

# TÜRKİYEDE PAMUK REKOLTE KESTİRİMİ İÇİN BÖLGESEL AYZ YORUMLANABİLİR LSTM MODELLERİNİN KARŞILAŞTIRILMASI

Furkan YARDIMCI<sup>1</sup>, Alp ERTÜRK<sup>2</sup>, Esra ERTEN<sup>3</sup>

<sup>1</sup> Kocaeli Üniversitesi, Elektronik ve Haberleşme Mühendisliği Bölümü, 41001, İzmit, Kocaeli, furkanyardimci1006@gmail.com

<sup>2</sup> Doç. Dr., Kocaeli Üniversitesi, Elektronik ve Haberleşme Mühendisliği Bölümü, 41001, İzmit, Kocaeli, alp.erturk@kocaeli.edu.tr

<sup>3</sup> Prof. Dr., İstanbul Teknik Üniversitesi, Geomatik Mühendisliği Bölümü, 34000, Maslak, İstanbul, erten@itu.edu.tr

## ÖZET

Bu çalışma, küresel tekstil endüstrisinin vazgeçilmez hammaddelerinden biri olan pamuğun derin öğrenme temelli rekolte kestirim süreçlerinin yorumlanabilirliğini artırmayı hedeflemektedir. Derin öğrenme yaklaşımları, özellikle Uzun Kısa Süreli Hafıza (Long Short Term Memory - LSTM) ağları, tarımsal rekolte kestiriminde yüksek başarı sağlamaktadır; ancak bu modellerin "kara kutu" yapısı, kestirim sürecini yönlendiren faktörlerin anlaşılmasını zorlaştırmaktadır. Bu çalışmada, LSTM modellerinin karar verme süreçlerini şeffaflaştırmak ve verideki her bir girdi özelliğinin rekolte kestirimi üzerindeki etkisini belirlemek için Açıklanabilir Yapay Zeka (AYZ) tekniklerinden biri olan Shapley Katkı Açıklamaları (SHapley Additive exPlanations - SHAP) yöntemi uygulanmıştır. Araştırmada, Türkiye'nin Ege, Akdeniz ve Güneydoğu Anadolu bölgelerinden toplanan iklim, toprak ve uzaktan algılama verilerinden oluşan bir veri kümesi ve yerleşim birimi seviyesindeki rekolte değerleri kullanılmıştır. Ege, Akdeniz ve Güneydoğu Anadolu bölgelerinin her biri için LSTM modelleri ayrı ayrı eğitilmiştir. Veri kümesi, ERA5 iklim verileri, Landsat-8 uydu verilerinden elde edilen bitki örtüsü indeksleri, SoilGrids'ten alınan toprak ölçümleri ve Sentinel-1 sentetik açıklık radar (SAR) polarimetri verilerini içermektedir. LSTM modelleri bölgesel verilerde eğitilmiş ve SHAP analizi ile modellerin yorumlanabilirliği sağlanmıştır. Yerleştirilmiş modeller yüksek başarımla rekolte kestirimi sağlarken, SHAP analizleri ise özneliklerin katkısının nicel olarak değerlendirilmesi sağlamaktadır. Elde edilen bulgular, farklı bölgelerde pamuk rekoltesi kestirimi üzerinde etkili olan özneliklerin önemini nicel olarak ortaya koymaktadır. Çalışmamız, Türkiye'de tarımsal üretkenliği ve sürdürülebilirliği artırmaya yönelik adaptif yönetim stratejilerinin geliştirilmesine katkı sağlamayı hedeflemektedir.

**Anahtar Sözcükler:** Açıklanabilir yapay zeka (AYZ), LSTM, pamuk, rekolte kestirimi, SHAP.

## ABSTRACT

### REGIONAL COMPARISON OF XAI-INTERPRETED LSTM MODELS FOR COTTON YIELD ESTIMATION IN TURKIYE

This study aims to enhance the interpretability of deep learning based yield estimation processes of cotton, a vital raw material for the global textile industry. Deep learning approaches, particularly Long Short-Term Memory (LSTM) networks, have achieved significant success in agricultural yield prediction; however, the "black-box" nature of these models makes it difficult to understand the factors driving their predictions. In this study, the SHapley Additive exPlanations (SHAP) method, an eXplainable Artificial Intelligence (XAI) method, is utilized to make the decision-making processes of LSTM models transparent and to determine the impact of each input feature on yield estimation. The research utilizes a dataset comprising of climate, soil, and remote sensing data collected over the Aegean, Mediterranean, and Southeastern Anatolia regions of Türkiye, along with commune level yield values. LSTM models are trained separately for each region. The dataset includes ERA5 climate data, vegetation indices derived from Landsat-8 satellite data, soil measurements from SoilGrids, and Sentinel-1 synthetic aperture radar (SAR) polarimetry data. The LSTM models are trained on regional data, and SHAP analysis is employed to achieve model interpretability. Localized models provide high-performance yield prediction, whereas SHAP analyses provide quantitative evaluation of the features' contributions. The findings reveal the quantitative importance of various features influencing cotton yield estimation across different regions. Our study aims to contribute to the development of adaptive management strategies that can enhance agricultural productivity and sustainability in Türkiye.

**Keywords:** Cotton, explainable artificial intelligence (XAI), LSTM, SHAP, yield estimation.

## 1. GİRİŞ

Tarım, ekonomik kalkınmanın temel taşlarından biridir. Tarımsal üretkenlik yalnızca gıda arzını güvence altına almakla kalmaz, aynı zamanda istihdam yaratır, sanayi sektörü için hammadde sağlar ve ihracat yoluyla döviz kazancı sağlar. Özellikle gelişmekte olan ülkelerde tarım, yoksulluğun azaltılması ve kırsal kalkınmanın sağlanmasında kritik bir rol oynar. Tarımsal büyüme, genel ekonomik büyümeyi teşvik ederek yaşam standartlarını yükseltir ve sürdürülebilir kalkınmaya katkıda bulunur (Timmer, 2002). Ancak, tarımsal üretim; iklim değişiklikleri, hastalıklar, zararlılar ve ekonomik dalgalanmalar gibi birçok belirsizlikten etkilenmektedir. Bu belirsizlikler, üretimde dalgalanmalara yol açarak ülke ekonomisinde istikrarsızlık yaratabilir (Nyéki ve Neményi, 2022). Rekolte kestirimi, tarımsal üretimdeki bu belirsizlikleri azaltmak ve üretimin sürdürülebilirliğini sağlamak için kritik bir araçtır. Doğru ve güvenilir rekolte tahminleri, çiftçilerden tarım politikası geliştiricilerine kadar birçok karar vericinin bilinçli ve etkin kararlar almasına olanak tanır (Nyéki ve Neményi, 2022).

