çözümler sunarak kentin planlanmasını sağlaması gerekmektedir. Örneğin geçmişte kentin dış çeperinde kalan bir sanayi alanı kentin yayılması sonucu kentin içerisinde kalmış ve çevresine zararlı etkiler oluşturuyor olabilir. Bu durum hava kalitesi değişiminin haritalandırılması ile tespit edilebilir ve sanayi alanının kentten desantrilizasyonu bir mekânsal değişim olarak önerilebilir. Benzer şekilde yıllara göre artan trafiğin etkisi ile hava kirliliği değişiminin kente etkisi tespit edilebilir ve bu bölgede yapılacak trafiği azaltmaya ilişkin tedbirler alınabilir ya da mekânsal stratejiler oluşturulabilir.

Elde edilen verilerin haritalandırılarak görselleştirilmesi ve bunun üzerinden analizler sağlanabilmesi bir sistemi gerektirmektedir. Coğrafi Bilgi Sistemleri uygulamaları modelleme çalışmaları için veri sağlamada veya modelleme çalışmalarından elde edilen sonuçların görselleştirilmesinde yaygın olarak kullanılmaktadır [44, 45]. Literatürde, CBS ile oluşturulan haritaların sürdürülebilir kaynak yönetimi, doğal kaynakların korunması ve ekolojik değerlendirmeler için kullanıldığı görülmektedir [46, 47]. CBS uygulamaları; yöneticiler, planlamacılar ve karar vericiler tarafından kentsel alanların planlanması ve yönetimine yönelik doğru karar verme kapasitesini artıran anahtar destek araçlarıdır [48, 49].

CBS ile hava kirliliği tahmini sonuçlarının mekânsal dağılımı çeşitli analiz teknikleri ile sağlanabilir. Bunlardan arazinin topografik yapısının ele alınması ya da alınmaması durumları söz konusu olabilir. Buna göre arazinin özelliğine uygun mekânsal dağılımı ortaya koyan enterpolasyon analizleri (kriging, IDW, vb.) ArcGIS kullanımı içerisinde yer almaktadır.

Bu çalışma ile kentsel çevrenin sürdürülebilirliği üzerinde önemli rolü olan hava kirliliğine ilişkin hangi parametre, hangi değişkenlerle, nasıl bir bölgede ve hangi yöntem kullanılarak incelenmesine ilişkin seçim karmaşasının çözümlenmesine yönelik bir fikir sunmak amaçlanmaktadır. Bu kapsamda öncelikle son yıllarda makine öğrenimi yöntemiyle yapılan literatür kapsamındaki çalışmalar incelenmiştir. Ardından bu çalışmalardaki sonuçların mekânsal dağılımına ilişkin kullanılan analiz yöntemleri ve değerlendirmeleri ortaya konulmuştur.

Literatürde, Tamas ve diğ., [50] yirmi dört saatlik ozon (O_3), azot dioksit (NO_2) ve partikül madde (PM_{10}) hava kalitesi konsantrasyonlarını tahmin etmek için yapay sinir ağları ve kirletici noktaları tespit etmeye yönelik kümeleme yöntemini hibrit bir şekilde kullanan bir model geliştirmiştir. Batı Akdeniz'de Korsika Adası'nın iki kent bölgesi incelenmiştir. Bu çalışmada sunulan modellerle Korsika'da hava kalitesi tahminleri için kullanılan operasyonel bir araç geliştirilmiştir.

Rybarczyk ve Zalakeviciute, [51] Ekvator'un Quito kentinde iki bölgede $PM_{2.5}$, rüzgar (hız ve yön) ve yağış parametrelerine göre ince parçacıkların seviyesini tahmin etmek için makine öğrenmesine dayanan bir model önermiştir. Kullanılan karar ağacı algoritması ile iki farklı model oluşturulmuştur. Her iki bölge için, $PM_{2.5}$ konsantrasyonlarının 15µg / m^3 eşiğine dayalı bir sınıflandırma ile öngörülmüş ve %65 doğruluk elde edilmiştir. Genel olarak bu bölgedeki ekvatorun oluşturduğu peyzaj özelliği ve meteorolojik verilerden kaynaklanan kirliliği tahmin etmedeki karmaşıklık göz önüne alındığında minimum parametrelere ve maksimum doğruluğa sahip bir model üretme konusundaki zorluklar oldukça fazla olduğu sonucuna varılmıştır. Buna göre %65 doğruluk oldukça yüksek bir doğruluk olarak kabul edilmiştir.

