Hiperspektral Sınıflandırma için SHAP ile Bant Tabanlı Yorumlanabilirlik
Band-Based Interpretability with SHAP for Hyperspectral Classification

İsmail Şahin*, Alp Ertürk*, Erchan Aptoula†
*Elektronik ve Haberleşme Müh., †Mühendislik ve Doğa Bilimleri Fakültesi (VPALab)
*Kocaeli Üniversitesi, †Sabancı Üniversitesi
*Kocaeli, Türkiye, †İstanbul, Türkiye
180208025@kocaeli.edu.tr, alp.erturk@kocaeli.edu.tr, erchan.aptoula@sabanciuniv.edu

***Özetçe* —Hiperspektral veriler, barındırdıkları yüksek miktardaki spektral bilgi ile sınıflandırmada yüksek başarım imkanı sağlarlar. Bununla birlikte, makine öğrenmesi yaklaşımlarının bütünüyle kara kutu veya en azından ara aşamaları kapalı olabilen yapıları sınıflandırma süreçlerini genellikle yorumlanması zor ve açıklanabilirliği düşük kılmaktadır. Bu çalışma kapsamında, bir açıklanabilir yapay zeka (AYZ) yöntemi olan SHAP, hiperspektral sınıflandırma sürecinin yorumlanabilirliği için kullanılmaktadır. Önerilen yaklaşım ile, eğitilen sınıflandırma modeli üzerinden, her bir sınıf için, bant-tabanlı SHAP değerleri elde edilerek, spektral bantların sınıf-tabanlı olarak sınıflandırma sürecindeki katkılarına dair nicel bir değerlendirme elde edilmektedir. Ön deneysel sonuçlar bu bildiride sunulmuştur ve yaklaşımın potansiyeline ışık tutmaktadır.**

Anahtar Kelimeler—açıklanabilir yapay zeka (AYZ), hiperspektral, SHAP, sınıflandırma, yorumlanabilirlik.

***Abstract*—Hyperspectral data enable high classification accuracies due to the large amount of spectral information they contain. However, the complete black box or at least partially opaque nature of machine learning approaches often make classification processes challenging to interpret and low in explainability. In this work, SHAP, which is an explainable artificial intelligence (XAI) method, is used for the interpretability of hyperspectral classification. For each class, band-based SHAP values are obtained over the trained classification model, providing a quantitative evaluation of the contribution of the spectral bands to the classification process. Preliminary experimental results are provided in this paper, and shed light on the proposed method's potential.**

Keywords—classification, explainable artificial intelligence (XAI), hyperspectal, interpretability, SHAP.

I. GİRİŞ

Yapay sinir ağlarının yaklaşık 80 yıl önce ilk ortaya çıkışından bu yana, makine öğrenimi alanında ciddi ilerlemeler kaydedilmiştir. Etkili öğrenme algoritmaları ve donanımsal hızlandırma desteği ile, ilk zamanlarda dillendirilen sinir ağlarına yönelik birçok eleştiri kaynağı ortadan kalkmıştır. Ancak biri hala kendinden söz ettirmektedir: açıklanabilirlik ve yorumlanabilirlik. Türünden bağımsız olarak, sığ veyan derin fark etmeksizin (örneğin evrişimsel sinir ağı, otokodlayıcı, vb.), tüm yapay sinir ağları birer "kara kutu" olarak betimlenmektedir ve gerçekleştirdikleri tahminlerin neden-sonuç bağıntılarının tespiti genelde güçtür. Açıklanabilir Yapay Zeka (AYZ) (*Explainable Artificial Intelligence* - XAI) bu gereksinime istinaden ortaya çıkmıştır ve çeşitli araçlar ile bir karar, tahmin veya kestirimin arkasında yer alan nedenleri ve işleyişi sergilemeyi ve/veya anlaşılabilir kılmayı hedefler.

