

Derin Öğrenme Modellerinde Çoklu Sınıflandırma İçin Loss Fonksiyonlarının Performans Analizi

Cüneyt ÖZDEMİR^{1*}

¹Bilgisayar Mühendisliği / Mühendislik Fakültesi, Siirt Üniversitesi, Türkiye

*(cozdemir@siirt.edu.tr)

Özet – Derin öğrenme, yapay zekâ alanındaki büyük ilerlemelerin bir alt dalıdır ve çok katmanlı yapay sinir ağları kullanarak karmaşık verilerden öğrenme yeteneği sunar. Loss fonksiyonları, modelin tahminleri ile gerçek etiketler arasındaki uyumsuzluğu ölçer ve modelin eğitimini optimize etmek için kullanılır. Bu çalışma, derin öğrenme modellerinin çoklu sınıflandırma görevlerinde kullanılan loss fonksiyonlarının performansını değerlendirmeyi amaçlamaktadır. 12 farklı loss fonksiyonu Cifar-10 veri seti üzerinde değerlendirilmiştir. Elde edilen sonuçlar, loss fonksiyonlarının doğruluk, precision, recall ve F1 Score gibi metrikler üzerinde farklı etkiler gösterdiğini ortaya koymaktadır.

Anahtar Kelimeler – Derin Öğrenme, Loss Fonksiyonları, Çoklu Sınıflandırma, Cifar-10

I. GİRİŞ

Derin öğrenme, son yıllarda büyük ilerlemeler kaydedilen yapay zekâ alanının bir alt dalıdır. Derin öğrenme, çok katmanlı yapay sinir ağları kullanarak karmaşık ve yüksek boyutlu verilerden öğrenme ve tahmin yapma yeteneğine sahiptir. Derin öğrenme modelleri, görüntü işleme, doğal dil işleme, ses tanıma, biyoenformatik, tıbbi görüntüleme, oyun oynama, robotik ve daha birçok alanda başarılı uygulamalar göstermiştir [1][2]. Derin öğrenme modellerinin performansını değerlendirmek ve iyileştirmek için, loss fonksiyonu adı verilen bir metrik kullanılır. Loss fonksiyonu, modelin tahminleri ile gerçek etiketler arasındaki uyumsuzluğu ölçer ve modelin ağırlıklarını güncellemek için optimizasyon algoritmasına geri bildirim sağlar [3]. Loss fonksiyonu seçimi, modelin başarısını etkileyen önemli faktörlerden biridir.

Derin öğrenme modellerinin performansının değerlendirilmesinde kullanılan loss fonksiyonları, sınıflandırma problemlerine özgü olarak seçilir. Sınıflandırma, bir veri noktasının belirli bir sınıfa veya kategoriye ait olup olmadığını belirlemek için kullanılan bir makine öğrenmesi yöntemidir.

Sınıflandırma problemleri, ikili sınıflandırma, çoklu sınıflandırma ve çok etiketli sınıflandırma olmak üzere üç ana türe ayrılabilir [4]. İkili sınıflandırma, bir veri noktasının iki olası sınıftan birine ait olup olmadığını belirlemek için kullanılır. Örneğin, bir e-postanın spam olup olmadığını, bir hastanın hasta olup olmadığını veya bir resmin kedi olup olmadığını belirlemek gibi. Çoklu sınıflandırma, bir veri noktasının birden fazla sınıftan yalnızca birine ait olup olmadığını belirlemek için kullanılır. Örneğin, bir resmin hangi hayvana ait olduğunu, bir metnin hangi konuya ilişkin olduğunu veya bir ses kaydının hangi dilde olduğunu belirlemek gibi. Çok etiketli sınıflandırma, bir veri noktasının birden fazla sınıfa ait olabileceğini belirlemek için kullanılır. Örneğin, bir resimdeki nesnelere tanımlamak, bir metindeki duyguları belirlemek veya bir ses kaydındaki konuşmacıları ayırt etmek gibi.

Derin öğrenme modellerinin eğitimi sırasında, modelin tahminlerini gerçek etiketlerle karşılaştırmak için kullanılan kayıp fonksiyonunun doğru seçilmesi, modelin doğruluğu, genelleme yeteneği ve eğitim sürecinin istikrarı üzerinde belirleyici bir etkiye sahiptir.

