

## Sigmoid-Gumbel: Yeni Bir Hibrit Aktivasyon Fonksiyonu

Mustafa KAYTAN<sup>1\*</sup>, İbrahim Berkan AYDİLEK<sup>2</sup>, Celaleddin YEROĞLU<sup>3</sup>,  
Ali KARCI<sup>4</sup>



<sup>1</sup>Bilgisayar Programcılığı Programı, Bilgisayar Teknolojileri Bölümü,  
Teknik Bilimler Meslek Yüksekokulu, Harran Üniversitesi, Şanlıurfa

<sup>2</sup>Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Harran Üniversitesi, Şanlıurfa

<sup>3</sup>Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, İnönü Üniversitesi, Malatya

<sup>4</sup>Yazılım Mühendisliği Bölümü, Mühendislik Fakültesi, İnönü Üniversitesi, Malatya

(ORCID: [0000-0002-8416-6520](https://orcid.org/0000-0002-8416-6520)) (ORCID: [0000-0001-8037-8625](https://orcid.org/0000-0001-8037-8625)) (ORCID: [0000-0002-6106-2374](https://orcid.org/0000-0002-6106-2374))

(ORCID: [0000-0002-8489-8617](https://orcid.org/0000-0002-8489-8617))

### Anahtar kelimeler:

Aktivasyon fonksiyonu,  
Derin sinir ağı, Derin  
öğrenme.

### Öz

Bu makalede daha önce sunulan aktivasyon fonksiyonlarının olumlu yanlarını birleştiren ve onlardan daha iyi başarımlar sağlayan ve Sigmoid-Gumbel (SG) olarak adlandırılan yeni bir hibrit aktivasyon fonksiyonu önerilmiştir. Önerilen fonksiyonun başarımlarını değerlendirmek için dört uygulama yapılmıştır. Yapılan uygulamalarda karşılaştırma fonksiyonları olarak Sigmoid, Gumbel, ReLU ve Adaptive Gumbel fonksiyonları kullanılmıştır. Uygulamalarda MLP ve CNN sinir ağı modelleri kullanılmıştır. MLP ağı derin öğrenmede ikili sınıflandırma sınıf dengesizliği problemi için kullanılmıştır. CNN ağı ise derin öğrenmede görüntü sınıflandırma uygulamaları yapmak üzere tercih edilmiştir. Birinci uygulamada, önerilen fonksiyonun etkinliğini göstermek için MLP ağına 25 dengesiz veri kümesi kullanılmıştır. En yüksek AUC ortalamasını 0.9013 değeri ile SG elde etmiştir. İkinci uygulamada, önerilen fonksiyon CNN ağına MNIST veri kümesi kullanılarak Sigmoid ve Gumbel fonksiyonlarıyla karşılaştırılmıştır. En yüksek ortalama doğruluk değerini 0.9921 ile SG elde etmiştir. Üçüncü uygulamada, önerilen fonksiyonun üç farklı versiyonu karşılaştırılmıştır. Bunun için Fashion-MNIST veri kümesi CNN ağı üzerinde denenmiştir. En yüksek doğruluğu 0.9351 ortalama değeri ile SGv3 elde etmiştir. Dördüncü uygulamada, önerilen fonksiyon CNN ağına MNIST veri kümesi kullanılarak ReLU ve Adaptive Gumbel fonksiyonlarıyla karşılaştırılmıştır. En yüksek başarımlarını 0.9926 değeri ile SG tarafından elde edilmiştir. Yapılan deney sonuçlarına bakıldığında önerilen aktivasyon fonksiyonunun genel olarak daha başarılı olduğu görülmektedir.

## Sigmoid-Gumbel: A New Hybrid Activation Function

**Keywords:** Activation function, Deep neural network, Deep learning.

### Abstract

In this article, a new hybrid activation function called Sigmoid-Gumbel (SG) is proposed, which combines the positive aspects of the previously presented activation functions and performs better than them. Four applications were made to evaluate the performance of the proposed function. In the applications, Sigmoid, Gumbel, ReLU and Adaptive Gumbel functions were used as comparison functions. MLP and CNN neural network models were used in the applications. MLP network is used for binary classification class imbalance problem in deep learning. CNN network is preferred to perform image classification applications in deep learning. In the first application, 25 unbalanced datasets are used in the MLP network to demonstrate the effectiveness of the proposed function. SG had the highest mean AUC with a value of 0.9013. In the second application, the proposed function is compared with the

\*Sorumlu yazar: [mkaytan@harran.edu.tr](mailto:mkaytan@harran.edu.tr)

Geliş Tarihi: 03.09.2021, Kabul Tarihi: 04.03.2022

Sigmoid and Gumbel functions using the MNIST dataset in the CNN network. SG obtained the highest average accuracy value of 0.9921. In the third application, three different versions of the proposed function are compared. For this, the Fashion-MNIST dataset has been tested on the CNN network. SGv3 achieved the highest accuracy with an average value of 0.9351. In the fourth application, the proposed function is compared with ReLU and Adaptive Gumbel functions using MNIST dataset in CNN network. The highest performance was obtained by SG with a value of 0.9926. Considering the experimental results, it is seen that the proposed activation function is more successful in general.

## 1. Giriş

Yapay Sinir Ağları (YSA) tahmin, örüntü tanıma, sınıflandırma, fonksiyon yaklaşımı gibi problemleri farklı disiplinlerde kullanan hesaplamalı modelleme araçlarıdır. YSA'lar, ayarlanabilir parametreleri olan ve nöron olarak adlandırılan basit işlem birimlerinden oluşurlar. Nöronlar, veri işlemede paralel hesaplamaların yapılmasını sağlarlar. Öğrenme, biyolojik nöronlar arasındaki sinaptik bağlantılarla gerçekleştirilir. Yapay öğrenme ise yapay nöronlar arasında yapılan bazı ayarlamalarla gerçekleştirilir. YSA, biyolojik bir sistemin, genelleştirilmiş bulanık mantık, işlem kabiliyeti, öğrenme, hata ve hata toleransı, sağlamlık, yüksek düzeyde paralellik, doğrusal olmama (nonlinear) gibi bilgi işleme özelliklerinden faydalanmaktadır. YSA, çeşitli parametreleri kullanarak ve verilen örneklerden yinelemeler yaparak öğrenen bir matematiksel modelleme yöntemidir.

