

DENTELNet: Teknik olmayan elektrik kayıplarının tespiti için yeni bir hibrit derin öğrenme yöntemi

Hamdullah KARAMOLLAOĞLU*

Elektrik Üretim Anonim Şirketi, Kumburgaz Eğitim Müdürlüğü, İstanbul, Türkiye

Geliş Tarihi (Received Date): 27.01.2025

Kabul Tarihi (Accepted Date): 10.08.2025

Öz

Elektrik enerjisi, modern toplumların temel ihtiyaçlarından biri olup, ekonomik büyüme ve toplumsal refah açısından kritik bir rol oynamaktadır. Ancak enerji üretiminden tüketime kadar geçen süreçte meydana gelen kayıplar, kaynakların verimsiz kullanımına ve ekonomik zararlara yol açmaktadır. Teknik olmayan kayıplar, özellikle usulsüz elektrik kullanımı, sayaç manipülasyonu ve kaçak bağlantılar gibi nedenlerle enerji sektörünün karşılaştığı en büyük zorluklardan biridir. Bu çalışmada, teknik olmayan kayıpların tespiti amacıyla yenilikçi bir hibrit derin öğrenme modeli olan DENTELNet önerilmiştir. Model, Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN), İki Yönlü Uzun Kısa Vadeli Bellek (Bidirectional Long Short-Term Memory – BiLSTM) ve Sıvı Sinir Ağı (Liquid Neural Network – LNN) bileşenlerinden oluşmaktadır. Öznitelik indirgeme amacıyla Zaman Serisi Ayırımı tabanlı Temel Bileşen Analizi yöntemi kullanılmış ve veri dengesizliği, Sınırdaki Sentetik Azınlık Aşırı Örneklemeye Tekniği (Borderline SMOTE – BSMOTE) ve Tomek Bağlantısı Yöntemi (Tomek Link – TLink) yaklaşımı ile giderilmiştir. Yapılan deneysel çalışmalar sonucunda, model %92.153 doğruluk, %90.402 kesinlik, %92.153 duyarlılık ve %90.397 F1-skoru elde etmiştir. ROC eğrisi analizi, yüksek AUC değerleri ile modelin ayırt edici gücünü ve dengeli performansını göstermektedir.

Anahtar kelimeler: Teknik olmayan kayıplar, elektrik sahtekârlığı tespiti, derin öğrenme, öznitelik indirgeme, veri dengeleme.

DENTELNet: A novel hybrid deep learning method for detecting non-technical electricity losses

Abstract

Electricity is one of the essential needs of modern societies, playing a critical role in economic growth and social welfare. However, losses occurring throughout the energy

*Hamdullah KARAMOLLAOĞLU, h.karamollaoglu@euas.gov.tr, <http://orcid.org/0000-0001-6419-2249>

production and consumption process lead to inefficient use of resources and economic damages. Non-technical losses, particularly due to unauthorized electricity consumption, meter tampering, and illegal connections, are among the most significant challenges faced by the energy sector. In this study, an innovative hybrid deep learning model named DENTELNet is proposed for detecting non-technical losses. The model consists of Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Liquid Neural Networks (LNN). To achieve feature reduction, a Time-Series Decomposition-based Principal Component Analysis method is used, and data imbalance is addressed using the Borderline Synthetic Minority Over-sampling Technique (BSMOTE) and Tomek Link (TLink) approach. Experimental results demonstrate an accuracy of 92.153%, precision of 90.402%, recall of 92.153%, and an F1-score of 90.397%. The ROC curve analysis, with high AUC values, reveals the model's discriminative power and balanced performance.

Keywords: *Non-technical losses, electricity fraud detection, deep learning, feature reduction, data balancing.*

1. Giriş

Elektrik enerjisi, modern toplumların temel ihtiyaçlarından biri olup günlük yaşamın hemen her alanında hayati bir rol oynamaktadır. Bu nedenle, elektrik enerjisinin güvenilir ve kesintisiz bir şekilde sağlanması, ekonomik büyüme ve toplumsal refah açısından kritik bir öneme sahiptir. Ancak, üretimden nihai tüketiciye ulaşana kadar geçen süreçte enerji kayıpları meydana gelmekte, bu durum hem enerji kaynaklarının verimsiz kullanılmasına hem de ciddi ekonomik kayıplara yol açmaktadır. Özellikle gelişmekte olan ülkelerde kayıp-kaçak oranlarının yüksekliği, enerji sektörünün çözmesi gereken önemli bir sorun olarak öne çıkmaktadır. Elektrik enerjisi kayıpları genel olarak teknik kayıplar ve teknik olmayan kayıplar olmak üzere iki ana başlıkta incelenmektedir. Teknik kayıplar, iletim hatları, transformatörler ve diğer şebeke bileşenlerinde yaşanan enerji kayıplarını ifade etmektedir. Buna karşın, teknik olmayan kayıplar, genellikle elektrik enerjisinin usulsüz kullanımı, sayaçlara müdahale, kaçak bağlantılar ve faturalandırma hataları gibi nedenlerden kaynaklanmaktadır.

Teknik olmayan kayıplar, enerji sektöründe çözülmesi daha karmaşık ve zorlu bir sorun olarak dikkat çekmektedir. Bu tür kayıplar çoğunlukla kasıtlı davranışların bir sonucu olduğundan, tespit edilmesi ve önlenmesi geleneksel yöntemlerle oldukça güçtür. Yaygın olarak görülen yöntemler arasında sayaçlara müdahale ederek endeks değerlerinin değiştirilmesi, sayaç çarpanlarının manipüle edilmesi ve kaçak bağlantılar yoluyla sayaçsız elektrik kullanımı yer almaktadır. Bu nedenle, teknik olmayan kayıpların önlenmesine yönelik daha yenilikçi ve ileri teknolojilerin geliştirilmesi büyük önem arz etmektedir. Son yıllarda, makine öğrenimi ve derin öğrenme yöntemleri tabanlı teknolojiler, enerji sektöründe teknik olmayan kayıpların tespiti ve önlenmesinde güçlü bir araç olarak öne çıkmaktadır. Bu teknolojiler, büyük veri setlerini analiz ederek anormal tüketim davranışlarını tespit etmekte ve kaçak kullanım şüphesi bulunan aboneleri belirlemekte etkili sonuçlar sağlamaktadır. Bunun yanı sıra, bu çözümler enerji şirketlerine daha etkin ve hassas bir kayıp yönetimi sağlamakta ve enerji verimliliğini artırmada kritik bir rol üstlenmektedir. Bu çalışmanın, teknik olmayan elektrik kayıplarının tespitine yönelik literatüre katkıları aşağıdaki gibi özetlenebilir:

- Teknik olmayan elektrik kayıplarının tespitine yönelik olarak, CNN, BiLSTM ve LNN modellerinin entegre edildiği yeni bir hibrit derin öğrenme yöntemi olan DENTELNet modeli geliştirilmiştir.
- Çalışmada kullanılan veri setine ilişkin zaman serisi verilerini analiz etmek ve boyutlarını optimize ederek sınıflandırma başarımını artırmak amacıyla TSD-PCA yöntemi uygulanmıştır.
- Verisetindeki sınıf dengesizliğini gidermek için aşırı örnekleme ve azınlık örnekleme tekniklerinin birleştirildiği, BSMOTE-TLink yöntemi kullanılarak sınıflar arasında veri dengesi sağlanmıştır.
- Önerilen DENTELNet modeli, geleneksel makine öğrenmesi yöntemleri ve literatürdeki çalışmalarla karşılaştırıldığında daha yüksek doğruluk oranı sağladığı görülmüştür.

Bu makalenin geri kalan kısmı şu şekilde yapılandırılmıştır: İkinci bölümde literatür taraması yapılmış; üçüncü bölümde çalışma kapsamında kullanılan materyal ve yöntemler açıklanmıştır. Dördüncü bölümde deneysel çalışmalar ve elde edilen sonuçlar detaylandırılmıştır. Son olarak, beşinci bölümde genel bir değerlendirme ve bulgulara dayalı sonuçlar sunulmuştur.

2. Literatür taraması

Elektrik sahtekârlığı ve teknik olmayan kayıpların tespiti, enerji sektöründe ekonomik kayıpların azaltılması ve şebeke verimliliğinin artırılması açısından kritik bir öneme sahiptir. Bu sorunu çözmeye yönelik literatürde birçok farklı yöntem önerilmiş olup, derin öğrenme ve hibrit yaklaşımlar, son yıllarda bu alanda elde edilen en başarılı sonuçları sunmaktadır. Çeşitli veri setleri ve algoritmalarla yapılan çalışmalar, elektrik sahtekârlığı tespitinde yeni yöntemlerin geliştirilmesine katkı sağlamıştır.

