

AKÜ FEMÜBİD 22 (2022) 045201 (789-797)

AKU J. Sci. Eng. 22 (2022) 045201 (789-797)

DOI: 10.35414/akufemubid.1106065

Araştırma Makalesi / Research Article

Lityum-iyon Bataryaların Katot Malzemelerindeki Kristal Yapının Kestirimi için Makine Öğrenimi Yöntemlerinin Uygulanması

Meltem SÜPÜRTÜLÜ^{1,3}, Çiğdem DİNDAR^{2,3}, Recep AKYÜZ³¹ Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Bursa.² Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Makine Mühendisliği Bölümü, Bursa.³ TOFAŞ Türk Otomobil Fabrikası A.Ş., AR-GE Merkezi, Bursa.Sorumlu yazar e-posta: ¹meltem.supurtulu@tofas.com.tr ORCID ID: <https://orcid.org/0000-0003-0473-7731>²cigdem.dindar@tofas.com.tr ORCID ID: <https://orcid.org/0000-0002-4597-906X>³recep.akyuz@tofas.com.tr ORCID ID: <https://orcid.org/0000-0003-3310-6437>

Geliş Tarihi: 20.04.2022

Kabul Tarihi: 07.08.2022

Öz

Lityum-iyon silikat katotların fiziksel ve kimyasal özelliklerinin kristal yapı sisteminden önemli ölçüde etkilendiği yaygın olarak kabul edilmektedir. Bu nedenle, kristal yapılarının kestirimi, batarya uygulamalarında katotların diğer birçok özelliğini tahmin etmek için hayati öneme sahiptir. Li-Si-(Mn,Fe,Co)-O bileşik yapılarına sahip silikat bazlı katotların üç ana kristal sistemi (monoklinik, ortorombik ve triklinik) makine öğrenimi yöntemlerinden çeşitli sınıflandırma teknikleri kullanılarak tahmin edilmiştir. Hesaplamalar, Malzeme Projesi'nden elde edilen yoğunluk fonksiyonel teorisi hesaplamalarının sonuçlarına dayanmaktadır. Kristal sistemi ile katotların diğer fiziksel özellikleri arasındaki güçlü korelasyon, istatistiksel modellerdeki özellik değerlendirmesine dayalı olarak doğrulanmıştır. Ayrıca, en iyi tahmin doğruluğunu elde etmek için çeşitli sınıflandırma yöntemlerinin parametreleri optimize edilmiştir. XGBoost ve Destek Vektör Sınıflandırıcı algoritmalarının çapraz doğrulama testlerinde çalışmada kullanılan diğer sınıflandırma yöntemleri arasında en yüksek tahmin doğruluğunu sağlamıştır.

Anahtar kelimeler

Lityum-iyon bataryalar;
Makine öğrenimi;
Optimizasyon;
Kristal yapı;
Sınıflandırma teknikleri

Implementation of Machine Learning Approaches for Crystal Structure Estimation in Lithium-ion Battery Cathode Materials

Abstract

It has commonly been assumed that the physical and chemical characteristics of lithium-ion silicate cathodes are influenced significantly by the crystal structure system. Because of this, crystal structure estimation has played a vital role in bringing about forecasting many other features of cathodes in battery applications. Using a variety of classification techniques in machine learning which three primary crystal structure (monoclinic, orthorhombic, and triclinic) of silicate-based cathode materials with compound systems of Li-Si-(Mn,Fe,Co)-O has been estimated. The computations are based on the Materials Project's density functional theory computations. In this study, it has been explained that based on property evaluation in statistical models, the considerable correlation between the crystal system and other physical characteristics of the cathodes was validated. Furthermore, the parameters of several categorization techniques have been tuned in order to achieve maximum prediction accuracy. This case has shown that the XGBoost and Support Vector Classifier algorithms carried out in this study the highest forecasting accuracy in this study along with many other classification methods in cross-validation tests.

Keywords

Lithium-ion batteries;
Machine learning;
Optimization;
Crystal structure;
Classification techniques

1. Giriş

Çok çeşitli, yeni ve karmaşık malzemelerin fiziksel ve kimyasal özelliklerinin hesaplanması, malzemelerin elektronik yapısını sayısal yöntemlerdeki gelişmelerin yanı sıra hesaplama gücündeki hızlı gelişmeler sayesinde mümkün olmuştur. Bu sayede, araştırmacılar malzemelerin tahmin edilen niteliklerine ilişkin büyük miktarda veriye erişebilmektedir. Örnek olarak Malzeme Projesi (Jain *et al.* 2013), elektronik yapı yoğunluğu fonksiyonel teorisinde (DFT, Density Functional Theory) malzemelerin belirlenen fiziksel ve kimyasal özellikleri halka açık veri seti arşivlerine sunulmuştur. DFT hesaplamaları, malzemelerdeki elektron yoğunluğunu ve bant yapısını kestirmek için etkili araçlardır. Li-iyon bataryalar da dahil olmak üzere birçok farklı malzeme türü için fiziksel özelliklerin kesin hesaplamaları, değişim-korelasyon potansiyelinin geliştirilmesindeki ilerlemeden kaynaklanmıştır (Yan *et al.* 2014).