Pamuk gibi küresel öneme sahip ürünlerde rekolte tahminlerinin doğruluğu hem üreticiler hem de pamukla ilgili endüstriyel ve ekonomik paydaşlar için büyük önem taşımaktadır. Pamuk rekoltesindeki dalgalanmalar, tekstil endüstrisinden diğer yan sektörlerle kadar birçok alanda etkili olabilir (Nyéki ve Neményi, 2022). Bu sebeple, pamuk rekoltesi tahminlerinin doğruluğu, üretim planlaması ve pazar stratejilerinin başarılı bir şekilde uygulanması için kritik öneme sahiptir.

Son yıllarda, uzaktan algılama verilerinin derin öğrenme temelli zaman serisi analizlerinin tarımsal rekolte kestirimlerinde büyük bir potansiyele sahip olduğu görülmüştür (Bhagat vd., 2022). Bu veriler ve analizler, tarım alanlarındaki mevsimsel değişiklikleri ve uzun vadeli trendleri izleyerek geleneksel yaklaşımlara göre daha doğru, daha geniş kapsamlı ve daha erken rekolte tahminleri yapılmasına olanak tanımaktadır (Joshi vd., 2023). Özellikle Uzun Kısa Süreli Hafıza (Long Short-Term Memory - LSTM) modelleri (Hochreiter, 1997) gibi derin öğrenme teknikleri, zaman serisi verilerindeki karmaşık zamansal ilişkileri yakalama yetenekleri (Van Houdt vd., 2020) sayesinde geleneksel yaklaşımlara göre daha yüksek performans sunmaktadır ve bu nedenle bu modellerin tarımsal verim tahminlerinde kullanım sıklıkları artmaktadır (Martinez-Ferrer vd., 2020; Tian vd., 2021). Ancak, derin öğrenme modellerinin “kara-kutu” yapısı (Fomin, 2022), tahminleri yönlendiren faktörlerin anlaşılabilirliği ve yorumlanabilirliğini sınırlamaktadır. Derin öğrenmenin bu eksikliğini gidermek için, Açıklanabilir Yapay Zeka (AYZ) yöntemleri geliştirilmiştir ve kullanılmaktadır (Taskin vd., 2024).

AYZ yöntemleri, son yıllarda, tarımsal yapay zekâ uygulamalarında da derin öğrenme modellerin yorumlanabilirliğini artırarak kullanıcıların yapay zekâ kararlarını daha iyi anlamalarını ve bu kararlara güvenlerinin artmasını sağlamak amacıyla kullanılmaya başlamıştır (Linheiro vd., 2023; Mateo-Sanchis vd., 2023, Celik vd., 2023a, , Celik vd., 2023b). Bu bağlamda özellikle Shapley değerlerine (Shapley, 1953) dayanan Shapley Katkı Açıklamaları (SHapley Additive exPlanations - SHAP) (Lundberg, 2017) AYZ yönteminin kullanımı öne çıkmaktadır (Jones vd., 2022; Isik vd., 2023; Zhou vd., 2024). SHAP, makine öğrenmesi ve yapay zeka yaklaşımları ile rekolte kestirimi için yorumlanabilirlik amacıyla kullanıldığında, her bir özneliğin kullanılan modelin çalışmasında etkisinin nicel olarak değerlendirilmesini sağlamaktadır.

Bu çalışma kapsamında, Türkiye'nin üç farklı bölgesinde derin öğrenme ile pamuk rekolte kestiriminde yorumlanabilirlik üzerinde çalışılmıştır. Çalışmada, üç bölgeden toplanan ERA5 iklim verileri, Landsat-8 multispektral uydu verilerinden elde edilen bitki örtüsü indeksleri, SoilGrids'ten (Poggio vd., 2021) alınan toprak ölçümleri ve Sentinel-1 sentetik açıklık radar (SAR) polarimetri verileri kullanılmıştır. Her bir bölge için LSTM derin öğrenme modeli ile ayrı ayrı rekolte kestirimi gerçekleştirilmiştir ve modellerin yorumlanabilirliği için her modele SHAP analizi uygulanmıştır. Türkiye'nin farklı tarım bölgelerinde yapılan pamuk üretimi, iklimsel ve çevresel faktörlerdeki değişiklikler ile bölgesel tarım uygulamaları nedeniyle büyük farklılıklar göstermektedir. SHAP analizi, her bir bölgedeki modelin hangi öznelilikleri rekolte tahminlerinde daha baskın olarak kullandığını ortaya koymuştur. Analizler bölgesel modellerde baskın öznelilikler arasında farklılıklar göstermektedir. Çalışma, yerel koşulların pamuk verimi üzerindeki etkilerini anlamak ve her bölgeye özgü tarımsal stratejiler geliştirilmesi için potansiyel ipuçları sunmaktadır. Su yönetimi, toprak özellikleri ve iklimsel değişkenlerin her bölgedeki rekolte kestirimleri üzerindeki etkileri takip eden detaylı çalışmalarda karşılaştırılarak, tarımsal verimliliği artırmaya yönelik bölgesel uygulamalar geliştirilebilir.

## 2. YÖNTEM

### 2.1 LSTM Ağları

LSTM ağları, geleneksel yinelenen sinir ağlarının (Recurrent Neural Networks - RNN) uzun vadeli bağımlılık ve gradyan sönümlenmesi sorunlarını aşmak üzere geliştirilmiş bir derin öğrenme mimarisidir. LSTM'ler, RNN'lerden farklı olarak hücre durumu adı verilen bir bellek yapısına sahiptir ve bu bellek, giriş ve çıkış dizileri arasındaki uzun süreli ilişkilerin korunmasını sağlayarak bilgi akışını daha verimli hale getirir (Hochreiter, 1997).