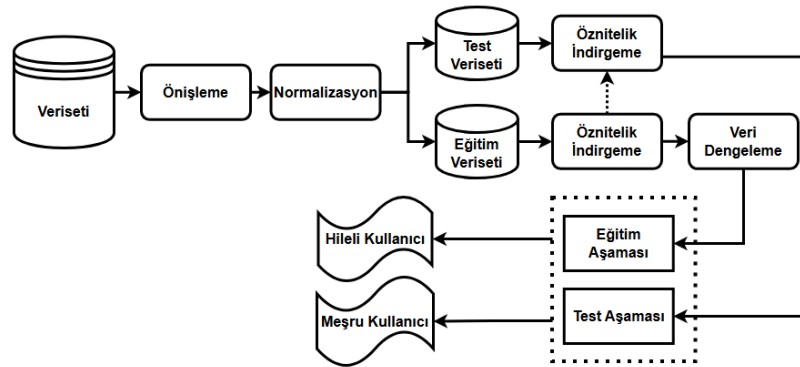
Saqib vd. [1], akıllı şebeke altyapısının bulunmadığı ortamlarda elektrik hırsızlığını tespit etmek amacıyla derin öğrenme tabanlı bir model önermiştir. Model, Çin Devlet Şebeke Şirketi (State Grid Corporation of China - SGCC)'nden alınan 42373 müşterinin tüketim verileriyle geliştirilmiştir. Girdi katmanından sonra üç gizli katman ve bir çıktı katmanına sahip derin öğrenme mimarisi, öznelik indirgeme teknikleri olan PCA, t-dağılımı Tabanlı Yakınsama Algoritması (t-Distributed Stochastic Neighbor Embedding – tSNE), Birleşik Veri Alanı Haritalama (Uniform Manifold Approximation and Projection – UMAP) ile desteklenmiştir. Ayrıca veri dengesizliğini gidermek amacıyla SMOTE, Rastgele Azaltma Örnekleycisi (Random Under Sampler – RUS) ve Rastgele Aşırı Örnekleme Yöntemi (Random Over Sampler – ROS) teknikleri kullanılmıştır. Khan vd. [2] elektrik hırsızlığının büyük mali kayıplara yol açtığını belirterek, akıllı sayaçlardan elde edilen verilerle elektrik tüketiminde anormallikleri tespit etmeyi amaçlamıştır. SGCC veri setinde 42.372 tüketici kaydı kullanarak AlexNet tabanlı bir derin öğrenme modeli geliştirilmiştir. Bu model, ön işleme teknikleri ve veri dengesizliğine karşı çeşitli stratejilerle desteklenmiştir. Souza vd. [3] ise İrlanda Enerji Düzenleme Komisyonu'ndan alınan elektrik tüketim verilerini kullanarak, teknik olmayan kayıpları tespit etmek amacıyla Çok Katmanlı Algılayıcı (Multi-Layer Perceptron – MLP) ve LSTM ağlarını kullanmış ve model performansını iyileştirmek için ek özellikler eklemiştir.

Öznitelik seçimi ve veri dengesizliği sorunları, elektrik sahtekârlığı tespitinin etkinliği üzerinde önemli bir etkiye sahiptir. Bu nedenle, meta-sezgisel yöntemlerin kullanımı da literatürde öne çıkan bir yaklaşımdır. Shehzad vd. [4] Pakistan'daki PRECON enerji bilgi grubu ile SGCC veri setini kullanarak, Genetik Algoritma ve Yapay Arı Kolonisi gibi meta-sezgisel yöntemlerle öznitelik seçimi gerçekleştirmiştir. Denoising autoencoder yöntemiyle yüksek varyansa sahip öznitelikler çıkarılmış ve sınıflandırma işlemi Destek Vektör Makinesi algoritması ile yapılmıştır. Tripathi vd. [5] ise Tunus'ta Société Tunisienne de l'Électricité et du Gaz (STEG) adlı kuruluştan elde edilen veri seti üzerinde, Gri Kurt Optimizasyonu ve Uyarlanabilir Sentetik Azınlık Aşırı Örneklemme Tekniği kullanarak öznitelik seçimi ve veri dengeleme gerçekleştirilmiş ardından Rastgele Orman algoritması ile sınıflandırma gerçekleştirmiştir. Shehzad vd. [6] bir diğer çalışmasında ise elektrik sahtekârlığı tespitine yönelik olarak iki aşamalı bir meta-öğrenici model geliştirilmiştir. İlk aşamada, makine öğrenmesi sınıflandırıcıları, verilerden bilgi çıkarmaktadır. İkinci aşamada ise MLP tabanlı meta-öğrenici, farklı sınıflandırıcılardan elde edilen tahminler arasındaki doğrusal olmayan ilişkileri öğrenerek daha doğru bir sınıflandırma gerçekleştirmiştir.

Hibrit model yaklaşımları, elektrik sahtekârlığı tespitinde yüksek performans gösteren diğer bir araştırma alanıdır. Nawaz vd. [7], CNN ve Aşırı Gradient Artırma yöntemlerini birleştirerek hibrit bir model önermiştir. Bu model, SGCC veri seti üzerinde test edilmiştir. Fei vd. [8] ise teknik olmayan kayıpları tespit etmek için ND-CP adı verdikleri bir model geliştirmiştir. Bu model, CNN ve Kapalı Tekrarlayan Birimler (Gated Recurrent Units – GRU) yöntemlerinin birleşiminden oluşmakta olup, modelin performansı SGCC veri seti üzerinde test edilmiştir.

3. Materyal ve metot

Çalışmada önerilen DENTELNet modeline ilişkin akış diyagramı Şekil 1'de görüldüğü gibidir.



Şekil 1. Çalışmada önerilen DENTELNet modeline ilişkin akış diyagramı.

Şekil 1'de görüldüğü gibi çalışmada önerilen model, Önişleme, normalizasyon, öznitelik indirgeme, veri dengeleme ve sınıflandırma aşamalarından oluşmaktadır.

3.1. Veri Seti

Bu çalışmada kullanılan veri seti, SGCC tarafından derlenen 1035 günlük gerçek elektrik tüketim verilerini içermekte olup meşru elektrik kullanımı ve hileli kullanım olmak üzere

iki ana sınıftan oluşmaktadır [9]. SGCC veri setinde yer alan hileli kullanıcı sınıfı, elektrik enerjisinin usulsüz yollarla tüketilmesine yönelik çeşitli yasa dışı eylemleri kapsamaktadır. Bu eylemler; sayaç devre dışı bırakma, fiziksel müdahaleler, kaçak hat çekme gibi doğrudan fiziksel manipülasyonları; akıllı sayaç yazılımının değiştirilmesi, manyetik alan kullanımı ve frekans/voltaj değerlerinin değiştirilmesi gibi teknolojik müdahaleleri içermektedir [10].

SGCC veri setinde, kullanıcıların %91.6'sını oluşturan 38757 meşru kullanıcı ve %8.4'ünü oluşturan 3615 hileli kullanıcı bulunmaktadır. Bu sınıf dağılımı, veri setindeki sınıflar arasında belirgin bir dengesizlik olduğunu ortaya koymaktadır.

3.2. Veri Önışleme

Veri önışleme aşamasında öncelikle veri kümesinde yer alan sabit sütunlar ve sabit satırlar tespit edilerek analiz dışı bırakılmıştır. Özellikle 10.03.2014 tarihli sütun, %99.96 oranında eksik veri içermesi nedeniyle çıkarılmıştır. Ayrıca, 2081 adet sabit satır da analiz dışı bırakılmış ve sadece bilgi taşıyan 40291 satır kullanıma alınmıştır. Eksik veri problemi, zaman serisi verileri için sık rastlanan bir durumdur. Bu çalışmada eksik veriler, lineer interpolasyon yöntemiyle hem ileri hem de geri yönlü doldurulmuştur. Bu yöntem, eksik veri noktalarının, etrafındaki bilinen değerlere göre tahmin edilmesine olanak tanımaktadır. İşlem, aşağıdaki Eşitlik 1'de tanımlanmıştır:

$$y(x) = y_1 + \frac{x-x_1}{x_2-x_1}(y_2 - y_1) \quad (1)$$

Burada, x_1 ve x_2 , bilinen veri noktalarının x -değerlerini; y_1 ve y_2 ise bu noktalarla ilişkili y -değerlerini temsil etmektedir. x , eksik verinin bulunduğu noktayı belirtirken, $y(x)$ ise bu eksik verinin lineer interpolasyon yöntemiyle tahmin edilen değeridir. Eksik verilerin tahmini, etrafındaki bilinen veri noktalarına dayanarak hesaplanmakta ve bu işlem hem ileri hem de geri yönlü interpolasyon ile gerçekleştirilebilmektedir.

3.4. Kodlama ve Veri Bölme

Önışleme aşamasının devamında, kategorik değişkenler etiket kodlama yöntemiyle sayısal formata dönüştürülmüştür. Ardından, veri seti tabakalı örnekleme yöntemiyle %80 eğitim ve %20 test alt veri setlerine bölünmüştür. Bu sayede sınıflar arası dengenin hem eğitim hem de test veri setlerinde korunması sağlanmıştır. Eğitim veri setinde 29342 meşru ve 2890 hileli kullanıcı örneği bulunurken, test veri seti 7337 meşru ve 722 hileli kullanıcı örneği içermektedir.

