

Predicting the Usable Life of Lithium-Ion Batteries Using NASA Battery Data

Batuhan Ayyıldız
Bilgisayar Mühendisliği
Celal Bayar üniversitesi
Manisa ,Türkiye
batin.1bey@gmail.com

Abstract—Bu rapor, NASA’nın lityum-iyon pil veri setlerini kullanarak bataryaların kullanılabilir ömürlerini tahmin etmeyi amaçlamaktadır. Bu çalışma, veri setlerinin işlenmesi, veri ön işleme, özellik çıkarımı ve görselleştirme adımlarını içermektedir.

I. GİRİŞ

Lityum-iyon piller, enerji depolama sistemlerinin önemli bileşenleridir. Bu çalışmada, NASA’nın sağladığı lityum-iyon pil veri setlerinden yararlanarak bataryaların ömürlerini tahmin etmeye çalıştık. Çalışmanın amacı, bataryaların State of Health (SOH) ve State of Charge (SOC) değerlerini hesaplayarak kullanılabilir ömürlerini belirlemektir.

II. VERİ SETİ VE ÖN İŞLEME

Veri seti, NASA’nın sağladığı .mat uzantılı dosyalar kullanılarak elde edilmiştir. Bu dosyalar, bataryaların performans verilerini içermekte olup, modelleme ve analiz çalışmaları için kullanılabilir hale getirilmesi gerekmektedir. Başlangıçta, .mat formatındaki veri dosyalarının daha yaygın kullanılan .csv formatına dönüştürülmesi gerekiyordu. Bu dönüşüm işlemi, verilerin analizi ve işlenebilirliği açısından önemli bir adımdır.

Veri dönüşüm süreci şu adımları kapsamaktadır:

- **Veri Setinin İndirilmesi:** İlk olarak, veri seti <https://phmdatasets.s3.amazonaws.com/NASA/5.+Battery+Data+Set.zip> bağlantısından indirilmiştir. İndirilen zip dosyası, dört farklı .mat dosyasını içermektedir. Bu dosyalar şunlardır: B0005.mat, B0006.mat, B0007.mat ve B0018.mat. Her bir dosya, bataryaların çeşitli test verilerini içermektedir.
- **Python ile Dönüştürme:** İndirilen .mat dosyalarını .csv formatına dönüştürmek için Python kullanılmıştır. Bu işlem, verilerin daha analiz edilebilir ve yaygın olarak desteklenen bir formatta sunulmasını sağlamaktadır. Python’da, veri dönüşümü için özel olarak hazırlanmış `NASA_Battery_Data_Extraction.ipynb` adlı bir dosya kullanılmıştır. Bu dosya, .mat dosyalarını okuyarak verileri uygun şekilde .csv formatına dönüştürmeyi başarmıştır.
- **Capacity Değerlerinin Hesaplanması:** .csv dosyalarına dönüştürme sürecinde, bataryaların kapasite değerleri de hesaplanmıştır. Capacity, bataryaların kalan ömrü hakkında bilgi sağlamak için kritik bir ölçüttür. Bu hesaplama, bataryaların kullanım ömrünün doğru bir

şekilde tahmin edilebilmesi için gereklidir. Böylece, her bir bataryanın kapasite değeri, ilgili .csv dosyasında yer alacak şekilde eklenmiştir.

- **Veri Setinin Düzenlenmesi:** Veri dönüşümü tamamlandıktan sonra, sadece discharge (boşaltma) verilerini içeren dosyalar kullanılmıştır. Literatür taramaları ve mevcut bilgiler doğrultusunda, tahminleme işlemlerinin discharge verileri üzerinde yapılması gerektiği belirlenmiştir. Bu nedenle, test amacıyla kullanılan B0007.mat dosyası da dahil olmak üzere, veriler dikkatlice seçilmiş ve yalnızca gerekli olan bilgiler içeren dosyalar kullanılmıştır.