LSTM yapısı, giriş kapısı, unutma kapısı ve çıkış kapısı olmak üzere üç temel bileşenden oluşur. Bu kapılar, ağına hangi bilgilere odaklanacağını, hangi bilgileri saklayacağını ve hangi bilgileri unutacağını belirleyerek öğrenme sürecini yönetir. Giriş kapısı, hücreye giren yeni bilginin ne kadarının saklanacağına karar verir. Unutma kapısı, önceki hücre durumundan hangi bilginin unutulacağına karar verir. Çıkış kapısı ise hangi bilginin çıktı olarak kullanılacağını belirler. Giriş, unutma ve çıkış kapıları için denklemler sırasıyla (1), (2) ve (3)'de paylaşılmıştır.

$$it = \sigma(Wi[ht - 1, xt] + bi) \quad (1)$$

$$ft = \sigma(Wf[ht - 1, xt] + bf) \quad (2)$$

$$ot = \sigma(Wo[ht - 1, xt] + bo) \quad (3)$$

Bu kapıların birleşik etkisiyle hücre durumu sürekli güncellenir ve ağın çıktısı, mevcut hücre durumu ile yeni bilginin işlenmesi sonucunda elde edilmektedir. Bu yapıyla LSTM, zaman serisi verilerinde uzun vadeli bağımlılıkların etkili bir şekilde öğrenilmesini sağlamakta ve bu nedenle, zaman serisi ile yapılan tahmin uygulamalarında sıklıkla kullanılmaktadır. Bu çalışma kapsamında da, LSTM ağları, zaman serisi analizi ile rekolte kestirimi için kullanılmaktadır.

## 2.2 SHAP

SHAP, oyun teorisinden türetilmiş olup, bir makine öğrenimi modelinde "oyuncular" olarak kabul edilen her bir özneliğin model tahmin sonucuna yaptığı katkının adil bir şekilde hesaplanması amacıyla geliştirilmiştir (Lundberg, 2017). SHAP, karmaşık bir model olan  $g(x)$ 'in daha anlaşılır ve yorumlanabilir bir biçimde  $f(x)$  ile temsil edilmesini sağlamaktadır. Bu süreç, denklem (4) ile ifade edilmektedir:

$$f(x) = \phi_0 + \sum_{i=1}^n \phi_i x_i \quad (4)$$

Bu denklemde  $\phi_0$ , modelin tüm olası öznelik konfigürasyonları üzerindeki ortalama tahminini,  $\phi_i$  ise her özneliğin tahmine yaptığı katkıyı göstermektedir. SHAP, bu katkıları hesaplarken öznelikler arasındaki etkileşimleri de dikkate alır.

SHAP değerlerinin büyüklüğü, o özneliğin modelin tahmini üzerindeki etkisinin gücünü gösterir; pozitif SHAP değerleri tahminin artmasına, negatif SHAP değerleri ise azalmasına neden olur. Özellikle tarımsal verim tahmini gibi regresyon problemlerinde, SHAP, özneliklerin model üzerindeki etkisini daha iyi anlamak için güçlü bir araç olarak öne çıkmaktadır. Bu çalışma kapsamında, SHAP, rekolte kestirimi için kullanılan bölgesel LSTM modellerinin yorumlanabilirliği için kullanılmıştır.

## 2.3 Model Değerlendirme Kriterleri

Model tahmin, bu çalışma kapsamında rekolte kestirimi, performansını değerlendirmek amacıyla, ortalama mutlak hata (Mean Absolute Error - MAE), kök ortalama karesel hata (Root Mean Square Error - RMSE) ve belirleme katsayısı ( $R^2$ ) ölçütleri kullanılmıştır. Bu ölçütlerin denklemleri sırasıyla (5), (6) ve (7)'de verilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (7)$$

Bu denklemlerde  $y_i$  gerçek pamuk verimini,  $\hat{y}_i$  modelin tahmin ettiği verimi,  $\bar{y}$  tarlaların gerçek pamuk veriminin ortalamasını ve  $n$  toplam veri sayısını ifade etmektedir.

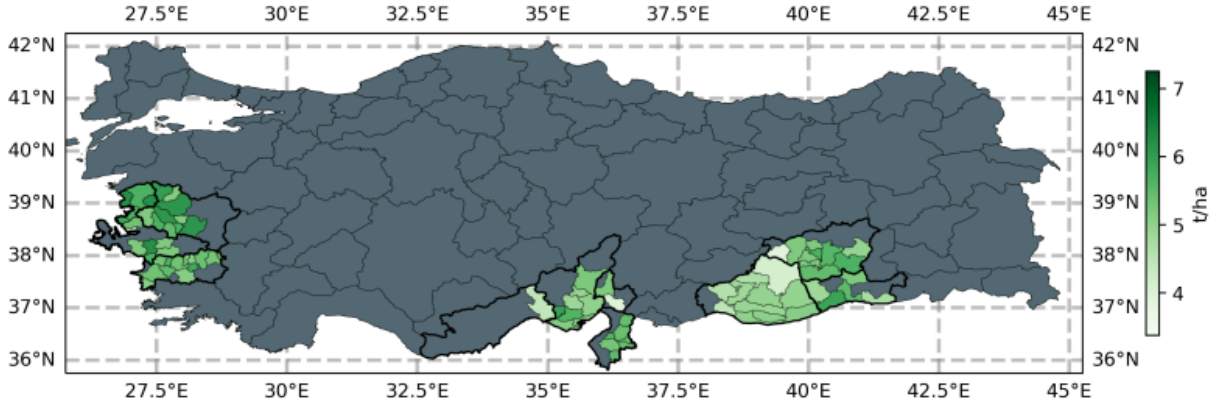
## 3. DENEYSEL SONUÇLAR

### 3.1 Veri Kümesi

Bu çalışmada kullanılan veri kümesi, Isik vd. (2023) yazarları tarafından üretilmiş olup, Türkiye'nin pamuk üretimi yapılan üç bölgesi, Ege, Akdeniz ve Güneydoğu Anadolu bölgelerinden elde edilen uzaktan algılama verilerini içermektedir. Veri kümesinde, Copernicus İklim Değişikliği Servisi ERA5'ten alınan saatlik iklimsel değişkenler bulunmaktadır. Bu değişkenler, yağış (precipitation - P), çiy noktası sıcaklığı (dew temperature - DT), maksimum,