Malzemelerin özellikleri hakkında daha fazla bilgi sahibi olabilmek için malzemelerle ilgili büyük miktardaki verinin analiz edilmesi gerekmektedir. Genel olarak, geleneksel istatistiksel modeller ile çeşitli fiziksel özellikler arasındaki karmaşık ilişkileri ortaya çıkarmak zor olmaktadır. Gelişmiş makine öğrenimi (ML, Machine Learning) algoritmaları ise kristal yapı ile diğer fiziksel ve kimyasal özellikler arasındaki karmaşık ilişkiyi ortaya çıkarma yeteneğine sahiptir. ML, birçok karmaşık sınıflandırma ve regresyon problemini çözmek için kullanılmıştır. Bu alanlardan biri de Li-iyon bataryalardır.

Li-Si-(Mn,Fe,Co)-O bileşimlerine sahip katot malzemeleri, Li-iyon bataryalardaki uygulamaları nedeniyle araştırmalar tarafından büyük ilgi görmektedir (Askanazi *et al.* 2021; Pan *et al.* 2021; Bartel *et al.* 2020; Park and Wolverton 2020; Cubuk *et al.* 2019). Shandiz ve Gauvin (2016), Malzeme Projesi veri setini ile ML yöntemleri olan lineer, kuadratik ve daralan diskriminant analizi, sinir ağları, destek vektör makineleri, k-en yakın komşular, rastgele ormanlar ve aşırı derecede rastgele ağaçları modelleri oluşturmada kullanmışlardır. Rastgele

ormanlar ve aşırı derecede rastgele ağaçlar, kristal sistemin sınıflandırılması için topluluk yöntemlerinin gücünü ve esnekliğini kanıtlayan diğer sınıflandırıcılar arasında en yüksek genel ortalama doğruluğu verdiğini belirtmiştir. Aşırı derece rastgele ağaçlarda özellik önem değerlendirmesine dayalı olarak, kristal hacmi ve site sayısı, veri setindeki kristal sistem tipini belirlemek için en yüksek etkileri göstermiştir. Ancak, silikat katotların oluşum enerjisi, gövde üzerindeki enerji ve bant aralığı gibi diğer özellikleri de kristal sisteminin belirlenmesi için önemli olduğu gösterilmiştir. Maphanga vd. (2020), Malzeme Projesi veri setini kullanarak girdi verisi DFT'den hesaplanan malzeme özellikleri ile batarya malzemesindeki gerilimleri tahmin etmede ML regresyon yöntemlerini araştırmışlardır. Önerilen model ile iyi bir doğruluk performansı elde edilmiştir. Jha vd. (2018), temel bileşiklerden malzeme özelliklerini kestirmeye yönelik derin sinir ağı modelinin tasarımını ve uygulamasını gerçekleştirmiştir. ElemNet, malzeme özelliklerini daha iyi doğruluk ve hızla tahmin etmesine olanak tanıyan yapay zekâ kullanarak farklı elementler arasındaki fiziksel ve kimyasal etkileşimleri ve benzerlikleri otomatik olarak yakalamaktadır. Katotların kristal yapısı, Lityum-iyon bataryaların özellikleri üzerinde önemli bir etkiye sahiptir (Longo *et al.* 2014). Katodun fiziksel ve kimyasal özelliklerinin daha iyi belirlenmesi için, tanımlama doğru hesaplamaları ve deneysel metodolojilerin araştırılması ve geliştirilmesi kritik öneme sahiptir.

Gerçekleştirilen çalışmada; Li-Si-(Mn,Fe,Co)-O bileşikleri ile katot malzemelerinin üç ana kristal sistemi (CS, Crystal System) türünü tahmin etmede Malzeme Projeleri'nden (Jain *et al.* 2013) elde edilen veriler ile çeşitli sınıflandırma yöntemleri incelenmiştir. Tahmin edilen veya bilinen katotlar için DFT sonuçlarının çoğu üç (monoklinik, ortorombik ve triklinik) sınıf için mevcuttur. Modelleri oluşturmak için kullanılan sınıflandırma yöntemleri; Derin Sinir Ağları (DNN, Deep Neural Networks), Destek Vektör Sınıflandırıcı (SVC, Support Vector Classifier), XGBoost Sınıflandırıcı (XGBoost, eXtreme Gradient Boosting Classifier) ve Karar Ağaçları (DT, Decision Tree) algoritmalarıdır.

Çizelge 1. Veri setinden seçilen bazı silikat katotlar (Shandiz and Gauvin 2016).

Formül	Uzay Grubu	E_f (eV)	E_H (eV)	E_g (eV)	N_s	ρ (g/cm ³)	V (Å ³)	Kristal Sistemi
Li ₂ MnSiO ₄	Pc	-2.699	0.006	3.462	16	2.993	178.513	Monoklinik
Li ₂ Mn ₂ (SiO ₃) ₃	P21/c	-2.769	0.077	3.188	64	2.517	929.064	Monoklinik
Li ₂ Co ₂ (SiO ₃) ₃	P21/c	-2.598	0.069	2.727	64	2.739	872.856	Monoklinik
Li ₂ FeSi ₃ O ₈	P21	-2.84	0.069	3.081	28	2.665	351.384	Monoklinik
LiMn(SiO ₃) ₂	Pbca	-2.824	0.036	0.037	80	3.343	850.626	Ortorombik
LiFeSiO ₄	Pn21a	-2.604	0.018	2.961	28	2.89	355.979	Ortorombik
Li ₂ Co ₂ Si ₂ O ₇	C2cm	-2.453	0.072	2.84	26	3.579	278.304	Ortorombik
Li ₇ Mn ₁₁ (Si ₃ O ₁₆) ₂	P1	-2.439	0.092	0.3611	56	3.909	566.407	Triklinik
LiFeSi ₂ O ₈	P1	-2.896	0.032	3.342	26	2.76	330.953	Triklinik
LiCo ₃ (SiO ₄) ₂	P1	-2.25	0.076	0.005	42	3.318	552.402	Triklinik

