



FINDIK ÜRETİMİ İÇİN ÖĞRENME TABANLI VERİM TAHMİNİ

Doğan YILDIZ^{1*}, Gülcan YILDIZ², Sercan DEMİRCİ²

¹Ondokuz Mayıs University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 55239, Samsun, Türkiye

²Ondokuz Mayıs University, Faculty of Engineering, Department of Computer Engineering, 55239, Samsun, Türkiye

Özet: Artan nüfus yoğunluğunun ve küresel ısınmadan kaynaklanan iklim değişikliklerinin tarımsal üretimin miktarı ve kalitesi üzerinde önemli etkileri bulunmaktadır. Bu nedenle dünya, gıda kıtlığı gibi ciddi problemlerle karşı karşıyadır. Dolayısıyla tarım, ülkelerin ekonomik ve sosyal kalkınma programlarında bulunan hayati bir faaliyettir. Sert kabuklu meyveler kategorisinde dünyada bademden sonra üretimi en fazla yapılan ürün fındıktır. Üretimi Türkiye, İtalya, İspanya, ABD gibi ülkelerde yoğunlaşmış olan fındığın dünyadaki talebinin yaklaşık %70'ini Türkiye karşılamaktadır. Bu nedenle, fındık üretiminde verime etki eden parametrelerin tespit edilmesi ve bu parametrelere bağlı olarak verimin tahmin edilmesi çok büyük önem arz etmektedir. Bu çalışmada, Sakarya, Düzce, Samsun, Giresun, Trabzon illerindeki ve bu illerin fındık üretimi yapılan ilçelerindeki, fındık üretimine etki eden parametreler incelenerek bir sonraki adımda elde edilecek verim tahmini yapılmıştır. Tahmin işlemi için gereken özniteliklerin eldesi için ham veri setine bazı ön işleme adımları uygulanmış, temel bileşenler analizi (Principal Component Analysis, PCA) ile öznitelik sayısı azaltılmıştır. Ön işlemeden geçen bu veriler girdi olarak, XGBoost algoritması, Yalın Uzun-Kısa Süreli Bellek (Long Short-Term Memory, LSTM) modeli, Yığılmış LSTM modeli, Konvolüsyonel Sinir Ağı (Convolutional Neural Network, CNN) modeli ve CNN-LSTM hibrit modeline verilerek tahmin işlemi gerçekleştirilmiştir. Son olarak bu yöntemlerin her biri için Ortalama Hataların Karekökü (Root Mean Square Error, RMSE) ile tahmin hatası elde edilmiş ve yöntem kıyaslamaları için kullanılmıştır. Bu sonuçlara göre, en düşük RMSE=1,32 değeriyle en iyi performans CNN-LSTM hibrit modeli ile elde edilmiştir.

Anahtar kelimeler: Fındık verim tahmini, CNN, LSTM, XGBoost algoritması, Temel bileşenler analizi

Learning Based Yield Forecasting for Hazelnut Production

Abstract: Increasing population density and climate changes resulting from global warming have significant effects on the amount and quality of agricultural production. Therefore, the world is facing severe problems, such as food shortages. Therefore, agriculture is a vital activity in countries' economic and social development programs. In the category of hard-shelled fruits, hazelnut is the most produced product in the world after almonds. Turkey meets approximately 70% of the world's demand for hazelnut, which is concentrated in countries such as Turkey, Italy, Spain, and the USA. For this reason, it is essential to determine the parameters affecting the yield in hazelnut production and to estimate the yield depending on these parameters. In this study, the parameters affecting the hazelnut production in the provinces of Sakarya, Düzce, Samsun, Giresun, Trabzon and the districts of these provinces where hazelnut production is made were examined, and the yield estimation to be obtained in the next step was made. In order to get the features required for the forecasting process, some pre-processing steps were applied to the raw data set, and the number of features was reduced by Principal Components Analysis (PCA). These pre-processed data were given as input to the XGBoost algorithm, Lean (Long Short-Term Memory, LSTM) model, Stacked LSTM model, (Convolutional Neural Network, CNN) model and CNN-LSTM hybrid model, and estimation was carried out. Finally, (Root Mean Square Error, RMSE) estimation error was obtained for each method and used for comparison. According to these results, the best performance with the lowest RMSE=1.32 value was obtained with the CNN-LSTM hybrid model.

Keywords: Hazelnut yield forecasting, CNN, LSTM, XGBoost algorithm, Principal component analysis

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 55239, Samsun, Türkiye

E mail: dogan.yildiz@omu.edu.tr (D. YILDIZ)

Doğan YILDIZ <https://orcid.org/0000-0001-9670-4173>

Gülcan YILDIZ <https://orcid.org/0000-0001-8631-8383>

Sercan DEMİRCİ <https://orcid.org/0000-0001-6739-7653>

Gönderi: 27 Şubat 2023

Kabul: 29 Mart 2023

Yayınlanma: 01 Nisan 2023

Received: February 27, 2023

Accepted: March 29, 2023

Published: April 01, 2023

Cite as: Yıldız D, Yıldız G, Demirci S. 2023. Learning based yield forecasting for hazelnut production. BSJ Eng Sci, 6(2): 117-126.

1. Giriş

Bilişim teknolojileri ve haberleşme gibi alanlardaki ilerlemeler, çevremizdeki verilerin toplanıp, saklanması ve sonrasında da bu verilerin işlenebilmesi konularında önümüze değerli bir fırsat sunmaktadır. Son zamanlarda yapay öğrenme tekniklerinde yaşanan gelişmeler, günümüz problemlerine yaklaşımımızı değiştirmeye başlamıştır. Ayrıca, gelişmekte olan bulut teknolojileri kompleks problemlere karşı her geçen gün daha basit ve

erişilebilir çözümler sunmaktadır. Bu çerçevede farklı rekolte tahmin uygulamalarında da yapay öğrenme yöntemleri bir süredir kullanılmaktadır (Cioffi ve ark., 2020).

Tahmin, temel olarak var olan bilgiler üzerinden çıkarım yapma işlemidir. Bu tahmin, bir nicel kavram için yapıyorsa genellikle kestirim adını alır. Kestirimde, bağımsız değişkenlerin yani gözlenmiş veya ölçülmüş değerlerin oluşturduğu veri kümesine dayanarak belirli bir bağımlı değişken kestirilmeye çalışılır. Bu doğrultuda,



tarım ürünlerinin rekolte tahmini için kullanılan yöntemlerin sınırlamaları da, tahmin yöntemlerinde kullanılan, ölçülebilen ve gözlenebilen bağımsız değişkenlerin ortaya koyacağı sınırlar çerçevesinde şekillenmektedir (Liu ve ark., 2001; Kent, 2020).

Tarım ürünlerinin rekolte tahmini, ülkelerin doğru ve etkili tarım politikaları belirlemelerinde önemli rol oynar. Böylece, üreticilerin elindeki kısıtlı kaynakların doğru kullanılması sağlanarak kazanç optimize edilmiş olur (Marko ve ark., 2016; Liliane ve Charles, 2020). Bunun yanı sıra, tedarik zincirinin sağlıklı bir şekilde inşa edilebilmesi, ilgilenilen ürünün ihracat ve ithalatının belirlenip planlanması da bu tahminlerden yararlanılarak yapılmaktadır. Sert kabuklu meyveler kategorisinde bademden sonra üretimi en yoğun yapılan ürün fındıktır. Fındık kültür çeşitlerinin yetiştiriciliği Türkiye, İtalya, İspanya ve ABD gibi ülkelerde yoğunlaşmıştır. Ülkemiz 600 bin tonu geçen üretim miktarıyla dünya fındık talebinin %70'ini karşılamaktadır. Türkiye'nin pazardaki bu payı doğru rekolte tahminlemelerinin önemini artırmaktadır (Romero ve ark., 2013; Üstün ve Karaosmanoğlu, 2017).