Son yıllarda AYZ'ye yönelik ilgi katlanarak artmıştır. Haliyle bu durum uzaktan algılama alanına da yansımış olup, yayımlanmış çalışmaların sayısı hızla artmaktadır. Uzaktan algılama bağlamındaki erken bir olgu-sonrası (*post hoc*) yorumlanabilirlik girişimine [1] çalışmasında rastlanmaktadır. Tarım ürünlerinin fenolojik modeller aracılığı ile sınıflandırılmasının yer aldığı bu örnekte, Sentinel-2 görüntüleri kullanılmış ve yapay sinir ağının etkinleştirmeleri bulutlu ve bulutsuz havalarda görselleştirilmiştir. Böylece ağın bulut süzgeçleme yetkinliği hakkında çıkarımlar elde edilmiştir. Bir diğer önemli çalışmada ise [2], yine Sentinel-2 görüntüleri ile yer kullanım haritası üretimi amacıyla çift-yönlü uzun kısa vadeli bellek ağları kullanılmıştır. Olgu sonrası çözümleme yaklaşımını benimseyen bu yayında, İspanya üzerindeki zaman serileri işlenmiş, ve ağ etkinleştirmeleri üzerinden hangi dalga boylarının uygulama için daha önemli ve yararlı olduğu araştırılmıştır. AYZ amacıyla evrişimsel sinir ağları ile birlikte kullanılan bir yorumlanabilirlik yaklaşımı olarak Sınıf Etkinleştirme Haritalama (*Class Activation Mapping* - CAM) özellikle öne çıkmaktadır [3]. Sınıflandırma esnasında sınıf başına önem haritalarının hesaplanması ile örneklerin uzamsal olarak nerelerinin katkı sağladığını belirlemekte yararlı olmuştur. Bir diğer araştırma istikametinde ise, modelin yerine tekil tahminlerin yorumlanabilirliğine odaklanılmaktadır. Bu bağlamda Shapley eklemeli açıklamalar (*SHapley Additive exPlanations* - SHAP) [4] her girdi öznitelik yöney boyutunun nihai sonuca olan katkısını hesaplamaya odaklıdır ve yaygın olarak kullanım bulmaktadır. Örneğin [5]'de, çok katmanlı bir yapay sinir ağının, 2018'de Endonezya'da yer alan deprem sonrası, WorldView-3 görüntülerindeki deprem kaynaklı bina hasarlarının tahminlerini yorumlamak için kullanılmıştır. SHAP'ın bir diğer kullanım örneğinde ise [6], optik görüntülerden kentsel bitki örtüsünün derin sinir ağları ile haritalandırılmasında renk uzayı ve doku öznitelikleri konu alınmıştır. SHAP yakın tarihte

979-8-3503-4355-7/23/$31.00 ©2023 IEEE

hiperspektral görüntü işlemede de kullanım bulmuştur, [7]'de tayfsal ışınımölçer ölçümleri ile farklı büyüme aşamalarındaki buğdayda azot içeriği kestiriminde dalgaboylarının öneminin tespiti amacıyla kullanılmıştır, [8]'de ise Çinghay-Tibet platosu üzerinden alınan hiperspektral ölçümlerden yer üstü biyokütle evirme uygulaması için mevsim, çimen uzunluğu ve spektral bantların öneminin tespit edilmesi amacıyla kullanılmıştır . [9]'de SHAP, hiperspektral ve LiDAR verilerden uzun kısa-soluklu bellek (*Long Short-Term Memory* - LSTM) ile mısır rekolte kestirimi için uzaktan algılama indekslerinin önemini ölçmek amacıyla kullanılmıştır. [10]'de ise Kırşehir üzerinden alınan 20 spektral bantlı Jilin1-GP01 verisi üzerinde elle etiketleme sonrası spektral bantların sınıflandırma başarımına etkisi SHAP ile yorumlanmıştır.

Bu çalışma kapsamında, SHAP yöntemi, hiperspektral verilerin sınıflandırılmasında bant-tabanlı yorumlanabilirlik amacıyla kullanılmaktadır. SHAP ile, spektral bantların hiperspektral veride sınıflandırma kararına katkısı sınıf tabanlı olarak elde edilmekte ve bu vesileyle, her bir sınıf için hangi spektal bantların sınıflandırma sürecinde daha yüksek önem taşıdığı konusunda nicel bir değerlendirme elde edilmektedir. SHAP ve benzeri AYZ yaklaşımlarının bant tabanlı yorumlanabilirlik için kullanılması, geleneksel bant seçimi yaklaşımları yerine, sınıflandırma modelinin de sürece dahil edilmesini sağlamakta ve bu sayede uçtan uca (*end-to-end*) bir yaklaşım sunmaktadır.