2. Materyal ve Metot

Veri seti, Malzeme Projesi'nden (Jain *et al.* 2013) alınan verileri kullanarak Li-Si-(Mn, Fe, Co)-O bileşimlerine sahip 339 katot malzemesi için DFT hesaplamalarının sonuçlarını içermektedir. Malzeme Projesi, yüksek verimli bir sürece dayanmaktadır. Malzeme Projesi'nin DFT hesaplamalarında kullanılan kristal yapıların çoğu, atom konumlarını ve kristal kafes özelliklerini içeren inorganik kristal yapısal veri tabanından (ICSD, Inorganic Crystal Structural Database) elde edilmiştir. Atomik konumların optimizasyonu, mevcut veya oluşturulan yapılar üzerinde de gerçekleştirilir. Hesaplamaların detayları hakkında daha fazla bilgi Jain vd. (2013) makalesinde bulunabilir.

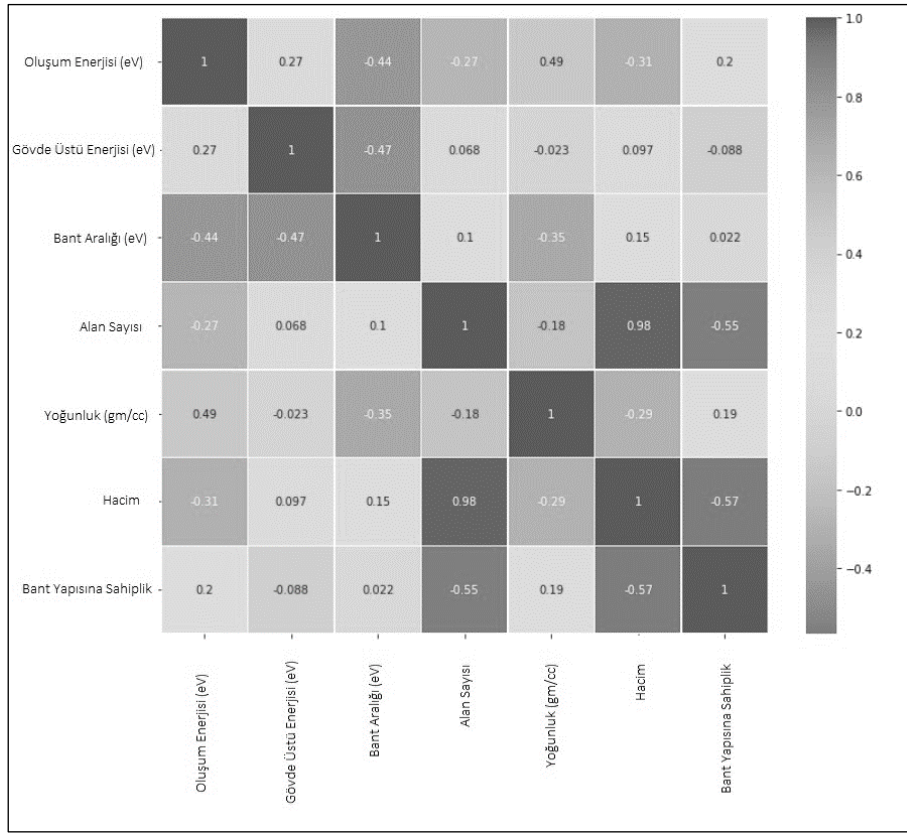
Veri seti, her katodun kristal sisteminde kimyasal formülü, uzay grubunu, oluşum enerjisini (E_f), gövde üzerindeki enerjiyi (E_H), bant aralığını (E_g), alan sayısını (N_s), yoğunluğunu (ρ) ve birim hücrenin hacmini (V) içermektedir. Veri setinde belirtilen özellikler Malzeme Projesi sözlüğüne göre N_s ve ρ sırasıyla kristalin birim hücresindeki atom sayıları ve toplu kristal malzemelerin yoğunluğudur. Makine öğrenimi ve derin öğrenme modellerini oluşturmak için, $V = M/\rho$ (M atomik kütedir) verildiğinde yalnızca V değişkeni kullanılmaktadır. Ayrıca E_H , malzemenin en kararlı olanlara ayrışma enerjisi olarak tanımlanır. Oluşum enerjisinin ve diğer özelliklerin hesaplanmasının OK sıcaklıkta ve ortam basıncındadır. Çizelge 1, veri setinden seçilen bazı silikat katotları için verileri göstermektedir.

Veri seti, çok çeşitli karmaşık yapılar ve çeşitli kimyasal bileşimler içermektedir. Şekil 1, veri kümesindeki silikat katotların özelliklerinin eş çizimlerini göstermektedir. Çapraz grafikler, katotların her bir özelliğinin dağılımı için histogram grafikleridir. Görülebileceği gibi, genellikle özellikler ile kristal sistemi arasında belirgin bir ilişki yoktur. Bu karmaşıklık, sınıflandırma probleminin geleneksel yöntemlerle çözülmesini zorlaştırmaktadır. Malzeme Projesi'ndeki hesaplama sonuçlarının, yeni optimizasyonlar yapılması veya yeni potansiyellerin kullanılması nedeniyle değişmeye eğilimli olduğu görülmektedir.