Fındık bitkisi, *Corylus* cinsinin *Betulaceae* familyası içindeki çok yıllık bitkilerden birisidir. Fındığın çıkış noktaları Orta Asya, Kafkasya ve Anadolu'dur. Anadolu'da fındık üretiminin yapıldığı bölge ise Orta ve Doğu Karadeniz bölgeleridir. Dünyanın en kaliteli fındık çeşitleri Anadolu'da üretilmektedir (Köksal, 2002). Fındık, ılıman ve nemli bölgelerde oldukça iyi bir gelişim göstermekte ve bol ürün vermektedir. Karadeniz bölgesi, bu sayede dünyanın en uygun fındık yetiştirme bölgelerinden biridir. 13 °C ile 16 °C ortalama sıcaklık aralığı, fındık için en uygun yetiştirme sıcaklık aralığıdır. Fındık için hava sıcaklığı -8 °C - 10 °C sıcaklık değerlerinin altına inmemeli ve 36 °C - 37 °C 'nin de üstüne çıkmamalıdır. Ayrıca, yıllık yağış toplamının 700 mm'nin altında olmaması ve yağışın yıl boyunca dengeli bir şekilde dağılım göstermesi gerekmektedir. Haziran ve Temmuz aylarındaki bağıl nemin %60'ın altına düşmemesi gereklidir (Tuncer ve ark., 2020). Fındığın oluşumuna etki eden en önemli faktörlerden biri de rüzgârdır. Rüzgâr, fındığın döllenenmesini sağlar. Bitkilerde optimum rüzgâr hızı 3-5 m/sn' dir. Rüzgârın hızının daha fazla olması bitki gelişimine olumsuz yönde etki edebilmektedir (Asar ve ark., 2007). Fındık oluşumuna etki eden bir başka faktör, toprağın pH değeridir. Fındık, 6-6,5 pH değerlerinde çok iyi derecede gelişim göstermektedir (Ali, 2010). Bir diğer önemli faktör ise yüksekliktir. Fındık yetiştirmek için en ideal olan yükseklik deniz seviyesinden 700-800 metre yükseklikleridir. 1500 metre civarları fındık yetiştiriciliği için elverişsiz yüksekliklerdir (Zaman, 2004).

(Şimşek ve ark., 2014) çalışmasındaki Türkiye Fenoloji Atlasında fenolojik dönemlerin ayrıştırılmasının kültür bitkisi yetiştiriciliğinde büyük öneme sahip olduğu belirtilmiş ve bu fenolojik dönemler çiçeklenme, meyve oluşumu ve olgunlaşma-hasat dönemi olarak üç ana

başlık altında ele alınmıştır. İl bazında minör farklılıklar bulunmasına rağmen, fındık bitkisi için şubat-mart-nisan ayları çiçeklenme, mayıs-haziran ayları meyve oluşumu, temmuz-ağustos ayları ise olgunlaşma ve hasat dönemi şeklinde ifade edilir. Yapılan çalışmalarda, mevsimsel ortalama değerlerin ve özellikle fenolojik dönem başlangıçlarının rekolte üzerine önemli etkilerinin olduğu görülmüştür. Fındık verimliliği, özellikle kış ve ilkbahar dönemlerinde düşük sıcaklıklardan önemli ölçüde etkilenmektedir. Yani, fenolojik dönemler ile sıcaklık ilişkisinin doğrudan verimi etkilediği söylenebilir. Böylece, bağıl nem, yağış ve ortalama sıcaklık gibi faktörler fındığın kış ve ilkbahar gelişme dönemlerini çeşitli yollarla etkileyerek fındık veriminin gözlemlenebilmesine olanak vermektedir (Beyhan ve Odabaş, 1996). Örneğin, İrdem (2021) çalışmasının araştırma alanında fındık verimi ile aylık ortalama sıcaklıkların korelasyon ve regresyon değerleri Tablo 1'de verilmiştir. Yıllık ortalama sıcaklık ile verim arasında Düzce ve Giresun illerinde negatif bir ilişki bulunmaktadır. Diğer üç ilde ise çok zayıf bir pozitif ilişki görülmektedir. Verim tahmininin, görüntü işleme tabanlı yapıldığı çalışmalar da mevcuttur (Zhao ve ark., 2019; Yadav ve ark., 2021).

Hasattan önce mahsul veriminin doğru ve zamanında tahmin edilmesi son derece önemlidir. Ancak araştırmacılar için verimi doğru bir şekilde tahmin etmek, farklı nitelikteki birçok faktörden dolayı her zaman zor olmuştur. Makine öğrenmesi alanındaki gelişmeler, bu alanda umut verici bir gelecek vaat etmektedir (Bali ve Singla, 2022). Bu nedenle, son yıllarda makine öğrenmesi temelli tahmin çalışmaları öne çıkmaktadır (Cedric ve ark., 2022; Fei ve ark., 2022; Paudel ve ark., 2022; Desloires ve ark., 2023). Kim ve Lee (2016) çalışmalarında, dört makine öğrenme tekniği, yani, destek vektör makinesi (SVM (Support Vector Machine), RF (Random Forest), ERT (Extremely Randomized Trees) and DL (Deep Learning)) kullanarak ve uzaktan algılama verilerini inceleyerek Iowa Eyaletindeki mısır mahsulünün verimini tahmin etmişlerdir. Sonuçlar incelendiğinde, özellikle derin öğrenmenin verimi tahmin etme aracı olarak diğer yöntemlere göre daha kararlı sonuçlar ürettiği görülmüştür (Kim ve Lee, 2016). Villanueva ve Salenga (2018) tarafından yapılan çalışmada, CNN yöntemi kullanılarak acı kavun veya acı kabak mahsulünün meyve verme yeteneği incelenmiştir. İyi ve kötü olarak sınıflandırılan yaprak görüntülerinin eğitimi; Keras, Tensor Flow ve Python aracılığıyla yapılmıştır. Sonuç olarak, tahmin için modele sunulan fazla sayıda görüntü, bir makinenin iyi ve kötü bir Ampalaya bitkisi arasındaki farkı öğrenmesini sağlayabilmektedir (Villanueva ve Salenga, 2018). Fei vd. (2022), insansız hava aracı (İHA) kaynaklı çoklu sensör verilerinin birleştirilmesi için makine öğrenmesi (Machine Learning, ML) yöntemlerinin kullanılmasının, mahsul veriminin tahmin doğruluğunu artırabileceğini savunmuşlardır.

Tablo 1. Araştırma alanında fındık verimi ile aylık ortalama sıcaklıkların korelasyon ve regresyon değerleri (İrdem, 2021)