Şekil 1 bu dönüşüm sürecindeki dosyaları göstermektedir:

B0005.csv	24.07.2024 14:44
B0005.mat	1.10.2008 15:50
B0006.csv	24.07.2024 14:46
B0006.mat	1.10.2008 15:50
B0007.csv	27.07.2024 14:35
B0007.mat	1.10.2008 15:50
B0018.csv	24.07.2024 14:46
B0018.mat	1.10.2008 15:50

Fig. 1. .mat ve .csv dosyaları

A. Veri Setinin Dönüştürülmesi

Veri setleri Python kullanılarak .csv formatına dönüştürüldü ve capacity değerleri hesaplandı. Dönüştürme işlemi sırasında literatür taramaları sonucu capacity değerlerinin de csv dosyalarında yer alması gerektiği anlaşıldı.

B. Veri Ön İşleme

Veri setleri üzerinde yapılan incelemelerde, eksik değerler tespit edilmedi. Bu durum, veri setlerinde herhangi bir eksiklik veya kayıp bilgi bulunmadığını göstermektedir. Eksik veri olmaması, veri ön işleme sürecinde doldurma işlemi gibi ek adımların gerekli olmadığını belirledi.

Zaman serisi verileri, sürekli ve sıralı veri noktalarını içerdiğinden, veri setlerinde yalnızca eksik değerler değil, aynı zamanda aykırı değerler veya anormal değişiklikler de

kontrol edilmiştir. Bu kontrol, verilerin bütünlüğünü sağlamak ve modelleme sürecinde güvenilir sonuçlar elde etmek için yapılmaktadır.

Bu bağlamda, veri setlerinde silme işlemi veya benzeri yapısal değişiklikler yapılmamıştır. Veri setinin doğal yapısının korunması, zaman serisi analizi için önemli bir adımdır çünkü bu tür veriler, batarya performansının gerçek zamanlı ve sürekli bir şekilde değerlendirilmesini sağlar.

Şekil 2 bu inceleme sürecinde elde edilen bulguları ve veri setlerinin durumunu görsel olarak sunmaktadır. Şekilde, veri setlerinin eksiksiz ve düzenli yapısı açıkça görülmektedir.

```
DataFrame df_5 bilgileri:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50285 entries, 0 to 50284
Data columns (total 12 columns):
#   Column                      Non-Null Count  Dtype
---  ---
0   cycle                       50285 non-null  int64
1   ambient_temperature        50285 non-null  int64
2   datetime                   50285 non-null  object
3   capacity                   50285 non-null  float64
4   voltage_measured           50285 non-null  float64
5   current_measured           50285 non-null  float64
6   temperature_measured       50285 non-null  float64
7   current_load                50285 non-null  float64
8   voltage_load                50285 non-null  float64
9   time                       50285 non-null  float64
10  SOH                        50285 non-null  float64
11  SoC                        50285 non-null  float64
dtypes: float64(9), int64(2), object(1)
```

Fig. 2. Veri setlerinin incelenmesi.

III. VERİ GÖRSELLEŞTİRME VE ANALİZ

Veri setleri incelendikten sonra, bataryaların kullanım ömürleri ve State of Health (SOH) değerlerine göre çeşitli grafikler oluşturulmuştur. SOH değerlerinin analizinde, threshold olarak 0.75 değeri belirlenmiştir. Bu eşik değeri altında kalan SOH değerleri, bataryanın remaining useful life (RUL) değerinin tükenmiş olduğunu göstermektedir.

Şekil 3, her bir bataryanın kullanım ömrünü görselleştirmektedir. Grafikte, bataryaların kullanım ömrü boyunca SOH değerlerinin değişimi ve bataryanın ömrünü tamamladığı cycle noktaları gösterilmektedir.

Grafiklere göre her bir bataryanın ömrü aşağıdaki şekilde belirlenmiştir:

- **B0005:** Bataryanın kullanılabilir ömrü 126. cycle'da tükenmiştir.
- **B0006:** Bataryanın kullanılabilir ömrü 70. cycle'da tükenmiştir.
- **B0007:** Bataryanın kullanılabilir ömrü 160. cycle'da tükenmiştir.
- **B0018:** Bataryanın kullanılabilir ömrü 99. cycle'da tükenmiştir.