3.5. Normalizasyon

Makine öğrenmesi algoritmalarının daha hızlı ve etkili öğrenbilmesi için tüm özellikler z-puanı standardizasyonu yöntemiyle normalize edilmiştir. Bu yöntem, her özelliğin sıfır ortalama ve birim varyansa sahip olmasını sağlamaktadır. Normalizasyon işlemi, Eşitlik 2'de gösterilmiştir:

$$\hat{x} = \frac{x - \text{ortalama}(x)}{\text{std}(x)} \quad (2)$$

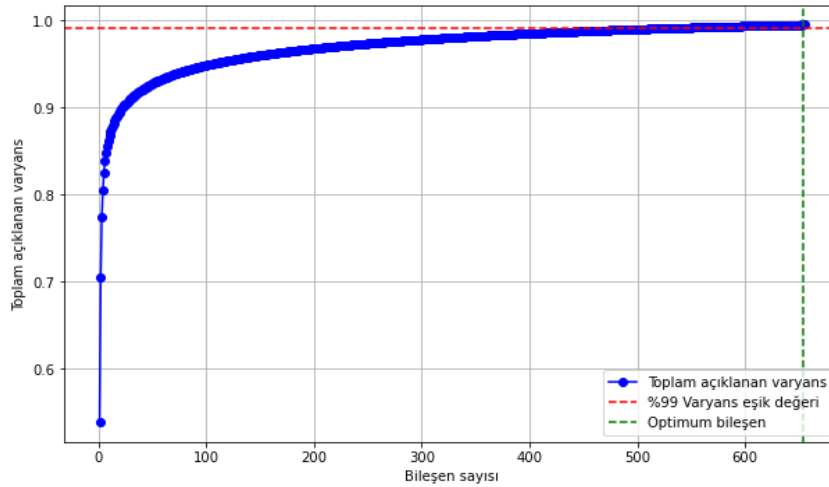
Burada, x , orijinal veri değerini, $\text{ortalama}(x)$, verinin ortalama değerini, $\text{std}(x)$, verinin standart sapma değerini, \hat{x} ise standartlaştırılmış (normalize edilmiş) veriyi temsil etmektedir.

3.6. Öznitelik İndirgeme

Veri kümesinin yüksek boyutluluğu, öğrenme algoritmalarının performansını olumsuz etkileyebilir. Bu nedenle, zaman serisi verilerine özel bir boyut indirgeme yöntemi olan TSD-PCA kullanılmıştır [11, 12].

TSD-PCA yöntemi, iki ana aşamadan oluşmaktadır. Bu yöntemle ilk aşamada, her bir veri özelliği için eğilim (trend), sezonluk (seasonal) ve artık (residual) bileşenler, statsmodels kütüphanesinin time series analysis (tsa) modülünde yer alan seasonal_decompose fonksiyonu kullanılarak ayrıştırılmıştır. Bu fonksiyon, zaman serisi verilerini eğilim, sezonluk ve artık bileşenlerine ayırmak amacıyla yaygın olarak kullanılmaktadır. Ayrıştırma işlemi, verinin yapısına uygun olarak additif model varsayımıyla gerçekleştirilmiştir. Sezonluk bileşenlerin periyodik yapısını belirlemek için periyot uzunluğu 20 olarak ayarlanmıştır. Eğilim bileşeni, zaman serisinin uzun vadeli genel hareketini ifade etmekte ve serinin ana eğilimlerini ortaya çıkarmaktadır. Sezonluk bileşen, verideki periyodik veya döngüsel desenleri yakalamakta ve serinin tekrar eden düzenlerini analiz etmek amacıyla kullanılmaktadır. Artık bileşen ise, eğilim ve sezonluk bileşenlerden arta kalan, rastgele dalgalanmaları veya gürültüyü temsil etmekte ve modelde açıklanamayan değişkenlikleri içermektedir [13]. Eksik veri olması durumunda, eğilim bileşeni doğrusal interpolasyon yöntemiyle tahmin edilerek tamamlanmıştır. Her özelliğin zaman serisi dinamiklerini detaylı bir şekilde analiz etmek amacıyla seasonal_decompose fonksiyonu kullanılmıştır. Bu ayrıştırma süreci, verilerin temel bileşenlerini analiz etmek ve bu bileşenlerin daha sonraki aşamalarda boyut indirgeme ve modelleme işlemlerine yönelik hazırlanmasını sağlamak için önemlidir.

İkinci aşamada, ayrıştırılmış bileşenler birleştirilmiş ve boyut indirgemesi gerçekleştirilmiştir. İlk aşamada gerçekleştirilen ayrıştırma işlemi sonucunda her bir özelliğe karşılık elde edilen bileşenler yatay bir şekilde birleştirilerek yeni bir veri matrisi oluşturulmuştur. Bu veri matrisi, orijinal özellik sayısının üç katı boyutunda bir matris olarak elde edilmiştir. Daha sonra, bu matris üzerinde PCA yöntemi uygulanmıştır. PCA işleminde, verinin toplam varyansının %99'unu açıklayacak şekilde bileşen seçimi yapılmıştır. PCA için scikit-learn kütüphanesindeki PCA sınıfı kullanılmış ve hesaplama yöntemi olarak svd_solver='full' parametresi tercih edilmiştir. PCA bileşenlerinin açıklanan varyans oranları Şekil 2'de görüldüğü gibidir.



Şekil 2. PCA bileşenlerinin açıklanan varyans oranı.

Şekil 2’de görüldüğü gibi PCA uygulanması sonucunda, orijinal 1034 öznitelik 655 özniteliğe indirgenmiştir. Bu boyut indirgeme işlemi sırasında toplam varyansın %99’unu açıklayan bir öznitelik alt kümesi elde edilmiştir. Bu durum, veri setindeki önemli bilgilerin büyük bir kısmının korunduğunu göstermektedir.

3.7. Veri Dengeleme

Bu çalışmada, sınıf dengesizliğinin model performansı üzerindeki olumsuz etkilerini azaltmak amacıyla, eğitim veri seti üzerinde veri dengeleme işlemleri uygulanmıştır. Bu kapsamda, iki aşamalı bir yaklaşım benimsenmiş ve BSMOTE ile TLink algoritmalarından yararlanılmıştır. Her iki yöntem, sınıf dengesini sağlamakla birlikte farklı stratejilere dayanmaktadır. İlk aşamada, azınlık sınıfa ait örnek sayısını artırmak için BSMOTE algoritması kullanılmış ve bu sınıfa yönelik sentetik örnekler üretilmiştir. Ancak BSMOTE sonrası, özellikle sınıf sınırlarında bulunan ve modele kararsızlık yaratabilecek örneklerin veri setinde kalma ihtimali bulunmaktadır. Bu durumu gidermek amacıyla, ikinci aşamada TLink algoritması uygulanmış; farklı sınıflara ait olup birbirine çok yakın olan, sınıf sınırında bulunan örnekler veri setinden çıkarılmıştır. Bu iki aşamalı dengeleme yaklaşımı sayesinde hem veri setindeki dengesizlik azaltılmış hem de sınıf sınırlarının daha belirgin hale getirilmesi sağlanarak modelin genel sınıflandırma performansına olumlu katkıda bulunulmuştur.

BSMOTE, azınlık sınıfına ait örneklerin yapay olarak çoğaltılmasında, yalnızca karar sınırına yakın olan (yani sınıflar arasındaki sınırdaki yer alan) örneklerin çoğaltılmasına odaklanarak, sınıf dengesizliğinin minimize edilmesine katkı sağlamaktadır. Bu yaklaşım, özellikle karar sınırlarında yer alan örneklerin daha fazla çoğaltılmasını sağlayarak, modelin daha iyi bir genelleme yapmasına olanak tanımaktadır. Yeni örnekler Eşitlik 3’e göre oluşturulmaktadır.

$$x_{new} = x_i + \lambda(x_j - x_i) \quad (3)$$

Burada, x_{new} yeni yapay örneği, x_i mevcut azınlık sınıf örneğini, x_j , x_i ’nin k en yakın komşularından birini, λ ise [0,1] aralığında rastgele bir değeri ifade etmektedir [14, 15].