Sinir ağlarının dikkat çeken bazı yetenekleri şu şekilde sıralanabilir:

- ✓ Genelme: daha önce görmediği veri kümelerinde uygulanabilme
- ✓ Yüksek paralellik: donanım hata toleransını ve hızlı veri işlemeyi sağlayabilme
- ✓ Uyumluluk ve öğrenme: değişken dış ortamlarda iç yapısını güncelleyebilme
- ✓ Doğrusal olmama: veri kümelerine daha iyi uyum sağlayabilme
- ✓ Gürültü duyarsızlığı: ölçüm hatalarını ve belirsiz verileri daha iyi tahmin etme

Modelin öğrenmesi, ağırlık olarak adlandırılan parametrelerin güncellenmesiyle gerçekleştirilir. Modelin önemli hiper parametrelerinden biri, her bir nöronun ayrı ayrı gerçekleştirdiği aktivasyon fonksiyonudur. Modelin doğrusal olmama özelliği aktivasyon fonksiyonu ile gerçekleştirilir. Aktivasyon fonksiyonu seçimi sinir ağının başarımını büyük oranda değiştirebilir.

Standart Derin Sinir Ağı (DSA) modelleri sınıflandırmada çeşitli uygulamalarda başarıyla uygulanmıştır [1]. Ancak, dengesiz verilerden oluşan ve çarpık sınıf dağılımına sahip veri kümeleri standart derin öğrenme mimarilerinde düşük sınıflandırma başarımına sebep olmaktadır [2], [3]. Dengesiz veriler, veri kümesinde azınlık olarak adlandırılan bazı sınıfların çoğunluk olarak adlandırılan diğer sınıflara göre çok az yer alması şeklinde tanımlanabilir. Standart derin öğrenme mimarileri, azınlık örnekleri nadir olduğundan, azınlık sınıfları çoğunluk sınıflara kıyasla yanlış sınıflandırabilmektedir [4], [5].

Sınıf dengesizliği problemi, web reklamcılığı, dolandırıcılık tespiti ve tıbbi teşhis gibi birçok uygulamada görülen yaygın bir problemdir. Standart derin öğrenme yöntemleri, dengeli sınıfların olduğu veri kümelerinde yüksek başarımlar elde edebilmektedir. Ancak bu yöntemlerin başarımlarının dengesiz veri kümeleri söz konusu olduğunda düşük olduğu görülmektedir.

Aktivasyon fonksiyonu seçimi oldukça karmaşık bir süreçtir. Bir DSA'nın doğruluğu ve başarımı üzerinde birçok etkiye sahiptir. Keşfe dayalı basit bir seçim süreci bile yakınsama için kapsamlı deneme yanılma analizine ihtiyaç duyar. Yapılacak analiz sonucu bize yüksek kaliteli bir çözüm bulmayı garanti edemez.

Bu nedenlerden dolayı çalışmamızdaki motivasyon öğrenme eğrisinden yararlı bilgiler çıkarmanın etkili bir yolunu bulmak ve geliştirmektir. Ayrıca daha iyi bir başarımlar oranı sağlayabilecek aktivasyon fonksiyonlarını araştırmak ve tanımlamaktır. Eğitim sırasında temel aktivasyon fonksiyonlarının kombinasyonlarını otomatik olarak öğrenebilen iki yaklaşım incelenmiştir. Konveks kombinasyonu veya afin kombinasyonu olarak adlandırılan yaklaşımlarla aktivasyon fonksiyonları için bir arama uzayının belirlenebilmesi [6] de çalışmamızdaki motivasyonlardandır.

Bu makalede Sigmoid-Gumbel (SG) adlı yeni bir hibrit aktivasyon fonksiyonu önerilmiştir. Önerilen aktivasyon fonksiyonu, çalışmamızın

etkisini görmek için 4 farklı deneyde kullanılmıştır. Birinci deneyde ikili sınıflandırma sınıf dengesizliği problemi için çeşitli veri kümeleriyle değerlendirmeler yapılmıştır. Bir çok katmanlı algılayıcı (Multi Layer Perceptron-MLP) ağında bazı veri kümeleriyle başarımlar elde edilmiştir. Elde edilen başarımların diğer fonksiyonların başarımlarından daha yüksek olduğu gözlenmiştir. İkinci deneyde bir CNN modelinde MNIST veri kümesi kullanılmıştır. Önerilen aktivasyon fonksiyonu, Sigmoid ve Gumbel ile karşılaştırılmış ve başarımının yüksek olduğu gözlenmiştir. Üçüncü deneyde aynı CNN modelinde Fashion-MNIST veri kümesi kullanılmıştır. Önerilen aktivasyon fonksiyonunun farklı versiyonları olan SGv2, SGv3 ve SGv4 karşılaştırılmıştır. Bazı versiyonların başarımlarının diğerlerine göre yüksek olduğu görülmüştür. Dördüncü deneyde aynı CNN modelinde MNIST veri kümesi kullanılmıştır. Önerilen aktivasyon fonksiyonu, ReLU ve Adaptive Gumbel ile karşılaştırılmış ve başarımının yüksek olduğu gözlenmiştir.

Çalışmanın geri kalan kısımları şu şekilde düzenlenmiştir: Bölüm 2’de önceki çalışmalara yer verilmiştir. Bölüm 3’te derin öğrenme ile ilgili bazı temel bilgiler sunulmuştur. Bölüm 4’te önerilen yeni hibrit SG aktivasyon fonksiyonu anlatılmıştır. Bölüm 5’te materyal ve metot açıklanmıştır. Bölüm 6’da bulgular ve tartışma sunulmuştur. Bölüm 7’de sonuç ve öneriler kısmına yer verilmiştir.