ortalama ve minimum hava sıcaklıkları (Tmax, Tmean, Tmin), toprak sıcaklığı (soil temperature - ST), evapotranspirasyon (ET), yüzey güneş radyasyonu (surface solar radiation - SSRD) ve toprak nemidir (soil moisture - SM). Veri kümesinde yer alan toprak parametreleri ise SoilGrids'ten temin edilmiş olup, bunlar toplam yoğunluk (bulk density - BD), kum, silt, kil, katyon değişim kapasitesi (cation exchange capacity - CEC), pH, azot (nitrogen - N), toprak organik karbon (soil organic carbon - SOC), iri parçacık (coarse fragment - COF), organik karbon yoğunluğu (organic carbon density - OCD) ve organik karbon stokudur (organic carbon stock - OCS). Veri kümesi ayrıca, Landsat-8 multispektral uydu verilerinden elde edilen geliştirilmiş bitki örtüsü indeksi (enhanced vegetation index - EVI) ve Sentinel-1 SAR uydu verilerinden elde edilen VV ve VH polarize ölçümlerini içermektedir. Tarımsal bölgeler için pamuk maskeleri, Sentinel-2 multispektral verileri ve Sentinel-1 SAR verileri bir arada kullanılarak elde edilmiştir (Isik vd., 2023).

Bu çalışma kapsamında yalnızca 2023 yılına ait veriler kullanılmıştır. Türk İstatistik Kurumu (TÜİK) tarafından sağlanan yerleşim birimi düzeyindeki pamuk rekolte değerleri, LSTM modellerinde yaşanabilecek eksik uyum (underfit) problemini aşmak amacıyla tarla seviyesine iz düşürülerek kullanılmıştır. Bu çalışmada, Şekil 1'de görülen üç farklı bölgeden elde edilen veri kümeleri üzerinde ayrı LSTM modelleri eğitilmiş ve bölge tabanlı pamuk rekolte kestirimi gerçekleştirilmiştir.



Şekil 1. Çalışmada pamuk rekolte kestirimi için veri toplanan bölgeler (Celik vd., 2023b).

### 3.2 Model Eğitimi ve Başarımlar

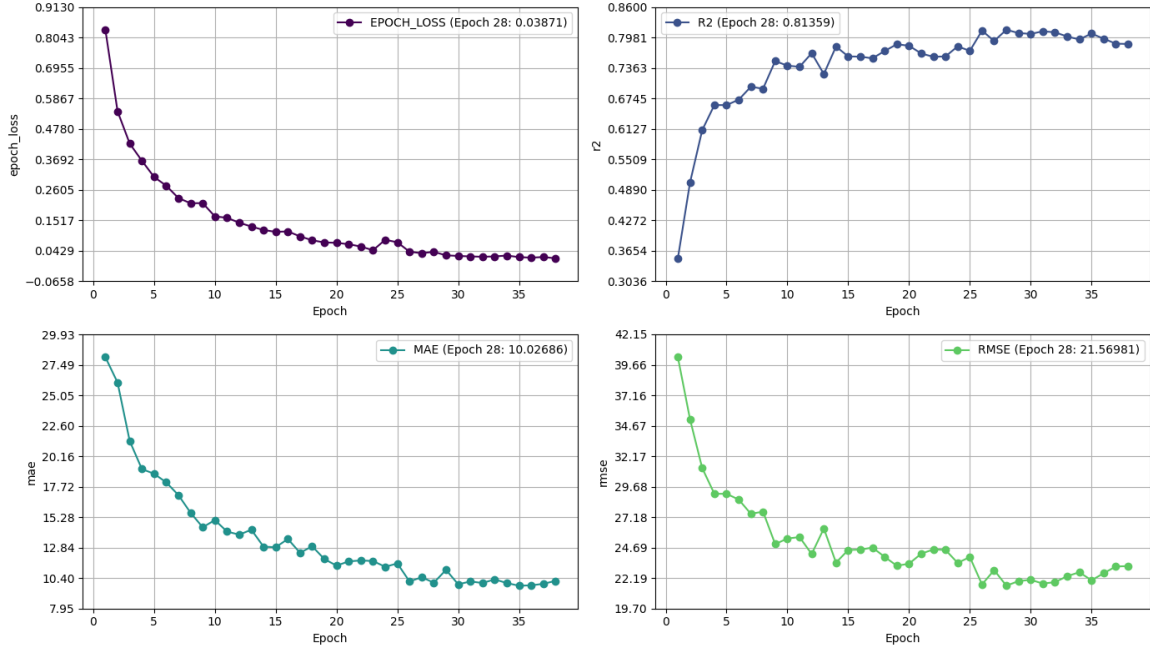
Zaman serisi analizi ile rekolte kestirimi için katman sayısı 2 ve gizli boyutu 64 olan LSTM ağları oluşturulmuştur. Regresyon için 16 katmanlı basit bir çok katmanlı algılayıcı (Multi-Layer Perceptron - MLP) yapısı kullanılmıştır. Model eğitimi, her bir bölge için, veri kümesinin %80'lik kısmı eğitim ve %20'lik kısmı test için ayrılarak başlatılmıştır. Her eğitimde bölümlenmenin tutarlılığını sağlamak için "random\_state" argümanı ile rastgele bölümlenme seed'i 12 olarak belirlenmiştir. Eğitim sürecinde, 0.001 öğrenme oranı, 'Adam' optimizasyon algoritması ve minimum karesel hata kaybı (Minimum Square Error Loss - MSELoss) kayıp fonksiyonu kullanılmıştır. Model, toplam 100 epoch boyunca eğitilmiş ve her epoch sonunda test kümesi üzerinde RMSE, MAE, R<sup>2</sup> ölçütleri ile başarımları değerlendirilmiştir. Aşırı öğrenmeyi (overfit) engellemek için erken durdurma (Early Stop) kullanılmış ve durma adımı olarak ise 10 epoch belirlenmiştir. Batch boyutu 32 olarak belirlenmiştir.