TLink yöntemi ise, çoğunluk sınıfındaki hatalı ve gereksiz örnekleri tespit edip çıkararak, veri setindeki gürültüyü azaltmakta ve sınıflar arasındaki ayrımın daha belirgin hale gelmesini sağlamaktadır. TLink, her iki sınıftan örnekler arasında minimum mesafeyi hesaplayarak, her bir örnek için en yakın komşusuyla olan mesafeyi değerlendirmektedir. İki örnek arasındaki mesafe Eşitlik 4’teki gibi hesaplanmaktadır.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2} \quad (4)$$

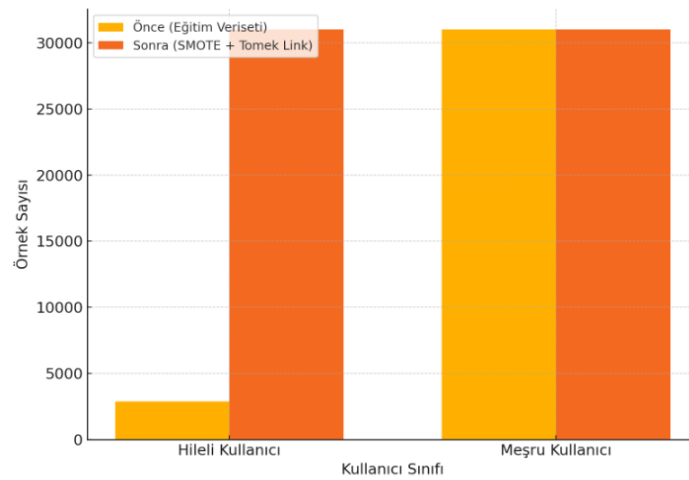
Burada x_i ve x_j iki örneği temsil etmekte $x_{i,k}$ ve $x_{j,k}$ ise x_i ve x_j ’nin k. özelliğindeki değerlerini ifade etmektedir. Eğer $d(x_i, x_j)$ mesafesi çok küçükse ve her iki örnek farklı sınıflara aitse, bu örnekler bir “TLink” olarak kabul edilmekte ve çoğunluk sınıfındaki örnek çıkarılmaktadır [16, 17].

Çalışmada, veri dengeleme işlemi, sınıf dengesizliğinin modelin öğrenme süreci üzerindeki olumsuz etkilerini azaltmak amacıyla yalnızca eğitim veri seti üzerinde uygulanmıştır. Sınıf dağılımındaki orantısızlık, modelin baskın sınıfa yönelme eğilimini artırmakta ve azınlık sınıfına ait örüntülerin yeterli düzeyde öğrenilememesine neden

olabilmektedir [18]. Bu nedenle, eğitim sürecinde uygulanan dengeleme yöntemleri ile sınıflar arası örnek sayıları eşitlenmiş; böylece modelin her iki sınıfı da dengeli biçimde öğrenmesi ve karar sınırlarının daha sağlıklı bir şekilde oluşturulması hedeflenmiştir.

Diğer yandan, modelin doğrulama ve test aşamalarında gerçekçi bir performans değerlendirmesi yapılabilmesi için test veri setine herhangi bir dengeleme işlemi uygulanmamıştır. Bu yaklaşım, test veri setinin yapay müdahalelerden arındırılmış şekilde, doğal sınıf dağılımını korumasını ve gerçek dünya koşullarını yansıtacak biçimde bağımsız kalmasını sağlamaktadır.

Veri dengeleme öncesi ve sonrası için eğitim veri setine ait örnek sayıları Şekil 3'te görüldüğü gibidir.



Şekil 3. Veri dengeleme öncesi ve sonrası sınıflara ait örnek sayıları dağılımı.

Şekil 3'te görüldüğü gibi eğitim veri setine veri dengeleme yöntemlerin uygulanmasının ardından, eğitim veri seti için her iki sınıfın örnek sayısı 31005 olarak eşitlenmiştir. Bu dengeleme süreci, modelin öğrenme sürecinde sınıf dengesizliğinden kaynaklanabilecek olumsuz etkilerin minimize edilmesine ve modelin genelleme yeteneğinin artırılmasına katkı sağlamaktadır.

Çalışmada önerilen DENTELNet modeli, farklı türden derin öğrenme katmanlarını bir araya getirerek hem uzamsal hem de zamansal bağımlılıkları etkin şekilde öğrenmeyi amaçlamaktadır. DENTELNet mimarisinde sırasıyla, iki adet CNN, iki adet BiLSTM katmanı ve bir adet LNN katmanı yer almaktadır. Modelin hiperparametrelerinin optimizasyonu için Rastgele Arama yöntemi tercih edilmiştir. Bu yöntem, geniş ve yüksek boyutlu hiperparametre uzaylarında rastgele örnekler seçerek modelin performansını değerlendirme esasına dayanmaktadır. Özellikle tüm olası kombinasyonların denenmesinin hesaplama maliyetinin yüksek olduğu durumlarda, Rastgele Arama yöntemi verimli ve etkili bir alternatif olarak öne çıkmaktadır.

3.8. CNN Katmanları

CNN, özellikle görüntü ve zaman serisi verilerinde yaygın olarak kullanılan derin öğrenme modelidir. CNN'de, giriş verisindeki yerel bağımlılıkları ve mekânsal hiyerarşileri öğrenmek için evrişim işlemi kullanılmaktadır. Bir giriş verisi X ve bir filtre F ile evrişim işlemi Eşitlik 5'teki gibi ifadelendirilmektedir.

$$Y(p, q) = \sum_i \sum_j X(p + i, q + j)F(i, j) \quad (5)$$

Burada $Y(p, q)$ özellik haritasının (p, q) noktasındaki değerini, $X(p + i, q + j)$ giriş matrisindeki bir değeri, $F(i, j)$ filtre ağırlıklarını temsil etmektedir. Evrişim katmanları CNN'in temel taşını oluşturmakta ve öğrenilebilir filtreler aracılığıyla belirli özelliklerin çıkarılmasını sağlamaktadır [19, 20]. Bu katmanlar, öğrenilebilir filtreler aracılığıyla giriş verisinden anlamlı öznitelikler çıkartılmasına olanak tanımaktadır.

3.9. BiLSTM Katmanları

BiLSTM, giriş dizisini hem ileri hem de geri yönde işleyerek geçmiş ve gelecek bağlamları aynı anda öğrenmeyi sağlamaktadır. Bu yapı sayesinde, daha kapsamlı bir zaman bağımlılığı modeli elde edilmektedir. BiLSTM katmanının çıktısı Eşitlik 6'da görüldüğü gibi ileri ve geri yönlü çıktılar birleştirilerek hesaplanmaktadır.

$$h_t^{BiLSTM} = [\vec{h}_t, \overleftarrow{h}_t] \quad (6)$$

Burada \vec{h}_t ileri yönlü, \overleftarrow{h}_t ise geri yönlü çıktı durumunu temsil etmektedir [21].

3.10. LNN Katmanı

LNN, zaman serilerine ait dinamik yapıları modellemek üzere tasarlanmış rezervuar tabanlı bir ağ mimarisidir. LNN, zamana bağlı verilerin işlenmesi ve dinamik ilişkilerin modellenmesi için kullanılan rezervuar tabanlı bir sinir ağı mimarisidir. Bu mimaride, girdiler yüksek boyutlu bir uzaya projelenecek doğrusal olmayan ilişkiler analiz edilmektedir. Rezervuar katmanının t anındaki durumu $h(t)$, önceki durum $h(t - 1)$ ve mevcut girdi $x(t)$ yardımıyla Eşitlik 7'deki gibi hesaplanmaktadır.

$$h(t) = (1 - \alpha)h(t - 1) + \alpha \tanh(W_{in}x(t) + W_{res}h(t - 1)) \quad (7)$$

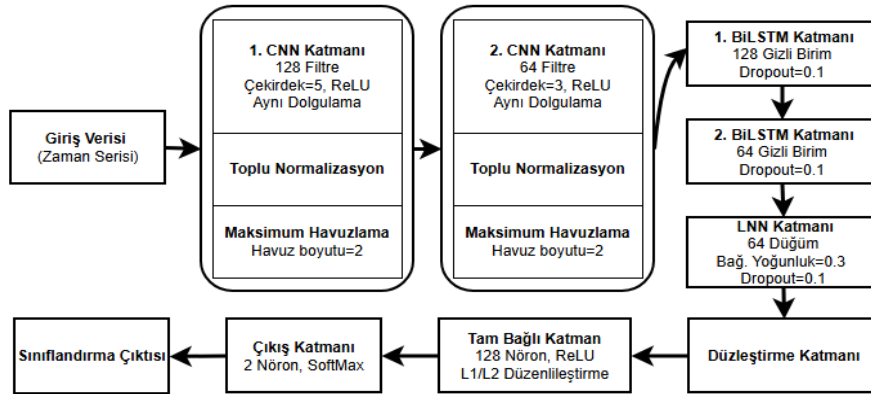
Burada, $h(t)$ rezervuarın t anındaki durumunu, α geçmiş bilgilerin korunmasını kontrol eden sızıntı oranını, W_{in} girdilerin rezervuara aktarılmasını sağlayan ağırlık matrisini, W_{res} rezervuar düğümleri arasındaki bağlantıları temsil eden ağırlık matrisini ifade etmektedir. Rezervuarın bağlantı yoğunluğu, düğümler arasındaki bağlantıların seyrekliğini ifade etmektedir. Rezervuarın bağlantı yoğunluğu Eşitlik 8 yardımıyla hesaplanmaktadır.

$$d = \frac{\text{Mevcut bağlantı sayısı}}{\text{Maksimum olası bağlantı sayısı}} \quad (8)$$

Burada, bağlantı yoğunluğu (d), rezervuar ağındaki mevcut bağlantıların sayısının, ağdaki maksimum olası bağlantı sayısına oranını ifade etmektedir. Bu oran ile rezervuarın ne kadar seyrek ya da yoğun bağlantılı olduğunu nicel bir şekilde değerlendirilebilmektedir [22, 23]. Düşük bağlantı yoğunluğu hesaplama verimliliğini artırırken aşırı öğrenmeyi önlemeye yardımcı olmaktadır.