B0005 : bataryasının kullanılabilir ömrü 126. cycle'da tükenmiştir
B0006 : bataryasının kullanılabilir ömrü 70. cycle'da tükenmiştir
B0007 : bataryasının kullanılabilir ömrü 160. cycle'da tükenmiştir
B0018 : bataryasının kullanılabilir ömrü 99. cycle'da tükenmiştir

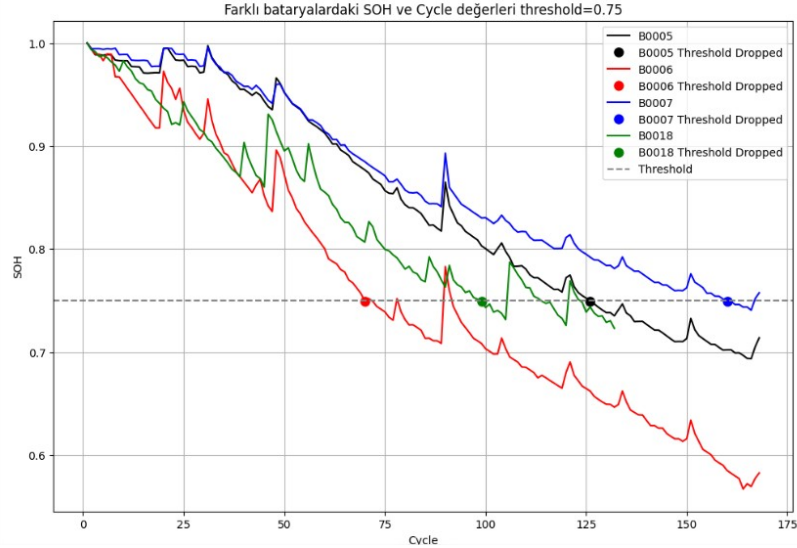


Fig. 3. Bataryaların kullanım ömrü. Şekil, SOH değerlerinin threshold değeri (0.75) altında kaldığı noktaları ve bataryaların kullanım ömürlerini göstermektedir.

IV. FEATURESLERİN ZAMANA GÖRE DEĞERLENDİRİLMESİ

2. görselleştirme yönteminde, her bir döngüde (cycle) batarya parametrelerinin zamanla nasıl değiştiğini gösteren grafikler oluşturulmuştur. Bu yöntem, 'df 5' veri çerçevesindeki her döngü için aşağıdaki parametreleri zaman ekseninde görselleştirir:

- **SoC (State of Charge):** Bataryanın mevcut yük durumunu zamanla gösterir. Bu grafik, bataryanın şarj durumunun döngüler boyunca nasıl değiştiğini görselleştirir.
- **Voltage Measured:** Bataryanın ölçülen voltajını zamanla gösterir. Voltaj değişimleri, bataryanın performansını ve sağlık durumunu analiz etmek için önemlidir.
- **Current Measured:** Bataryanın ölçülen akımını zamanla gösterir. Bu, bataryanın yük altında nasıl davrandığını ve enerji tüketimini anlamaya yardımcı olur.
- **Temperature Measured:** Bataryanın ölçülen sıcaklığını zamanla gösterir. Sıcaklık değişimleri, bataryanın çalışma koşulları ve güvenliği hakkında bilgi sağlar.
- **Voltage Load:** Bataryanın yük altındaki voltajını zamanla gösterir. Bu, bataryanın yük altında nasıl performans gösterdiğini analiz etmek için kullanılır.
- **Current Load:** Bataryanın yük altındaki akımını zamanla gösterir. Bu, bataryanın yük altındaki performansını ve kapasitesini değerlendirmeye yardımcı olur.

Her bir döngü için bu parametrelerin zamanla nasıl değiştiğini gösteren altı farklı grafik oluşturulmuştur. Bu grafikler, bataryanın çeşitli döngülerdeki performansını detaylı bir şekilde analiz etmeyi sağlar ve bataryanın zaman içindeki sağlık durumunu görselleştirir.