Bu çalışma, derin öğrenme alanındaki öne çıkan zorluklardan biri olan çoklu sınıflandırma problemlerinde kullanılan loss fonksiyonlarına odaklanmaktadır. Çoklu sınıflandırma problemlerinde, loss fonksiyonu seçimi, modelin başarısını değerlendirmede ve yakınsama sürecini etkilemede kritik bir rol oynamaktadır. Farklı loss fonksiyonları, çeşitli özelliklere, avantajlara ve dezavantajlara sahip olduğundan, doğru bir seçim modelin performansını belirleme açısından kritiktir. Bu çalışmada, çoklu sınıflandırma görevlerinde kullanılan Weighted Categorical Crossentropy, Smooth L1 Loss, Center Loss, Bi-Tempered Logistic Loss, Generalized Cross Entropy Loss, PHuber Cross Entropy Loss, Symmetric Cross Entropy Loss, Categorical Focal Loss, Categorical Crossentropy, Kullback-Leibler Divergence ve Categorical Hinge Loss fonksiyonları incelenecek ve bu fonksiyonların performansları Cifar-10 veri seti üzerinde karşılaştırılacaktır. Bu fonksiyonlar, farklı öğrenme paradigmalarına sahip olup, modellerin çeşitli yönlerini etkileyebilen önemli unsurları temsil etmektedirler.

Çalışma, her loss fonksiyonunun teorik temelini açıklamayı ve bu fonksiyonların Cifar-10 veri seti üzerinde elde edilen sonuçları karşılaştırarak performanslarını değerlendirmeyi hedeflemektedir. Bu analiz, derin öğrenme modellerinin çoklu sınıflandırma görevlerinde hangi loss fonksiyonlarının daha etkili olduğuna dair değerli bir anlayış sağlamayı amaçlamaktadır. Çalışmanın geri kalanında şu bölümler yer almaktadır: Bölüm 2’de, loss fonksiyonlarının tanımlarını, formüllerini, avantajlarını ve dezavantajları sunulmuştur. Bölüm 3’te, materyal ve metot açıklanmıştır. Bölüm 4’te, deney sonuçlarını, karşılaştırmaları ve analizleri sunulmuştur. Bölüm 5’te ise makalenin özetini, sonuçlarını ve gelecek çalışmalar için öneriler sunulmuştur.

II. LOSS FONKSİYONLARI

Bu bölümde, çoklu sınıflandırma modellerinde kullanılan loss fonksiyonlarının tanımları, formülleri, avantajları ve dezavantajları sunulacaktır. Çalışma kapsamında olan, Weighted Categorical Crossentropy, Smooth L1 Loss, Center Loss, Bi-Tempered Logistic Loss, Generalized Cross Entropy Loss, PHuber Cross Entropy Loss, Symmetric Cross Entropy Loss, Categorical Focal Loss, Categorical Crossentropy, Kullback-Leibler

Divergence, Categorical Hinge Loss fonksiyonları ele alınmıştır. Bu loss fonksiyonları, softmax tabanlı ve softmax dışı loss fonksiyonları olarak iki kategoriye ayrılabilir. Softmax tabanlı olanlar:

- Weighted Categorical Crossentropy
- Categorical Crossentropy
- Kullback-Leibler Divergence
- Categorical Focal Loss
- Generalized Cross Entropy Loss
- Symmetric Cross Entropy Loss
- Bi-Tempered Logistic Loss

Softmax dışı olanlar ise şunlardır:

- Smooth L1 Loss
- Center Loss
- PHuber Cross Entropy Loss
- Categorical Hinge Loss

Aşağıda, her bir loss fonksiyonu için ayrıntılı bir açıklama verilmiştir.