Çalışmada önerilen DENTELNet modeli, CNN, BiLSTM ve LNN katmanlarının birleşimiyle hem uzamsal hem de zamansal özelliklerin etkili bir şekilde işlenmesini sağlamaktadır. Bu yapı, CNN ile uzamsal, BiLSTM ile zamansal özellikleri çıkarırken, LNN dinamik ilişkileri modeller. Ayrıca, normalizasyon ve düzenleme teknikleri kullanılarak modelin kararlılığı artırılmış, aşırı öğrenme riski azaltılmıştır.

DENTELNet modelinde kullanılan hibrit derin öğrenme mimarisinin genel yapısı, Şekil 4'te ayrıntılı biçimde gösterilmektedir.



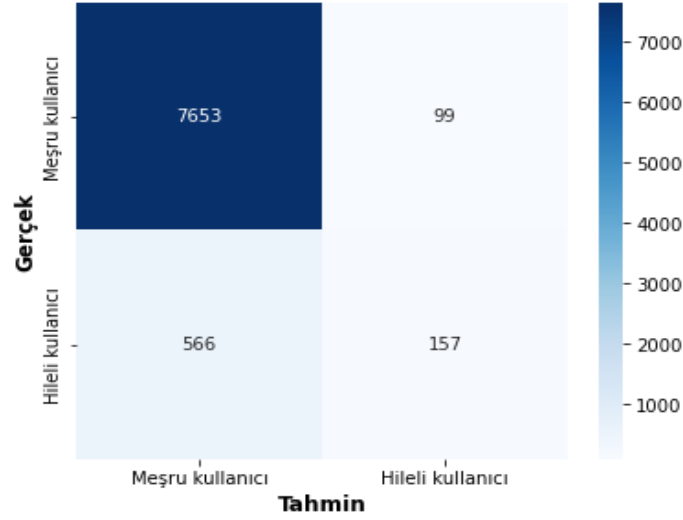
Şekil 4. DENTELNet modeline ait hibrit derin öğrenme mimarisinin genel yapısı.

Şekil 4'te görüldüğü gibi, DENTELNet modeli, zaman serisi verilerini işlemek üzere tasarlanmış hibrit bir derin öğrenme mimarisi olarak sunulmaktadır. Modelde, CNN, BiLSTM ve LNN gibi farklı yapılar bir araya getirilerek her birinin avantajlarından yararlanılmıştır. İlk aşamada, uzamsal özelliklerin çıkarılması için CNN katmanları kullanılmış; geniş ve küçük çekirdek boyutlarıyla hem temel hem de detaylı özelliklerin elde edilmesi sağlanmıştır. Ardından, toplu normalizasyon ve maksimum havuzlama işlemleri uygulanarak model boyutlarının küçültülmesi ve öğrenme sürecinin stabil hale getirilmesi amaçlanmıştır. CNN'in ardından zamansal bağımlılıkların modellenmesi amacıyla BiLSTM katmanlarından yararlanılmış ve aşırı öğrenmenin önlenmesi için dropout tekniği kullanılmıştır. Dinamik örüntülerin yakalanmasında LNN katmanına başvurulmuş ve modelin esnekliği artırılmıştır. Son olarak, çıkarılan özellikler düzeltilerek tam bağlı katmana aktarılmış ve L1/L2 düzenleme yöntemleriyle modelin genelleme performansının güçlendirilmesi hedeflenmiştir. Modelin son katmanında ise, iki sınıflı bir sınıflandırma problemini çözmek için 2 nöronlu ve Softmax aktivasyon fonksiyonu kullanılan bir çıkış katmanı bulunmaktadır. Modelin optimizasyonu için Adam algoritması seçilmiş ve kayıp fonksiyonu olarak “categorical_crossentropy” kullanılmıştır.

4. Deneysel çalışma

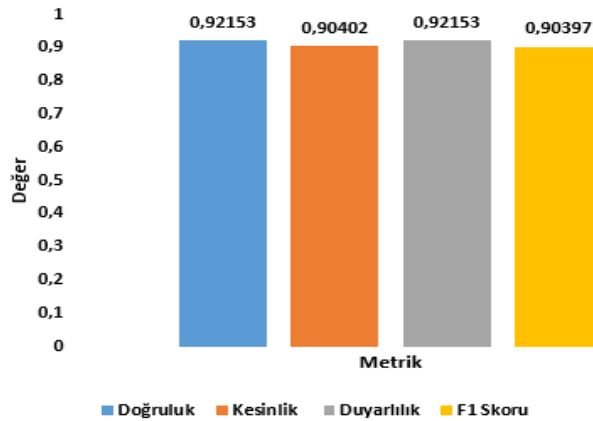
Çalışmanın eğitim aşamasında, modelin genel başarısını artırmak ve aşırı öğrenmeyi engellemek amacıyla çeşitli teknikler uygulanmıştır. Bu teknikler, modelin eğitim sürecinin daha verimli hale getirilmesini ve daha güvenilir sonuçların elde edilmesini sağlamaktadır. Eğitim süreci, 64'lük bir batch size ile 200 epoch boyunca gerçekleştirilmiş ve erken durdurma (early stopping) ile modelin aşırı öğrenme sorunuyla karşılaşması engellenmiştir. Erken durdurma, modelin doğrulama kaybı üzerinde gözle görülür bir iyileşme sağlanmadığı durumlarda eğitim sürecini sonlandırmayı amaçlar. Bu mekanizma, modelin en iyi doğrulama performansını elde ettiği noktada eğitim sürecini sonlandırarak, modelin aşırı öğrenmesini engellemektedir. Patience parametresi, doğrulama kaybında iyileşme gözlemlenmediği sürenin belirlenmesinde kullanılmış, burada 10 epoch'luk bir süre tanınmıştır. Ayrıca, modelin en iyi ağırlıklarını geri yüklemek için restore_best_weights parametresi etkinleştirilmiştir. Bir diğer önemli

teknik ise Öğrenme Oranı Azaltma (ReduceLRonPlateau) olmuştur. Bu teknik, doğrulama kaybında bir iyileşme gözlemlenmediğinde, öğrenme oranını azaltarak modelin daha hassas bir şekilde öğrenmesini sağlamaktadır. Factor parametresi ile öğrenme oranının %50 oranında düşürülmesi sağlanmış, patience parametresi ile doğrulama kaybında iyileşme olmadığı durumdaki epoch sayısı 5 olarak belirlenmiştir. Min_lr parametresi ise öğrenme oranının alt sınırını $1e-6$ olarak belirlemiştir. Bu tekniklerin kombinasyonu, modelin eğitim sürecinde daha güvenilir ve verimli bir öğrenme süreci elde edilmesini sağlamış, aynı zamanda genel performansın artırılmasına yardımcı olmuştur. Test veri seti üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen karmaşıklık matrisi, Şekil 5'te sunulmuştur.



Şekil 5. Karmaşıklık matrisi.

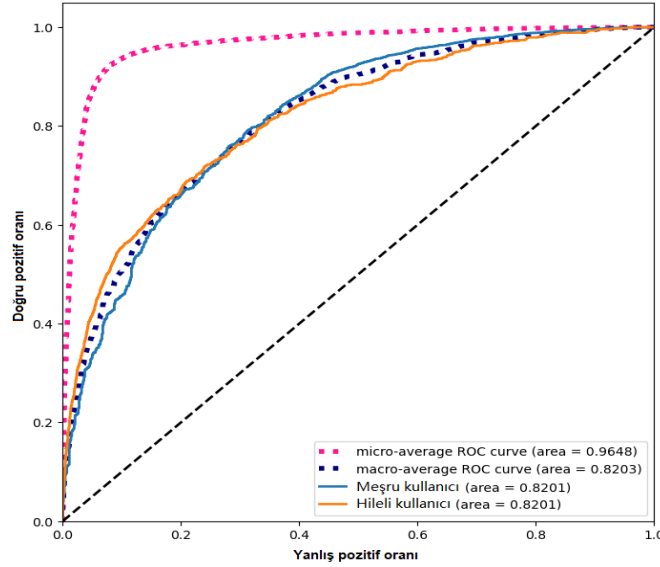
Şekil 5'te sunulan karmaşıklık matrisinden, modelin test veri seti üzerinde meşru ve hileli kullanıcıları başarılı şekilde ayırt edebildiği görülmektedir. Model, meşru kullanıcıları büyük ölçüde doğru sınıflandırmış, hileli kullanıcıların da önemli bir kısmını doğru şekilde tespit etmiştir. Bu matris temelinde hesaplanan performans metrikleri ise Şekil 6'da ayrıntılı olarak sunulmuştur.



Şekil 6. Sınıflandırma performans değerleri.