İller	İstatistik Unsurlar	Eylül	Ekim	Kasım	Aralık	Ocak	Şubat	Mart	Nisan	Mayıs	Haz.	Tem.	Ağus.	Yıllık
Sakarya	Korelasyon	-0,16	-0,28	-0,22	0,11	-0,26	-0,02	0,20	0,25	-0,130,04	0,31	0,13	0,09	
	Anlamlılık	0,41	0,16	0,27	0,58	0,19	0,94	0,32	0,21	0,510,84	0,12	0,52	0,64	
	R ²	0,03	0,08	0,05	0,01	0,07	0,00	0,04	0,06	0,020,00	0,09	0,02	0,00	
Düzce	Korelasyon	-0,46*	-0,18	-0,36	0,08	0,05	-0,16	-0,04	0,22	-0,060,08	0,06	0,23	-0,10	
	Anlamlılık	0,02	0,36	0,07	0,67	0,79	0,43	0,85	0,28	0,760,69	0,78	0,25	0,63	
	R ²	0,21	0,03	0,13	0,00	0,00	0,03	0,00	0,05	0,000,00	0,00	0,05	0,01	
Samsun	Korelasyon	-0,01	-0,11	-0,16	0,14	-0,21	-0,08	0,18	0,36	-0,160,07	0,20	0,40*	0,02	
	Anlamlılık	0,95	0,59	0,42	0,48	0,29	0,71	0,38	0,06	0,430,73	0,32	0,04	0,92	
	R ²	0,00	0,01	0,03	0,02	0,04	0,00	0,03	0,13	0,030,00	0,04	0,16	0,00	
Ordu	Korelasyon	0,05	0,07	-0,27	0,08	-0,18	-0,15	0,15	0,21	-0,14-0,18	0,16	0,11	-0,02	
	Anlamlılık	0,83	0,77	0,21	0,72	0,40	0,50	0,49	0,33	0,520,42	0,46	0,60	0,92	
	R ²	0,00	0,00	0,07	0,00	0,03	0,02	0,02	0,05	0,020,03	0,03	0,01	0,00	
Giresun	Korelasyon	-0,24	-0,26	-0,35	-0,01	-0,36	-0,34	-0,01	0,06	-0,28-0,30	0,06	-0,13	-0,32	
	Anlamlılık	0,22	0,18	0,07	0,96	0,07	0,08	0,96	0,75	0,150,12	0,75	0,52	0,10	
	R ²	0,06	0,07	0,13	0,00	0,13	0,12	0,00	0,00	0,080,09	0,00	0,02	0,10	
Trabzon	Korelasyon	0,03	-0,27	-0,05	0,02	-0,13	0,10	0,38*	0,31	-0,07-0,09	0,18	0,12	0,16	
	Anlamlılık	-0,16	-0,28	-0,22	0,11	-0,26	-0,02	0,20	0,25	-0,130,04	0,31	0,13	0,09	
	R ²	0,00	0,08	0,00	0,00	0,02	0,01	0,14	0,09	0,000,01	0,03	0,01	0,03	

Bunun için Kübist, SVM, derin sinir ağı (Deep Neural Network, DNN), ridge regresyon (RR) ve rastgele orman (Random Forest, RF) gibi beş ML algoritması, çoklu sensör veri birleştirme ve buğdayda tahıl verimi tahmininde topluluk öğrenmesi için kullanılmıştır. Sonuçlar, düşük irtifalı İHA tabanlı çoklu sensör verilerinin, veri füzyonu ve yüksek doğrulukta bir topluluk öğrenme çerçevesi kullanılarak erken tahıl verimi tahmini için kullanılabileceğini kanıtlamıştır (Fei ve ark., 2022).

Yapılan çalışmalar incelendiğinde, çoğunlukla fındık ürünü dışındaki ürünlerin verilerinden yararlandığı görülmektedir. Fındık verisi özelinde yapılan örnek bir çalışma An ve ark. tarafından 2020 yılında yapılmıştır. Çalışmalarında RCP8.5 (The Representative Concentration Pathway) kapsamında, bölgesel iklim modeli RegCM4.4'ten 10 km yatay çözünürlüklü ızgaralı iklim verilerine dayanarak, 2021–2050 dönemi için verim değişimini araştırmak amacıyla çoklu regresyon analizi yapmışlardır. Toplam 88 farklı lokasyon ve üç fenolojik büyüme aşaması incelenmiştir. Nemli-ılıman Karadeniz ve yarı-nemli Marmara bölgelerinde fındık veriminin fenolojik, alt bölge ve mikro iklim koşullarına bağlı olarak önemli farklılıklar gösterdiği gözlemlenmiştir (An ve ark., 2020). Fakat bu çalışmada DL temelli yaklaşımlar kullanılmamıştır.

Bu çalışmada ise, Sakarya, Düzce, Samsun, Giresun, Trabzon illerindeki ve bu illerin fındık üretimi yapılan ilçelerindeki, fındık üretimine etki eden parametreler incelenerek bir sonraki adımda elde edilecek verim tahmini ML ve DL modelleri üzerinde yapılmıştır. Bu çalışmanın geri kalanı ise şu şekilde organize edilmiştir:

Bölüm 2'de, bu çalışmada kullanılan materyal ve yöntemlere değinilmiştir. Bölüm 3'te tasarlanan modeller ile elde edilen tahmin sonuçlarından bulgular verilmiştir. Dördüncü bölümde ise, elde edilen sonuçlar tartışılmış ve gelecekte yapılabilecek muhtemel çalışmalar özetlenmiştir.

2. Materyal ve Yöntem

2.1. Veri Seti

Bu çalışmada kullanılan veri seti yapısı, Türkiye'deki fındık üretimine en çok katkıda bulunan Sakarya, Düzce, Samsun, Giresun, Trabzon gibi illerdeki ve bu illerin fındık üretimi yapılan ilçelerindeki yıllık rekolte verilerinden oluşturulmuştur. Bu veriler, Türkiye İstatistik Kurumu (TÜİK) görevlileri tarafından ilgili yerlerde fındık verimliliğine etki eden faktörlerin gözlemlenmesi ile oluşturulmuş, TÜİK/Biruni platformunda erişilebilir verilerdir. Model eğitiminde kullanılacak veri setinde toplam 34 değişken belirlenmiştir. 5 ile ve ayrı ayrı ilçelerine ait iklimsel veriler ise bölgesel olarak Amerikan Ulusal Havacılık ve Uzay Dairesi (National Aeronautics and Space Administration, NASA) Power platformu aracılığıyla elde edilmiştir. İklimsel verilerin elde edilmesine ait örnek görüntü Şekil 1'de gösterilmiştir.

NASA Power platformu üzerinden sunulan agro-klimatoloji verileri arasından seçilen 7 temel parametre şunlardır:

1. Yağış (mm)
2. Dünya yüzeyinden 2 metre yükseklikte günlük ortalama sıcaklık (°C).

3. Dünya yüzeyinden 2 metre yükseklikte günlük ortalama rüzgâr hızı (m/s).
4. Dünya yüzeyinden 2 metre yükseklikte günlük ortalama bağıl nem (%).
5. Dünya yüzeyinden 2 metre yükseklikte günlük ortalama çiğ/işba noktası (°C).
6. Dünya yüzeyinde yatay bir yüzeye gelen toplam güneş ışınımının günlük ortalama miktarı. (MJ/m²/gün)
7. Dünya yüzey sıcaklığının günlük ortalaması (°C).

Günlük olarak toplanmış bu verilerin yıllık aralıklarla ifade edilebilmesi amacıyla, ilgili yıla ait bütün verilerin mevsimsel ortalamaları alınmıştır. Yağış değişkenine ait mevsimsel ortalama değerleri değişken olarak veri setinde yer almıştır.

TÜİK üzerinden elde edilen verilerde, meyve veren yaştaki ağaç sayısı verisi ve meyve vermeyen yaştaki ağaç sayısı verisi bulunmaktadır. Bu çalışma çerçevesinde, sadece meyve veren yaştaki ağaç sayısı bir değişken olarak kullanılmış ve verim hesaplanırken kg/ağaç yerine kg/dekar birimi dikkate alınmıştır. TÜİK platformunda kırılım seçimine ait görüntü Şekil 2 ile verilmiştir.

Veri setinin genel yapısı, TÜİK ve NASA Power platformlarından sağlanan verilerle ücretsiz bir şekilde oluşturulmuştur. Ayrıca çalışmada LSTM tabanlı bir

model kullanılması öngörüldüğünden veri setine bir değişken olarak zaman damgası da eklenmiştir.