A. Weighted Categorical Crossentropy

Weighted Categorical Crossentropy (WCC), softmax tabanlı bir loss fonksiyonudur. WCC, Categorical Crossentropy (CCE) loss fonksiyonunun genelleştirilmiş bir halidir. CCE, modelin tahmin ettiği olasılık dağılımı ile gerçek etiketlerin olasılık dağılımı arasındaki çapraz entropiyi hesaplar. CCE, çoklu sınıflandırma problemlerinde standart bir loss fonksiyonu olarak kabul edilir ve genellikle iyi performans gösterir [3]. Ancak, CCE, sınıf dengesizliği gibi durumlarda etkili olmayabilir. Sınıf dengesizliği, veri setindeki sınıfların eşit sayıda örnek içermediği durumdur. Bu durumda, CCE, az sayıda örneği olan sınıfları göz ardı edebilir ve modelin doğruluğunu düşürebilir. WCC, sınıf dengesizliği sorununu çözmek için CCE’ye bir ağırlık vektörü ekler. Ağırlık vektörü, her bir sınıf için bir ağırlık değeri içerir. Ağırlık değeri, sınıfın önemini veya zorluğunu belirtir. Ağırlık değeri yüksek olan sınıflar, loss fonksiyonunda daha fazla etkiye sahip olur. Ağırlık değeri düşük olan sınıflar, loss fonksiyonunda daha az etkiye sahip olur. Böylece, WCC, sınıflar arasında bir denge sağlar ve modelin tüm sınıfları eşit şekilde öğrenmesini sağlar [5],[6].

WCC'nin matematiksel formülü şöyledir:

$$WCC(y, \hat{y}) = - \sum_{i=1}^N \sum_{j=1}^C w_j y_{ij} \log(\hat{y}_{ij})$$

Burada, N veri setindeki örnek sayısı, C sınıf sayısı, w_j j'inci sınıf için ağırlık değeri, y_{ij} i'inci örneğin j'inci sınıf için gerçek etiketi, \hat{y}_{ij} i'inci örneğin j'inci sınıf için modelin tahminini ifade eder.

B. Smooth L1 Loss

Smooth L1 Loss, mutlak hata değerinin bir eşik değerinden küçük olup olmadığına göre farklı davranan bir kayıp fonksiyonudur. Eşik değerinden küçük olan hatalar için karesel bir terim, eşik değerinden büyük olan hatalar için doğrusal bir terim kullanır. Böylece, Smooth L1 Loss, hem MSE Loss hem de L1 Loss'un avantajlarını birleştirir. MSE Loss, küçük hataları daha iyi öğrenir, ancak büyük hatalara karşı hassastır. L1 Loss, büyük hatalara karşı dayanıklıdır, ancak küçük hataları daha az öğrenir. Smooth L1 Loss, hem küçük hem de büyük hataları dengeli bir şekilde öğrenir [7] - [9].

Smooth L1 Loss matematiksel formülü şöyledir:

$$SmoothL1Loss(y, \hat{y}) = \begin{cases} 0.5(y - \hat{y})^2, & \text{if } |y - \hat{y}| < \delta \\ (\delta |y - \hat{y}| - 0.5\delta), & \text{otherwise} \end{cases}$$

Burada, y gerçek değer, \hat{y} modelin tahmini, δ eşik değeri ifade eder.

Smooth L1 Loss'un avantajları şunlardır:

- Küçük hatalarda MSE Loss gibi davranarak, modelin daha hassas olmasını sağlar.
- Büyük hatalarda L1 Loss gibi davranarak, modelin daha dayanıklı olmasını sağlar.
- Büyük hatalarda karesel terim yerine doğrusal terim kullanarak, modelin aşırı değerlere karşı daha az duyarlı olmasını sağlar.
- Küçük hatalarda doğrusal terim yerine karesel terim kullanarak, modelin kayıp fonksiyonunun türevinin sifıra yaklaşmasını önler.

Smooth L1 Loss'un dezavantajları şunlardır:

- Eşik değerinin belirlenmesi zor olabilir ve deneme-yanılma gerektirebilir.
- Eşik değerinin seçimi, modelin performansını etkileyebilir.

C. Center Loss

Center Loss, softmax dışı bir loss fonksiyonudur. Center Loss, modelin tahminleri ile gerçek etiketler arasındaki mesafeyi değil, modelin tahminleri ile

sınıf merkezleri arasındaki mesafeyi minimize etmeye çalışan bir loss fonksiyonudur. Sınıf merkezi, bir sınıfa ait örneklerin ortalaması olarak tanımlanır. Center Loss, sınıf içi dağılımı azaltırken, sınıflar arası dağılımı arttırmayı amaçlar. Böylece, Center Loss, modelin sınıfları daha iyi ayırt etmesini sağlar. Center Loss, softmax loss ile birlikte kullanılır. Softmax loss, modelin tahminleri ile gerçek etiketler arasındaki çapraz entropiyi hesaplar. Softmax loss, çoklu sınıflandırma problemlerinde standart bir loss fonksiyonu olarak kabul edilir ve genellikle iyi performans gösterir. Ancak, softmax loss, sınıf içi ve sınıflar arası dağılımları dikkate almaz. Bu, modelin sınıfları yeterince ayırt edememesine veya overfitting'e neden olabilir. Center loss, softmax loss'un bu eksikliğini gidermek için kullanılır. Center loss, softmax loss ile birlikte, modelin hem doğruluğunu hem de ayırt ediciliğini artırır [10] - [13].