II. YÖNTEM

A. SHAP

SHAP [4], oyun teorisi bağlamında 1953'de ortaya çıkan ve bir oyundaki tüm oyuncuların adil bir şekilde oyun sonrası ödüllerini alması ile ilgili olan Shapley değerlerine dayanır. Makine öğrenimi bağlamında ise, oyun, model tarafından gerçekleştirilen tahmin/kestirim, oyuncular, ona katkı sağlayan girdi öznitelikleri, ödül ise her özniteliğin tahmine/kestirime olan katkı düzeyidir.

SHAP toplamsal bir yöntem olup, karmaşık bir makine öğrenimi modeli $g(x)$ için, açıklanabilir yaklaşık bir $f(x)$ modeli sunmayı amaçlar:

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

Bu denklemde, ϕ_0 önyargı terimidir ve makine öğrenimi modelinin tüm olası öznitelik konfigürasyonları arasındaki ortalama tahminini ifade eder; ϕ_i ise, i. özniteliğin Shapley değeridir. Denklem 1'e göre, SHAP modelin tahminini tüm özniteliklerden gelen katkıların toplamı olarak ifade eder. SHAP her özniteliğe, aralarındaki etkileşimleri hesaba katacak şekilde eşsiz bir önem yöneltir. Shapley değeri, her özniteliğin model sonucuna hem kendi tekil etkisini, hem de diğer öznitelikleri ile birlikte ortaya çıkan etkisini hesaba katar. Shapley değerleri özniteliğin sonuca olan katkısına bağlı olarak pozitif veya negatif olabilir. Pozitif Shapley değerleri, söz konusu özniteliğin tahmin değerini artırdığını, negatif Shapley değerleri ise onu azalttığını ifade eder. Shapley değerinin mutlak değeri ise, ait olduğu özniteliğin sonucu hangi düzeyde etkilediği anlamını taşır. Ayrıca, farklı özniteliklere ait değerler karşılaştırılarak hangi özniteliğin model çıktısına etkisinin daha fazla olduğu konusunda bilgi edinilir.

B. Rastgele orman

Karar ağaçları, belirli kararlar yönelten olay serilerinden oluşur. Bu ağaçların her düğümü bir özniteliğin sınamasından ibarettir ve her dalı o sınamanın bir sonucunu ifade eder. Gözetimli eğitim esnasında, etiketli veri ile eğitilen karar ağacının, asgari sayıda soru ile başarılı sonuçlar vermesi talep edilir. Bunu başarmak için de enformasyon kazancı ilkesine göre özniteliklerin sınanma sırası belirlenir. Rastgele orman (*Random Forest* - RF) [11] ise, tek bir karar ağacı yerine, birbirinden farklı çok sayıda karar ağacı üretir ve tahminlerini kaynaştırır. Tek veri kümesinden birden fazla karar ağacı elde etmek için iki yöntem kullanılır. İlki, torbalamadır (*bagging*), ve eğitim kümesinden rastgele alt-kümelerin seçilip, her biri ile ayrı bir karar ağacının oluşturulmasından ibarettir. Diğeri ise öznitelik torbalama (*feature bagging*) olarak bilinir, ve onda ise, her karar ağacının özniteliklerin rastgele bir alt-kümesi ile eğitilmesi öngörülür. Bu işlemler sonucunda ortaya çıkan karar ağaçlar topluluğu veya ormanının tahminler kümesi çeşitli yöntemler ile kaynaştırılabilir. Örneğin en yaygın olanı çoğunluğunun tahmininin seçilmesidir.