Zhan ve diğ., [52] günlük $PM_{2.5}$ konsantrasyonlarını Çin genelinde mekânsal ve zamansal olarak makine öğrenme algoritması kullanarak tahmin etmiştir. Bu çalışmada, Coğrafi Ağırlıklı Gradyan Artırma Makinesi (GW-GBM) olan yeni bir makine öğrenme algoritması kullanılmıştır. Aerosol optik derinliği (AOD) ve meteorolojik koşullar gibi tahmin değişkenleri ve $PM_{2.5}$ konsantrasyonları arasındaki mekânsal değişim incelenmiştir. Sonuçta, Çin nüfusunun %95'inin ortalama yıllık $PM_{2.5}$ konsantrasyonunun 35 mg / m^3'ten yüksek olduğu bölgelerde yaşadığı ve nüfusun % 45'inin 100 gün boyunca $PM_{2.5}$> 75 mg / m^3'e maruz kaldığı tahmin edilmiştir. Çin'deki akut insan sağlığı etkilerini değerlendirmek için GW-GBM algoritmasının günlük $PM_{2.5}$ konsantrasyonlarını doğru olarak tahmin ettiği belirlenmiştir.

Deters ve diğ., [53] makine öğrenmesi yöntemini ve seçilmiş meteorolojik parametreleri kullanarak $PM_{2.5}$ konsantrasyonlarını tahmin eden kentsel kirlilik modellenmesini gerçekleştirmiştir. Makine öğrenmesi yöntemlerinden Artırma Ağaçları (Boosted Trees – BTs) ve Lineer Destek Vektör Makinesi (L-SVM) algoritmaları kullanılmıştır. Bu çalışma, meteorolojik verilerden $PM_{2.5}$ konsantrasyonlarını öngörmede makine öğrenmesine dayanan istatistiksel modellerin kullanılmasının önemli olduğunu göstermektedir.

García Nieto ve diğ., [54] İspanya'nın kuzeyinde bulunan bir şehirde Ocak 2017 ile Temmuz 2017 arasında yaptıkları çalışmada 4 farklı istatiksel ve makine öğrenmesi tabanlı regresyon modeli (Vektör Otoregresif Hareketli Ortalamalar – Vector Autoregressive Moving Average (VARMA), Otoregresif Entegre Hareketli Ortalamalar – Autoregressive Integrated Moving Average (ARIMA), Çok Katmanlı Algılayıcı Yapay Sinir Ağları – Multilayer Perceptron (MLP) Neural Networks ve Destek Vektör Makinası (SVM)) kullanarak PM_{10} konsantrasyon verilerini değerlendirmiştir. Sonuç olarak çalışma SVM modelinin bir ay öncesinden ve sonraki yedi ay boyunca tahmin yaparken diğer modellerden daha iyi performans gösterdiğini belirtmiştir.

Huang ve Kuo, [55] $PM_{2.5}$ konsantrasyonunu izlemek ve tahmin etmek için Evrişimli Sinir Ağı (Convolutional Neural Network - CNN) ve Uzun-Kısa Süreli Bellek (Long-Short Term Memory - LSTM) algoritmalarını birleştirmiştir. Her algoritmanın genel performansını karşılaştırmak için, bu çalışmada yapılan deneylere dört ölçüm endeksi, Ortalama Mutlak Hata (Mean Average Error - MAE), Kare Ortalamalarının Karekökü Hatası (Root Mean Square Error - RMSE) Pearson korelasyon katsayısı ve Sözleşme Endeksi (IA – Index of Agreement) uygulanmıştır. Makine öğrenimi yöntemleriyle karşılaştırıldığında bu çalışmada önerilen CNN-LSTM modelinin tahmin doğruluğunun en yüksek olduğu tespit edilmiştir.

Chen ve diğ., [56] yaptıkları çalışmada Çin'de 2014-2016 yılları arasında $PM_{2,5}$ konsantrasyon verilerini alarak makine öğrenme algoritmalarından rastgele orman algoritması ve iki geleneksel regresyon modeli kullanmıştır. Daha sonra bu verileri kullanarak 2005-2016 yıllarındaki hava konsantrasyon değerleri tahmin edilmiştir. Çalışmalarının neticesinde tahmin yüzdesi en iyi model makine öğrenmesi algoritması olan rastgele orman algoritması olmuştur.