2.1 Makine Öğrenimi için Sınıflandırma Yöntemleri

Sınıflandırma, veri setini belirli sınıflara bölmek için makine öğrenmesinde kullanılan bir yöntemdir. Bu çalışmadaki sınıflandırma probleminde kristal sistemi (monoklinik, ortorombik ve triklinik) belirtildiği için denetimli öğrenme yöntemleri seçilmiştir. Ayrıca, sınıflandırmanın doğruluğu, sınıfların doğru tahmin edilen kısmı olarak tanımlanmaktadır.

Denetimli sınıflandırma için, $n \times m$ boyutlu özellik matrisi X ve farklı sınıflarda n ve K uzunluklarına sahip tek boyutlu bir matris olarak cevap matrisi Y kullanılmıştır. Burada n, gözlem (örnek) sayısı ve m, öznitelik sayısıdır. Bu çalışma için n, m ve K sırasıyla 339, 5 ve 3'tür.



Şekil 1. Malzeme Projesi'nden alınan verilere ait Li- (Mn, Fe,Co)-Si-O katotlarının farklı özelliklerinin korelasyon (sayısal) grafiği.

Kristal sistemi, diğer değişkenlere bağlı bir fonksiyon olarak tanımlanabilir: $CS = f(V, E_g, N_s, E_f, E_H)$. Buradaki, oluşum enerjisini (E_f), gövde üzerindeki enerjiyi (E_H), bant aralığını (E_g), alan sayısını (N_s), yoğunluğunu (ρ) ve birim hücrenin hacmi (V)

değişkenlerine dayanarak kristal sınıfı, makine öğrenimi yöntemleri ile tahmin edilebilmektedir. Bu bölümde, modelleri oluşturmak için veri kümesi üzerinde uygulanan sınıflandırma yöntemleri kısaca tanıtılmaktadır (Çizelge 2).

Çizelge 2. Kullanılan yöntemlere ait temel ifadeler.

Yöntem	İfade	Açıklama
Destek Vektör Sınıflandırıcı	$\min(w, b, \xi): \frac{1}{2} w^T w + C \sum_{i=0}^{N-1} \xi_i$ $\max: \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$ $f(x) = \text{sgn} \left(\sum_{i=0}^{N-1} y_i \alpha_i K(x_i, x) + b \right)$	$y_i (w^T \phi(x_i) + b) \geq (1 - \xi_i); \xi_i \geq 0$ $\sum_i \alpha_i y_i = 0; 0 \leq \alpha_i \leq C, i = 0, \dots, N - 1$ $K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$
Karar Ağaçları	$\text{Gini İndeksi, } G = 1 - \sum_{i=1}^C (p_i)^2$	<p>p, başarı; q, başarısızlık</p> $(p^2 + q^2)$
XGBoost	$\sum_{i=1}^n \frac{1}{2} h_i (f_i(x_i) - \frac{g_i}{h_i})^2 + \Omega(f_i) + \text{sabit}$	<p>h_i, ağırlıklar;</p> <p>$\frac{g_i}{h_i}$, etiketlerle tam ağırlıklı kare kaybı</p>
Derin Sinir Ağları	$\hat{y}_i = f_2 \left\{ \sum_{j=1}^R w_{kj}^{(2)} \left[f_1 \left(\sum_{j=1}^R w_{kj}^{(1)} x_j + b_k^{(1)} \right) \right] + b_k^{(2)} \right\}$	<p>f_1, f_2: Aktivasyon fonksiyonu;</p> <p>w_{kj}: Ağırlıklar; x_j: Girdiler; b_k: Bias</p>

2.1.1 Destek Vektör Makineleri

Destek vektör makineleri (SVM, Support Vector Machine) yöntemi, genellikle sınıflandırma problemlerinde kullanılan denetimli öğrenme yöntemlerinden biridir. Bir doğru yardımı ile bir düzlem üzerine yerleştirilen noktaları ayırmaktadır. Bu doğrunun, iki sınıfın noktaları için de maksimum marjlı hiper düzlemleri bulmaya dayalı olarak eğitim verilerini bölme fikri üzerinde çalışmaktadır. Karmaşık ancak küçük ve orta ölçekteki veri setleri için uygundur. Hiper düzlem, $w^T \phi(x_i) + b$ şeklinde tanımlanmaktadır. w hiper düzlemin normal vektörüdür, x_i eğitim veri kümesidir ve $\phi(x_i)$ eğitim verilerini özellik uzayıyla eşleştirir. Maksimum marjın optimal çözümünü bulmak için optimizasyon problemi çözülmelidir (Chang 2011, Int Kyn. 1).

2.1.2 Karar Ağaçları Algoritmaları

Karar ağacı algoritmaları, sınıflandırma algoritmalarından biridir. Genellikle ölçeklendirme ve çok fazla ön işleme süreçlerine gerek duymamaktadır. Bir karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır. Önceden tanımlanmış bir hedef değişkene sahiptirler.