Center Loss'un matematiksel formülü şöyledir:

$$CenterLoss(y, \hat{y}) = \frac{1}{2} \sum_{i=1}^N \|\hat{y}_i - c_{y_i}\|^2$$

Burada, N veri setindeki örnek sayısı, y_i i'inci örneğin gerçek etiketi, \hat{y}_i i'inci örneğin modelin tahmini, c_{y_i} y_i 'inci sınıfın merkezini ifade eder. Sınıf merkezleri, modelin ağırlıkları ile birlikte güncellenir.

Center Loss'un avantajları şunlardır:

- Modelin tahminleri ile gerçek etiketler arasındaki çapraz entropi kaybı ile birlikte kullanılarak, modelin hem doğruluğunu hem de ayırt ediciliğini artırır.
- Modelin tahminlerinin sınıf merkezlerine yaklaşmasını sağlayarak, modelin aşırı güven yapmasını önler ve daha iyi kalibre edilmiş olasılıklar üretir.
- Modelin sınıf merkezlerini dinamik olarak güncelleyerek, modelin öğrenme sürecine uyum sağlamasını sağlar.

Center Loss'un dezavantajları şunlardır:

- Sınıf merkezlerinin hesaplanması ve güncellenmesi ekstra bellek ve işlem gerektirir.
- Düzenlilik parametresinin belirlenmesi zor olabilir ve deneme-yanılma gerektirebilir.

- Sınıf merkezleri, veri setinin dağılımını tam olarak yansıtmayabilir veya sınıflar arasında örtüşebilir.

D. P-Huber Cross Entropy Loss

P-Huber Cross Entropy Loss, softmax tabanlı bir loss fonksiyonudur. P-Huber Cross Entropy Loss, CCE loss fonksiyonunun bir varyasyonudur. CCE, modelin tahmin ettiği olasılık dağılımı ile gerçek etiketlerin olasılık dağılımı arasındaki çapraz entropiyi hesaplar. CCE, çoklu sınıflandırma problemlerinde standart bir loss fonksiyonu olarak kabul edilir ve genellikle iyi performans gösterir. Ancak, CCE, modelin overfitting olmasına veya sınıflar arasında dengesizlik olmasına neden olabilir. P-Huber Cross Entropy Loss, CCE'ye bir düzenlilik terimi ekler. Düzenlilik terimi, Huber Loss ile ilgilidir. Huber Loss, mutlak hata değerinin bir eşik değerinden küçük olup olmadığına göre farklı davranan bir loss fonksiyonudur. Eşik değerinden küçük olan hatalar için karesel bir terim, eşik değerinden büyük olan hatalar için doğrusal bir terim kullanır. Böylece, Huber Loss, hem MSE Loss hem de L1 Loss'un avantajlarını birleştirir. MSE Loss, küçük hataları daha iyi öğrenir, ancak büyük hatalara karşı hassastır. L1 Loss, büyük hatalara karşı dayanıklıdır, ancak küçük hataları daha az öğrenir. Huber Loss, hem küçük hem de büyük hataları dengeli bir şekilde öğrenir. P-Huber Cross Entropy Loss, CCE'nin yanı sıra, modelin tahminleri ile sınıf merkezleri arasındaki Huber Loss'u da hesaplar. Sınıf merkezi, bir sınıfa ait örneklerin ortalaması olarak tanımlanır. P-Huber Cross Entropy Loss, CCE ile birlikte, modelin doğruluğunu hem de ayırt ediciliğini arttırmayı amaçlar. Ayrıca, modelin aşırı güvenini azaltır ve daha iyi kalibre edilmiş olasılıklar üretir [9] – [11].