Xu ve diğ., [57] yer seviyesindeki $PM_{2.5}$'un aylık konsantrasyonlarını tahmin etmek için makine öğrenme tekniklerini çoklu uzaktan algılama veri setleri ile değerlendirmiştir. Bu çalışmada, yer seviyesinde $PM_{2.5}$ konsantrasyonunun tahmini için makine öğrenimine ilişkin sekiz farklı algoritmanın (MLR, BRNN, SVM, LASSO, MARS, RF, XGBoost, Cubist) performansı, uydudan türetilmiş Aerosol Optik Derinlik (Aerosol Optical Depth – AOD) verileri alınarak çoklu uzaktan algılama veri setlerinin kullanımıyla karşılaştırılmıştır. Elde edilen sonuçlara göre, Kübist, rastgele orman ve xGBoosting algoritmalarının bu algoritmalar içinde en iyi performansa sahip oldukları belirlenmiştir. Sonuç olarak, uygun makine öğrenme algoritmaları seçiminin özellikle $PM_{2.5}$ ile karmaşık arazilerin neden olduğu tahmin ediciler arasında doğrusal olmayan ilişkileri olan alanlar için zemin seviyesinde $PM_{2.5}$ tahminini iyileştirebilir olduğu tespit edilmiştir.

Chen ve diğ., [58] CO_2, TVOC ve HCHO parametrelerini kullanarak kapalı bir alanda makine öğrenme yönteminin dört farklı algoritmasını (SVM, GP, BPNN, M5P) kullanarak test etmiştir. Genel olarak, SVM algoritması, tahmin doğruluğunun değerlendirilmesinde en yüksek puanı almıştır.

Requia ve diğ., [59] $PM_{2.5}$ bileşeninin (Al, Cu, Fe, K, Ni, Pb, S, Ti, V ve Zn) tahmin edilmesi için basit geoistatistiksel enterpolasyon, hibrit enterpolasyon ve makine öğrenme (rasgele orman algoritması – random forest algorithm) yöntemlerini karşılaştırmıştır. Bu veriler Amerika Birleşik Devletleri'ndeki Doğu Massachusetts'te bulunan 2002 ve 2010 yılları arasında elde edilmiştir. Tüm bileşenlerin karşılaştırılan yöntemlere göre mekânsal dağılımı haritalandırılmıştır. Partikül bileşenlerinin çoğu için rastgele orman modeli en iyi performansı sağlamıştır.

Suleimana ve diğ., [60] trafik ile ilişkili partikül maddelerin (PM_{10} ve $PM_{2.5}$) kentsel konsantrasyonlarını yönetmede makine öğrenme yöntemlerini (ANN, BRT ve SVM) uygulamıştır. Bölgedeki partikül konsantrasyonlarını makine öğrenme yöntemiyle tahmin edebilmek için olarak 2007 ve 2012 yılları arasında Londra'daki on dokuz Hava Kalitesi İzleme sahasından kirlilik (PM_{10} ve $PM_{2.5}$), meteorolojik (rüzgar hızları, rüzgar yönü, güneş radyasyonu, bağıl nem ve ortam sıcaklığıdır) ve trafik (hacmi, ses seviyesi ve hızları) verileri toplanarak kullanılmıştır. Sonuçta ANN ve BRT modellerinin yol kenarındaki PM_10 ve $PM_{2.5}$ konsantrasyonlarının tahminlerinde başarıyla uygulanabilir olduğu tespit edilmiştir.

Lim ve diğ., [61] Seul, Güney Kore'de düşük maliyetli sensörler ile arazi kullanım (yollar, binalar, ulaşım noktaları, kamu binaları ve su alanları) üzerinde ince partikül madde ($PM_{2.5}$) konsantrasyon seviyelerini mobil örnekleme yaparak elde etmiş ve makine öğrenimi yöntemiyle kentsel hava kalitesi haritası oluşturmuştur. Veriler on gönüllü tarafından beş rota boyunca yaklaşık üç haftalık uzun bir sürede 169 saat düşük maliyetli akıllı telefon tabanlı partikül sayacı ile coğrafi olarak konumsal verileri açık bir veri tabanı üzerinden toplanmıştır. Makine öğrenme algoritmaları (Lineer Regresyon – Linear Regression (LR), Rastgele Orman - Random Forest (RF), ve İstifli Topluluk – Stacked Ensemble (SE)) verilere uygulanmış ve $PM_{2,5}$ tahmin haritaları oluşturulmuştur. Sonuçta SE algoritması ile en başarılı tahmin ve yayılım haritası oluşturulmuştur.

Maa, ve diğ., (2019)'da (TL-BLSTM) modelini transfer eden derin öğrenmeye dayalı bir yöntem kullanarak büyük ölçekteki zaman serilerine bağlı verilere dayalı $PM_{2,5}$ hava kalitesi tahmin doğruluğunu iyileştirmek amaçlı Çin'in Guangdong bölgesinde bir çalışma yapmıştır. Sonuçlar önerilen TL-BLSTM modelinin, özellikle daha büyük zamansal çözünürlükler için daha küçük hatalara sahip olduğunu göstermiştir.