Şekil 6'da görüldüğü üzere, test veri seti üzerinde gerçekleştirilen sınıflandırma sonucunda, önerilen DENTELNet modeli yüksek doğruluk ve tutarlı sınıflandırma

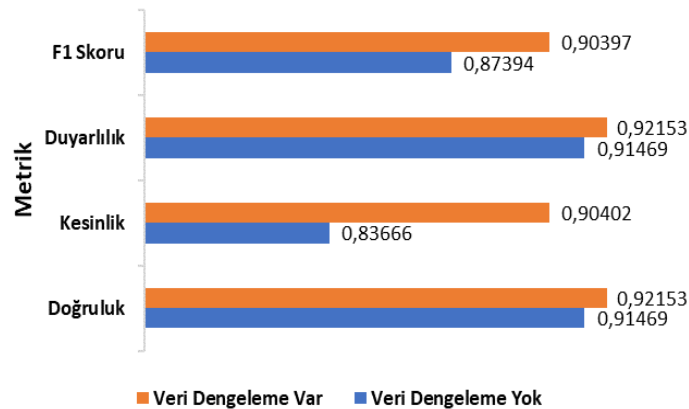
performansı sergilemiştir. Elde edilen ağırlıklı doğruluk, kesinlik, duyarlılık ve F1-skoru değerleri, modelin genel başarısını ortaya koymaktadır. Test işlemi sonucunda elde edilen modelin ROC eğrisi Şekil 7’de sunulmuştur. ROC eğrisi, modelin farklı sınıflandırma eşiklerinde pozitif ve negatif örnekleri ayırt etme başarısını görselleştirirken, eğrinin altındaki alan (AUC) değeri modelin genel ayırt edici gücünü değerlendirmek için kullanılmaktadır.



Şekil 7. ROC eğrisi.

Şekil 7’de görüldüğü üzere, micro-average ROC eğrisi için AUC değeri 0.9648, macro-average ROC eğrisi için AUC değeri 0.8203 olarak hesaplanmıştır. Ayrıca, meşru kullanıcılar ve hileli kullanıcılar için AUC değerleri eşit olup 0.8201 olarak elde edilmiştir. Bu sonuçlar, modelin genel olarak yüksek bir ayırt etme kapasitesine sahip olduğunu ve her iki sınıf için dengeli bir performans sergilediğini ortaya koymaktadır.

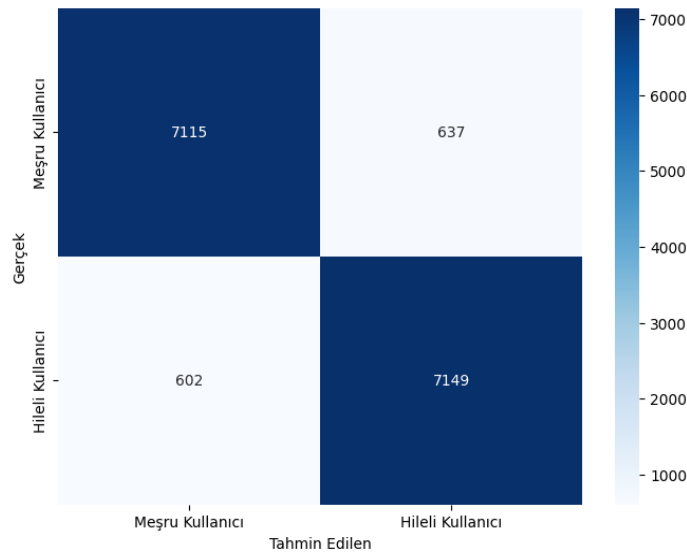
Bu çalışmada ayrıca, veri dengesizliğinin model performansı üzerindeki etkisini incelemek amacıyla, veri dengeleme uygulanmamış veri seti ile BSMOTE-TLink yöntemiyle dengelenmiş veri seti üzerinde model performansı karşılaştırılmıştır. Elde edilen sonuçlar Şekil 8’de sunulmuştur.



Şekil 8. Veri dengelemesinin model performansına etkisi.

Şekil 8’de görüldüğü üzere, veri dengelemesi sonrasında kesinlik değeri 0,83666’dan 0,90402’ye yükseltilmiştir. Bu anlamlı artış, modelin yanlış pozitif tahminlerinin önemli ölçüde azaltıldığını göstermektedir. Özellikle hileli işlemlerin tespitinde yanlış alarm oranının azaldığı tespit edilmiştir. F1 skorundaki yükselişle birlikte, modelin sınıflar arasında daha dengeli bir performans sergilediği belirlenmiştir. Azınlık sınıfı olan hileli kullanıcılar için duyarlılığın korunduğu ve kesinlik değerinin optimize edildiği gözlemlenmiştir. Veri dengelemesi uygulanmadan elde edilen 0,91469’luk doğruluk değerinin, çoğunluk sınıfına yönelik yanlılıktan dolayı yanıltıcı olabileceği değerlendirilmiştir. BSMOTE-TLink yönteminin uygulanmasıyla doğruluk oranı 0,92153’e yükseltilmiş ve modelin genel başarısında iyileşme sağlanmıştır.

Çalışmada ayrıca, veri dengeleme işleminin sınıflandırma performansı üzerindeki etkilerini daha kapsamlı bir şekilde analiz edebilmek amacıyla, alternatif bir veri ayırma yöntemi uygulanmıştır. Bu kapsamda, veri setinin tamamı üzerinde sınıf dağılımı dengelenmiş ve ardından veriler %80 eğitim, %20 test olacak şekilde ayrılmıştır. Böylece hem eğitim hem de test aşamalarında sınıflar arası örnek sayıları eşitlenmiş ve modelin performansı dengeli veri yapısı altında değerlendirilmiştir. Bu yaklaşım, özellikle sınıf dengesizliğinin model başarısını üzerindeki etkisini ortadan kaldırarak, sınıflandırıcının ideal koşullardaki genel performansını gözlemlemeye olanak sağlamaktadır. Test veri seti üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen karmaşıklık matrisi, Şekil 9’da sunulmuştur.



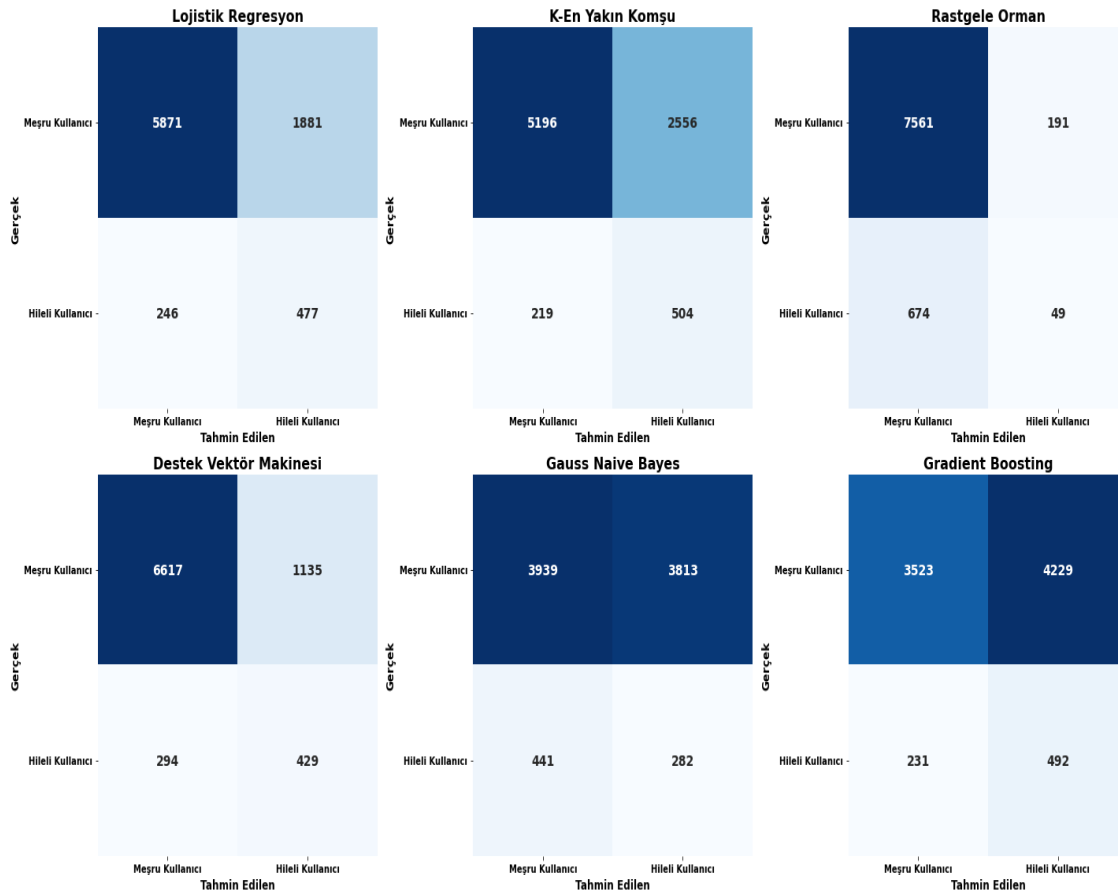
Şekil 9. Karmaşıklık matrisi.