## 2. Önceki Çalışmalar

Apicella ve arkadaşları [7] tarafından yapılan bir çalışmada yeni bir aktivasyon fonksiyonu oluşturmak için bir ağı az sayıda nöronla yerel alt ağlar ekleyen bir yöntem önerilmiştir. Önerilen yöntemin farklı bir aktivasyon fonksiyonuna ve daha fazla parametreye gerek duyulmadan daha iyi sonuçları olduğu görülmüştür.

Ding ve arkadaşları [8] tarafından yapılan bir çalışmayla yaygın olarak kullanılan aktivasyon fonksiyonlarının durumu ve gelişmeleri incelenmiştir. Aktivasyon fonksiyonlarının sinir ağları üzerindeki etkileri, avantajları ve dezavantajları tartışılmıştır. Farklı aktivasyon fonksiyonlarının başarımını karşılaştırmak için MNIST veri kümesi kullanılmıştır.

Babu ve Edla [9] tarafından sunulan bir çalışmada yeni bir parametrik cebirsel

aktivasyon (Parametric Algebraic Activation-PAA) fonksiyonu önerilmiştir. Önerilen fonksiyon, Elliott aktivasyon fonksiyonunun özel bir durumu olan genelleştirilmiş bir fonksiyon olarak düşünülmüştür. Önerilen fonksiyon, S-şekilli bir eğriler ailesi olan ve aktivasyon fonksiyonlarının tüm önemli özelliklerini karşılayan bir fonksiyon gibi tasarlanmıştır. Önerilen fonksiyonunun başarımı, çeşitli veri kümeleri kullanılarak yaygın olarak kullanılan aktivasyon fonksiyonlarıyla eğitim tur (epoch) sayısı ve test hatası açısından karşılaştırılmıştır. Yapılan çalışmalar sonucunda önerilen fonksiyonun başarılı olduğu gösterilmiştir.

Ramachandran ve arkadaşları [10] tarafından sunulan bir çalışmayla Swish adlı yeni bir aktivasyon fonksiyonu önerilmiştir. Yapılan deneylerle Swish’in bazı veri kümeleri üzerinde daha derin modellerde ReLU’dan daha iyi olduğu gösterilmiştir. ImageNet sınıflandırma doğruluğu için, ReLU birimleri Swish birimleriyle değiştirilip Mobile NASNet-A ve Inception-ResNet-v2’de test edilmiştir. Swish’in basitliği ve ReLU’ya olan benzerliğinden dolayı, herhangi bir sinir ağında ReLU birimlerinin Swish birimleriyle değiştirilmesiyle etkili sonuçların alındığı görülmüştür.

Xu ve arkadaşları [11] tarafından yapılan diğer bir çalışmada yaygın olarak kullanılan iki doymuş fonksiyon olan lojistik sigmoid ve tanh incelenmiştir. Lojistik fonksiyonun orijine yakın olan eğimi ve sıfır merkezli olmayan özelliği DSA eğitimini zorlaştırmıştır. Lojistik sigmoid doğru bir şekilde yeniden ölçeklendirilerek tanh ile karşılaştırılmıştır. Tanh, aynı yöntem takip edilerek ve negatif kısım cezalandırılarak geliştirilmiştir. Cezalandırılmış tanh’ın derin CNN ağlarında ReLU ve leaky ReLU gibi doymamış fonksiyonlardan daha iyi başarımlar sağladığı gösterilmiştir.

Agostinelli ve arkadaşları [12] tarafından yapılan diğer bir çalışmayla gradyan iniş kullanılarak her nöron için bağımsız öğrenen yeni bir parçalı doğrusal aktivasyon fonksiyonu tasarlanmıştır. Sunulan yeni aktivasyon fonksiyonu ile farklı DSA mimarileri geliştirilmiştir. CIFAR-10 ve CIFAR-100 üzerinde çok iyi başarımlar elde edilmiştir.

Zhou ve arkadaşları [13] tarafından sunulan farklı bir çalışmada Shape Autotuning Activation Function (SAAF) adlı yeni bir aktivasyon fonksiyonu önerilmiştir. Önerilen aktivasyon fonksiyonu, sigmoid ve tanh gibi pürüzsüz aktivasyon fonksiyonlarının ve ReLU ve çeşitleri gibi parçalı aktivasyon fonksiyonlarının avantajlarını kullanmış ve eksikliklerini gidermiştir. Sunulan aktivasyon fonksiyonu ile negatif bilgileri

yakalanmış ve sıfıra yakın bir ortalama çıktı sağlanmıştır. Bunun için bir uyarlamalı bağımsız eğitilebilir çift parametre ayarlanmıştır. Bu şekilde daha iyi bir başarımla elde edilmiş ve daha hızlı bir öğrenme gerçekleştirilmiştir. Önerilen aktivasyon fonksiyonu, ağ eğitimi sırasında çıktının daha kararlı bir dağılımını sağlamak için sınırlı çıktılar sunmuştur. Sunulan aktivasyon fonksiyonu, görüntü sınıflandırma, makine çevirisi ve üretken modelleme gibi çeşitli görevlerde uygulanan DSA'larda değerlendirilmiştir. Önerilen aktivasyon fonksiyonunun yapılan kapsamlı karşılaştırma çalışmasıyla son teknoloji aktivasyon fonksiyonlarından daha üstün olduğu gösterilmiştir.

Farhadî ve arkadaşları [14] tarafından sunulan farklı bir çalışmayla sinir ağının bir şekil parametresi üzerinden eğitilmesi önerilmiştir. Bu sayede her nöronun kendi aktivasyon fonksiyonunu ayarlaması ve nöron eğiliminin daha iyi bir tahmine uyarlanması sağlanmıştır. Bu değişiklik her nöron için geri yayılıma sadece bir denklem daha eklenmiştir. Aktivasyon fonksiyonları ile kümülatif dağılım fonksiyonları arasındaki ilişki incelenmiştir. Aktivasyon fonksiyonlarının eğriliğini ve pürüzsüzlüğünü incelemek için kapsamlı bir aktivasyon fonksiyonu sınıfı araştırması yapılmıştır. Bunun için Gumbel ve Sigmoid arasında bir köprü olarak Adaptive Gumbel aktivasyon fonksiyonu sunulmuştur. Sunulan aktivasyon fonksiyonu, ReLU'nun pürüzsüz bir türünü oluşturmak için önerilmiştir. Önerilen fonksiyon diğer bazı aktivasyon fonksiyonlarıyla karşılaştırılmış ve tahminlerin iyileştirilmesi sağlanmıştır.