Zhong, ve diğ., [62] Çin'in Pekin bölgesindeki 2016/17 ve 2017/18 kış mevsiminde $PM_{2.5}$ kitlesel değişiminin meteorolojik geri bildirim etkisini tahmin etmek için gözlemsel veriler ve makine öğrenimi yöntemini kullanmıştır. En ilişkili üç değişkene sahip makine öğrenme algoritmaları kullanarak, kirli günlerde, 2017/18 kış mevsimindeki meteorolojik geri bildirime bağlı olarak $PM_{2,5}$ artışının kış 2016/17'deki oranın yalnızca % 49 olduğu tespit edilmiştir. Etkili şekilde yapılan kirlilik kontrolü ve uygun meteorolojik koşullar $PM_{2,5}$ azaltımında fayda sağladığı tespit edilmiştir.

Li ve Zhang, [63] Pekin'in NanjinHebei bölgesindeki, 2015-2017 döneminde yer seviyesinde $PM_{2.5}$ konsantrasyonlarını tahmin etmek için uzaktan algılama (aerosol optik derinliği) ve makine öğreniminin (Rastgele Orman) hibrit bir yaklaşımını oluşturularak kullanmıştır. Bu hibrit metedoloji; gerçek zamanlı mekânsal ve zamansal veri dağılımının incelenmesinde, ıssız yerlerde ve karmaşık yüzeylerde yeterli karakterizasyonun sağlanmasında ve çok sayıda meteorolojik, çevresel ve hava kirletici faktör arasındaki karmaşık doğrusal olmayan ilişkilerin ele alınmasında fayda sağlamaktadır. Önerilen hibrit model, yıllık ve mevsimsel $PM_{2.5}$ konsantrasyonlarının mekânsal-zamansal varyasyonlarını karakterize etmede iyi bir performans göstermiştir.

Liu ve diğ., [64] 2016 yılında meteorolojik parametrelerle entegre edilmiş uydu tabanlı atmosfer üstü yansımalarından doğrudan elde edilen $PM_{2.5}$ konsantrasyonunu makine öğrenme yöntemi (Rastgele Orman algoritması – Random Forest

algorithm) kullanarak tahmin etmiştir. Model, saatlik $PM_{2.5}$ konsantrasyon tahmini için, 0.86 doğruluk oranı (R^2) ve saatlik ortalama $PM_{2.5}$ konsantrasyon tahmini için 17.3 umg −3'lük bir kare ortalamalarının karekökü hatası (Root Mean Square Error - RMSE) içermektedir. Bu çalışmada geliştirilen modelin nispeten güçlü tahmin edilebilirliği, $PM_{2.5}$ konsantrasyonlarının günlük döngüsü ve bölgesel kirlilik olaylarının süreçlerinin izlenmesine yardımcı olabileceği sonucuna varılmıştır.

Ayturan, [1] Ankara ili Keçiören ilçesinde geçmişteki saatlik PM konsantrasyon verilerini, meteorolojik faktörler (Sıcaklık, basınç, Rüzgar hızı, rüzgar yönü, UVB ve UVA radyasyonu) ve kirlilik konsantrasyon verileri ile kullanarak gelecekteki saatlerde $PM_{2.5}$ konsantrasyonlarını tahmini kısa vadeli olarak belirlemiştir. Bunun için 17 giriş parametresi kapsamında derin öğrenme; Kapılı Tekrarlayan Ünite (Gated Recurrent Unit - GRU) ve Yinelemeli Sinir Ağları (Recurrent Neural Network – RNN) modelinde değerlendirilerek kısa süreli tahmin performansı yaklaşık %83 bir başarı ile elde edilmiştir.

Irmak ve Aydilek [65] kükürt dioksit (SO_2), azot dioksit (NO_2), ozon (O_3), karbon monoksit (CO) ve toz parçacıkları (PM_{10}) gibi hava kirletici gazların ölçüm verilerine farklı makine öğrenme algoritmaları uygulayarak hava kalite indeksinin tahmin başarısını artırılmasına ilişkin bir çalışma yapmıştır. Kullanılan algoritmalardan rastgele orman algoritmasının karar ağacı, destek vektör makinası, k-en yakın komşu, doğrusal, yapay sinir ağı, yığın, uyumlu artırıcı, eğimli artırıcı ve örneklemeli toplam regresyonuna göre hava kalite indeksini en iyi tahmin edebilen algoritma olduğu tespit edilmiştir.

Ayrıca literatürde Makine öğrenim yöntemleri ile yapılan hava kalitesi tahminlerinin mekânsal dağılımlarının uygulandığı analiz teknikleri incelendiğinde enterpolasyon analizleri (Reid ve diğ., 2015; Lim, ve diğ, 2019; Requia ve diğ., 2019) en çok kullanılan tekniklerdir. Bunlar Coğrafi Bilgi Sistemleri yardımıyla ArcGIS yazılımı kullanılarak basemap ya da open Street map olarak ifade edilen açık kullanıma sahip haritalar üzerinden gerçekleştirilir. Elde edilen hava kalitesi tahmini konsantrasyon değerlerinin sınıflandırılarak görselleştirilmesi ile sonuç harita elde edilmektedir.