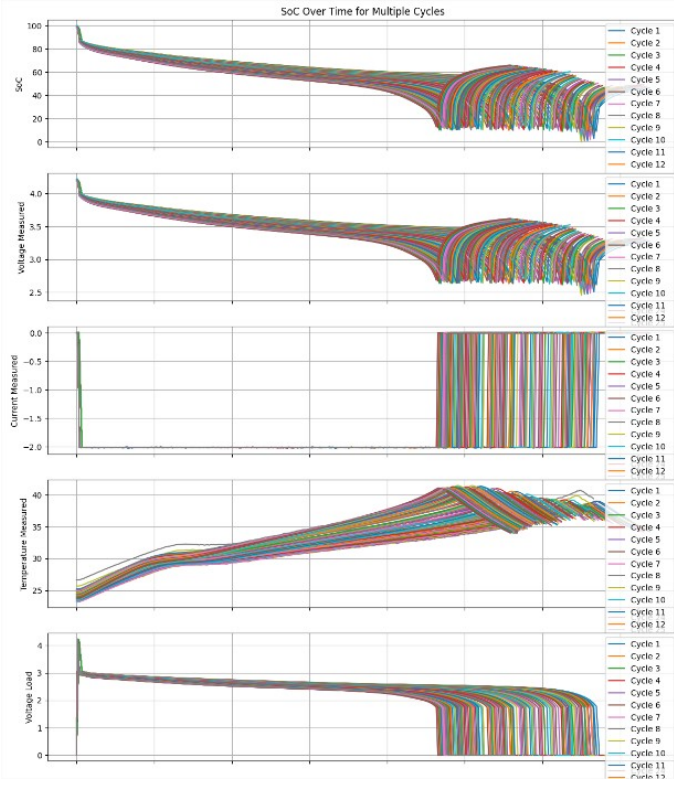


Fig. 4. Her döngüde batarya parametrelerinin zamanla nasıl değiştiğini gösteren grafikler. Bu görseller, SoC, ölçülen voltaj, ölçülen akım, ölçülen sıcaklık, yük altındaki voltaj ve yük altındaki akımı zaman ekseninde göstermektedir.

V. KORELASYON GRAFIKLERİ

3. görselleştirme yönteminde, batarya özellikleri arasındaki ilişkileri gösteren korelasyon grafikleri kullanılmıştır. Korelasyon katsayıları, iki değişken arasındaki ilişkinin gücünü ve yönünü gösterir ve -1 ile 1 arasında değerler alır. Bu katsayıların anlamları şu şekildedir:

- **1:** Mükemmel pozitif korelasyon, iki değişkenin birlikte artış gösterdiğini veya azalış gösterdiğini belirtir.
- **-1:** Mükemmel negatif korelasyon, bir değişken artarken diğerinin azaldığını gösterir.
- **0:** Korelasyon yokluğu, iki değişken arasında herhangi bir ilişki bulunmadığını ifade eder.

Bu grafikler, batarya özelliklerinin birbirleriyle olan ilişkilerini detaylı bir şekilde incelememize olanak tanır. Aşağıda grafikte yer alan bazı önemli ilişkiler özetlenmiştir:

• Güçlü Pozitif Korelasyonlar:

- **Cycle ve Capacity (0.92):** Cycle sayısı arttıkça kapasitenin de arttığını gösterir. Bu, bataryanın kullanılabilir ömrü boyunca kapasitesinin nasıl değiştiğini anlamamıza yardımcı olabilir.
- **SOH ve Capacity (0.88):** Bataryanın sağlığı (SOH) ile kapasite arasında güçlü bir pozitif ilişki vardır. Bu, bataryanın sağlığı arttıkça kapasitesinin de arttığını gösterir.

- **Current Measured ve Current Load (0.75):** Bu iki akım ölçümü arasında güçlü bir pozitif korelasyon vardır. Bu, bataryanın yük altındaki akımının ve ölçülen akımın genellikle aynı yönde hareket ettiğini gösterir.