Cai ve arkadaşları [15] tarafından yapılan bir çalışmada bazı yeni tür aktivasyon fonksiyonları önerilmiştir. Sunulan aktivasyon fonksiyonlarına sahip DSA'ların, geleneksel stokastik gradyan iniş (Stochastic Gradient Descent-SGD) ve ortalama normalleştirilmiş SGD ile eğitilebilecekleri belirtilmiştir. Önerilen aktivasyon fonksiyonlarının sınıflandırma sorunlarını çözümedeki başarımları incelenmiştir. TIMIT veri kümesiyle yapılan deneylerde, konuşma tanıma görevlerinde daha iyi başarımla sağlanmış ve fonem hata oranlarında iyileşme elde edilmiştir. Büyük eğitim kümeleriyle yapılan diğer deneylerde, ortalama normalleştirilmiş SGD ile eğitim süreçlerinin kolaylaştığı görülmüştür. Modeller, eğitim kümesi yeterince büyük olduğunda, ön eğitim

olmadan da doğrudan eğitilebilmiştir. Bunun sonucunda kelime hata oranlarında önemli bir iyileşme sağlanmıştır.

Koçak ve Şiray [16] tarafından yapılan bir çalışmayla bilinen aktivasyon fonksiyonlarının avantajlarını birleştiren ve onlardan daha iyi başarımla gösteren bazı yeni aktivasyon fonksiyonları tanıtılmıştır. Bu amaçla, generalized swish, mean-swish, ReLU-swish, triple-state swish, sigmoid-algebraic, triple-state sigmoid, exponential swish, sinc-sigmoid olarak adlandırılan ve sigmoid'in türevi olan bazı yeni aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonları iyi bilinen ve yakın zamanda önerilen bazı aktivasyon fonksiyonlarıyla karşılaştırılmıştır. Bu aktivasyon fonksiyonlarının başarımlarını araştırmak için simüle edilmiş veriler (simulated data), optik ara bağlantı ağı verileri (optical interconnection network data), arabaların özellikleri verileri (specifications of cars data) ve ortalama ev maliyetleri verileri (average house costs data) olmak üzere dört farklı veri kümesi kullanılmıştır.

### 3. Derin Öğrenme

Derin öğrenme, makine öğreniminin bir evrimi olan YSA'dan türemiştir. Derin ağ modelleri farklı katmanlardan oluşur. Derin öğrenme, karmaşık kavramları ve ilişkileri modellemede kullanılan ikiden fazla katmana sahip bir makine öğrenme algoritmasıdır [17]. Derin öğrenme, sosyal ağ filtreleme, görüntü tanıma, konuşma tanıma ve doğal dil işleme gibi farklı kategorilerde araştırılmaktadır. Derin öğrenme algoritmaları, farklı kaynaklardan elde edilen büyük veriler arasında korelasyonlar bulma yeteneğine sahiptir. Ancak, özellik öğrenme ve sınıflandırma veya kümelemeyi eşzamanlı olarak yapma yeteneklerinde farklıdır [18]. Derin öğrenme mimarileri, üretken (generative), ayırt edici (discriminative) ve hibrit (hybrid) mimariler olmak üzere üç ana kategoriye ayrılabilir [19].

#### 3.1. Aktivasyon Fonksiyonları

Biyolojik sinir ağında bir nöronun

$$z_i = g(x, w_i)$$

şeklinde yaptığı işlem bir aktivasyon fonksiyonudur. Bu işlemin çıktısı olan ve diğer bir nörona aktarılan genel aktivite ise

$$y_i = f(z_i)$$

şeklinde ifade edilebilir. Bir nöronun ateşleme oranına göre iki durumu vardır. Nöronu ateşleyebilir veya ateşlemeyebilir. YSA'da bir nöronun önemli bir bileşeni olan aktivasyon fonksiyonu, ağıın başarımını etkiler ve nöronun çıktısını değiştirir. Step, Tanh ve Sigmoid en çok kullanılan aktivasyon fonksiyonlarıdır. Çok katmanlı ileri beslemeli bir ağda farklı aktivasyon fonksiyonlarının bir karışımı uygulanabilir. Literatürde YSA'da çok farklı aktivasyon fonksiyonlarının kullanıldığı görülmektedir.

### 3.1.1. Sigmoid Aktivasyon Fonksiyonu

Lojistik aktivasyon fonksiyonu olarak da bilinir. YSA'da bir eşik fonksiyonu olarak kullanılır. S şeklinde bir eğri oluşturur. Lojistik regresyon, Gauss olasılık dağılımına bir yaklaşım olarak ve bulanık mantık teorisinde bir üyelik fonksiyonu olarak kullanılır. 0 ile 1 arasında sayılar üretir.

Matematiksel model Denklem (1)'de gösterilmiştir.

$$y = \frac{1}{1 + e^{-x}} \quad (1)$$

Denklem (2)'de Denklem (1)'in türevi gösterilmiştir.

$$\frac{d}{dx}y = \frac{e^x}{(1 + e^x)^2} \quad (2)$$

### 3.1.2. Gumbel Aktivasyon Fonksiyonu

Gumbel dağılımı, genelleştirilmiş aşırı değer (Generalized Extreme Value-GEV) dağılımı Tip-I olarak da adlandırılır. Farklı türdeki dağılımların aşırı örneklerinin dağılımını tasarlamak için kullanılır. Yaygın bir kullanım alanına sahiptir [20]. Gumbel fonksiyonu, ana dağılım veya asimptotik bir yaklaşım olarak geniş bir kullanım alanına sahiptir. Örnek olarak; aşırı rüzgar hızları, seller, ölüm yaşı, minimum veya maksimum sıcaklık, malzemelerin elektriksel dayanımı, finansal risk değerlendirmesi, jeolojik problemler verilebilir [21]. Gumbel dağılım fonksiyonu sürekli olarak türevlenebilme özelliğine sahiptir. Bu özelliğinden dolayı SGD optimizasyonu ile sinir ağı modellerinde bir aktivasyon fonksiyonu gibi kullanılabilir. Asimetriden dolayı yanlış sınıflandırmayı her iki sınıfta da farklı şekilde cezalandırmaktadır [22].