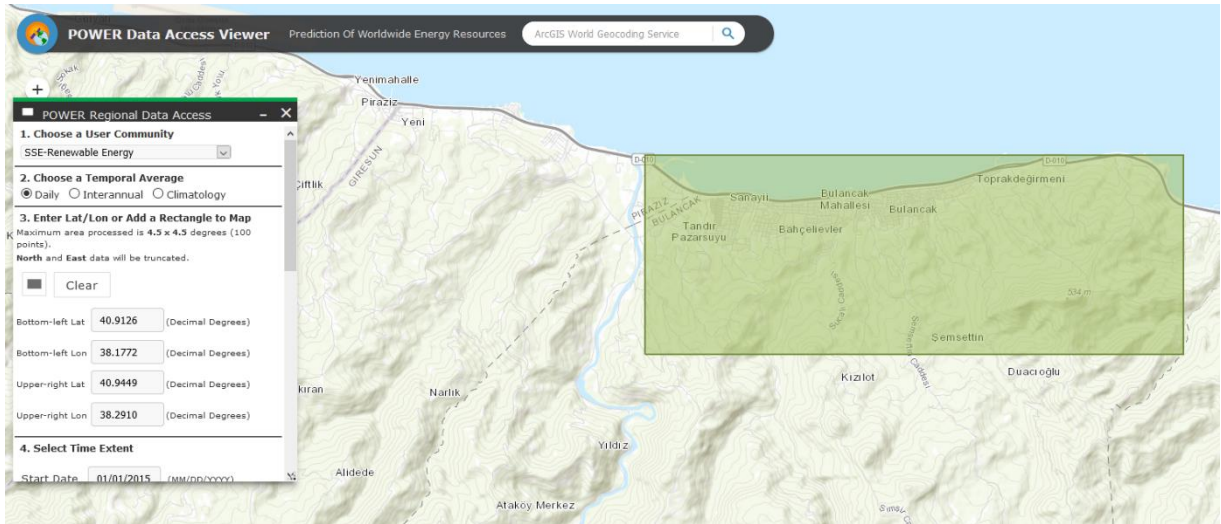
2.2. Veri Ön İşleme

TÜİK veri tabanı üzerinden elde edilen en eski tarihli veri 2004 yılına aittir. Veri ön işleme aşamasında, veri setine temel olarak bütün ilçelere ait kaydedilen verilerde tutarlılığı gözetmek adına, 2004-2020 yılları arasında eksik veri bulunan ilçeler dahil edilmemiştir. Sakarya'nın Adapazarı ilçesine ait eksik veri örneği bu duruma örnek olarak Şekil 3 ile verilmiştir. Burada 2004-2007 arasındaki verilerin işlenmediği görülmektedir.

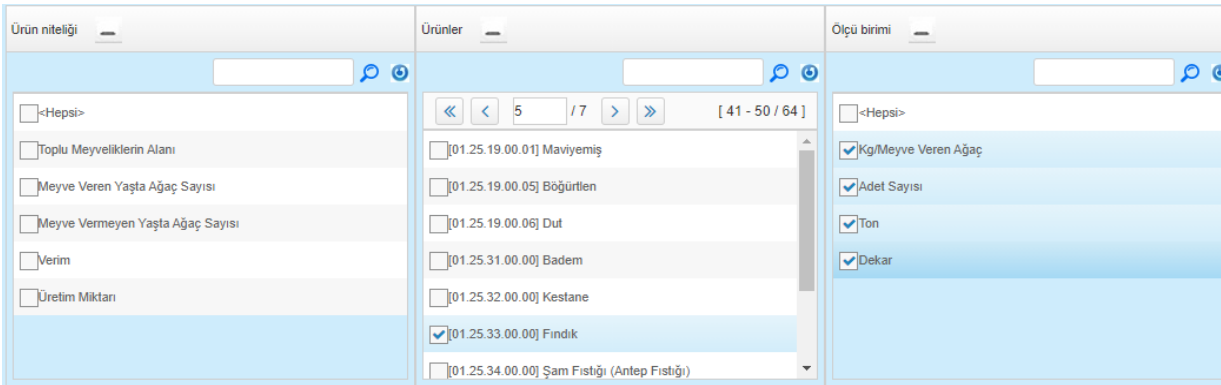
TÜİK/Biruni platformu üzerinden elde edilen parametreler aşağıda sıralanmıştır:

1. Meyve veren yaştaki ağaç sayısı
2. Toplu Meyveliklerin Alanı (Dekar)
3. Üretim Miktarı (Kg)
4. Verim (Kg/Dekar)

Burada üretim miktarı sadece verim hesaplamasında kullanılmış olup modelin girdisi olarak kullanılmamıştır. İlk veri ön işleme aşamasına ilk olarak verimin kg/dekar şeklinde hesaplanmasıyla başlanmıştır. Tespit edilen aykırı gözlem değerleri o ilçeye ait ortalama değerlerle tekrar düzenlenmiştir ve başka bir ön işleme yapılmamıştır.



Şekil 1. Giresun Bulancak ilçesine ait iklimsel verilerin elde edilmesi.



Şekil 2. TÜİK platformu kırılım seçilimi.

2008	503320	7583	1856000	3687.5149	244.7580113
2009	303320	7583	1213000	3999.0769	159.9630753
2010	303320	7580	910000	3000.1319	120.0527704
2011	303320	7580	607000	2001.1869	80.07915567
2012	298000	7450	745000	2500.0000	100
2013	177500	3550	516000	2907.0423	145.3521127
2014	177500	3550	710000	4000.0000	200
2015	177550	3550	568000	3199.0988	160
2016	190000	4500	632000	3326.3158	140.4444444
2017	213750	4750	737000	3447.9532	155.1578947
2018	236250	5250	336000	1422.2222	64
2019	238600	5300	452000	1894.3839	85.28301887
2020	280000	6220	513000	1832.1429	82.47588424

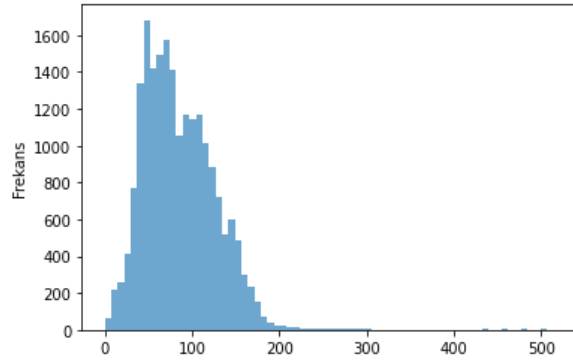
Şekil 3. Veri setine dahil edilmeyen ilçe bazında eksik veri örneği.

2.3. Sentetik Veri Üretimi

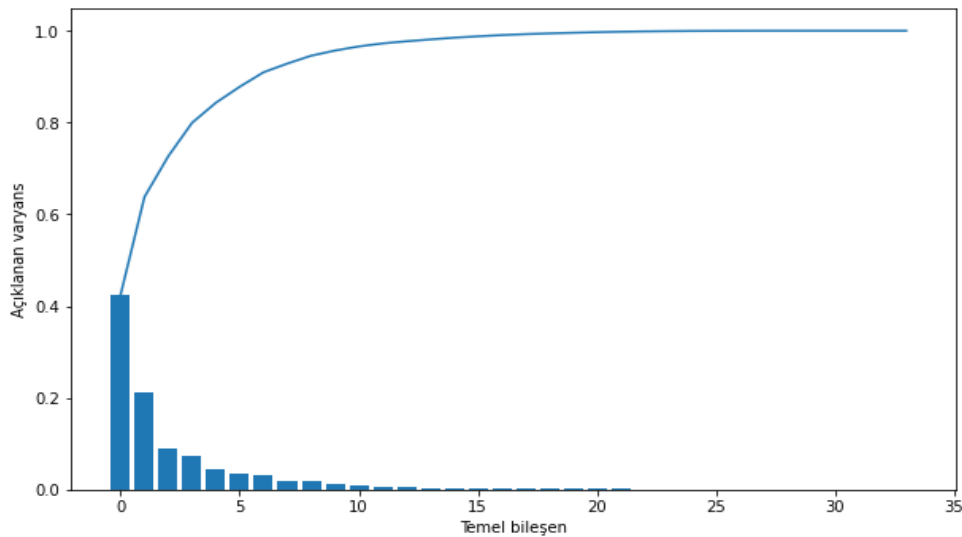
Rekoltenin yıllık olarak elde edilmesi nispeten kısıtlı bir veri seti oluşturulmasına neden olmaktadır. Bu sorunun çözümü için lineer interpolasyon yöntemine başvurulmuştur. Kısaca bu yöntem, lineer polinomlar yardımıyla bilinen veri noktalarından yeni verilerin üretilebilmesine olanak sağlayan bir yöntemdir (Box ve ark., 2015). Lineer interpolasyon yöntemi ile yeni veriler oluşturulmuş ve bu sayede derin öğrenme modeli için yeterli seviyede örnek sayısına ulaşılmıştır. Daha sonra bu işlemin doğru bir şekilde gerçekleşip gerçekleşmediğini kontrol etmek amacıyla tanımsız değerler ve yinelenen satırlar aranmıştır. İşlem sonucunda elde edilen final verisinin verim parametresi dağılımı Şekil 4'te verilmiştir.