• Güçlü Negatif Korelasyonlar:

- **Cycle ve SOH (-0.92):** Cycle sayısı arttıkça bataryanın sağlığının (SOH) azaldığını gösterir. Bu, bataryanın kullanım ömrü boyunca yıprandığını ve kapasitesinin azaldığını ifade eder.
- **Voltage Measured ve Current Measured (-0.88):** Ölçülen voltaj ile ölçülen akım arasında güçlü bir negatif ilişki vardır. Bu, batarya deşarj olurken voltajın azalması ve akımın artması nedeniyle beklenen bir durumdur.

• Zayıf veya Anlamsız Korelasyonlar:

- **Ambient Temperature ve Diğer Özellikler:** Ortam sıcaklığı ile diğer özellikler arasında genellikle zayıf veya anlamsız korelasyonlar görülmektedir. Bu, ortam sıcaklığının batarya performansı üzerindeki doğrudan etkisinin sınırlı olduğunu gösterebilir.
- **Voltage Load ve Diğer Özellikler:** Yük voltajı ile diğer özellikler arasında da zayıf korelasyonlar bulunmaktadır. Bu, yük altındaki voltajın diğer ölçümlerle doğrudan bir ilişkisinin olmadığını gösterebilir.

• Orta Düzey Korelasyonlar:

- **Current Load ve Capacity (0.37):** Bu orta düzeyde pozitif korelasyon, bataryanın yük altındaki akımı ile kapasite arasında bir ilişki olduğunu, ancak bu ilişkinin güçlü olmadığını gösterir.
- **Voltage Measured ve Capacity (0.15):** Ölçülen voltaj ile kapasite arasında çok zayıf bir pozitif korelasyon vardır. Bu, voltaj ve kapasite arasındaki ilişkinin minimal olduğunu gösterir.

Genel olarak, korelasyon matrisi bataryaların ömrü ve performansını tahmin ederken hangi özelliklerin önemli olduğunu ve hangi özelliklerin birbiriyle ilişkili olduğunu anlamamıza yardımcı olur. Özellikle cycle sayısı, kapasite ve SOH gibi özellikler, batarya ömrü ve performansını anlamak için kritik öneme sahiptir.

VI. LSTM İLE MODELLEME

Bu bölümde, batarya sağlığı (SOH) tahminleri için kullanılan LSTM (Long Short-Term Memory) modelinin işleyişi ve sonuçları detaylandırılmaktadır. Model, zaman serisi verilerini kullanarak bataryanın SOH değerlerini tahmin etmek amacıyla eğitilmiştir. Aşağıda, modelin nasıl hazırlandığı, eğitildiği ve değerlendirildiği açıklanmıştır.

A. Veri Hazırlığı

Veri setleri, zaman serisi analizi için uygun hale getirilmiştir. Bu süreç, aşağıdaki adımları içermektedir:

- **Zaman Pencerelemesi Oluşturma:** `create_sequences` fonksiyonu kullanılarak belirli bir uzunluktaki

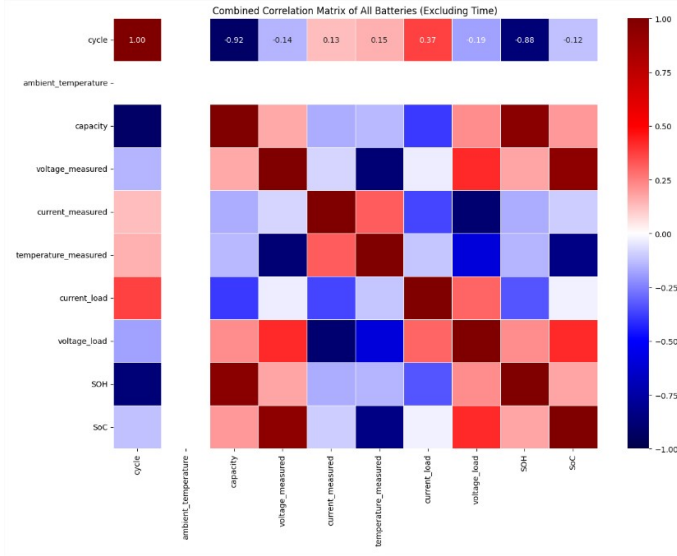


Fig. 5. Korelasyon grafiği.