Denklem (3)'te Gumbel için CDF gösterilmiştir.

$$y = e^{-e^{-x}} \quad (3)$$

Denklem (4)'te Denklem (3)'ün türevi gösterilmiştir.

$$\frac{d}{dx}y = e^{-e^{-x}-x} \quad (4)$$

## 4. Önerilen Yeni Bir Hibrit Sigmoid-Gumbel Aktivasyon Fonksiyonu

Bu çalışmada, derin sinir ağlarında ilgi odağı haline gelen kaybolan gradyan probleminin (vanishing gradient problem) üstesinden gelmek için yeni bir hibrit Sigmoid-Gumbel aktivasyon fonksiyonu önerilmiştir. Derin sinir ağlarında, kaybolan gradyan problemi nedeniyle ağlar çok az güncellenir veya kullanılamaz hale gelir. Bu süreç sinir ağıının derinleşmesini ve öğrenmesini engeller. Modelin eğitim için çok fazla hesaplama gücü kullanmasına neden olabilir. Dolayısıyla bu, derin öğrenmede yerel minimumlara düşmeye neden olabilir [23]. Kaybolan gradyan problemi Hochreiter [24] tarafından keşfedilmiştir. Bu sorun, kademeli derin sinir ağlarında her işlem adımında çok sayıda türev alındığında meydana gelir. Derin sinir ağı derinleşmeye başladıkça türevler küçüldükçe gradyan kaybolur. Bu da modelin öğrenme sürecinde bir düşüşe neden olur. Örneğin derin sinir ağlarında yaygın olarak kullanılan sigmoid ve tanjant aktivasyon fonksiyonlarını kullanan modellerde kaybolan gradyan problemi gözlenmiştir [25]. Kaybolan gradyan probleminin üstesinden gelebilmek için türevlenebilir fonksiyonlarla değiştirilmesi gerekir. Böylece sürekli bir öğrenme süreci sağlanır ve hatalar geri yayılır. Kaybolan gradyan probleminin üstesinden gelmek için çok sayıda çalışma yapılmıştır.

Doğrusal olmayan farklı tipte birçok fonksiyon önerilmiştir. Sigmoid, bilinen en iyi doğrusal olmayan fonksiyondur. Girdi gerçek değerini 0 ile 1 arasında normalleştirir. Gizli katmanlardan giriş katmanlarına doğru geri yayılım aşamasında sıfır merkezli olmayan çıktı, yavaş yakınsama ve gradyan doygunluğu gibi bazı dezavantajlara sahiptir. Bu nedenlerden dolayı gradyan güncellemeleri farklı yönlerde yayılır. Bu da istenmeyen bir durumdur. Hiperbolik Tanjant (tanh), gerçek değerli bir sayıyı -1 ile 1 arasındaki bir aralığa sıkıştırır. Sigmoid'teki gibi kaybolan gradyan problemine sahiptir. Hesaplama sırasında bazı ölü nöronlar üretir. Ölü nöron üretilmesi,

aktivasyon ağırlığının sıfır gradyan sonucu nadiren kullanıldığı olumsuz bir durum olduğunu gösterir. Sigmoid'in aksine çıktısı sıfır merkezlidir ve pratikte sigmoid'e tercih edilir. Rectified Linear Unit (ReLU) [26], [27], [28], Sigmoid'e kıyasla kolayca aşırı uyum (overfitting) gösteren bir sınırlamaya sahiptir. Eğitim sırasında bazen kırılabilir olabilmektedir. Bundan dolayı bazı gradyanların ölmesine neden olmaktadır. Bu da önemli bir sınırlamadır. Bu sınırlama bazı nöronların ölmesine neden olur. Ölü nöronlardan dolayı ağırlık güncellemeleri gelecekteki veri noktalarında etkinleştirilmez. Bu şekilde sıfır aktivasyon sağlanmış olduğundan öğrenme engellenmiş olur [29]. Maxout, ReLU'nun bir genellemesi olarak düşünülmüş ve önerilmiştir [30], [31]. Tüm nöronlarda kullanılan parametreleri iki katına çıkararak ve bundan dolayı hesaplama açısından pahalı olan olumsuz bir özelliğe sahiptir. Bundan dolayı hesaplanacak parametrelerin sayısı artar. Ampirik çalışmalar, doğrusal olmayan durumlar için, optimizasyonu daha kolay ve yakınsaması daha hızlı doğrusal olmayan fonksiyonların aktivasyon fonksiyonu olarak tercih edildiğini kanıtlamıştır. Şekil 1'de doğrusal olmayan bazı fonksiyonlara ait grafikler gösterilmektedir.

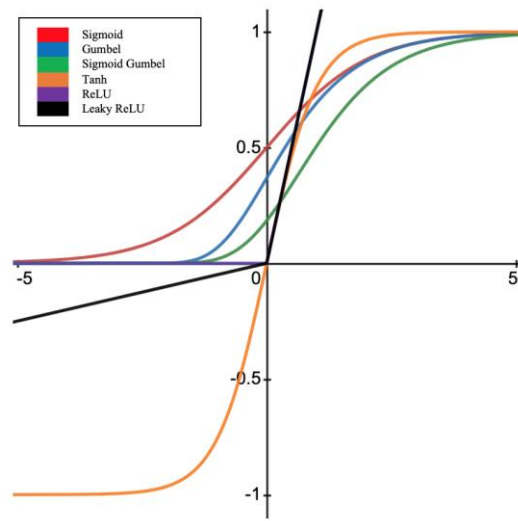
Sigmoid fonksiyonunun sınıf dengesizliği probleminde azınlık sınıf olasılığını ihmal etmesi ve simetrik bir fonksiyon olması gibi bazı olumsuz yanları vardır. Sigmoid'in bu olumsuz yanlarından dolayı Gumbel dağılımı bir aktivasyon fonksiyonu olarak düşünülmüştür [32]. Gumbel dağılımı, ikili sınıflandırma sınıf dengesizliği problemi başarımını iyileştirmek için bir aktivasyon fonksiyonu olarak düşünülmüştür.