2.4. Temel Bileşenler Analizi

Temel bileşenler analizi (Principal Component Analysis, PCA), veri seti boyutunu sıkıştırarak hesaplama maliyetini düşüren, bağımlı değişken kümesindeki varyansın büyük çoğunluğunu açıklayabilecek değişkenlerin seçilmesini sağlayan ve model performansını iyileştirerek daha iyi tahminleme sonuçlarının elde edilmesine yardımcı olan bir yöntem olarak kullanılmaktadır. Temel bileşenlerdeki birim değişimin varyans üzerindeki açıklayıcılık dağılımı Şekil 5'te görülmektedir. Bu analiz sonucunda en yüksek açıklayıcılığa sahip beş değişken belirlenmiş ve model girdisinde bu beş temel bileşen etkili olmuştur.



Şekil 4. Verim parametresinin histogram grafiği.



Şekil 5. Temel bileşenler analizi açıklanan varyans grafiği.

3. Bulgular ve Tartışma

Veri setinde bulunan 34 değişken, PCA ile 5 temel bileşene indirilmiş ve öznelik olarak kullanılmıştır. Zaman serisi olarak geçmiş tarih uzunluğu 24 olarak seçilmiştir. Elde edilen veri setinin %60'ı eğitim, %20'si validasyon ve geriye kalan %20'si de test aşamasında kullanılmak üzere ayrılmıştır.

3.1. Model Tasarımı

Bu çalışma doğrultusunda, birisi denek model olmak üzere toplamda 5 yapay öğrenme modeli geliştirilmiş ve belirlenen metrikler üzerinden bu modellere ait kıyaslamalar yapılmıştır. Naive modelleme, XGBoost yapısı ile gerçekleştirilmiş ve bu modelin denek model olarak temel kıyaslamada kullanılması öngörülmüştür. Bunun yanı sıra, geliştirilen modeller CNN veya LSTM yapısını kullanmaktadır. Ayrıca, bu iki ağı hibrit şekilde tasarımıyla elde edilen model de eğitilmiştir. Temel kıyaslamada ise belirlenen metrik, tahminleyicinin tahmin ettiği değerler ile gerçek değerleri arasındaki uzaklığın bulunmasında sıklıkla kullanılan, hatanın büyüklüğünü ölçen kuadratik bir metrik olan kök ortalama kare hata (Root Mean Square Error, RMSE)'dir. Temelde tahmin hatalarının (kalıntıların) standart sapmasına işaret etmektedir. Bunun dışında başarımlarını sağlamak ve hesaplama maliyetini azaltmak adına *Eğitimi Erken Durdurma (Early Stopping)* yöntemi kullanılmıştır. Buna ek olarak Dropout işlemi de bu doğrultuda kullanılan bir diğer yöntemdir.

3.2. Deneysel Sonuçlar

3.2.1. XGBoost yapısı

Bu yöntem, 2015 yılında Chen ve arkadaşları tarafından yapılan araştırmadaki (Chen ve ark., 2015) gradient boosting algoritmasının optimize edilmiş bir versiyonu olarak ifade edilebilir. Çok hızlı olması ve yüksek doğruluklu tahminlemelere olanak sağlaması nedeniyle çoğu tahminleme probleminde ilk başvurulmuş yöntemlerden birisi olarak kabul edilebilmektedir. Bu doğrultuda model, maksimum derinlik 180 olacak şekilde eğitilmiştir. Bir adım sonraki verim tahminine ait RMSE değeri 8,882 kg/dekar olarak elde edilmiştir. Bu RMSE değeri, tahminlemede kullanılan modeller arasında üretilen en yüksek değer olarak karşımıza çıkmaktadır.

3.2.2. Yalın LSTM modeli

LSTM ağırları 1997'de Hochreiter ve Schmidhuber tarafından ortaya atılan bir tekrarlayan sinir ağı (Recurrent Neural Network, RNN) olarak tasarlanmıştır (Hochreiter ve Schmidhuber, 1997). LSTM'ler, uzun vadeli bağımlılık sorununa engel olmak amacıyla geliştirilmiştir. Geleneksel RNN sinir ağlarının eğitilmesi süresince yaşanan sorunlar LSTM'de tamamen giderilmiştir. LSTM, hafıza geçişli mekanizmasının

varlığından dolayı uzun vadeli bağımlılıkları öğrenebilir olmasıyla zaman serisi ya da sıralı problemlerde yaygın bir biçimde kullanılmaktadır (Ahmadi ve ark., 2019).

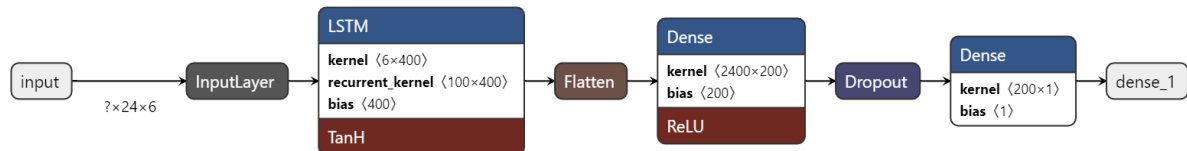
Bu çalışmada kullanılan LSTM modeli incelendiğinde, öncelikle giriş özneliklerinin LSTM katmanına verildiği söylenebilir. Daha sonra bu verilere düzleştirme katmanı uygulanarak elde edilen veriler tam bağlantılı katmana verilmiştir. Aşırı öğrenmeden kaçınmak için *Dropout* katmanı eklenmiştir. Son olarak verim tahminini ifade eden tek skaler değer için tek nörona sahip çıkış katmanı eklenmiştir. Model ayrıntıları Şekil 6'daki blok diyagram ile gösterilmiştir. Ayrıca eğitim sırasında elde edilen hata ve kayıp değerlerindeki değişime ait grafik Şekil 7'de verilmiştir. Bu grafikte, validasyon verisindeki hatanın genel anlamda eğitim verisine ait hatadan daha düşük seyrettiği görülmektedir. Modelde herhangi bir aşırı öğrenme görülmemekle birlikte, Şekil 7'den eğitim verisi üzerindeki kayıp düşüşünün validasyon verisi üzerindeki kayıp düşüşüyle paralel olduğu görülmektedir. Elde edilen RMSE skoru 4,384 kg/dekar olarak hesaplanmıştır.