P-Huber Cross Entropy Loss'un matematiksel formülü şöyledir:

$$PHuberCrossEntropyLoss(y, \hat{y}) = CCE(y, \hat{y}) + \lambda \sum_{i=1}^N \sum_{j=1}^C H(y_{ij}, y_{ij}^{\wedge} - c_{y_i})$$

Burada, N veri setindeki örnek sayısı, C sınıf sayısı, y_{ij} i'inci örneğin j'inci sınıf için gerçek etiketi, y_{ij}^{\wedge} i'inci örneğin j'inci sınıf için modelin tahmini, c_{y_i} yi'inci sınıfın merkezi, CCE kategorik çapraz entropi fonksiyonu, H Huber Loss fonksiyonu, λ düzenlilik parametresi ifade eder.

E. Bi-Tempered Logistic Loss

Bu loss fonksiyonu, çapraz entropi ve Huber Loss'un bir kombinasyonudur. Çapraz entropi, modelin doğruluğunu arttırmayı amaçlarken, Huber Loss, modelin büyük hatalara karşı daha dayanıklı olmasını sağlar. Bu loss fonksiyonu, modelin hem sınıf merkezleri arasındaki mesafeyi hem de sınıf içi varyansı minimize etmeyi hedefler. Çoklu sınıflandırma problemleri için uygun bir loss fonksiyonudur [14].

F. Generalized Cross Entropy Loss

Bu loss fonksiyonu, MAE ve CCE loss fonksiyonlarının bir genelleştirmesidir. MAE ve CCE, gürültülü etiketlere karşı hassas olabilirler. Bu loss fonksiyonu, gürültülü etiketlere karşı daha dayanıklı olmak için bir parametre olan q'yu kullanır. q=1 olduğunda, bu loss fonksiyonu MAE'ye eşit olur. q=0'a yaklaştığında, bu loss fonksiyonu CCE'ye yaklaşır. Bu loss fonksiyonu, gürültülü etiketlerin olduğu derin sinir ağları için kullanılabilir [15].

G. Symmetric Cross Entropy Loss

Bu loss fonksiyonu, CCE ve RCE (Reverse Cross Entropy) loss fonksiyonlarının birleşimidir. CCE, modelin tahminleri ile gerçek etiketler arasındaki farkı ölçerken, RCE, modelin tahminleri ile tahmin edilen olasılık dağılımı arasındaki farkı ölçer. Bu loss fonksiyonu, CCE'nin hem alt öğrenme hem de aşırı uyum problemlerini çözmek için RCE ile simetrik olarak desteklenir. Bu loss fonksiyonu, gürültülü etiketlerin olduğu çoklu sınıflandırma problemleri için uygundur [16].

H. Categorical Focal Loss

Bu loss fonksiyonu, CCE loss fonksiyonuna bir modülasyon terimi ekler. Bu terim, zor sınıflandırılan örnekleri daha ağır cezalandırmak ve kolay sınıflandırılan örnekleri daha az cezalandırmak için kullanılır. Bu loss fonksiyonu, sınıf dengesizliğinin olduğu çoklu sınıflandırma problemleri için uygundur. Bu loss fonksiyonu, bir odaklama parametresi olan gamma ve bir sınıf dengesi parametresi olan alpha kullanır.

I. Categorical Crossentropy

Bu loss fonksiyonu, modelin tahminleri ile gerçek etiketler arasındaki çapraz entropiyi hesaplar. Bu loss fonksiyonu, çoklu sınıflandırma problemleri için standart bir loss fonksiyonudur.

Bu loss fonksiyonu, modelin tahminlerinin gerçek olasılık dağılımına ne kadar yakın olduğunu ölçer.

J. Kullback-Leibler Divergence

Bu loss fonksiyonu, iki olasılık dağılımı arasındaki farkı ölçer. Bu loss fonksiyonu, bir referans dağılımı Q ile karşılaştırılan bir dağılım P'nin Q'dan ne kadar sapma gösterdiğini belirtir. Bu loss fonksiyonu, bilgi teorisinde ve istatistiksel çıkarımda önemli bir rol oynar. Bu loss fonksiyonu, modelin tahminleri ile gerçek olasılık dağılımı arasındaki çapraz entropiyi hesaplamak için kullanılabilir.