Ege Bölgesi için eğitilen model, erken durdurma yöntemiyle en iyi performansı 28. epoch'ta elde etmiş olup, eğitim süreci 38. epoch'ta sonlanmıştır. Eğitim sürecinde modelin kayıp değeri (epoch\_loss) hızla azalmış ve 0.03871 seviyesine ulaşmıştır. Modelin belirleme katsayısı (R<sup>2</sup>) 0.81359 olarak hesaplanmış olup, bu değer modelin pamuk rekoltesi tahmininde oldukça başarılı olduğunu göstermektedir. MAE değeri 10.02686 kg/da, RMSE ise 21.56981 kg/da olarak elde edilmiştir. Genel olarak bu sonuçlar, modelin Ege Bölgesi'ndeki pamuk rekoltesi tahminlerinde yüksek bir performans sergilediğini ve tahminlerinin isabetli olduğunu göstermektedir. Ege Bölgesi için başarımları ölçütü grafikleri Şekil 2'de sunulmuştur.

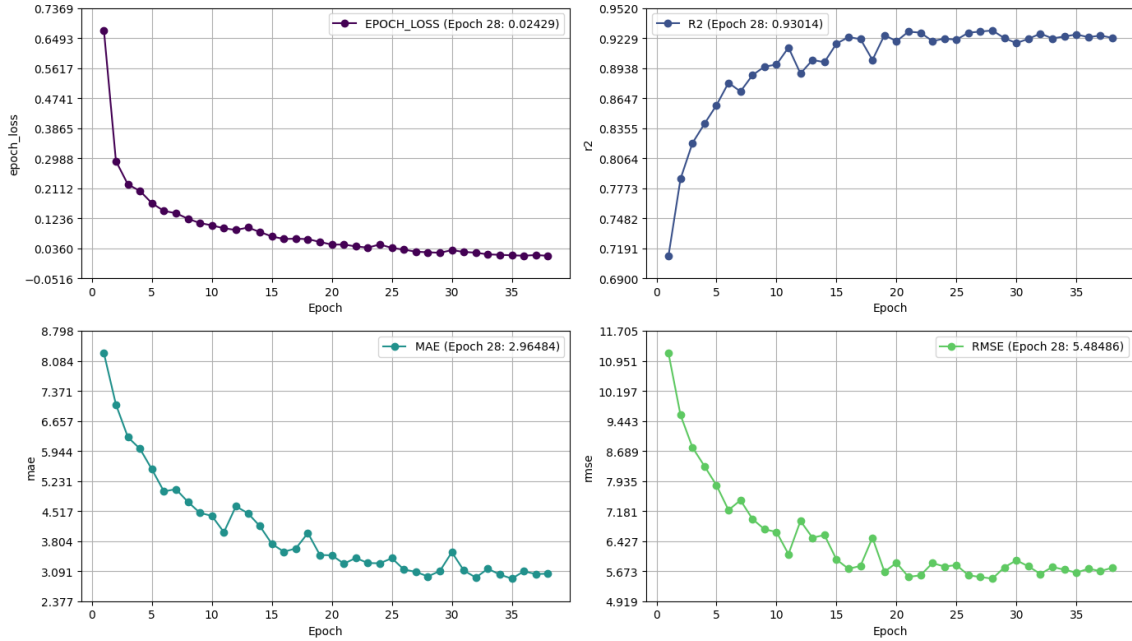
Güneydoğu Anadolu Bölgesi için eğitilen model, erken durdurma yöntemiyle en iyi performansı 28. epoch'ta elde etmiş olup, eğitim süreci 38. epoch'ta sonlanmıştır. Eğitim sürecinde modelin kayıp değeri hızla azalmış ve 0.02429 seviyesine ulaşmıştır. Modelin belirleme katsayısı (R<sup>2</sup>) 0.93014 olarak hesaplanmış olup, modelin pamuk rekoltesi tahmininde başarılı sonuçlar verdiğini göstermektedir. MAE değeri 2.96484 kg/da, RMSE ise 5.48486 kg/da olarak elde edilmiştir. Bu sonuçlar, modelin Güneydoğu Anadolu Bölgesi'nde pamuk rekoltesi tahminlerinde yüksek bir doğruluk sağladığını ve düşük hata oranlarıyla başarılı olduğunu göstermektedir. Güneydoğu Anadolu Bölgesi için

başarımı ölçütü grafikleri Şekil 3'te sunulmuştur.

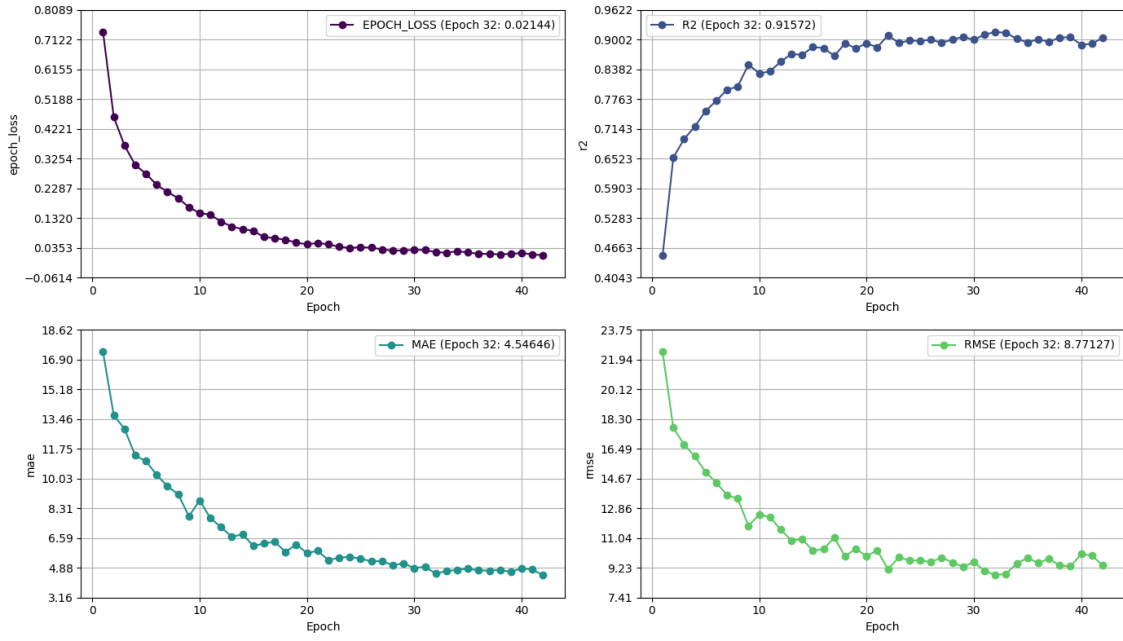
Akdeniz Bölgesi için eğitilen model, erken durdurma yöntemiyle en iyi performansı 32. epoch'ta elde etmiş olup, eğitim süreci 42. epoch'ta sonlanmıştır. Eğitim sürecinde modelin kayıp değeri hızla azalmış ve 0.02144 seviyesine ulaşmıştır. Modelin belirleme katsayısı ( $R^2$ ) 0.91572 olarak hesaplanmış olup, modelin pamuk rekoltesi tahmininde başarılı sonuçlar verdiğini göstermektedir. MAE değeri 4.54646 kg/da, RMSE ise 8.77127 kg/da olarak elde edilmiştir. Bu sonuçlar, modelin Akdeniz Bölgesi'nde pamuk rekoltesi tahminlerinde iyi bir performans sağladığını göstermektedir. Akdeniz Bölgesi için başarımlı ölçütü grafikleri Şekil 4'te sunulmuştur.