(seq_length) zaman pencereleri oluşturulmuştur. Bu fonksiyon, girdi özelliklerinden oluşan sekanslar ve hedef değişken için karşılık gelen değerler üretmiştir.

- **Özelliklerin ve Hedef Değişkenin Belirlenmesi:** Modelde kullanılan özellikler şunlardır: cycle, ambient_temperature, voltage_measured, current_measured, temperature_measured, current_load, voltage_load, SoC. Hedef değişken ise SOH (State of Health) olarak belirlenmiştir.

B. Model Eğitimi

LSTM modelinin eğitimi, üç farklı veri seti üzerinde gerçekleştirilmiştir. Eğitim süreci şu adımları kapsamaktadır:

- **Veri Ön İşleme:** Girdi ve hedef değişkenler MinMaxScaler kullanılarak ölçeklendirilmiş ve zaman pencerelerine dönüştürülmüştür.
- **LSTM Modeli:** LSTM katmanı ve bir Dense katmanından oluşan bir model oluşturulmuştur. Model, adam optimizasyon algoritması ve MSE (Mean Squared Error) kayıp fonksiyonu kullanılarak eğitilmiştir. Eğitim sırasında kayıp değerlerinin azaldığı gözlemlenmiştir, bu da modelin başarılı bir şekilde öğrendiğini göstermektedir.

C. Model Tahminleri ve Birleştirme

Modelin tahmin aşaması aşağıdaki adımları içermektedir:

- **Test Verisi:** Test verisi özellikleri aynı şekilde ölçeklendirilmiş ve zaman pencerelerine dönüştürülmüştür.
- **Model Tahminleri:** Her model, test verisi üzerinde tahminlerde bulunmuştur. Üç modelin tahminleri ağırlıklı ortalama yöntemiyle birleştirilmiştir. Ağırlıklar: İlk iki model için 0.375, üçüncü model için 0.25 datasetlerin oranına göre böyle bir ağırlıklandırma seçtim

- **Blending Modeli:** Ridge regresyon modeli, üç modelin tahminlerini kullanarak nihai tahminler yapmıştır.

D. Sonuçlar ve Değerlendirme

- **Tahminler:** Blending modeli tarafından yapılan tahminler ve gerçek SOH değerleri karşılaştırılmıştır. Eşik değeri (0.75) üzerinde ilk tahmin edilen ve gerçek cycle numaraları belirlenmiştir.
- **Gerçek ve Tahmin Edilen Cycle:** Gerçek RUL Cycle: 160, Tahmin RUL Cycle: 153.
- **Model Performansı:** Mean Squared Error (MSE): 5.124117147158386e-05, R-squared (R^2): 0.992144136227721.

E. Grafiklerin Yorumlanması

Aşağıdaki grafik, tahmin edilen ve gerçek SOH değerlerini karşılaştırmaktadır. Kırmızı çizgiler tahmin edilen değerleri, mavi çizgiler ise gerçek değerleri göstermektedir.

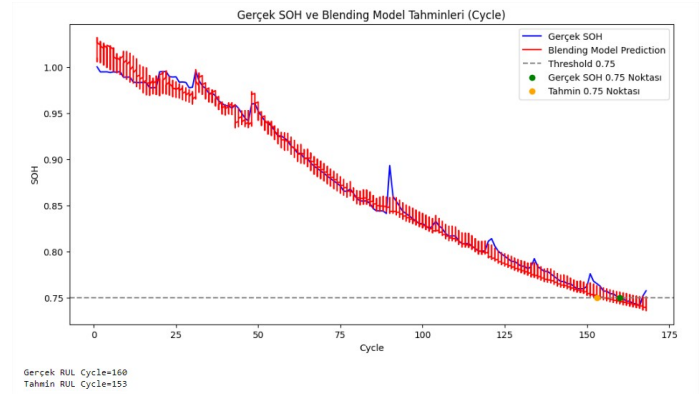


Fig. 6. Tahmin Edilen ve Gerçek SOH Değerlerinin Karşılaştırması. Kırmızı çizgiler tahmin edilen değerleri, mavi çizgiler ise gerçek değerleri göstermektedir.