K. Categorical Hinge Loss

Bu loss fonksiyonu, çoklu sınıflandırma problemleri için maksimum marjın sınıflandırmasını gerçekleştirir. Bu loss fonksiyonu, doğru sınıfın tahmin değerinin yanlış sınıfların tahmin değerlerinden ne kadar büyük olduğunu ölçer. Bu loss fonksiyonu, destek vektör makineleri (SVM) gibi marjın önemli olduğu sınıflandırıcılar için kullanılabilir.

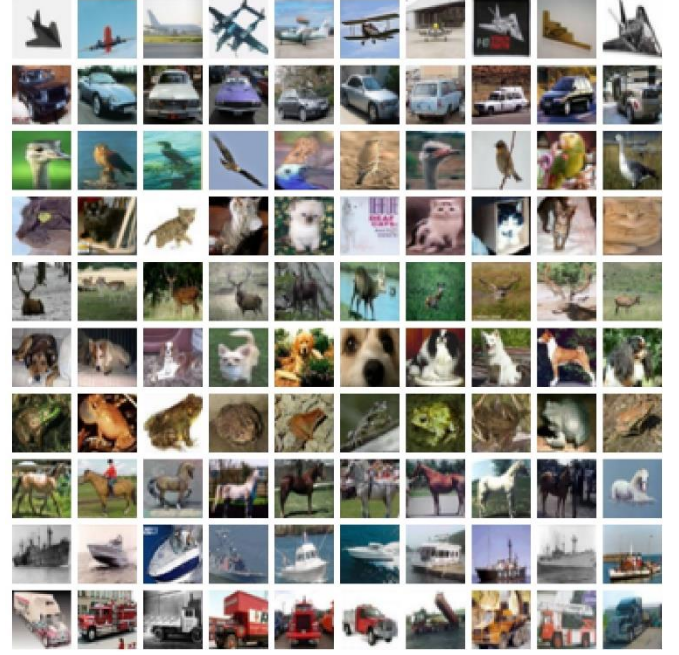
L. Mean Squared Error

Bu loss fonksiyonu, modelin tahminleri ile gerçek değerler arasındaki ortalama kare farkını hesaplar. Bu loss fonksiyonu, regresyon problemleri için standart bir loss fonksiyonudur. Bu loss fonksiyonu, modelin hata varyansını minimize etmeyi amaçlar.

III. MATERYAL VE METOT

A. VERİ SETİ

Veri seti olarak CIFAR-101 kullanılmıştır. CIFAR-10, 10 farklı sınıftan ve 3 kanallı 32x32 pikselden oluşan 60.000 renkli görüntü içeren bir veri setidir. Her sınıf 6.000 resim içermektedir. Eğitim seti 50.000 görüntü, test seti 10.000 görüntü içermektedir. 10 farklı sınıf şunlardır: uçak, otomobil, kuş, kedi, geyik, köpek, kurbağa, at, gemi ve kamyon. CIFAR-10, makine öğrenmesi ve bilgisayar görüşü algoritmalarını eğitmek ve test etmek için yaygın olarak kullanılan bir veri setidir. CIFAR-10 veri kümesinde her sınıfa ait örnek görüntüleri şekil 1'de gösterilmiştir.



Şekil 1. CIFAR-10 veri setine ait örnek görüntüler

B. Metot

Model olarak, Keras kütüphanesi ile oluşturulmuş basit bir evrişimli sinir ağı (CNN) kullanılmıştır. Model, giriş katmanı, bir evrişimli katman, bir max pooling katmanı, bir düzleştirme katmanı, bir tam bağlı katman ve çıkış katmanından oluşmaktadır. Evrişimli katman, 32 adet 3x3 boyutunda filtre kullanarak görüntüleri işler ve ReLU aktivasyon fonksiyonu uygular. Max pooling katmanı, 2x2 boyutunda bir pencere ile görüntüleri küçültür ve öznelikleri seçer. Düzleştirme katmanı, çok boyutlu tensörleri tek boyutlu bir vektöre dönüştürür. Tam bağlı katman, 64 adet nöron içerir ve ReLU aktivasyon fonksiyonu uygular. Çıkış katmanı, 10 adet nöron içerir ve softmax aktivasyon fonksiyonu ile 10 sınıf için olasılık dağılımı üretir. Model, 'adam' optimizasyon fonksiyonu ile derlenmiştir ve kayıp fonksiyonu olarak farklı seçenekler denenmiştir. Modelin performansını ölçmek için metrik olarak yaygın olarak kullanılan Accuracy, Precision, Recall ve F1 Score kullanılmıştır.