Masih [66] atmosferik kirleticilerin konsantrasyonu ile emisyon kaynakları arasındaki ilişkiyi anlamak için yapılan deterministik modeller yerine makine öğrenimi yaklaşımlarına dayalı istatistiksel modellemedeki son gelişmelerin bu ilişkinin çözümü için ortaya çıktığını belirtmiştir. Yaptığı çalışmada son 6 yılda makine öğrenimi teknikleri uygulayan 38 en alakalı çalışmayı incelemiştir. Yapılan inceleme, girdi doğruluğunun tahmin doğruluğunu geliştirmedeki rolü, bu çalışmaların yapıldığı coğrafya, kirletici konsantrasyonu tahmini veya tahmini için uygulanan başlıca teknikler ve bu tekniklerin Doğrusal Regresyon, Sinir Ağı, Destek Vektör Makinesi veya Topluluk öğrenme algoritmalarına dayalı olup olmadığıdır. Çalışma neticesinde, makine öğrenme tekniklerinin ağırlıklı olarak kıta Avrupa ve Amerika'da uygulandığını belirtmiştir. Ayrıca, kirlilik tahmininin genellikle topluluk öğrenme ve doğrusal regresyon tabanlı yaklaşımlar kullanılarak gerçekleştirildiğini, tahminin ise sinir ağları uygulama ve vektör makineleri tabanlı algoritmaları destekleme eğiliminde olduğunu belirtmiştir.

5. TARTIŞMA VE SONUÇ

Sanayileşme, kontrolsüz kentleşme ve iklim değişikliklerinin oluşturduğu olumsuz etkiler kentlerin yaşanabilirlik düzeyinin ve insan sağlığına etkilerinin incelenmesi gerekliliğini ortaya çıkarmaktadır. Bu açıdan son yıllarda kentler üzerinde insan sağlığına direk etkileri bulunan hava kirliliği üzerine pek çok araştırma yapılmıştır. Yapılan çalışmalarda artan teknolojinin de etkisiyle ölçüm tekniklerinin değişmesi, yapılan ölçüm analizlerinin ve değerlendirme yöntemlerinin de değişmesine neden olmuştur. Bu nedenle bu çalışmada son yıllarda gelişmiş teknolojiler ile belirlenen hava kalitesi değerlerinin tahminine ilişkin yöntem arayışları ortaya konulmuştur.

Çalışma kapsamında öncelikle hava kalitesi parametreleri neler oldukları, etkileri, insan sağlığı için önemi hakkında bilgi verilmiştir. Ardından son yıllarda yapılan çalışmalarda bilgisayar teknolojilerinin insan beynini taklit eden nitelikteki yapay zeka tekniklerinin neler olduğu bu tekniklerin genel çalışma prensipleri incelenmiştir. Son olarak makine öğrenimi yöntemleri kapsamında bir çok değişkeni ve çeşitli kısıtları (veri hacmi, incelenen bölge, meterolojik değişkenler ve süre) bulunan hava kirliliği parametrelerine yönelik yapılan çalışmalar literatür çerçevesinde incelenmiştir. Bu çalışmaların neler olduğu ve hangi algoritmalar kullanılarak hangi uygulama alanlarına ilişkin sorunları çözdüğüne yönelik bilgiler verilmiştir. Çalışma çerçevesinde literatüre göre hava kalitesi parametrelerinin tahmin edilmesinde dikkat edilecek hususlar şöyle özetlenebilir:

- Veri hacmi seçilen yöntemin başarısını etkilemektedir. Bu nedenle seçilen yöntemin veri hacmine yönelik belirlenmesi gerekmektedir.
- Uygulama alanına göre incelenecek uygulama alanını tehdit eden kirlilik düzeyine ilişkin tehdit oluşturan hava kalitesi parametresi ((PM_{10}, $PM_{2.5}$, NO, NO_X, CO, O_3, SO_2) belirlenmelidir.
- Belirlenen parametre ve onun mekânsal dağılımına ilişkin yeter sayıda istasyon olmasına yönelik konumsal tespitler yapılmalıdır.
- Belirlenen parametreye etki edecek ve iklimsel etkiler ile değişebilen meteorolojik değişkenler tespit edilmelidir.
- Değişkenlerin günlük, haftalık, aylık ve yıllık olarak nasıl sürelerde toplanması gerekliliği kullanılacak yöntemin başarısı için oldukça önem taşımaktadır. Bu nedenle veri düzenliliği, kalitesi ve hassasiyetine önem verilmelidir.
- Elde edilen tahmini konsantrasyon miktarlarının mekânsal dağılımının kentsel alan ya da kırsal alan olma durumuna, topografyanın tanımlanması gerekliliğine ve arazi kullanım türüne göre analiz tekniği belirlenmelidir.