Çalışma prensibi, yapıları itibarıyla en tepeden aşağı inen bir strateji ile verileri ayrı sınıflara bölerek ile "En İyi" olup olmadığını belirlemek için herhangi bir düğümde test edilecek öznelilikleri ayırmak için kullanılmaktadır. Her dalda elde edilen ayırım mümkün olduğu kadar saf olmalıdır, çünkü bölme kriterlerinin aynı olması gerekmektedir. Veriyi iyi bir şekilde açıklamayan aşırı karmaşık ağaçlar üretilebilir. Bu durumda ağaç dallanması takip edilememektedir. Ezbere öğrenme, aşırı uyum (overfitting) veya öğrenememe (underfitting) gibi sorunlarla karşılaşıldığında hiper parametre değerlerinin değiştirilmesi gerekmektedir (Jandav and Channe 2016, Int Kyn. 2).

Karar ağaçları, bir düğümü iki veya daha fazla alt düğüme bölünmesine karar vermesinde birden fazla algoritma kullanılmaktadır. Alt düğümlerin

oluşturulması, alt düğümlerin homojenliğini arttırmaktadır. Başka bir ifadeyle, düğümün saflığının hedef değişkenlere göre artmasıdır. Algoritma seçimi, hedef değişkenin tipine bağlıdır. Karar ağaçlarında en sık kullanılan algoritmalar; kategorik değişkenler için Entropi, Gini, Sınıflandırma Hatası; sürekli değişkenler için ise En Küçük Kareler yöntemi şeklindedir. Bu çalışmada Gini İndeksi kullanılmıştır. Entropi, verilerimizle ilgili belirsizliğin bir ölçüsüdür. Sezgisel olarak, bir veri kümesinin yalnızca bir etiketi varsa, daha düşük bir entropiye sahiptir. Dolayısıyla verilerdeki entropiyi en aza indirecek bir şekilde bölünmesi gerekmektedir (Int Kyn. 2).

2.1.3 XGBoost Algoritmaları

Ölçeklenebilir bir ağaç arttırma sistemi olan Aşırı Gradyan Arttırma (XGBoost, eXtreme Gradient Boosting) algoritmaları, "zayıf" tahminleyicileri olarak karar ağaçlarını kullanan bir gradyan arttırma algoritmasıdır. Ayrıca, yöntem en iyi performans ve hız için özel olarak tasarlanmıştır. XGBoost, kullanım kolaylığı ve tahmin gücü özellikleriyle regresyon ve sınıflandırma görevleri için kullanılan denetimli öğrenme algoritmasıdır. Seyrek verileri işlemek için seyrekliğe duyarlı yeni bir algoritma ve yaklaşık öğrenme için teorik olarak doğrulanmış ağırlıklı bir nicel çizim olan yöntem, önbellek erişim kalıplarının, veri sıkıştırmanın ve parçalamanın, karar ağaçlarının güçlendirilmesinde ölçeklenebilir bir uçtan uca sistem oluşturmak için temel unsurlar olduğunu göstermektedir. İç görüleri birleştirerek XGBoost, minimum miktarda kaynak kullanılarak gerçek dünya ölçeğindeki sorunları çözebilmektedir. (Chen and Guestrin 2016, Int Kyn. 3).

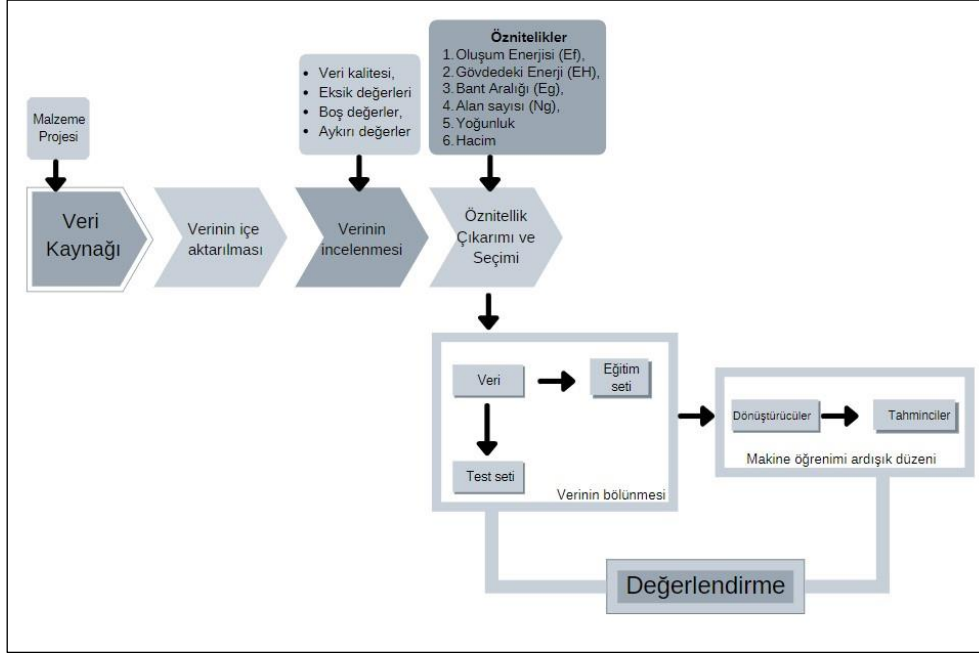
2.1.4 Derin Sinir Ağları

Yapay Sinir Ağları (ANN, Artificial Neural Network), sınıflandırma ve regresyon problemleri için makine öğrenimindeki çok yönlü ve etkili yöntemlerden biridir (Ripley, 1996). Ancak, yüksek derecede esneklik özelliği, parametrelerin optimal değerlerini bulma sürecini zorlaştırabilir.