IV. BULGULAR

Bu çalışma, derin öğrenme modellerinde çoklu sınıflandırma görevlerinde kullanılan loss fonksiyonlarının performansını değerlendirmeyi amaçlamaktadır. Analizleri güçlendirmek ve sonuçları daha güvenilir hale getirmek adına, her

loss fonksiyonu için model ve veri seti, 5 k-fold çapraz doğrulama işlemine tabi tutulmuştur. Bu yöntem, modelin genelleme yeteneğini değerlendirmek, overfitting azaltmak ve sonuçların güvenilirliğini artırmak için kullanılmıştır. Çapraz doğrulama işlemi, veri setinin farklı alt kümeleri üzerinde modelin eğitilip değerlendirilmesini sağlayarak istatistiksel güvenilirliği artırır. Her bir çapraz doğrulama katmanı, modelin farklı veri alt kümeleri üzerinde performansını değerlendirir ve bu sayede genelleme yeteneğini daha iyi değerlendirilmesine olanak tanır.

Elde edilen sonuçlar, çeşitli loss fonksiyonlarının performansını gösteren Tablo 1’de özetlenmiştir. Her bir loss fonksiyonu için elde edilen accuracy, precision, recall ve F1 Score değerleri, bu çalışmanın odak noktasını oluşturmaktadır. Sonuçlar, loss fonksiyonlarının çoklu sınıflandırma görevlerindeki etkilerini anlamamıza ve en etkili olanları belirlememize yardımcı olacaktır.

Tablo 1. Loss fonksiyonlarına göre model performansı

Loss Function	Accuracy	Precision	Recall	F1 Score
Weighted Categorical Crossentropy	0,617	0,629	0,617	0,615
Smooth L1 Loss	0,610	0,620	0,610	0,608
Center Loss	0,604	0,619	0,604	0,603
Bi-Tempered Logistic Loss	0,099	0,035	0,099	0,033
Generalized Cross Entropy Loss	0,019	0,018	0,019	0,015
P-Huber Cross Entropy Loss	0,618	0,623	0,618	0,614
Symmetric Cross Entropy Loss	0,600	0,617	0,600	0,604
Categorical Focal Loss	0,606	0,621	0,606	0,605
Categorical Crossentropy	0,611	0,626	0,611	0,608
Kullback-Leibler Divergence	0,615	0,626	0,615	0,611
Categorical Hinge Loss	0,114	0,026	0,114	0,034
Mean Squared Error	0,606	0,624	0,606	0,605

Tablo 1’e göre, Bi-Tempered Logistic Loss, Generalized Cross Entropy Loss, Symmetric Cross Entropy Loss ve Categorical Hinge Loss fonksiyonlarının düşük doğruluk oranları dikkat çekmektedir. Bu loss fonksiyonlarının, modelin öğrenme sürecini olumsuz etkilediği ve düşük performans gösterdiği söylenebilir. Bunlar dışında kalan loss fonksiyonlarının birbirine yakın yüksek

doğruluk oranlarına sahip olduğu gözlemlenmektedir. Bu loss fonksiyonlarının yüksek doğruluk oranları, modelin sınıfları doğru bir şekilde sınıflandırmada etkili olduğunu gösterir. Precision, recall ve F1 Score değerlerine bakıldığında, özellikle Bi-Tempered Logistic Loss, Generalized Cross Entropy Loss ve Categorical Hinge Loss fonksiyonlarının düşük değerlere sahip olduğu görülmektedir. Bu durum, bu loss fonksiyonlarının sınıflar arası ayrımı yapma konusunda zayıf olduğunu ve modelin dengesiz sınıflarla başa çıkma yeteneğini düşürdüğünü göstermektedir.

Sonuç olarak, Cifar-10 veri seti üzerinde yapılan deneyler, loss fonksiyonlarının seçiminin model performansını önemli ölçüde etkilediğini ortaya koymaktadır. Bu sonuçlar, modelin spesifik görevine uygun loss fonksiyonunun seçilmesinin önemini vurgulamaktadır.