Bir aktivasyon fonksiyonunun aşağıdaki özelliklere sahip olması istenir:

- ✓ Fonksiyon her noktada veya hemen hemen her noktada sürekli ve türevlenebilir olmalıdır.
- ✓ Fonksiyonun türevi doymuş olmamalıdır. Bu, beklenen giriş aralığı üzerinde sıfıra doğru çok küçük bir eğilime sahip olma demektir. Çok küçük türevler öğrenme sürecini durdurma eğilimindedir. Bu durum, kaybolan gradyan problemi (vanishing gradient problem) olarak bilinir.

- ✓ Fonksiyonun türevi patlamamalıdır. Bu, sayısal istikrarsızlık sorunlarına yol açan sonsuza doğru çok büyük bir eğilim anlamına gelmektedir. Bu durum, patlayan gradyan problemi (exploding gradient problem) olarak bilinir.

Yukarıdaki nedenlerden dolayı bu makalede Sigmoid-Gumbel olarak adlandırılan yeni bir aktivasyon fonksiyonu önerilmiştir. Önerilen aktivasyon fonksiyonunun yukarıdaki özelliklere sahip olduğu ve yapılan çeşitli deneylerdeki başarımlarını aşağıda devam eden bölümlerde anlatılmıştır. Şekil 1'de doğrusal olmayan bazı fonksiyonlara ait grafikler gösterilmektedir.



Şekil 1. Doğrusal olmayan bazı fonksiyonlara ait grafikler

#### 4.1. Sigmoid-Gumbel Aktivasyon Fonksiyonu

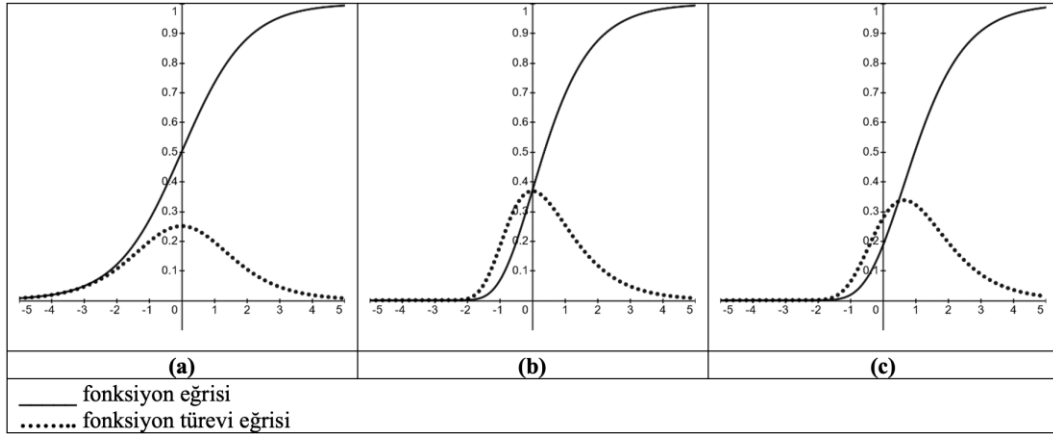
Matematiksel model Denklem (5)'te gösterilmiştir.

$$y = \left( \frac{1}{1 + e^{-x}} \right) \left( e^{-e^{-x}} \right) \quad (5)$$

Denklem (6)'da Denklem (5)'in türevi gösterilmiştir.

$$\frac{d}{dx} y = \frac{(1 + 2e^x)e^{-e^{-x}}}{(1 + e^x)^2} \quad (6)$$

Şekil 2'de (a) Sigmoid, (b) Gumbel ve (c) Sigmoid-Gumbel fonksiyonlarının ve türevlerinin eğrileri gösterilmiştir.



Şekil 2. (a) Sigmoid, (b) Gumbel ve (c) Sigmoid-Gumbel fonksiyonlarının ve türevlerinin eğrileri.

## 5. Materyal ve Metot

Bu çalışma kapsamında gerçekleştirilen deneyler 2,8 GHz Quad-Core Intel Core i7 işlemciye, 16 GB 1600 MHz DDR3 belleğe ve AMD Radeon R9 M370X 2 GB ekran kartına sahip bir makinede gerçekleştirilmiştir.

Ayrıca, çalışma kapsamında kullanılan otomatik kodlayıcının amacı, azınlık sınıfını açıklayabilen verimli özellikler elde etmektir. Bunun için yeniden yapılandırılmış girdiyi kullanır. Zaman serileri aşırı olay tahmini [33], denetimsiz anomali tespiti için verimli özellikler üretme [34] gibi çeşitli çalışmalarda başarıyla uygulanmıştır. Bir  $x$  örneği verildiğinde, yeniden oluşturulmuş  $x'$  girişini  $z_e = f_e(x; \theta_e)$  ve  $x' = f_d(z_e; \theta_d)$  ile hesaplar.  $z_e$ , kodlayıcı tarafından öğrenilen azaltılmış düşük boyutlu kodlanmış özelliklerdir.  $x'$ , yeniden yapılandırılmış girdidir.  $f_e$  kodlama ağıdır.  $f_d$  kod çözme ağıdır.  $\theta_e$ ,  $\theta_d$  otomatik kodlayıcının ağırlık parametreleridir. Ardından, orijinal ve yeniden yapılandırılmış girdiler kullanılarak azınlık sınıfı için bazı verimli özellikler  $z_d = d(x, x')$  ile üretilmiştir.  $z_d$ , orijinal ve yeniden yapılandırılmış girdiden oluşturulan özelliklerdir.  $d(*)$ , verimli özelliklerin hesaplanmasında mesafe fonksiyonudur. Çalışmada Öklid ve kosinüs mesafeleri kullanılmıştır [32].