Derin sinir ağları (DNN, Deep Neural Network) algoritmaları ise yapay sinir ağlarının çok katmanlı

biçimidir. Bu çalışmada çok katmanlı algılayıcılar (multi-layer perceptron) kullanılmıştır. DNN, girdi katmanı (input layer), gizli katman (hidden layer) ve çıktı katmanı (output layer) olmak üzere üç katmandan oluşmaktadır. Aktivasyon fonksiyonu,

doğrultulmuş doğrusal (RELU, Rectified Linear Unit) fonksiyonu ile uygunluk kriteri olarak maksimum koşullu olabilirlik kullanılmıştır (Venables and Ripley, 2002).



Şekil 2. Malzeme Projesi'nden alınan verilerin makine öğrenimi yöntemleri ile analiz edilmesine ilişkin ardışıl düzeni.

3. Bulgular

Çalışmada, veri setine makine öğrenimi sınıflandırma yöntemlerinden olan Destek Vektör Sınıflandırıcı (SVC, Support Vector Classifier), Karar Ağaçları (DC, Decision Tree), XGBoost ve Derin Sinir Ağları (DNN, Deep Neural Network) uygulanmıştır. Tüm bu yöntemler Kaggle (Int Kyn. 5) ve Google Colab (Int Kyn. 4) ortamında hazırlanmıştır.

Ayrıca, ham verilerin alınmasından makine öğrenimi yöntemlerinin uygulanıp değerlendirilmesine kadar gerçekleştirilen süreç, Şekil 2'de yer almaktadır.

Çalışmada kullanılan veri seti, sınırlı sayıda veri içerdiğinden makine öğrenimi ve derin öğrenme algoritmalarının performansının daha verimli olarak ölçülebilmesi için veri setinin %80'i eğitim ve %20'si test için kullanılmıştır. Doğrulama için k-katlı çapraz doğrulama (k-fold cross-validation) yöntemi uygulanmıştır. k=1 ve bölütleme değeri (fold) 10 olarak seçilmiştir.

Derin öğrenmede iki veya daha fazla sınıflandırma modellerinin performansının ölçümünde genellikle hata matrisi kullanılarak doğruluk (accuracy), kesinlik (precision), geri çağırma (recall) ve F1-puanı (F1 score) gibi parametreler hesaplanmaktadır. Hata matrisi, gerçek ve tahmin sınıfının dört farklı durumunu içeren bir tablodur. Doğru pozitif ve negatif, yanlış pozitif ve negatif gibi dört farklı hasta matrisi durumları Çizelge 4'te gösterilmiştir. Hata matrisi durumlarına göre doğruluk Eşitlik (1), kesinlik Eşitlik (2), geri çağırma Eşitlik (3) ve F1-puanı Eşitlik (4) ile hesaplanmaktadır (Demirezen, 2021).

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (2)$$

$$\text{Geri Çağırma} = \frac{DP}{DP+YN} \quad (3)$$

$$\text{F1-puanı} = \frac{2 \times (\text{Geri Çağırma} \times \text{Kesinlik})}{\text{Geri Çağırma} + \text{Kesinlik}} \quad (4)$$

Çizelge 3. Hata matrisi durumları

Tanım	Açıklama	Literatürdeki Karşılığı (ing)
<i>Doğru Pozitif (DP)</i>	Doğru olarak belirlenen pozitif tahmin sayısı	TP, True Positive
<i>Doğru Negatif (DN)</i>	Doğru olarak belirlenen negatif tahmin sayısı	TN, True Negative
<i>Yanlış Pozitif (YP)</i>	Yanlış olarak belirlenen pozitif tahmin sayısı	FP, False Positive
<i>Yanlış Negatif (YN)</i>	Yanlış olarak belirlenen negatif tahmin sayısı	FN, False Negative

3.1 Hiper Parametre Seçimi

Hiper parametre optimizasyonu için, her modelde optimal parametreleri seçmek üzere bir ızgara arama çapraz doğrulama gerçekleştirilmiştir. Her model, Python Scikit-Learn kitaplığı kullanılarak eğitilmiştir. Çizelge 4 ile listelenenler dışında çoğu hiper parametre varsayılan konfigürasyonda tutulmuştur.

Sınıflandırma yöntemlerinden ilk olarak Destek Vektör Sınıflandırıcısı algoritması seçilmiştir. Oluşturulan model üzerinde en büyük etkiye sahip olan parametreler ızgara araması ile bulunmuştur.

Doğrusal olmayan destek vektör makinesi algoritması kernel yöntemlerinden radyal tabanlı fonksiyon ve C=10000 parametre değerleri ile çalışılmış ve en yüksek başarı oranına ulaşılmıştır. Karar ağacı parametreleri; Gini Bilgi Kriteri, maksimum derinlik 3, ve karmaşıklık parametresi 0,0 olarak seçilmiştir. Aşırı Gradyan Arttırma algoritması; çoklu sınıf problemi olduğundan softmax, maksimum derinlik 5, öğrenme oranı (eta) 0,15 ile en yüksek başarı oranına ulaşılmıştır. Bu çalışmada önerilen derin öğrenme mimarisi Çizelge 4'te yer almaktadır.