Şekil 2. Ege Bölgesi için başarımlı ölçütü grafikleri.



Şekil 3. Güneydoğu Anadolu Bölgesi için başarımlı ölçütü grafikleri.



Şekil 4. Akdeniz Bölgesi için başarımlar ölçütü grafikleri.

### 3.3 SHAP Analizleri

LSTM modelleri ile bölgesel pamuk rekoltesi kestirimleri gerçekleştirildikten sonra, rekolte kestirim modellerine yorumlanabilirlik kazandırmak amacıyla SHAP yöntemi ile analizler gerçekleştirilmiştir. Bölgesel test kümeleri üzerinde gerçekleştirilen SHAP analizlerinde, her bölge için farklı girdi özniteliklerinin pamuk rekoltesi tahminine olan etkisi nicel olarak ve zamana göre elde edilmiştir.

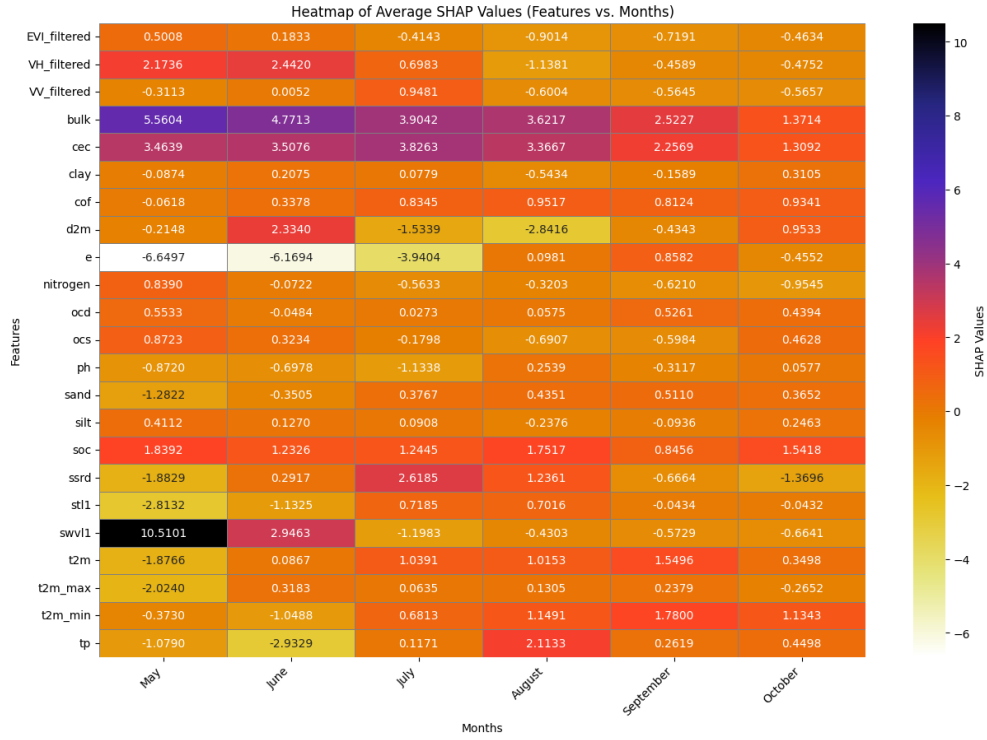
Ege Bölgesi test verisinde yapılan SHAP analizi, pamuk rekoltesi tahmininde öne çıkan özniteliklerin toprak nemi (swv11), kation değişim kapasitesi (CEC) ve toprak yoğunluğu (BD) olduğunu göstermiştir. Bu özniteliklerin baskın katkısı, toprak ve su yönetimi stratejilerinin pamuk veriminde kritik önemini göstermektedir. Şekil 5'te Ege Bölgesi için SHAP analizi sonuçları görülmektedir.

Güneydoğu Anadolu Bölgesi test verisinde yapılan SHAP analizi, pamuk rekoltesi tahmininde öne çıkan özniteliklerin minimum 2 metre sıcaklık (t2m\_min), toprak dokusu (kum) ve toprak nemi (swv11) olduğunu göstermiştir. Bu özniteliklerin sulama yönetimi ve su koruma uygulamalarının bölgedeki pamuk verimliliği üzerindeki kritik önemini vurguladığı görülmektedir. Şekil 6'da Güneydoğu Anadolu bölgesi için SHAP analizi sonuçları görülmektedir.