Bu şekilde seçilen yöntemin başarısı ve hava kalitesi tahminine yönelik doğruluk düzeyi artacaktır. Bu çalışma gelecek çalışmalara hava kalitesi tahminine yönelik doğruluk düzeyinin artırılmasında makine öğrenimi ve algoritmalarına ilişkin uygun alan, yöntem ve değişkelerinin belirlenmesinde literatürden yararlanarak bir yol çizilmesinde yardımcı olmaktadır.

KAYNAKLAR

[1] Y. A. Ayturan, "Derin Öğrenme İle Havadaki Partikül Madde Konsantrasyonu Tahmini", *KTO Karatay Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi*, Konya, 2019.

[2] S. Yeongkwon, A. R. Osornio-Vargas, M. S. O'Neill, P. Hystade, J. L. Texcalac-Sangrador, P. Ohman-Strickland, Q. Menga, S. Schwander, "Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters", *Science of the Total Environment,* 639, 40–48, 2018.

[3] H. Chu, J. Wei, W. Wu, "Streamflow prediction using LASSO-FCM-DBN approach based on hydrometeorological condition classification", *Journal of Hydrology*, 2020.

[4] G. Xua, X. Ren, K. Xiong, L. Li, X. Bi, Q. Wu, "Analysis of the driving factors of PM2.5 concentration in the air: A case study of the Yangtze River Delta, China", *Ecological Indicators,* 110, 105889, 2020.

[5] S. Zhu, X. Lian, H. Liu, J. Hu, Y. Wang, J. Che, "Daily air quality index forecasting with hybrid models: A case in China", *Environmental Pollution,* 231, 1232-1244, 2017.

[6] S. Zhu, X. Lian, L. Wei, J. Che, X. Shena, L. Yang, X. Qiu, X. Liu, W. Gao, X. Ren, J. Li, "PM2.5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors", *Atmospheric Environment*, 183, 20–32, 2018.

[7] W. Fan, F. Si, S. Ren, C. Yu, Y. Cui, P. Wang, "Integration of continuous restricted Boltzmann machine and SVR in NOx emissions prediction of a tangential firing boiler", *Chemometrics and Intelligent Laboratory Systems*, 195, 103870, 2019.

[8] J. Murillo-Escobara, J.P. Sepulveda-Suescun, M.A. Correa, D. Orrego-Metaute, "Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia", *Urban Climate* 29, 100473, 2019.

[9] H. Sun, D. Gui, B. Yan, Y. Liu, W. Liao, Y. Zhu, C. Lu, N. Zhao, "Assessing the potential of random forest method for estimating solar radiation using air pollution index", *Energy Conversion and Management,* 119, 121–129, 2016.

[10] Rubal, D. Kumar, "Evolving Differential evolution method with random forest for prediction of Air Pollution" , *Procedia Computer Science,* 132, 824–833, 2018.

[11] J. A. Kaminska, "The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in Wrocław", *Journal of Environmental Management,* 217, 164-174, 2018.

[12] Y. Wang, Y. Du, J. Wang, T. Li, "Calibration of a low-cost PM2.5 monitor using a random forest model", *Environment International* 133, 105161, 2019.

[13] Y. Fan, L. Hou, K. X. Yan, "On the density estimation of air pollution in Beijing", *Economics Letters,* 163, 110–113, 2018.

[14] C. Wen, S. Liu, X. Yao, L. Penga, X. Li, Y. Hua, T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction", *Science of the Total Environment,* 654, 1091–1099, 2019.

[15] E. Espositoa, S. De Vitoa, M. Salvatoa, V. Brightb, R.L. Jonesb, O. Popoola, "Dynamic neural network architectures for on field stochasticcalibration of indicative low cost air quality sensing systems", *Sensors and Actuators,* B 231, 701–713, 2016.

[16] Q. Song, M. R. Zhao, X. H. Zhou, Y. Xue, Y. J. Zheng, "Predicting gastrointestinal infection morbidity based on environmental pollutants: Deep learning versus traditional models", *Ecological Indicators,* 82, 76–81, 2017.

[17] J. Ma, Y. Ding, J. C.P. Cheng, F. Jiang, Y. Tan, V. J.L. Gan, Z. Wan, "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques", *Journal of Cleaner Production*, 244, 2020

[18] Y. Zhou, F. J. Chang, L. C. Chang, I-F. Kao, Y. S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts", *Journal of Cleaner Production*, 209, 134-145, 2019.