Çizelge 4. Hiper-parametre seçimi

Yöntem	Hiper Parametreler	Açıklama
<i>Destek Vektör Sınıflandırıcı</i>	{'C': 10000, 'gamma': 0.0001, 'kernel': 'rbf'}	Izgara Arama
<i>Karar Ağaçları</i>	{max_depth=3, random_state=0}	Gini indeksi
<i>XGBoost</i>	{objective='multi:softmax', n_estimators=200, seed=123, max_depth=5, learning_rate=0.15, colsample_bytree=1, subsample=1}	Varsayılan ayarlar
<i>Derin Sinir Ağları</i>	1024, activation='relu', kernel_regularizer=regularizers.l2(0.001) 512, activation='relu', kernel_regularizer=regularizers.l2(0.001) 256, activation='relu', kernel_regularizer=regularizers.l2(0.001) 128, activation='softsign', kernel_regularizer=regularizers.l2(0.001)	Keras paketleri

4. Tartışma ve Sonuç

Gerçekleştirilen çalışmada; Li-Si- (Mn, Fe, Co)-O bileşikleri ile silikat katotların üç ana kristal yapı sistemini (monoklinik, ortorombik ve triklinik) belirlemek için çok çeşitli makine öğrenimi sınıflandırma yöntemleri başarıyla kullanılmıştır.

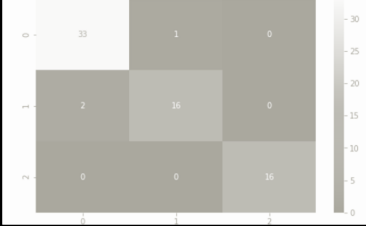
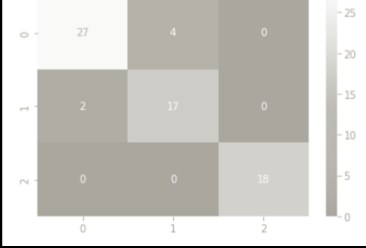
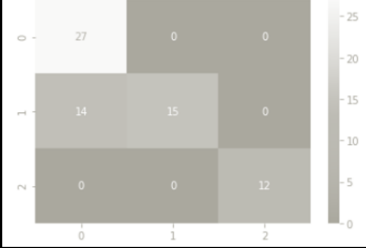

Shandiz ve Gauvin (2016), Materyal Projesi'nden alınan verileri sinir ağları, destek vektör makineleri, k-en yakın komşular, rastgele ormanlar ve aşırı derecede rastgele ağaçları yöntemlerinin optimizasyon parametrelerini değiştirerek etkilerini doğruluk ölçümü metriği kullanılarak gerçekleştirmişlerdir. Bu çalışmada ölçüm sonuçları F-1 skoru kullanılarak hesaplanmıştır. Çalışmada

kullanılan yöntemlere ait sonuçlar karmaşıklık matrisi ile oluşturulmuştur (Çizelge 5).

Doğrulama aşamasında, k-katlı çapraz doğrulama (k-fold cross-validation) yöntemi uygulanmıştır. k=1 ve bölütleme değeri (fold) 10 olarak seçilmiştir. Her sınıflandırma yönteminin parametrelerinin optimizasyonunun, genel ortalama doğruluk üzerinde önemli bir etkiye sahip olduğu doğrulanmıştır.

Materyal Projesi'nde alınan veri seti 339 satırlık veri içerdiğinden makine öğrenimi modelleri oluşturmak için eğitim verilerinin yüzdesini artırmak, genel ortalama tahmin doğruluğunun sürekli olarak artışa neden olduğu gözlenmiştir.

Çizelge 5. Makine Öğrenimi Yöntemleri ile kristal sınıfının belirlenmesine dair kestirim sonuçları

Yöntem	XGBoost	Sınıf	Doğruluk	Kesinlik	Geri Çağırma	F-1 Puanı	F1 Puanı Sonucu (%)
	Monoklinik	0.955882	0.942857	0.970588	0.956522	95.6522	
	Ortorombik	0.955882	0.941176	0.888889	0.914286	91.4286	
	Triklinik	0.955882	0.941176	0.888889	0.914286	91.4286	
	Ortalama	0.955882	0.941737	0.916122	0.914286	92.8364	
Yöntem	Destek Vektör Sınıflandırıcı	Sınıf	Doğruluk	Kesinlik	Geri Çağırma	F-1 Puanı	F1 Puanı Sonucu (%)
	Monoklinik	0.911765	0.870968	0.931034	0.900000	90.0000	
	Ortorombik	0.911765	0.809524	0.894737	0.850000	85.0000	
	Triklinik	0.911765	0.818182	0.900000	0.857143	85.7143	
	Ortalama	0.911765	0.832891	0.908590	0.869048	86.9048	
Yöntem	Karar Ağacı	Sınıf	Doğruluk	Kesinlik	Geri Çağırma	F-1 Puanı	F1 Puanı Sonucu (%)
	Monoklinik	0.794118	1.000000	0.658537	0.794118	79.4118	
	Ortorombik	0.794118	1.000000	0.517241	0.681818	68.1818	
	Triklinik	0.794118	1.000000	0.461538	0.631579	63.1579	
	Ortalama	0.794118	1.000000	0.545772	0.702505	70.2505	
Yöntem	Derin Sinir Ağı	Sınıf	Doğruluk	Kesinlik	Geri Çağırma	F-1 Puanı	F1 Puanı Sonucu (%)
	Monoklinik	0.720588	0.758621	0.647059	0.698413	69.8413	
	Ortorombik	0.720588	0.708333	0.586207	0.641509	64.1509	
	Triklinik	0.720588	0.588235	0.454545	0.512821	51.2821	
	Ortalama	0.720588	0.685063	0.562604	0.617581	61.7581	