Şekil 9’daki karmaşıklık matrisi incelendiğinde, modelin meşru kullanıcıları 7115 örnekte doğru şekilde sınıflandırdığı, 637 meşru kullanıcıyı ise yanlışlıkla hileli olarak etiketlediği görülmektedir. Benzer şekilde, 7149 hileli kullanıcı doğru şekilde tespit edilmiş, 602 hileli kullanıcı ise yanlışlıkla meşru olarak sınıflandırılmıştır. Bu değerlere göre modelin genel sınıflandırma Tablo 1’de sunulmuştur. Tablo 1’de sunulan sonuçlar incelendiğinde, veri dengeleme işleminin sınıflandırma modelinin performansını önemli ölçüde iyileştirdiği gözlemlenmektedir. Özellikle kesinlik metriğinde belirgin bir artış ($\approx\%10$) kaydedilmiştir. Bu durum, dengeli veri seti kullanıldığında modelin yanlış pozitif (FP) oranını düşürerek, hileli olmayan kullanıcıları daha doğru şekilde tanımlayabildiğini göstermektedir.

Tablo 1. Performans karşılaştırması.

Metrik	Veri dengeleme var	Veri dengeleme yok
Doğruluk	0.920080	0.91469
Kesinlik	0.920088	0.83666
Duyarlılık	0.920081	0.91469
F1-Skoru	0.920080	0.87394

Çalışmada önerilen DENTELNet modelin hesaplama karmaşıklığı, kullanılan katman türlerine göre farklılık göstermektedir. LNN katmanı, giriş uzunluğu (G) ve nöron sayısı (N) ile $O(G \cdot N^2)$ karmaşıklığa yol açmaktadır. LNN katmanında, her nöron önceki nöronlarla etkileşimde bulunmakta ve bu etkileşim her nöron için hesaplanmaktadır. Bu da karmaşıklığın N^2 olmasına neden olmaktadır. BiLSTM katmanları, çift yönlü olduklarından, hesaplama maliyeti $O(2 \cdot G \cdot N^2)$ olarak belirlenmektedir. CNN katmanları ise her bir filtre, giriş verisini baştan sona kaydırarak işlemekte, bu nedenle karmaşıklık $O(G \cdot N)$ olmaktadır. Dense katmanları ve çıktı katmanı ise $O(B \cdot C)$ hesaplama maliyeti eklemektedir. Burada, C çıktı boyutunu ifade etmektedir. Sonuç olarak, toplam karmaşıklık $O(G \cdot N^2 + 2 \cdot G \cdot N^2 + N \cdot C)$ olarak hesaplanmaktadır.



Şekil 10. Makine öğrenmesi yöntemlerine ilişkin karmaşıklık matrisleri.

Çalışmada, önerilen DENTELNet modelinin performansı, farklı makine öğrenmesi yöntemleri ile karşılaştırılmıştır. Bu kapsamda sırasıyla Lojistik Regresyon, K-En Yakın Komşu, Rastgele Orman, Destek Vektör Makineleri, Gauss Naive Bayes ve Gradient

Boosting algoritmaları kullanılarak sınıflandırma performansı değerlendirilmiştir. Kullanılan makine öğrenmesi yöntemlerinin performans değerlendirmeleri, Scikit-learn kütüphanesindeki varsayılan hiperparametre değerleri ile gerçekleştirilmiştir. Her bir makine öğrenmesi yöntemine ait test sonuçları, Şekil 10'da sunulan karmaşıklık matrisleri aracılığıyla detaylandırılmıştır.

Şekil 10'da görüldüğü gibi, Rastgele Orman algoritması, 7561 meşru kullanıcıyı doğru sınıflandırırken, yalnızca 674'ünü yanlış sınıflandırmıştır. Ancak, hileli kullanıcıların yalnızca 49'unu doğru tespit edebilmiş, 674'ünü meşru kullanıcı olarak değerlendirmiştir. Bu durum, genel doğruluk oranının yüksek olmasına rağmen, azınlık sınıfında düşük performans sergilendiğini göstermektedir. Destek Vektör Makinesi, 6617 meşru kullanıcıyı doğru sınıflandırırken, 1135'ini yanlış değerlendirmiştir. Hileli kullanıcı sınıfında ise 429 doğru tespit ve 294 yanlış sınıflandırma yapmıştır. Bu modelin hileli kullanıcıları tespit etme yeteneği, diğer modellere kıyasla daha başarılı olmuştur. Lojistik Regresyon, 5871 meşru kullanıcıyı doğru sınıflandırmış ancak 1881 yanlış sınıflandırma yapmıştır. Hileli kullanıcıların 477'si doğru tespit edilirken, 246'sı meşru kullanıcı olarak değerlendirilmiştir. Bu modelin dengeli bir performans sergilediği gözlemlenmiştir. Öte yandan, Gradient Boosting ve Gauss Naive Bayes algoritmaları, özellikle meşru ve hileli kullanıcıları ayırt etme konusunda diğer algoritmalara kıyasla daha düşük performans göstermiştir. Söz konusu yöntemlere ait doğruluk, kesinlik, duyarlılık ve F1-skoru performans metriklerinin karşılaştırmalı sonuçları, Tablo 2'de görüldüğü gibidir.

Tablo 2. Makine öğrenimi modellerinin performans değerleri.

Model/Metrik	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
Lojistik Regresyon	0.74903	0.89516	0.74903	0.80083
K-En Yakın Komşu	0.67257	0.89175	0.67257	0.74465
Rastgele Orman	0.89794	0.85724	0.89794	0.87388
Destek Vektör Makinesi	0.83139	0.89918	0.83139	0.85755
Gauss Naive Bayes	0.49805	0.82847	0.49805	0.60395
Gradient Boosting	0.47375	0.86730	0.47375	0.57555
DENTELNet	0.92153	0.90402	0.92153	0.90397

Tablo 2'de görüldüğü üzere, önerilen DENTELNet modeli, makine öğrenmesi yöntemleri ile karşılaştırıldığında belirgin bir üstünlük göstermektedir. Doğruluk metriği açısından DENTELNet modeli, 0.92153 ile en yüksek değere ulaşmış olup, bu sonuç en yakın rakibi olan Rastgele Orman modelinin 0.89794'lük doğruluk değerinden anlamlı bir şekilde yüksektir. Destek Vektör Makinesi ise 0.83139 doğruluk değeri ile üçüncü sırada yer almıştır. Bununla birlikte, K-En Yakın Komşu (0.67257), Lojistik Regresyon (0.74903), Gauss Naive Bayes (0.49805) ve Gradient Boosting (0.47375) modellerinin doğruluk değerleri, DENTELNet modelinden oldukça düşüktür. Kesinlik metriği incelendiğinde, DENTELNet modelinin 0.90402 kesinlik değeriyle hem Rastgele Orman (0.85724) hem de Destek Vektör Makinesi (0.89918) modellerini geride bıraktığı görülmektedir. Lojistik Regresyon (0.89516) ve K-En Yakın Komşu (0.89175) modelleri DENTELNet modeline yakın değerler göstermiş olsa da, bu iki modelin doğruluk ve duyarlılık metriklerindeki düşük performansı dikkate alındığında genel etkinlik açısından geride kaldıkları söylenebilir. Duyarlılık metriği açısından DENTELNet modeli, 0.92153 ile diğer tüm modellerden üstün performans sergilemiştir. Rastgele Orman modeli 0.89794 ile ikinci sırada yer alırken, Destek Vektör Makinesi 0.83139 değeriyle üçüncü sıradadır.

Lojistik Regresyon (0.74903), K-En Yakın Komşu (0.67257) ve diğer modeller ise daha düşük değerler almıştır. F1-skoru açısından da DENTELNet modeli 0.90397 ile en yüksek değeri elde etmiştir. Bu sonuç, modelin hem duyarlılık hem de kesinlik metriklerinde dengeli bir performans gösterdiğini kanıtlamaktadır. Rastgele Orman (0.87388) ve Destek Vektör Makinesi (0.85755) modelleri sırasıyla ikinci ve üçüncü sırada yer alırken, diğer modellerin F1-skorları belirgin şekilde daha düşüktür. Özellikle, Gauss Naive Bayes (0,60395) ve Gradient Boosting (0.57555) modelleri, DENTELNet modelinin performansına kıyasla oldukça geride kalmıştır. Sonuç olarak, DENTELNet modelinin doğruluk, kesinlik, duyarlılık ve F1-skoru gibi temel değerlendirme metriklerinde diğer modellere kıyasla tutarlı bir şekilde üstün performans sergilediği açıkça görülmektedir. Bu durum, DENTELNet modelinin hem sınıflandırma doğruluğu hem de tahmin başarısı açısından daha etkili bir çözüm sunduğunu ortaya koymaktadır.

Çalışmada ayrıca, literatürde SGCC veri seti kullanılarak gerçekleştirilen çalışmalar ile DENTELNet modelinin performansı karşılaştırılmıştır. Aşağıda, çeşitli araştırmalar ve kullanılan yöntemlere ait başarı oranları Tablo 3'te sunulmuştur.

Tablo 3. Literatürdeki modellerin doğruluk karşılaştırması.