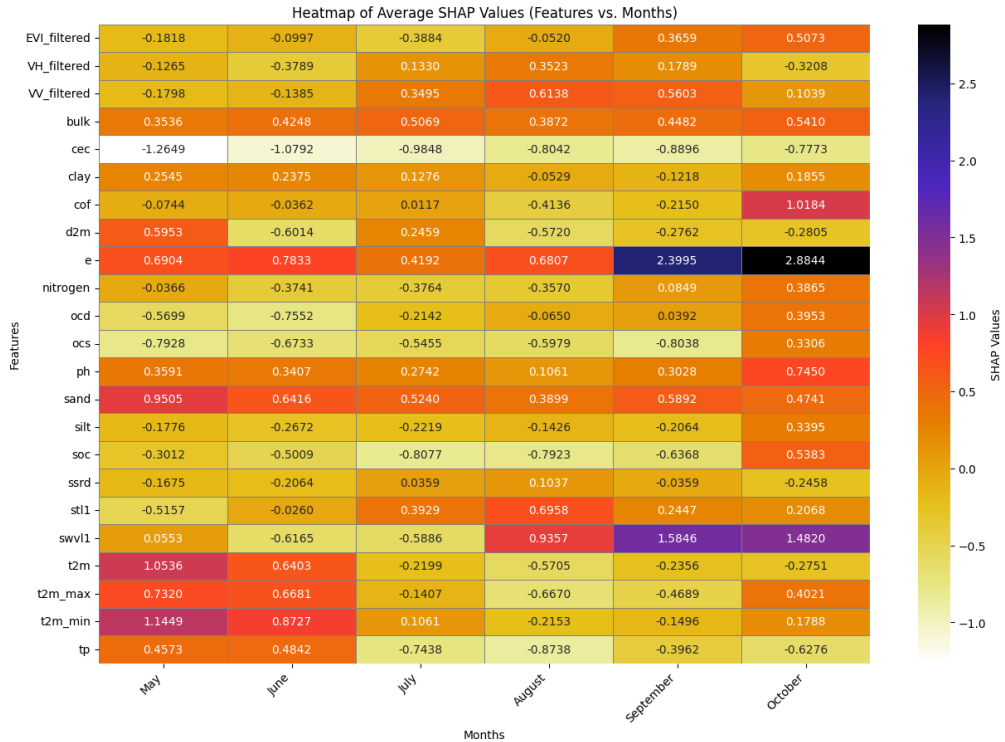
Akdeniz Bölgesi test verisinde yapılan SHAP analizi, pamuk rekoltesi tahmininde öne çıkan özniteliklerin organik karbon yoğunluğu (OCD), pH seviyeleri, toprak yoğunluğu (BD) ve toplam yağış (total precipitation - TP) olduğunu göstermiştir. Bu özniteliklerin baskın katkısı, besin maddesi bulunabilirliği ve yüzey neminin korunmasının pamuk verimliliği üzerinde önemli bir rol oynadığını ortaya koymaktadır. Şekil 7'de Akdeniz Bölgesi için SHAP analizi sonuçları görülmektedir.

Sonuçlardan görüldüğü üzere, Türkiye'nin Ege, Akdeniz ve Güneydoğu Anadolu bölgelerinden toplanan bölgesel veri kümeleri kullanılarak eğitilen LSTM modelleri, SHAP yöntemi ile detaylı bir şekilde incelenmiştir. Yapılan analizler, her bir bölgenin kendine özgü iklimsel ve toprak özelliklerinin pamuk rekoltesi üzerindeki etkilerini ortaya koymuştur. Bölgesel farklılıkların belirgin olduğu bu çalışmada, Ege Bölgesi'nde toprak nemi ve kation değişim kapasitesi öne çıkarken, Akdeniz Bölgesi'nde organik karbon yoğunluğu ve pH seviyeleri gibi toprak özellikleri pamuk verimini doğrudan etkilemiştir. Güneydoğu Anadolu Bölgesi'nde ise minimum sıcaklık ve toprak dokusu gibi özellikler, sulama ve su yönetiminin önemini vurgulamaktadır. Sonuçlar, bölgesel olarak farklılaşan özelliklerin pamuk rekoltesi kestiriminde önemli bir rol oynadığını ve bu nedenle yerleştirilmiş modellerin kullanılması gerektiğini göstermektedir.

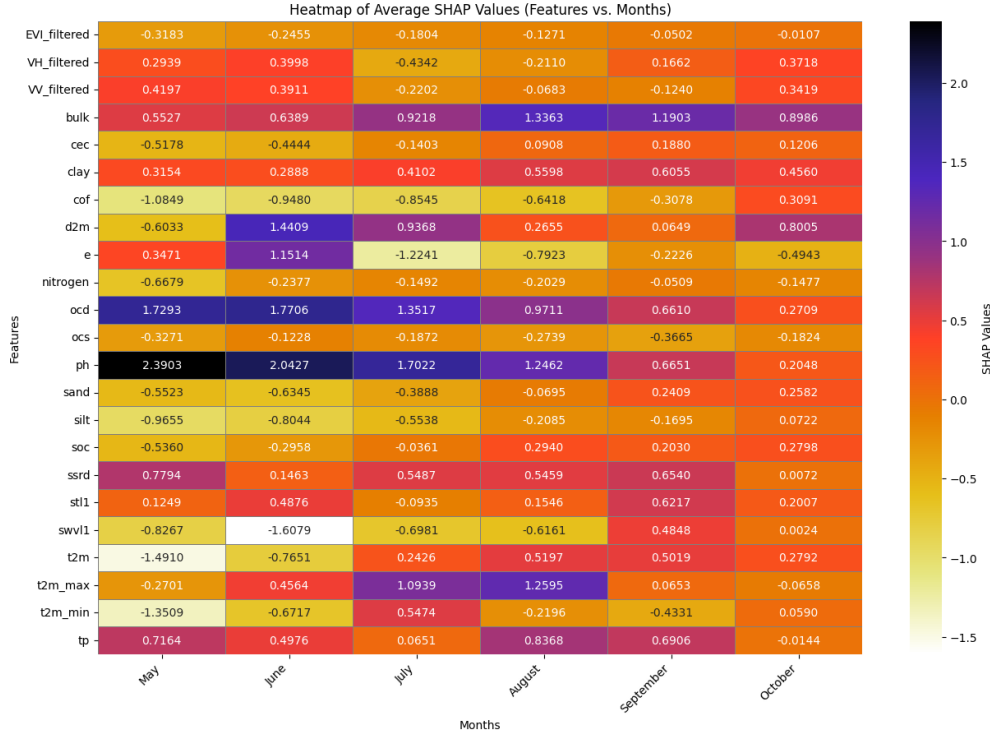
## F. Yardımcı vd.: Türkiye'de Pamuk Rekolte Kestirimi İçin Bölgesel Yorumlanabilir LSTM Modelleri