Aşırı Gradyan Arttırma ve Destek Vektör Sınıflandırıcı, kristal sistemlerin sınıflandırılması için diğer sınıflandırıcılar arasında en yüksek genel ortalama başarıma ulaşmıştır. Ayrıca, sınıfların bireysel olarak kestirilmesine dayanan değerlendirme de gerçekleştirilmiştir. Öznitelik önem değerlendirmesine bağlı olarak Aşırı Gradyan Arttırma yöntemi, kristal hacmi ve site sayısı, veri setindeki kristal sistem tipini belirlemek için en yüksek etkileri göstermiştir. Ancak, kristal sistemin belirlenmesi için silikat katotların oluşum enerjisi, gövde üzerindeki enerji ve bant aralığı gibi diğer özellikleri de dikkate alınarak hesaplanmaktadır.

Çıkar Çatışması

Yazarlar, bilinen herhangi bir çıkar çatışması veya herhangi bir kurum/kuruluş ya da kişi ile ortak çıkar bulunmadığını onaylamaktadırlar.

5. Kaynaklar

- Askanazi, E.M., Yadav, S., Grinberg, I., 2021. Prediction of Curie temperatures of ferroelectric solid solutions using machine learning methods. *Computational Materials Science*, **2021 (199)**, 113-121.
- Bartel, C.J., Trewartha, A., Wang, Q., Dunn, A., Jain, A., Ceder, G., 2020. A critical examination of compound stability predictions from machine-learned formation

- energies, energies. *Npj Computational Materials*, **6(1)**,1067-1078.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: A Library for Support Vector Machines, *Transactions on Intelligent Systems and Technology*, **2(3)**, 1-27.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*, **2016**, 785-794.
- Cubuk, E.D., Sendek, A.D., Reed, E.J., 2019. Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data. *The Journal of Chemical Physics*, **150(21)**, 214701-1 – 214701-7.
- Demirezen, M.U., Civrizoglu Buz, A., Yavanoglu, U, 2021. Time series- image transformation- based new approaches in detecting underwater objects with machine learning methods. *Journal of the Faculty of Engineering and Architecture of Gazi University*, **36(3)**, 1399-1415.
- Jadhav, S.D., Channe, H.P., 2016. Efficient recommendation system using decision tree classifier and collaborative filtering. *Int. Res. J. Eng. Technol.* 2016, **3(8)**, 2113-2118.
- Jain, A., Ong, S.P., et al., 2013. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, **1**, 011002-1-11.
- Jha, D., Ward, L., Paul, A., Liao, W., Choudhary, A., Wolverton, C., Agrawal, A., 2018. ElemNet: Deep Learning the Chemistry of Materials from Only Elemental Composition. *nature: scientificreports*, **8(1)**, 17593-1-13.
- Longo, R.C., Xiong, K., Santosh, K.C., 2013. Crystal structure and multicomponent effects in Tetrahedral Silicate Cathode Materials for Rechargeable Li-ion Batteries. *Electrochimica Acta*, **121(1)**, 434-442.
- Maphanga, R.R., Mokoena, T., Ratsoma, M., 2020. Estimating DFT calculated voltage using machine learning regression models. *Materials Today: Proceedings*, **38(2)**, 773- 778.
- Pan, S., Wang, Y., Yu, J., Yang, M., Zhang Y., Wei, H., Yuechao C., Wu, J., Han, J., Wang, C., Liu, X., 2021. Accelerated discovery of high performance Cu-Ni-Co-Si alloys through machine learning, *Materials & Design*, **209(2021)**, 109929-1-16.
- Park, C.W., Wolverton, C., 2020. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *American Physical Society-Physical Review Materials*, **4**, 063801-1-26.
- Ripley, B.D., 1996. Pattern Recognition and Neural Networks, Cambridge University Press, 143-181.
- Shandiz, M.A, Gauvin, R., 2016. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. *Computational Materials Science*, **117(2016)**: 270-278.
- Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed., Springer, 243-249.
- Yan, L.M, Su, J.M., Sun, C., Yue, B.H., 2014. Review of the first principles calculations and the design of cathode materials for Li-ion batteries. *Advances in Manufacturing*, **2**: 358-368.

İnternet kaynakları

- 1-[https://www.veribilimiokulu.com/xgboost-nasil-calisir/XGBoost\(eXtremeGradientBoosting\) \(18.04.2022\).](https://www.veribilimiokulu.com/xgboost-nasil-calisir/XGBoost(eXtremeGradientBoosting) (18.04.2022).)
- 2-[https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agaclari \(18.04.2022\).](https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agaclari (18.04.2022).)
- 3- [https://medium.com/deep-learning-turkiye/nedir-bu-destek-vektor-makineleri-makine-ogrenmesi-serisi-2 \(18.04.2022\).](https://medium.com/deep-learning-turkiye/nedir-bu-destek-vektor-makineleri-makine-ogrenmesi-serisi-2 (18.04.2022).)
- 4- [https://colab.research.google.com \(04.04.2022\).](https://colab.research.google.com (04.04.2022).)
- 5- [https://www.kaggle.com \(02.04.2022\).](https://www.kaggle.com (02.04.2022).)