C. Önerilen yaklaşım

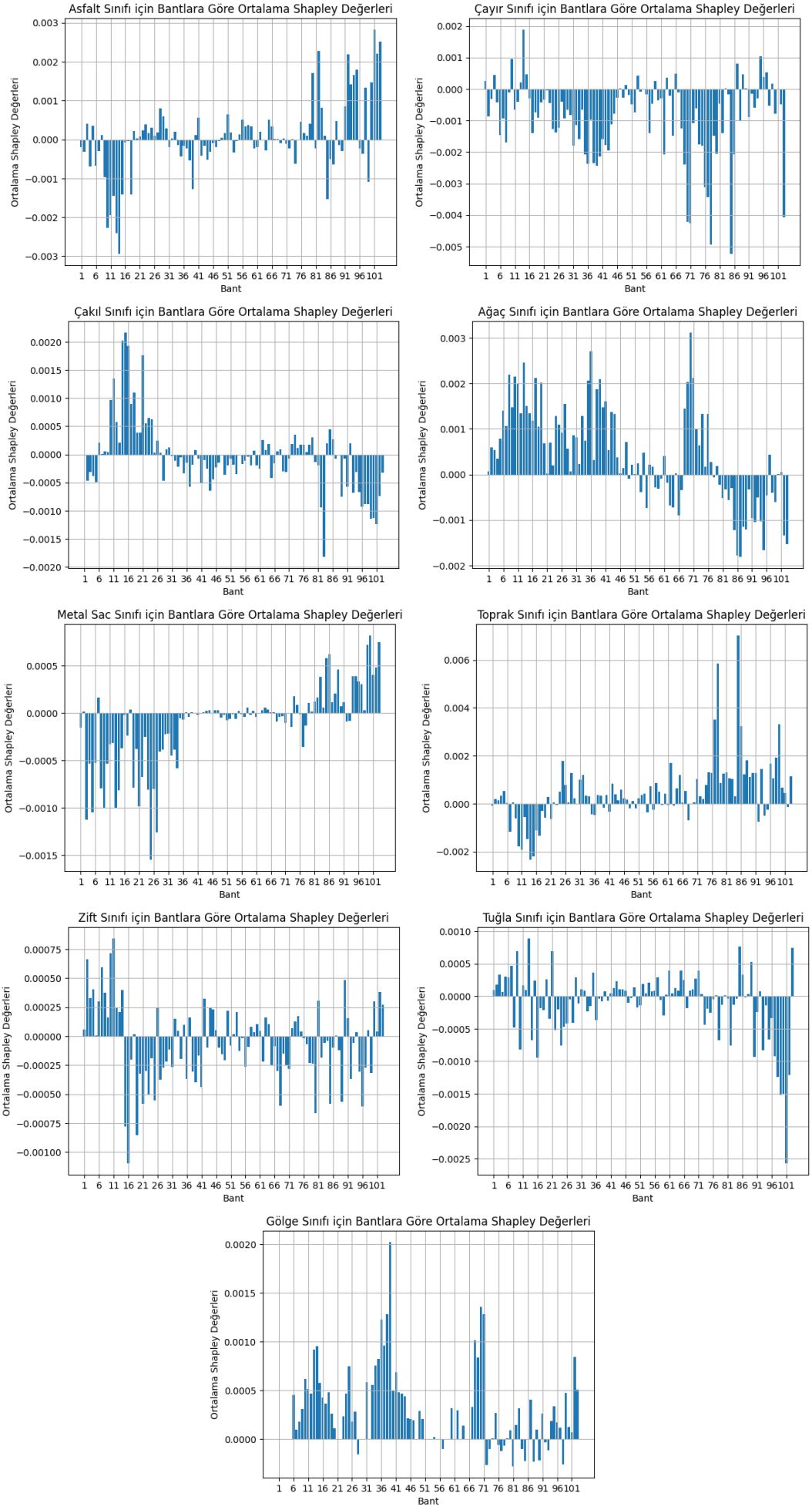
Bu çalışmada hiperspektral verilerin sınıflandırılmasında spektral bant-tabanlı yorumlanabilirliğin elde edilmesi amaçlanmaktadır. Yorumlanabilirlik, özellikle kara kutu niteliğindeki derin ağlar için kritik önemde olmakla birlikte, her makine öğrenme yaklaşımı için yaklaşımın çalışma sürecinin ve sonuçlarının anlamlandırılması ve açıklanabilmesi açısından önem taşımaktadır.