3.2.3. Yığılmış LSTM modeli

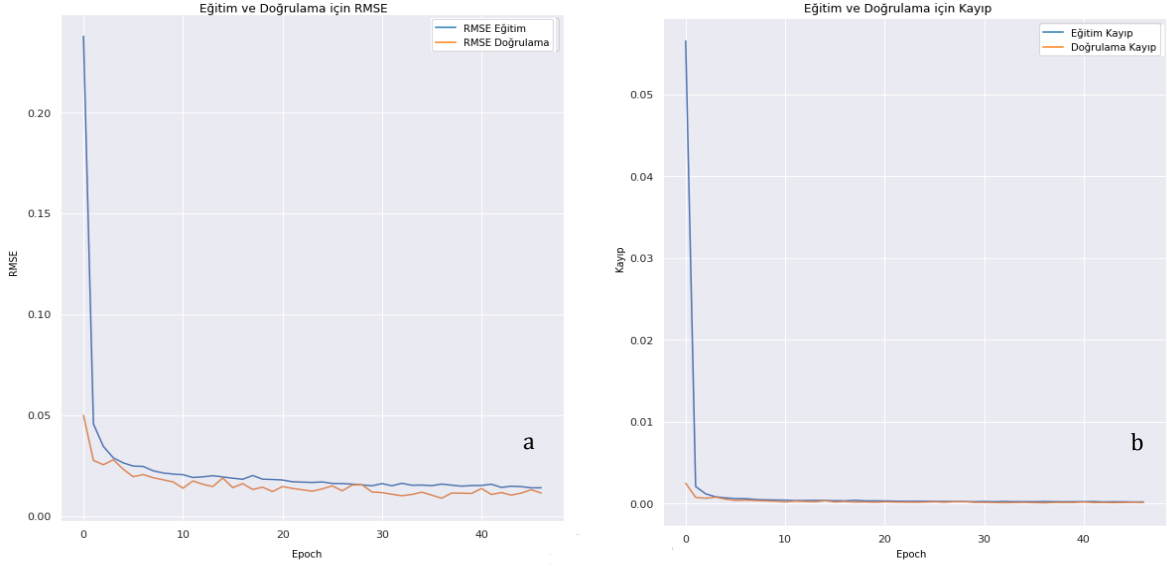
Yığılmış LSTM modelinde model, yalnız LSTM'den farklı olarak LSTM modüllerinin peşpeşe kullanılmasıyla oluşturulmuştur. Model yapısı Şekil 8, hata ve kayıp grafikleri Şekil 9'da verilmiştir. Grafikler incelendiğinde, Yığılmış LSTM modeli için de herhangi bir aşırı öğrenme durumunun söz konusu olmadığı görülebilmektedir. Modelin öğrenmesi için veri seti büyüklüğünün yeterli olduğu gözlemlenmiştir. RMSE değeri ise 2,11 kg/dekar olarak sonuçlanmıştır. RMSE değerine bakıldığında, LSTM'de modül sayısının artırılması ile hata değerinin bir önceki modele göre yarı yarıya düştüğü söylenebilir.

3.2.4. CNN modeli

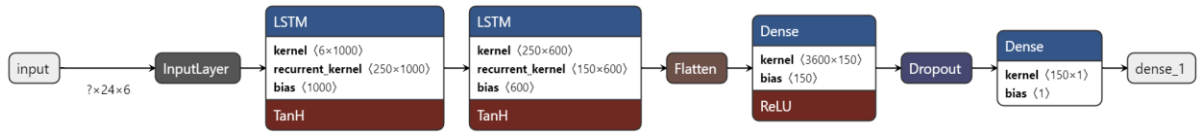
CNN modelinde, LSTM modelinde kullanılan geçmiş 24 zaman verisi girdi olarak kullanılmıştır. Model ayrıntıları Şekil 10 ile gösterilmiştir. Giriş verileri tek boyutlu konvolüsyon katmanına verilmiştir. Daha sonra LSTM modellerinde de bulunan sırasıyla düzleştirme, tam bağlantılı katman ve çıkış katmanı ile model hazır hale getirilmiştir. Eğitim sırasında erken durdurma işlemi yapıldığından tüm epochlar gerçekleştirilmeden eğitim tamamlanmıştır. Eğitime ait RMSE ve kayıp grafikleri Şekil 11 ile verilmiştir. CNN modeline ait parametre değişimleri incelendiğinde, zaman zaman doğrulama hata skoru eğitim skorunun üzerinde dalgalanmalar gösterse bile, bu durumdan aşırı öğrenme meydana geldiği sonucu çıkarılamayacaktır. CNN modeli eğitiminden sonra test verisi üzerinde elde edilen RMSE değeri 1,657 kg/dekar seviyesine gerilemiştir.



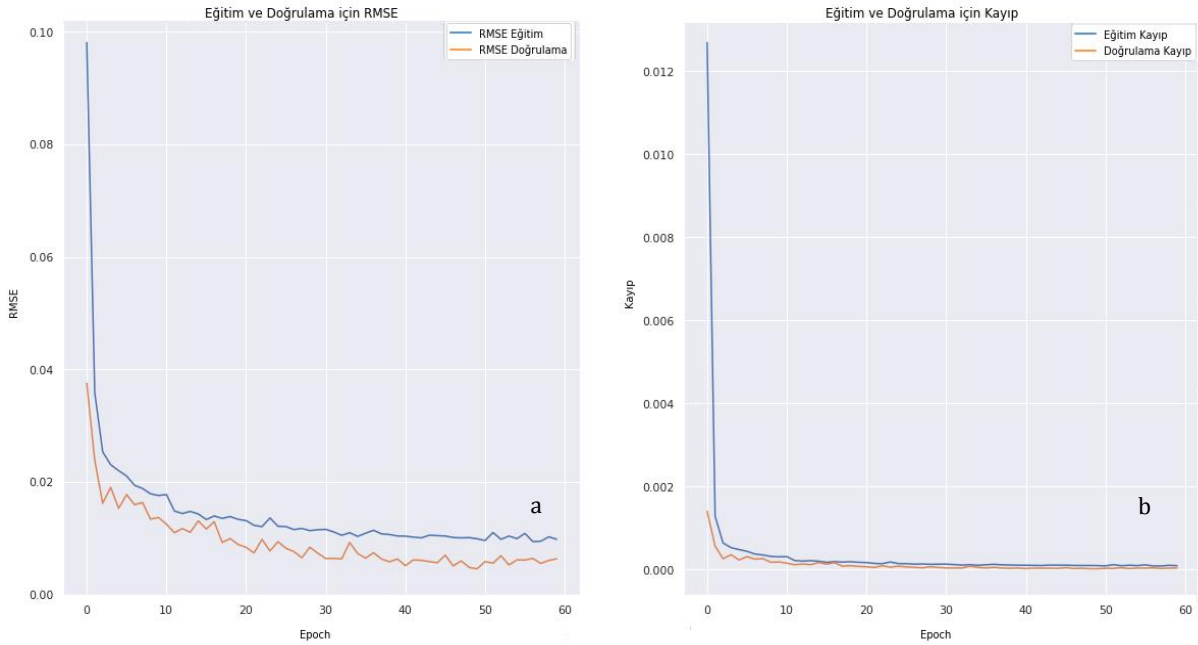
Şekil 6. Yalın LSTM yapısı.



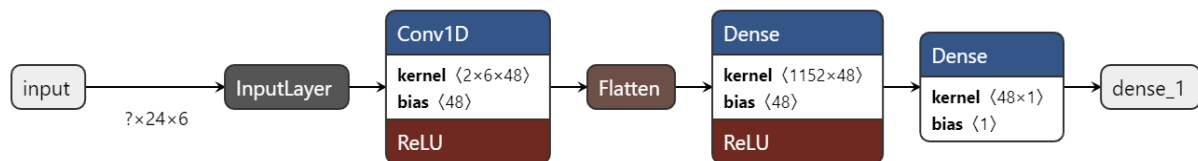
Şekil 7. Yalın LSTM modeli eğitim grafikleri (a) RMSE, (b) Kayıp.