Çalışma	Yöntem	Doğruluk
Fei vd. [8]	CNN,GRU	%77.0
Shehzad vd. [4]	GA, ABC, Autoencoder, SVM	%80.0
Depuru vd.[24]	SVM	%80.0
Khan vd. [2]	AlexNet	%86.0
Zheng vd.[9]	CNN	%88.9
Hasan vd. [25]	CNN, LSTM	%89.0
Munawar vd.[26]	LSTM	%89.4
A. Souza vd. [3]	MLP ve LSTM	%88.5
Saqib vd. [1]	DNN	%91.0
Pereira vd. [27]	CNN	%91.6
DENTELNet	CNN-BiLSTM-LNN	%92.2

Tablo 3'te görüldüğü gibi, çalışmada önerilen DENTELNet modeli literatürdeki diğer çalışmalar karşılaştırıldığında daha yüksek bir doğruluk başarımları sergilemektedir. Bu sonuçlar, DENTELNet'in SGCC veri seti üzerindeki etkinliğini ve genel performans avantajını ortaya koymaktadır.

5. Sonuç

Elektrik enerjisi, günümüz toplumlarının temel gereksinimlerinden biri olup, ekonomik kalkınma ve sosyal refah üzerinde doğrudan etkili bir role sahiptir. Bununla birlikte, enerji üretiminden nihai tüketime kadar olan süreçte ortaya çıkan kayıplar, kaynakların etkin kullanılmasını engelleyerek ekonomik maliyetlere neden olmaktadır. Özellikle teknik olmayan kayıplar, kaçak elektrik kullanımı, sayaçlarla oynama ve izinsiz bağlantılar gibi sorunlardan kaynaklanmakta ve enerji sektöründe ciddi problemlere yol açmaktadır Bu çalışmada, elektrik enerjisindeki teknik olmayan kayıpların tespiti için hibrit bir derin öğrenme yaklaşımı olan CNN-BiLSTM-LNN modeli sunulmuştur. Hibrit yapının avantajı, farklı katmanların güçlü yönlerini birleştirerek daha kapsamlı ve doğru

sonuçlar elde etmesidir. Modelin hesaplama karmaşıklığı, kullanılan katman türlerine göre değişkenlik göstermektedir. Özellikle, LNN ve BiLSTM katmanları daha yüksek hesaplama gereksinimleri ile daha derin özellikler öğrenirken, CNN katmanları daha düşük bir hesaplama maliyeti ile işlemektedir. Öznitelik indirgemedeki kullanılan Zaman Serisi Ayırışımı tabanlı PCA ve veri dengelemede kullanılan BSMOTE-TLink yöntemleri, modelin performansını optimize etmede kritik bir katkı sağlamıştır. Model, test veri seti üzerinde yüksek doğruluk ve duyarlılık oranlarıyla teknik olmayan kayıpların tespiti için etkili bir çözüm sunmuştur. Önerilen DENTELNet modeli, hem zamansal ilişkileri hem de karmaşık veri özelliklerini başarılı bir şekilde öğrenerek sektördeki mevcut yöntemlere kıyasla daha üstün bir performans sergilemiştir.

Kaynaklar

- [1] Saqib, S. M., Mazhar, T., Iqbal, M., Shahazad, T., Almogren, A., Ouahada, K. ve Hamam, H., Deep learning-based electricity theft prediction in non-smart grid environments, **Heliyon**, 10, 15, e35167, (2024).
- [2] Khan, N., Shahid, Z., Alam, M. M., Sajak, A. A. B., Nazar, M. ve Mazliham, M. S., A novel deep learning technique to detect electricity theft in smart grids using AlexNet, **IET Renewable Power Generation**, 18, 6, 941-958, (2024).
- [3] Souza, M. A., Gouveia, H. T., Ferreira, A. A., de Lima Neta, R. M., Nóbrega Neto, O., da Silva Lira, M. M., Torres, G. L. ve Aquino, R. R., Detection of Non-Technical Losses on a Smart Distribution Grid Based on Artificial Intelligence Models, **Energies**, 17, 7, 1729, (2024).
- [4] Shehzad, F., Javaid, N., Aslam, S. ve Javed, M. U., Electricity theft detection using big data and genetic algorithm in electric power systems, **Electric Power Systems Research**, 209, 107975, (2022).
- [5] Tripathi, A. K., Pandey, A. C. ve Sharma, N., A new electricity theft detection method using hybrid adaptive sampling and pipeline machine learning, **Multimedia Tools and Applications**, 83, 18, 54521-54544, (2024).
- [6] Shehzad, F., Ullah, Z., Alhussain, M., Aurangzeb, K. ve Aslam, S., Deep learning-based meta-learner strategy for electricity theft detection, **Frontiers in Energy Research**, 11, 1232930, (2023).
- [7] Nawaz, A., Ali, T., Mustafa, G., Rehman, S. U. ve Rashid, M. R., A novel technique for detecting electricity theft in secure smart grids using CNN and XG-boost, **Intelligent Systems with Applications**, 17, 200168, (2023).
- [8] Fei, K., Li, Q., Zhu, C., Dong, M. ve Li, Y., Electricity frauds detection in Low-voltage networks with contrastive predictive coding, **International Journal of Electrical Power and Energy Systems**, 137, 107715, (2022).
- [9] Zheng, Z., Yang, Y., Niu, X., Dai, H.-N., ve Zhou, Y., Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids, **IEEE Transactions on Industrial Informatics**, 14(4), 1606-1615, (2017).
- [10] SGCC dataset: Analysis on Abnormal Behavior of Electricity Customers. <https://www.datafountain.cn/competitions/241>, (06.05.2025).
- [11] Wen, Q., Gao, J., Song, X., Sun, L., Xu, H. ve Zhu, S., RobustSTL: A robust seasonal-trend decomposition algorithm for long time series, **Proceedings of the AAAI Conference on Artificial Intelligence**, 33, 01, 5409-5416, (2019).
- [12] Kherif, F. ve Latypova, A., **Principal component analysis** in Mechelli, A. ve Sandra, V., *Machine learning*, Elsevier, 209-225, (2019).

- [13] Sulandari, W. ve Utami, H., Forecasting time series with trend and seasonal patterns based on SSA, **3rd International Conference on Science in Information Technology**, 648-653, (2017).
- [14] Al Majzoub, H., Elgedawy, I., Akaydın, Ö. ve Köse Ulukök, M., HCAB-SMOTE: A hybrid clustered affinitive borderline SMOTE approach for imbalanced data binary classification, **Arabian Journal for Science and Engineering**, 45, 4, 3205-3222, (2020).
- [15] Smiti, S. ve Soui, M., Bankruptcy prediction using deep learning approach based on borderline SMOTE, **Information Systems Frontiers**, 22, 5, 1067-1083, (2020).
- [16] Pereira, R. M., Costa, Y. M. ve Silla Jr, C. N., MLTL: A multi-label approach for the Tomek Link undersampling algorithm, **Neurocomputing**, 383, 95-105, (2020).
- [17] Swana, E. F., Doorsamy, W. ve Bokoro, P., Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset, **Sensors**, 22, 9, 3246, (2022).
- [18] He, H. ve Garcia, E. A., Learning from imbalanced data, **IEEE Transactions on knowledge and data engineering**, 21, 9, 1263-1284, (2009).
- [19] Kabakus, A. T., DroidMalwareDetector: A novel Android malware detection framework based on convolutional neural network, **Expert Systems with Applications**, 206, 117833, (2022).
- [20] Riyaz, B. ve Ganapathy, S., A deep learning approach for effective intrusion detection in wireless networks using CNN, **Soft Computing**, 24, 22, 17265-17278, (2020).
- [21] Sri vidhya, G. ve Nagarajan, R., A novel bidirectional LSTM model for network intrusion detection in SDN-IoT network, **Computing**, 106, 2613-2642, (2024).
- [22] Chahine, M., Hasani, R., Kao, P., Ray, A., Shubert, R., Lechner, M., Amini, A. ve Rus, D., Robust flight navigation out of distribution with liquid neural networks, **Science Robotics**, 8, 77, eadc8892, (2023).
- [23] Kumar, K., Verma, A., Gupta, N. ve Yadav, A., Liquid Neural Networks: A Novel Approach to Dynamic Information Processing, **International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)**, 725-730, (2023).
- [24] Depuru, S. S. S. R., Wang, L. ve Devabhaktuni, V., Support vector machine based data classification for detection of electricity theft, **Power Systems Conference and Exposition**, 1-8, (2011).
- [25] Hasan, M. N., Toma, R. N., Nahid, A. A., Islam, M. M. ve Kim, J.-M., Electricity theft detection in smart grid systems: A CNN-LSTM based approach, **Energies**, 12, 17, 3310, (2019).
- [26] Munawar, S., Asif, M., Kabir, B., Pamir, Ullah, A. ve Javaid, N., Electricity theft detection in smart meters using a hybrid Bi-directional GRU Bi-directional LSTM model, **15th International Conference on Complex, Intelligent and Software Intensive Systems**, 297-308, (2021).
- [27] Pereira, J., ve Saraiva, F., Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques, **International Journal of Electrical Power and Energy Systems**, 131, 107085, (2021).