Bu çalışma kapsamında sınıflandırma yöntemi olarak yaygın kullanılan bir sığ yaklaşım olan rastgele orman yöntemi ve yorumlanabilirlik için SHAP yönteminin karar ağaçları ve rastgele orman gibi topluluk yöntemleri için geliştirilmiş versiyonu olan TreeSHAP [12] kullanılmaktadır. Eğitilen sınıflandırma modeli TreeSHAP yöntemine girdi olarak verilmekte ve bant tabanlı SHAP değerleri elde edilmektedir. Bu sayede, verinin spektral bantlarının sınıflandırılma sürecine katkısı konusunda sayısal değerler ile ifade edilen nicel bir değerlendirme elde edilmektedir. Bant tabanlı bu değerler spektral bantlar ve sınıflandırma yöntemi arasındaki ilişkinin ve sınıflandırma sürecinin daha yorumlanabilir / açıklanabilir olması amaçlarıyla kullanılabileceği gibi, yüksek SHAP değerleri veren spektral bantların korunması ve diğerlerinin elenmesi ile SHAP tabanlı bant seçimi ve boyut azaltımı amacıyla da kullanılabilir.

III. DENEYSEL SONUÇLAR

Bu çalışmada deneysel sonuçlar ROSIS alıcısı ile alınmış Pavia Üniversitesi hiperspektral verisi üzerinden elde edilmiştir. Veri 610 × 340 piksel boyutlarındadır ve uzamsal çözünürlüğü 1.3 metredir. Spektral aralığı 0.43 - 0.86 µm olan veri, 103 spektral banttan oluşmaktadır ve yer doğrusu 9 sınıftan oluşmaktadır.

Çalışma kapsamında Pavia Üniversitesi verisinde rastgele orman yöntemi %10 eğitim oranında eğitim sonrası sınıflandırma gerçekleştirilmiştir. Gini kriterli rastgele orman modelinin parametreleri, ızgara arama çapraz-doğrulaması (*grid search cross-validation*) ile belirlenmiştir. Buna göre ağaç sayısı 400, en fazla derinlik 40, düğüm ayrıştırmak için en küçük örnek sayısı 2, yaprak düğümü olmak için gereken en



Şekil 1: SHAP değerleri

küçük örnek sayısı 1 olarak belirlenmiştir. Bu parametreler ile eğitilen rastgele orman modeli üzerinden treeSHAP yöntemi ile SHAP değerleri elde edilmiştir. Sınıf tabanlı ortalama SHAP değerleri Şekil 1'de gösterilmektedir. Bu grafiklerde x ekseni spektral bantları, y ekseni ise o sınıf ile ilişkili piksel vektörleri üzerinden hesaplanan ortalama SHAP değerlerini göstermektedir. Beklendiği üzere, her sınıf için birbirinden oldukça farklı grafikler elde edilmiştir.

Şekil 1'deki pozitif SHAP değerleri, ilgili spektral bantların model tahminine pozitif katkı sağladığını ifade etmektedir. Başka bir deyişle, Şekil 1'de her bir sınıf için, kullanılan sınıflandırma yönteminin eğitilen modeli için SHAP yönteminin çalışma mantığına göre o sınıf için anlamlı, o sınıfı ayrıştırma yeteneği yüksek, spektral bantları ifade etmektedir. Şekil 2'de asfalt sınıfı için elde edilen SHAP değerleri normalize edildikten sonra iki farklı bant için görselleştirilmiştir. Şekil 1'de görüldüğü üzere, asfalt sınıfı için, 103. bant yüksek ortalama SHAP değeri elde edilen, 55. bant ise göreceli olarak düşük bir ortalama SHAP değeri elde edilen bant olmuştur. Şekil 2'de sunulmuş olan görsellerde asfalt sınıfının ayrıştırılabilirlik durumları bu sonuçları desteklemektedir. Şekil 3'te ise zift sınıfı için normalize edilmiş SHAP değerleri, 11. bant ve 103. bant için görselleştirilmiştir. Zift sınıfı, bu sınıf için ortalama SHAP değerleri daha yüksek olan 11. bant görselinde, 103. bant görseline göre daha ayrışır yapıda görünmektedir.