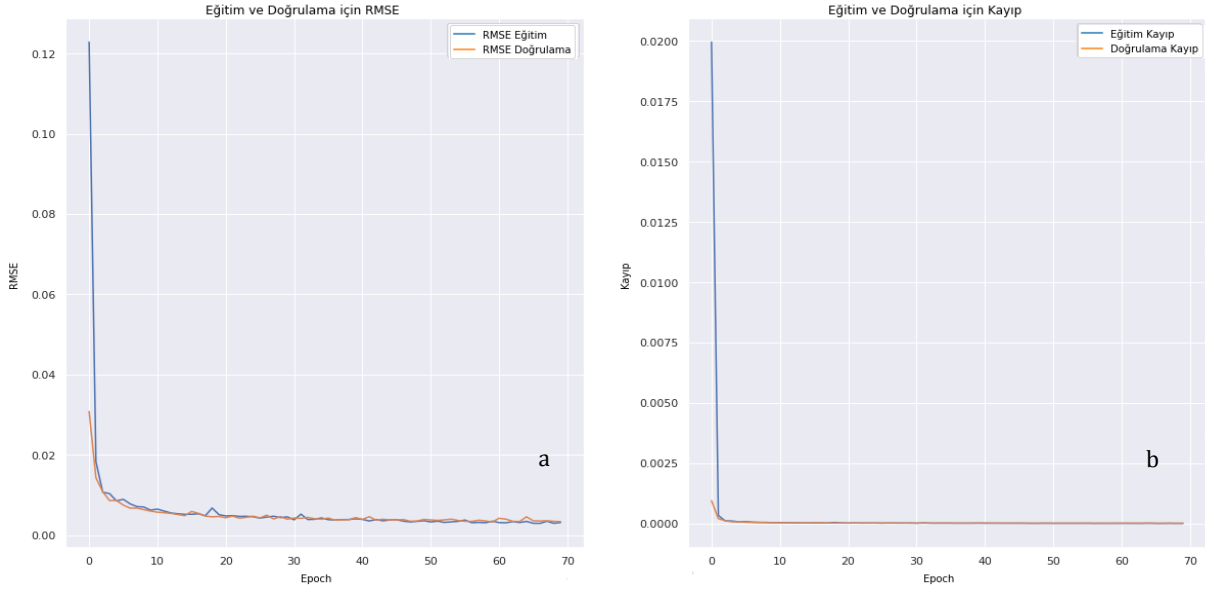
Şekil 8. Yığılmış LSTM model yapısı.



Şekil 9. Yığılmış LSTM modeli eğitim grafikleri (a) RMSE, (b) Kayıp.



Şekil 10. CNN model yapısı.



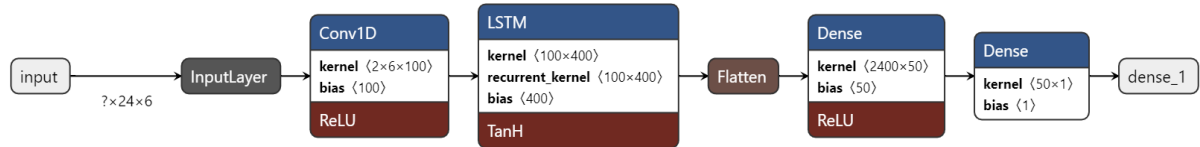
Şekil 11. CNN modeli eğitim grafikleri (a) RMSE, (b) Kayıp.

3.2.5. CNN-LSTM hibrit model

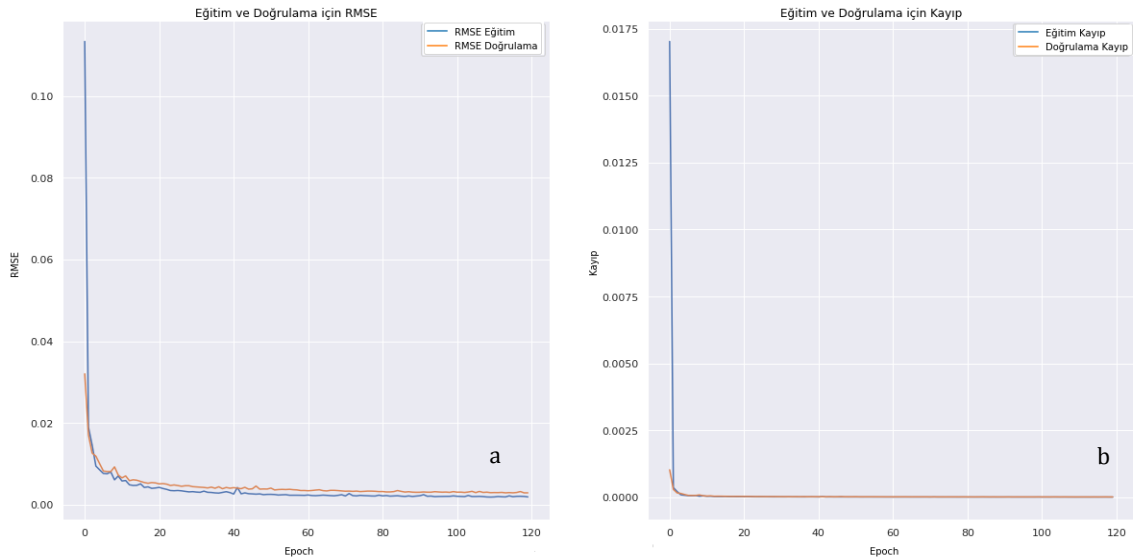
Çalışmada kullanılan son model, CNN modelindeki konvolüsyon katmanı ve LSTM katmanının sırasıyla kullanılarak hibrit edilmesiyle oluşturulan CNN-LSTM modelidir. Bu modelde, giriş verileri konvolüsyon katmanına verilmiştir ve bu katmanı LSTM katmanı takip etmiştir. Bu noktadan sonraki katmanlar CNN modeli ile benzerlik göstermektedir. Katmanlara ait parametre değerleri deneysel olarak seçilmiştir. Model ayrıntıları Şekil 12'deki blok diyagram ile verilmiştir. Ayrıca, RMSE değerinin değişim grafiği ve model eğitimi sırasındaki

kayıp değeri grafiği sırasıyla Şekil 13 (a) ve Şekil 13 (b) ile verilmiştir. Grafikler incelendiğinde, doğrulama hatasının çoğu zaman eğitim hatası ile paralellik gösterdiği görülmüştür. CNN-LSTM modeli ile RMSE değeri, diğer modellerle kıyaslandığında en düşük seviyeye gerilemiştir. Modelin test verisi üzerindeki RMSE değeri 1,321 kg/dekardır.

Son olarak tüm modeller ve hata değerleri karşılaştırmaları Tablo 2 ile gösterilmiştir. Sonuçlar incelendiğinde en az hata değerine CNN-LSTM Hibrit modeli ile ulaşıldığı görülmektedir.



Şekil 12. CNN-LSTM modeli yapısı.



Şekil 13. CNN - LSTM modeli eğitim grafikleri (a) RMSE, (b) Kayıp.