### 5.1. Deney-1

Bu deneyde sinir ağı mimarisi olarak MLP kullanılmıştır. Tahmin doğruluğunu değerlendirmek ve karşılaştırmak için 5 kat çapraz doğrulama yapılmıştır. Tüm veri kümeleri için aynı ağ modeli kullanılmıştır. Sinir ağı mimarisinde 28 nöronlu altı gizli katman kullanılmıştır. Hiper parametre ayarları

için öğrenme oranı 0.001, eğitim tur (epoch) sayısı 500, yığın boyutu (batch size) 8, kayıp ağırlık (loss weight- $\lambda$ ) 0,05 olarak belirlenmiştir. İkili sınıflandırma sınıf dengesizliği problemi için dengesiz veri kümeleri kullanılmıştır. Karşılaştırma fonksiyonları olarak Sigmoid ve Gumbel fonksiyonları değerlendirilmiştir. Referans çalışmaya [32] uyum için seçilen parametreler ve değerleri aynen kullanılmıştır. Burada değerlendirme ölçütü olarak en çok kullanılan ölçüt olan Area Under Curve (AUC) [35] kullanılmıştır.

### 5.2. Deney-2

Derin öğrenmede aşırı uyum (overfitting) ciddi bir sorundur. Test sırasında birçok farklı ağı tahminlerini birleştirerek aşırı uyum ile başa çıkması zorlaşır. Bundan dolayı derin ağların kullanımının yavaşlaması sorunu oluşur. Düşürme (dropout), bu sorunu çözmek için kullanılan bir yöntemdir. Bu yöntemde eğitim sırasında birimler bağlantılarıyla birlikte ağdan rastgele düşürülür. Böylece birimlerin aşırı uyumu önlenir. Düşürme, sinir ağlarının bilgisayarlı görme, konuşma tanıma, belge sınıflandırma ve hesaplamalı biyolojideki denetimli öğrenme görevlerinde başarılı bir şekilde uygulanmıştır. Ayrıca birçok kıyaslama veri kümesinde en gelişmiş sonuçları da elde etmiştir [36]. Bu nedenlerden dolayı bu deneydeki modelde düşürme (dropout) yöntemi kullanılmıştır.

Ayrıca MNIST veri kümesi kullanılarak ağı başarımı incelenmiştir. Şekil 3'te kullanılan CNN modeli özeti gösterilmiştir.

<b>Amaç:</b> El yazısı rakamların tanınması, Moda ürün görüntülerinin tanınması
<b>Veri kümesi:</b> MNIST Fashion-MNIST
<b>Algoritma:</b> Evrişimli Sinir Ağları (Convolutional Neural Networks)
<b>Döngü (epoch):</b> 20
<b>Ağ mimarisi:</b> - 32 x 3 x 3 Convolution - 64 x 3 x 3 Convolution - 2 x 2 Max Pooling - Dropout (%25) - 128 Fully Connected - Dropout (%50) - 10 Fully Connected

Şekil 3. CNN modeli özeti

### 5.3. Deney-3

Şekil 3'teki ağ mimarisi Fashion-MNIST veri kümesi üzerinde uygulanmıştır. Sunulan bir çalışmada [15] DSA'lar için çok durumlu aktivasyon fonksiyonları önerilmiştir. Önerilen aktivasyon fonksiyonlarının ikiden fazla durumu temsil edebilen yeni tür aktivasyon fonksiyonları olduğu söylenmiştir. Yapılan bir çalışmada [16] daha önceden tanımlı bazı aktivasyon fonksiyonlarının avantajlarını birleştiren ve onlardan daha iyi başarımlar sağlayan bazı yeni aktivasyon fonksiyonları önerilmiştir. Bu çalışmada ise SG için SGv2, SGv3 ve SGv4 versiyonları türetilmiştir.

Türevlenebilirlik ve süreklilik ile ilgili olarak bir fonksiyon bir noktada türevlenebilirse orada da süreklidir denebilir. Tersine bir fonksiyon bir noktada sürekli değilse orada türevlenemeyeceği söylenebilir. Bu bize bir fonksiyonun türevlenemediği yerleri tanımlamamızı sağlar. Farklı süreksizlik türleri

olsa da süreksizlik bir fonksiyonun türevlenemez olmasına neden olur. Bir  $f(x)$  fonksiyonunun  $x=x_0$  noktasında türevlenememesi ile ilgili üç durum olabilir.

- ✓  $f(x)$ ,  $x = x_0$ 'da sürekli değildir.
- ✓  $f(x)$ ,  $x = x_0$ 'da keskin bir noktası vardır.
- ✓  $f(x)$ ,  $x = x_0$ 'da dikey bir teğet çizgisine sahiptir.

Burada türetilen aktivasyon fonksiyonlarındaki katsayılar ve değerler arama uzayından örnek olarak seçilmiştir. Bunun amacı farklı parametre değerleriyle daha iyi bir başarımlar sağlanıp sağlanmadığını göstermektir. Ayrıca bu alanda çalışacak araştırmacılara farklı bakış açıları sunmaktır. Dengesizlik oranı farklı veri kümelerinde parametrelerin belirlenmesiyle ilgili daha ayrıntılı bir çalışmayı bir sonraki çalışmada yapmayı düşünüyoruz.