Önerilen yaklaşım sonucu elde edilen nicel değerler ile, kullanılan sınıflandırma modeline bağlı olarak, sınıf tabanlı bant seçimi ve boyut azaltımı yapılabileceği gibi, sınıflar ile ilgili o sınıflara özel yeni özniteliklerin elde edilmesine ışık tutulması imkanı da bulunmaktadır.



Şekil 2: Asfalt SHAP değer görselleri a) 103. bant, b) 55. bant

IV. VARGI

Bu çalışmada yorumlanabilir hiperspektral sınıflandırma için SHAP yöntemi kullanılması önerilmektedir. Sınıflandırma modeline uygulanan treeSHAP yöntemi ile spektral bantların farklı sınıflar için sınıflandırma modelinin çıktısına katkısı hakkında nicel bir değerlendirme elde edilebilmektedir. Önerilen yöntem, her sınıf için, sınıflandırmaya bant-tabanlı yorumlanabilirlik sağlamaktadır. Ön deneysel sonuçlar Pavia Üniversitesi hiperspektral verisi ve rastgele orman yöntemi için, ortalama SHAP değerleri görselleri cinsinden bu bildiride sunulmuştur. Çalışmanın ilerletilmesi ile hem elde edilen sonuçların spektral imzalardaki ayırt edici öznitelikler ile bant



Şekil 3: Zift SHAP değer görselleri a) 11. bant, b) 103. bant

tabanlı uyumunun, hem de yorumlanabilirlik tabanlı özgün bir bant seçimi ve boyut azaltımı yaklaşımının sınıflandırma başarımında etkisinin denetlenmesi planlanmaktadır. Olası gelecek çalışmalar arasında hiperspektral verilerin evrişimli sinir ağları gibi derin öğrenme yaklaşımları ile sınıflandırılması sürecinde SHAP tabanlı analizlerin gerek spektral bant gerek ise çıkarılan öznitelik uzayları üzerinden gerçekleştirilmesi sayılabilir.

KAYNAKLAR

[1] M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS International Journal of Geo-Information*, vol. 7, no. 4, p. 129, 2018.

[2] M. Campos-Taberner, F. J. García-Haro, B. Martínez, E. Izquierdo-Verdiguier, C. Atzberger, G. Camps-Valls, and M. A. Gilabert, "Understanding deep learning in land use classification based on sentinel-2 time series," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[3] R. Yang, X. Xu, Z. Xu, C. Ding, and F. Pu, "A class activation mapping guided adversarial training method for land-use classification and object detection," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 9474–9477.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[5] S. S. Matin and B. Pradhan, "Earthquake-induced building-damage mapping using Explainable AI (XAI)," *Sensors*, vol. 21, no. 13, p. 4489, 2021.

[6] A. Abdollahi and B. Pradhan, "Urban vegetation mapping from aerial imagery using Explainable AI (XAI)," *Sensors*, vol. 21, no. 14, p. 4738, 2021.

[7] H. Singh, A. Roy, R. Setia, and B. Pateriya, "Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (xai) approach," *Modeling Earth Systems and Environment*, pp. 1–7, 2021.

[8] W. Huang, W. Li, J. Xu, X. Ma, C. Li, and C. Liu, "Hyperspectral monitoring driven by machine learning methods for grassland aboveground biomass," *Remote Sensing*, vol. 14, no. 9, p. 2086, 2022.

[9] C. A. Toledo, M. Crawford, and T. Vyn, "Maize yield prediction based on multi-modality remote sensing and lstm models in nitrogen management practice trials," in *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2022, pp. 1–7.

[10] E. O. Yilmaz and T. Kavzoglu, "Classification of jilin-1 gp01 hyperspectral image using machine learning techniques with explainable artificial intelligence," *Intercontinental Geoinformation Days*, vol. 5, pp. 145–148, 2022.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.