Tablo 2. Kullanılan bütün modellerin RMSE performansı

Model Adı	RMSE değeri (kg/dekar)
XGBoost yapısı	8,882
Yalın LSTM model	4,384
Yığılmış LSTM model	2,11
CNN model	1,657
CNN-LSTM hibrit model	1,321

4. Sonuç

Mahsul verim tahmin çalışmalarının önemi gün geçtikçe artmaktadır. Bu durum, ilgili alanda çalışan kişilerin karar verme sürecini etkileyebilmektedir. Böylece, verim artımı için gerekli ön hazırlıkların sağlıklı bir şekilde yapılması sağlanabilmektedir. Literatür incelendiğinde, son zamanlarda veri tahmini amacıyla ML yöntemlerine bir eğilimin olduğu görülmektedir. Fakat fındık ürünü özelinde yapılmış DL tabanlı çalışmaya rastlanmamıştır. Bu çalışmada ise, fındık verim tahmini için ML modeli, DL tabanlı yaklaşımlar ve bunlara ait hibrit modeller önerilmiştir. Modele giriş olarak fındık üretimi yapılan 5 il ve ilçelerinden elde edilen öznitelikler kullanılmıştır. Bu öznitelikler, TÜİK ve NASA Power platformu tarafından elde edilen sıcaklık, bağıl nem, yağış, rüzgar hızı, meyve veren yaştaki ağaç sayısı, alan vb. gibi 34 farklı veriden oluşmaktadır. Model eğitimi için öncelikle PCA analizi uygulanarak 34 öznitelik 5 öznitelige indirgenmiştir. XGBoost yapısı, Yalın LSTM modeli, Yığılmış LSTM modeli, CNN modeli ve CNN-LSTM hibrit modeli olmak üzere 5 farklı model kullanılmıştır. Bu modeller eğitildikten sonra test verisi üzerinde yapılan tahminler incelenmiştir. Model performans karşılaştırma işlemi, test üzerindeki RMSE değerleri üzerinden yapılmıştır. Karşılaştırma sonuçları incelendiğinde, sırasıyla en iyi ve en kötü RMSE değerlerinin CNN-LSTM hibrit modeli ve XGBoost modeli ile elde edildiği görülmektedir.

Katkı Oranı Beyanı

Yazar(lar)ın katkı yüzdesi aşağıda verilmiştir. Tüm yazarlar makaleyi incelemiş ve onaylamıştır.

	D.Y.	G.Y.	S.D.
K	30	30	40
T	25	50	25
Y			100
VTI	20	40	40
VAY	40	40	20
KT	60	20	20
YZ	60	20	20
KI	40	30	30
GR	100		
PY	50		50
FA			

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi, FA= fon alımı.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

- Ahmadi N, Constandinou TG, Bouganis CS. 2019. Decoding hand kinematics from local field potentials using long short-term memory (LSTM) network. 9th International IEEE/EMBS Conference on Neural Engineering (NER), March 20-23, 2019, San Francisco, US, pp: 415-419.
- Ali I. 2010. The quest for active carbon adsorbent substitutes: inexpensive adsorbents for toxic metal ions removal from wastewater. *Separ Purificat Rev*, 39(3-4): 95-171.
- An N, Turp MT, Türkeş M, Kurnaz ML. 2020. Mid-term impact of climate change on hazelnut yield. *Agriculture*, 10: 159.
- Asar M, Yalçın S, Yücel G, Nadaroğlu Y, Erciyas H. 2007. Zirai meteoroloji. Çevre ve Orman Bakanlığı Devlet Meteoroloji İşleri Genel Müdürlüğü Yayınları, Ankara, Türkiye, pp: 189.
- Bali N, Singla A. 2022. Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. *Archives Comput Methods Eng*, 29: 95-112.
- Beyhan N, Odabaş F. 1996. İklimsel faktörlerin fındıkta verimlilik üzerine etkileri ve yetiştiricilik açısından önemi. *Ondokuz Mayıs Üniv Zir Fak Derg*, 11(1): 177-188.
- Box GE, Jenkins GM, Reinsel GC, Ljung GM. 2015. Time series analysis: forecasting and control. John Wiley & Sons, New York, US, pp: 712.
- Cedric LS, Adoni WYH, Aworka R, Zoueu JT, Mutombo FK, Krichen M, Kimpolo CLM. 2022. Crops yield prediction based on machine learning models: case of west african countries. *Smart Agri Tech*, 2: 100049.
- Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T. 2015. Xgboost: extreme gradient boosting. URL: <http://mirrors.nic.cz/R/web/packages/xgboost/xgboost.pdf> (erişim tarihi: 10 Ekim 2022).
- Cioffi R, Travaglioni M, Piscitelli G, Petrillo A, De Felice F. 2020. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, 12(2): 492.
- Desloires J, Ienco D, Botrel A. 2023. Out-of-year corn yield prediction at field-scale using sentinel-2 satellite imagery and

- machine learning methods. Available at SSRN 4329505. DOI: 10.2139/ssrn.4329505.
- Fei S, Hassan MA, Xiao Y, Su X, Chen Z, Cheng Q, Duan F, Chen R, Ma Y. 2022. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Prec Agri*, 1: 26.
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Comput*, 9(8): 1735-1780.
- İrdem C. 2021. Türkiye fındık verimi üzerinde sıcaklık ve yağışın etkileri. *Coğrafi Bil Derg*, 19(1): 242-262.
- Kent R. 2020. Data construction and data analysis for survey research. Bloomsbury Publishing, New York, US, pp: 248.
- Kim N, Lee YW. 2016. Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. *J Korean Soc Survey Geodesy Photogrammetry Cartography*, 34(4): 383-390.
- Köksal İ. 2002. Türk fındık çeşitleri. Fındık tanıtım Grubu Yayınları, Ankara, Türkiye, pp: 136.
- Liliane TN, Charles MS. 2020. Factors affecting yield of crops. In: Amanullah Dr, editör: *Agronomy-climate change & food security*, IntechOpen, Londoni UK, pp: 9.
- Liu J, Goering C, Tian L. 2001. A neural network for setting target corn yields. *Transact ASAE*, 44(3): 705.
- Marko O, Brdar S, Panic M, Lugonja P, Crnojevic V. 2016. Soybean varieties portfolio optimisation based on yield prediction. *Comput Electron Agri*, 127: 467-474.
- Paudel D, Boogaard H, de Wit A, van der Velde M, Claverie M, Nisini L, Janssen S, Osinga S, Athanasiadis IN. 2022. Machine learning for regional crop yield forecasting in Europe. *Field Crops Res*, 276: 108377.
- Romero JR, Roncallo PF, Akkiraju PC, Ponzoni I, Echenique VC, Carballido JA. 2013. Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. *Comput Electron Agri*, 96: 173-179.
- Şimşek O, Nadaroğlu Y, Yücel G, Dokuyucu O, Gökdağ A. 2014. Türkiye fenoloji atlası. Orman ve Su İşleri Bakanlığı, Ziraat Meteoroloji Şube Müdürlüğü, Ankara, Türkiye, pp: 104.
- Tuncer C, Özdemir İO, Kushiyev R. 2019. New pest species in Turkish hazelnut orchards: *Xylosandrus germanus* Blandford (Coleoptera: Curculionidae: Scolytinae), *Metcalfa pruinosa* Say (Hemiptera: Flatidae), *Croesus septentrionalis* Linnaeus (Hymenoptera: Tenthredinidae) and *Anoplophora chinensis* Forster (Coleoptera: Cerambycidae). *BSJ Agri*, 3(1): 74-81.
- Üstün NŞ, Karaosmanoğlu H. 2017. Sert kabuklu meyveler ve fonksiyonel özellikleri. *Meyve Bilimi*, 2: 142-148.
- Villanueva MB, Salenga MLM. 2018. Bitter melon crop yield prediction using machine learning algorithm. *International J Adv Comput Sci Appl*, 9: 3.
- Yadav SSP, Reddy R, Dharun D, Niveditha A, Hema N. 2021. Crop yield prediction using image processing. *Nveo-Natural Volats Essen Oils J*, 2021: 11981-11992.
- Zaman M. 2004. Geographical Distribution and Production of Hazelnut in Turkey. *Doğu Coğrafya Derg*, 9(11): 49-92.
- Zhao S, Zheng H, Chi M, Chai X, Liu Y. 2019. Rapid yield prediction in paddy fields based on 2D image modelling of rice panicles. *Comput Electron Agri*, 162: 759-766.