[19] Y. Park, B. Kwon, J. Heo, X. Hu, Y. Liu, T. Moon, "Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks", *Environmental Pollution*, 256: 113395, 2020.

[20] Y. Qi, Q. Li, H. Karimiana, D. Liu, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory", *Science of the Total Environment,* 664, 1–10, 2019.

[21] N. Dong, J. F. Chang, A. G. Wu, Z. K. Gao, "A novel convolutional neural network framework based solar irradiance prediction method", *International Journal of Electrical Power & Energy Systems*, 114: 105411, 2020.

[22] A. Sayeed, Y. Choi, E. Eslami, Y. Lops, A. Roy, J. Jung, "Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance", *Neural Networks,*121, 396–408, 2020.

[23] R. Feng, H. Zheng, H. Gao, A. Zhang, C. Huang, J. Zhang, K. Luo, J. Fan, "Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China", *Journal of Cleaner Production*, 231, 1005-1015, 2019.

[24] Z. Liu, C. J. Sullivan, "Prediction of weather induced background radiation fluctuation with recurrent neural networks", *Radiation Physics and Chemistry*, 155, 275–280, 2019.

[25] H.S. Peavy, D.R. Rowe, G. Tchobanoglous, Mcgraw, "Environmental Engineering", *Hill Inc*., Singapore, 1985.

[26] T. Godish, "Air Quality 4 Edition", *Lewis Publishers*, A CRC Press Company, p 32-185,2004.

[27] EC,Cleaner air fol all, Available: http://ec.europa.eu/environment/air/cleaner_air [Acsessed April, 2017]

[28] G. E. Fan, M. Qıu, H. Hu, H. Tıan, L. Wang, X. Xu, G. Weı, X. "Ambient Levels Associated With Reduced Risk of Initial Outpatient Visits For Tuberculosis: A Population Based Time Series Analysis", *Environmental Pollution*, 28, 408-415, 2017.

[29] A. Barbulescu, L. Barbes, "Mathematical Modeling Of Sulfur Dioxide Concentration in The Western Part Of Romania", *Journal of Environmental Management*, 204 Part 3, 825-830, 2017.

[30] D. Grano, "Clean Air Act requirements: Effect on Emissions of NO, from stationary sources. In: Ozkan, U.S., Agarwal, SK. And Marcelin, G. (Eds.), Reduction of Nitrogen Oxide Emissions, ACS Symposium Series 587", *American Chemical Society*, Washington, DC, pp. 14-31, 1995.

[31] Tosun, E., "Türkiye'nin 2009-2016 Yılları Arasındaki Hava Kalitesi Verilerinin Değerlendirilmesi", *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi,* Ankara, 2017.

[32] A. Müezzinoğlu, "Hava Kirliliği ve Kontrolünün Esasları, *Dokuz Eylül Yayınları*, İzmir, 2005.

[33] M. Türkeş, U. M. Sümer, G. Çetiner, "Küresel İklim Değişikliği Ve Olası Etkileri", *Çevre Bakanlığı, Birleşmiş Milletler İklim Değişikliği Çerçeve Sözleşmesi Seminer Notları*, 7-24, Çkök Gn. Md., Ankara, 2000.

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society*: Series B (Methodological), 58(1): 267-288, 1996.

[35] V. N. Vapnik, "The Nature of Statistical Learning Theory", *Springer Science & Business Media, 1st Edition*, 1995

[36] H. Drucker, C.J. Burges, L. Kaufman, A.J. Smola, V. Vapnik, "Support vector regression machines", *Advances in Neural Information Processing Systems* (pp. 155-161), 1997.

[37] L. Breiman, "Random forests", *Machine Learning*, 45(1): 5-32, 2001.

[38] Z. Yao, L.R. Walter. "A regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data", *BMC Bioinformatics*, 7(1): 1-11, 2006.

[39] Y. Bengio, "Learning Deep Architectures for AI", *Foundations and Trends in Machine Learning*, 2(1): 1-127, 2009.

[40] D.E. Rumelhart, G.E. Hinton, R.J. Williams, "Learning representations by back-propagating errors", *Nature*, 323, 533–536, 1986.

[41] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, 9(8): 1735-1780, 1997.

[42] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", *MIT Press*, p. 326, 2016.

[43] J. Ma, J.C.P. Cheng, C. Lin, Y. Tan, ve J. Zhang, "Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques", *Atmospheric Environment*, 214, 2019.

[44] Ertürk, A., Ekdal, A., Gurel, M., Yuceil, K., ve Tanik, A. "Use of mathematical models to estimate the effect of nutrient loadings on small streams", *Fresen. Environ. Bull. 13(11b)*, 1361-1370, 2004.