V. SONUÇLAR VE TARTIŞMA

Bu çalışmada, çoklu sınıflandırma görevlerinde kullanılan farklı loss fonksiyonlarının performansını değerlendirmek amacıyla Cifar-10 veri seti üzerinde 5 k-fold çapraz doğrulama işlemi gerçekleştirmiştir. Elde edilen sonuçlar, her bir loss fonksiyonunun model performansına olan etkisini göstermektedir.

Modelde kullanılan loss fonksiyonlarına göre Weighted Categorical Crossentropy, Smooth L1 Loss, Center Loss, P-Huber Cross Entropy Loss, Symmetric Cross Entropy Loss, Categorical Focal Loss, Categorical Crossentropy, Kullback-Leibler Divergence, Mean Squared Error loss fonksiyonları, yüksek doğruluk oranlarına sahip olmuştur. Bu sonuçlar, bu loss fonksiyonlarının Cifar-10 veri kümesi için çoklu sınıflandırma görevlerinde etkili olduğunu ve modelin sınıfları doğru bir şekilde sınıflandırmada başarılı olduğunu göstermektedir. Öte yandan, Bi-Tempered Logistic Loss, Generalized Cross Entropy Loss, Symmetric Cross Entropy Loss ve Categorical Hinge Loss fonksiyonları düşük doğruluk, Precision, recall ve F1 skor oranlarına sahiptir. Bu durum, bu loss fonksiyonlarının modelin öğrenme sürecini olumsuz etkilediğini ve düşük performans gösterdiğini göstermektedir. Bu durum, bu loss fonksiyonlarının sınıflar arası ayrımı yapma konusunda zayıf olduğunu ve modelin performansını olumsuz etkilediğini göstermektedir.

DeneySEL sonuçlar, loss fonksiyonu seçiminin model performansı üzerinde kritik bir etkiye sahip olduğunu vurgulamaktadır. Bu bağlamda, belirli bir görev için en uygun loss fonksiyonunun seçilmesi, modelin doğruluğunu artırabilir ve genelleme yeteneğini güçlendirebilir. Ancak, bu sonuçlar genel bir çerçevede değerlendirilmelidir. Çünkü her bir loss fonksiyonunun avantajları ve dezavantajları, kullanıldığı bağlam ve veri setine bağlı olarak değişebilir. Bu çalışmanın kapsamı dahilinde elde edilen sonuçlar, derin öğrenme modellerinin çoklu sınıflandırma görevlerinde loss fonksiyonu seçiminin karmaşıklığını anlamamıza katkı sağlamaktadır. Bir çoklu sınıflandırma modeli için en iyi optimizasyon fonksiyonunun seçilmesi verilerin özelliklerine, model mimarisine ve değerlendirme metriğine bağlıdır. Gelecekteki çalışmalarda daha geniş veri setleri ve farklı görevler üzerinde benzer analizlerin yapılması, loss fonksiyonlarının genel performansını daha kapsamlı bir şekilde değerlendirebilir.

KAYNAKLAR

- [1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- [3] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [4] Zhang, M. L., & Zhou, Z. H. (2014). A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8), 1819-1837.
- [5] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106, 249-259.
- [6] Cui, Y., Jia, M., Lin, T. Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9268-9277).
- [7] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
- [8] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).
- [9] Huber, P. J. (1964). Robust estimation of a location parameter. *The annals of mathematical statistics*, 35(1), 73-101.
- [10] Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In *European conference on computer vision* (pp. 499-515). Springer, Cham.
- [11] Zhang, Y., Wang, L., Qi, J., & Wang, D. (2019). P-huber loss for deep face recognition. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 2891-2895). IEEE.
- [12] Liu, W., & Wang, Z. (2018). Fast R-CNN with softmax loss for multi-label object detection. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 1799-1807).
- [13] Li, H., Xiong, P., An, J., & Wang, L. (2018). Faster R-CNN with a softmax loss for multi-label object detection. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 760-765). IEEE.
- [14] Amid, E., Warmuth, M. K., Anil, R., & Koren, T. (2019). Robust bi-tempered logistic loss based on bregman divergences. *Advances in Neural Information Processing Systems*, 32.
- [15] Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31.
- [16] Wang, Y., Ma, X., Chen, Z., Luo, Y., Yi, J., & Bailey, J. (2019). Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 322-330).