F. Genel Yorum

Model, çok yüksek bir R^2 skoruna sahip (0.992), bu da modelin gerçek SOH değerlerini oldukça iyi tahmin ettiğini göstermektedir. Düşük MSE değeri (5.124e-05) tahminlerin gerçek değerlere oldukça yakın olduğunu ifade etmektedir. Gerçek ve tahmin edilen cycle numaraları arasındaki farkın sadece 7 olması, modelin bataryanın ömrünü tahmin etmede yüzde 95 başarılı olduğunu göstermektedir. Blending yöntemi, birden fazla modelin tahminlerini birleştirerek daha sağlam ve doğru sonuçlar elde etmeyi sağlar.

Sonuç olarak, bu model batarya sağlığını (SOH) başarılı bir şekilde tahmin etmekte ve bataryanın ömrü hakkında güvenilir öngörülerde bulunmaktadır. Modelin düşük MSE ve yüksek R^2 değerleri, uygulamanın başarısını ortaya koymaktadır ve batarya yönetimi ile bakım stratejilerinde önemli bir rol oynayabilir.

VII. SONUÇ

Bu çalışmada, NASA'nın lityum-iyon pil veri setleri kullanılarak bataryaların kullanılabilir ömürleri tahmin edilmiştir. SOH ve SOC hesaplamaları, veri ön işleme, özellik çıkarımı ve görselleştirme adımları detaylı olarak açıklanmıştır. LSTM modeli, batarya sağlığını (SOH) tahmin etmekte yüksek doğruluk göstermiştir. Modelin düşük MSE ve yüksek R^2 değerleri, bataryaların kullanım ömrünü tahmin etmede güvenilir sonuçlar verdiğini ortaya koymaktadır. Bu nedenle, gelecekteki çalışmalar için LSTM tabanlı modellemelerin daha da geliştirilmesi ve farklı veri setleri üzerinde test edilmesi önerilmektedir.

REFERENCES

- [1] Mullapudi, B., & Shanmugasundaram, R. (2022). Unified Approach for Estimation of State of Charge and Remaining Useful Life of Li-Ion batteries using Deep Learning. In *2022 IEEE 19th India Council International Conference (INDICON)*. IEEE.
- [2] Chaturvedi, K. M., et al. (2023). Theoretical Comparison and Machine Learning Based Predictions on Li-Ion Battery's Health Using NASA-Battery Dataset. In *The International Conference on Metallurgical Engineering and Centenary Celebration*. Singapore: Springer Nature Singapore, pp. 108-118.
- [3] Sun, H., et al. (2022). Prediction of Li-ion battery state of health based on data-driven algorithm. *Energy Reports*, 8, 442-449.
- [4] Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. *Neural Computation*, 9(8), 1735-1780. This paper lays the foundation for Long Short-Term Memory (LSTM) networks and details their ability to learn long-term dependencies in time series data. It provides comprehensive information on the theory and applications of LSTM networks. <https://www.bioinf.jku.at/publications/older/2604.pdf>
- [5] Zhang, X., & Yang, H. (2019). *A Review of State-of-Charge and State-of-Health Estimation Techniques for Lithium-Ion Batteries*. *Journal of Power Sources*, 435, 226823. This study provides a comprehensive review of methods for estimating State-of-Health (SOH) and State-of-Charge (SoC) for lithium-ion batteries. It offers detailed information on techniques and methods used in battery health prediction. <https://doi.org/10.1016/j.jpowsour.2019.226823>
- [6] Cheng, C., & Ding, J. (2020). *Time Series Forecasting with Long Short-Term Memory Networks: A Review*. *Computers, Materials & Continua*, 66(2), 1547-1568. This paper reviews recent research and methods for time series forecasting using Long Short-Term Memory (LSTM) networks. It offers an extensive literature review on LSTM modeling and time series prediction. <https://doi.org/10.32604/cmc.2020.011749>