Şekil 5. Ege Bölgesi SHAP analizi sonuçları



Şekil 6. Güneydoğu Anadolu Bölgesi SHAP analizi sonuçları



Şekil 7. Akdeniz Bölgesi SHAP analizi sonuçları

### 3. SONUÇLAR

Bu çalışmada, Türkiye'nin farklı tarımsal bölgelerinde pamuk rekoltesi tahmini için bölge bazında LSTM modelleri kullanılarak bir analiz gerçekleştirilmiş ve bu modellerin yorumlanabilirliğini sağlamak için SHAP yöntemi uygulanmıştır. Ege, Akdeniz ve Güneydoğu Anadolu bölgelerine özgü verilerle eğitilen LSTM modelleri, her bölge için veri kümesinde yer alan ölçümleri ve uzaktan algılama verilerini kullanarak başarılı rekolte kestirimi sonuçları vermiştir. Her bölge için yapılan ayrı SHAP analizleri, tarımsal verim tahmininde bölgeye özgü stratejilerin geliştirilmesi gerektiğine ışık tutmaktadır. Ege Bölgesi'nde toprak nemi, katyon değişim kapasitesi ve toprak yoğunluğu en etkili öznitelikler olarak öne çıkarken, Akdeniz Bölgesi'nde organik karbon yoğunluğu, pH seviyeleri ve toplam yağış gibi faktörler ön plana çıkmıştır. Güneydoğu Anadolu Bölgesi'nde ise minimum 2 metre sıcaklık (t2m\_min), toprak dokusu ve toprak nemi özniteliklerinin rekolte kestiriminde kritik önemde olduğu görülmüştür. Bu bulgular, her bölgenin kendine özgü çevresel dinamiklerinin pamuk rekoltesi tahminini önemli ölçüde etkilediğini göstermektedir. Çalışmamız, bölgesel farklılıkların göz önünde bulundurulmasının pamuk rekoltesi tahmininde daha doğru sonuçlar elde edilmesi için gerekli olduğunu vurgulamaktadır. AYZ yöntemlerinin kullanımı, kullanılan yapay zekâ modellerinin çalışma mekanizmalarının ve girdi-çıkış ilişkilerinin yorumlanabilirliğini artırarak, tarımsal karar süreçlerinde daha bilinçli ve etkili stratejiler geliştirilmesine katkı sağlama potansiyeline sahiptir. Gelecekteki çalışmalar, daha fazla bölgeyi ve zamanı kapsayacak şekilde ve birden fazla derin öğrenme modeline genişletilerek, çevresel ve iklimsel özniteliklerin etkilerinin daha hassas ve daha gürbüz bir şekilde ölçülmesini ve rekolte kestirimleri için bölgeye özgü yüksek başarılı ve yorumlanabilir çözümler üretilmesini sağlayabilir.

### 4. TEŞEKKÜR

Bu çalışma, Climate Change AI ve Canada Hub of Future Earth tarafından desteklenmiştir.

### KAYNAKLAR

**Bhagat, P. R., Naz, F., Magda, R., 2022.** Artificial Intelligence Solutions Enabling Sustainable Agriculture: A Bibliometric Analysis. *PloS one*, 17(6), e0268989.

**Celik, M. F., Isik, M. S., Taskin, G., Erten, E., Camps-Valls, G., 2023a.** Explainable Artificial Intelligence for Cotton Yield Prediction with Multisource Data. *IEEE Geoscience and Remote Sensing Letters*, 20, 8500905.

- 
- Celik, M. F., Isik, M. S., Erten, E. and Taskin, G.,** 2023b. Informative Earth Observation Variables for Cotton Yield Prediction Using Explainable Boosting Machine," *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 3542-3545.
- Fomin, V. V., Astromskis, P.,** 2022. The Black Box Problem. *Future Law, Ethics, and Smart Technologies: The Future of Legal Education*, 393, 112.
- Hochreiter, S.,** 1997. Long Short-term Memory. *Neural Computation MIT-Press*.
- Isik, M. S., Celik, M. F., Erten, E.,** 2023. Interpretable Cotton Yield Prediction Model Using Earth Observation Time Series. *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 3442-3445.
- Jones, E. J., Bishop, T. F., Malone, B. P., Hulme, P. J., Whelan, B. M., Filippi, P.,** 2022. Identifying causes of crop yield variability with interpretive machine learning. *Computers and Electronics in Agriculture*, 192, 106632.
- Joshi, A., Pradhan, B., Gite, S., & Chakraborty, S.,** 2023. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sensing*, 15(8).
- Linheiro, E. S. M., Shinde, G. R., Mahalle, P. N., Mirajkar, R.,** 2023. Explainable AI (XAI) for Agriculture. *Industry 4.0 Convergence with AI, IoT, Big Data and Cloud Computing: Fundamentals, Challenges and Applications*, 161.
- Lundberg, S.,** 2017. A Unified Approach to Interpreting Model Predictions. *arXiv preprint arXiv:1705.07874*.
- Martínez-Ferrer, L., Piles, M., Camps-Valls, G.,** 2020. Crop Yield Estimation and Interpretability with Gaussian Processes. *IEEE Geoscience and Remote Sensing Letters*, 18(12), 2043-2047.
- Nyéki, A., & Neményi, M.,** 2022. Crop Yield Prediction in Precision Agriculture. *Agronomy*, 12(10), 2460.
- Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B., Kempen, B., Ribeiro, E., Rossiter, D.,** 2021. SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil*, 7(1), 217-240.
- Shapley, L. S.,** 1953. A Value for n-Person Games. *Contribution to the Theory of Games*, 2.
- Taskin, G., Aptoula, E., Ertürk, A.,** 2024. Explainable AI for Earth Observation: Current Methods, Open Challenges, and Opportunities. *Advances in Machine Learning and Image Analysis for GeoAI*, 115-152.
- Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., Li, H.,** 2021. An LSTM Neural Network for Improving Wheat Yield Estimates by Integrating Remote Sensing Data and Meteorological Data in the Guanzhong Plain, PR China. *Agricultural and Forest Meteorology*, 310, 108629.
- Timmer, C. P.,** 2002. Agriculture and Economic Development. *Handbook of Agricultural Economics*, 2, 1487-1546.
- Van Houdt, G., Mosquera, C., Nápoles, G.,** 2020. A Review on the Long Short-Term Memory Model. *Artificial Intelligence Review*, 53(8), 5929-5955.
- Zhou, W., Yan, Z., Zhang, L.,** 2024. A Comparative Study of 11 Non-Linear Regression Models Highlighting Autoencoder, DBN, and SVR, Enhanced by SHAP Importance Analysis in Soybean Branching Prediction. *Scientific Reports*, 14(1), 5905.