[45] Ataol, M. "Burdur Gölü'nde seviye değişimleri", *Coğrafi Bilimler Dergisi*. 8(1), 77-92, 2010.

[46] Liu, H., Wang, L., Yang, J., Nakagoshi, N., Liang C., Wang, W. ve Lv, Y. "Predictive modeling of the potential naturel vegetation pattern in Northeast China. Ecological Research", *Ecological research*, 24.6: 1313-1321, 2009.

[47] Ozkan, K. ve Senturk, Ö. "The applicatıon of group discrimination techniques to predict the potential distribution of turpentine tree", *In Proceedings of International Scientific Conference People Buildings and Environment, 7-9 October*, Lednice, Czech Republic, 728-736, 2012.

[48] Javadian, M., Shamskooshki, H. ve Momeni, M. "Application of sustainable urban development in environmental suitability analysis of educational land use by using AHP and GIS in Tehran", *Procedia Engineering*, 21, 72-80, 2011.

[49] Malczewski, J. "GIS-based land-use suitability analysis: a critical overview", *Progress in planning*, Elsevier, 62-65, 2004.

[50] Tamas, W., Notton, G., Paoli, C., Nivet, M-L. ve Voyant, C. "Hybridization of Air Quality Frecasting Models Using Machine Learning and Clustering: An Orginal Approach to Detect Pollutant Peaks", *Aerosol and Air Qaulity Research*, 16, 405-416, 2016.

[51] Rybarczyk, Y. ve Zalakeviciute, R., "Machine Learning Approach to Forecasting Urban Pollution", *IEEE Ecuador Technical Chapter Meetings*, 2016.

[52] Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M. L., Shen, X., Zhu, L. ve Zhang, M. "Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm", *Atmospheric Environment*, 155, 129-139, 2017.

[53] Deters, J. K., Zalakeviciute, R., Gonzalez, M. ve Rybarczyk, Y."Modeling PM2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters", *Journal of Electrical and Computer Engineering*, Makale ID 5106045, 14, 2017

[54] García Nieto, P.J., Sánchez Lasheras, F., García-Gonzalo, E. ve de Cos Juez, F.J. "PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM,MLP, VARMA and ARIMA: A case study", *Science of the Total Environment*, 621,753–761, 2018.

[55] Huang, C-J., ve Kuo, P-H. "A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities", *Sensors*, 18 (2220), 2018.

[56] Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J. ve Guo, Y. "A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information". *Science of the Total Environment*, 636, 52–60, 2018.

[57] Xu, Y., Chak, Ho, H., Wong, M.S., Deng, C., Shi, Y., Chan, T-C. ve Knudby, A. "Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of groundlevel PM2.5", *Environmental Pollution*, 242, 1417- 1426, 2018.

[58] Chen, S., Mihara, K. ve Wen, "Time series prediction of CO2, TVOC and HCHO based on machine learning at different sampling points", *Building and Environment*, 146, 238–246, 2018.

[59] Requia, W.J., Coull, B. A. ve Koutrakis, P. "Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM2.5 constituents over space", *Environmental Research*, 175, 421–433, 2019.

[60] Suleiman, A., Tight, M.R. ve Quinn, A.D. "Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM10 and PM2.5)", *Atmospheric Pollution Research*, 10, 134–144, 2019.

[61] Lim, C.C., Kim, H., Ruzmyn Vilcassim, M.J., Thurston, G.D., Gordon, T., Chen, L-C., Lee, K., Heimbinder, M. ve Kim, S-Y. "Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea", *Environment International*, 131, 105022, 2019.

[62] Zhong, J., Zhang, X. ve Wang, Y. "Relatively weak meteorological feedback effect on PM2.5 mass change in Winter 2017/18 in the Beijing area: Observational evidence and machine-learning estimations",*Science of the Total Environment*, 664, 140–147, 2019.

[63] Li, X. ve Zhang, X. "Predicting ground-level PM2.5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning", *Approach. Environmental Pollution*, 249, 735-749, 2019.

[64] Liu, J., Weng, F. ve Li, Z. "Satellite-based PM2.5 estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm", *Atmospheric Environment*, 208, 113–122, 2019.

[65] Irmak, M. E. ve Aydilek, İ.B. "Hava Kalite İndeksinin Tahmin Başarısının Artırılması için Topluluk Regresyon Algoritmalarının Kullanılması", *Academic Platform Journal of Engineering and Science*, 7-3, 507-514, 2019.

[66] Masih, A. "Machine learning algorithms in air quality modeling", *Global J. Environ. Sci. Manage*, 5, 515-534, 2019.