Denklem (7)'de SGv2 için matematiksel model gösterilmiştir.

$$y = \left( \frac{1}{1 + e^{-x}} \right) \left( e^{-e^{-(x-2.7)}} \right) \quad (7)$$

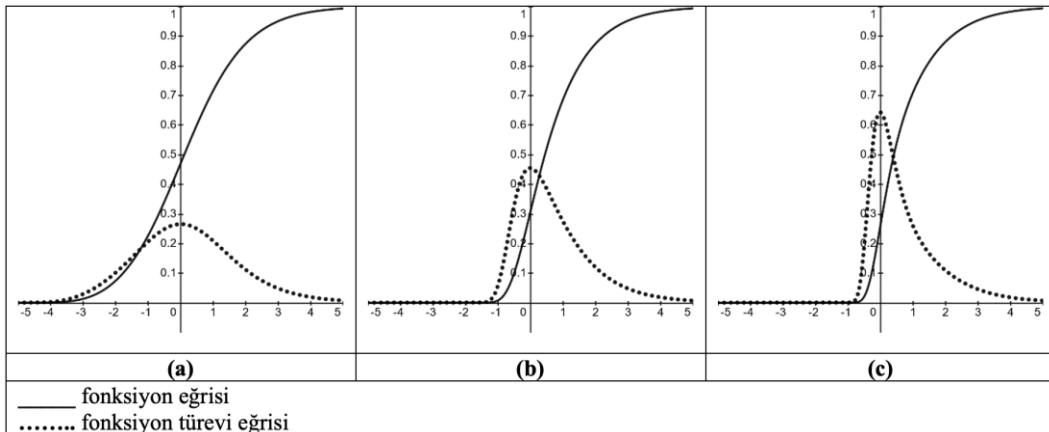
Denklem (8)'de SGv3 için matematiksel model gösterilmiştir.

$$y = \left( \frac{1}{1 + e^{-x}} \right) \left( e^{-e^{-(2x-0.7)}} \right) \quad (8)$$

Denklem (9)'da SGv4 için matematiksel model gösterilmiştir.

$$y = \left( \frac{1}{1 + e^{-x}} \right) \left( e^{-e^{-(3x-0.4)}} \right) \quad (9)$$

Şekil 4'te (a) SGv2, (b) SGv3 ve (c) SGv4 fonksiyonlarının ve türevlerinin eğrileri gösterilmiştir.



Şekil 4. (a) SGv2, (b) SGv3 ve (c) SGv4 fonksiyonlarının ve türevlerinin eğrileri.



#### 5.4. Deney-4

Yapılan bir çalışmada [13] belirtildiği gibi aktivasyon fonksiyonu seçimi, son teknoloji ürünü sinir ağı modelleri oluşturmak için çok önemli bir konudur. Şu anda etkin olan ve en yaygın kullanılan aktivasyon fonksiyonu ReLU'dur. Ancak ReLU'nun, sıfır olmayan ortalama, negatif kayıp ve sınırsız çıktı gibi bazı zayıf yönleri vardır. Bu nedenlerden dolayı optimizasyon sürecinde potansiyel dezavantajları görülmüştür.

Yapılan bir çalışmada [14] Gumbel ve Sigmoid arasında bir köprü olarak düşünülen ve Adaptive Gumbel olarak adlandırılan bir aktivasyon fonksiyonu sunulmuştur. Denklem (10)'da Adaptive Gumbel için matematiksel model gösterilmiştir.

$$\sigma_{\alpha}(x) = 1 - \{1 + \alpha \exp(x)\}^{-\frac{1}{\alpha}} \quad (10)$$

$$\alpha \in \mathbb{R}^+, x \in \mathbb{R}$$

Bu çalışmada önerilen SG'nin etkinliğini görmek için ReLU ve Adaptive Gumbel ile karşılaştırma yapılmıştır. Denklem (10)'da  $\alpha=1$  olarak alınmıştır. Şekil 3'te yer alan

ağ modeli MNIST veri kümesi üzerinde uygulanmıştır.

#### 5.5. Dengesiz Veri Kümeleri

Dengesiz veri kümeleri [37], sınıf dağılımının sınıflar arasında tekdüze olmadığı sınıflandırma problemi için özel bir durumdur. Tipik olarak çoğunluk (negatif) sınıfı ve azınlık (pozitif) sınıfı olmak üzere iki sınıftan oluşurlar. Standart sınıflandırma algoritmaları genellikle dengeli bir eğitim kümesini göz önünde bulundurur. Bu durum çoğunluk sınıfına karşı bir eğilim oluşturur. Bundan dolayı bu tür kümeler veri madenciliği için yeni ve zorlu bir problem olduğunu varsayar. Sunulan tüm dengesiz veri kümeleri, 5 katlı katmanlı çapraz doğrulama kullanılarak bölünmüştür. Test bölümlerinde yeterli miktarda azınlık sınıfı örneğinin atılması için veri kümesi 5 kate bölünmüştür. Bu şekilde, test bölümü örneklerinin, temel bilgileri daha iyi temsil etmesi sağlanmıştır. SG'yi değerlendirmek ve karşılaştırmak için Keel veri kümesi havuzundan alınan 25 gerçek dünya karşılaştırmalı dengesiz veri kümesi kullanılmıştır. Dengesiz veri kümeleri, değişken sayısı ve dengesizlik oranı farklı olan çeşitli örneklerden oluşmaktadır. Tablo 1'de dengesiz veri kümeleriyle ilgili özet bilgiler gösterilmiştir.

**Tablo 1.** Dengesiz veri kümelerinin özet bilgileri

Veri kümeleri	Örnek sayısı	Değişken sayısı	Dengesizlik oranı
ecoli0vs1	220	7	1.86
ecoli0137vs26	281	7	39.14
ecoli0146vs5	280	6	13
ecoli0147vs2356	336	7	10.59
ecoli0234vs5	202	7	9.1
ecoli034vs5	200	7	9
ecoli0346vs5	205	7	9.25
ecoli067vs5	220	6	10
ecoli1	336	7	3.36
ecoli2	336	7	5.46
ecoli3	336	7	8.6
haberman	306	3	2.78
iris0	150	4	2
led7digit02456789vs1	443	7	10.97
pima	768	8	1.87
wisconsin	683	9	1.86
yeast0256vs3789	1004	8	9.14
yeast02579vs368	1004	8	9.14
yeast0359vs78	506	8	9.12
yeast05679vs4	528	8	9.35
yeast1	1484	8	2.46
yeast2vs4	514	8	9.08
yeast2vs8	482	8	23.1
yeast3	1484	8	8.1
yeast5	1484	8	32.73

## 5.6. MNIST Veri Kümesi

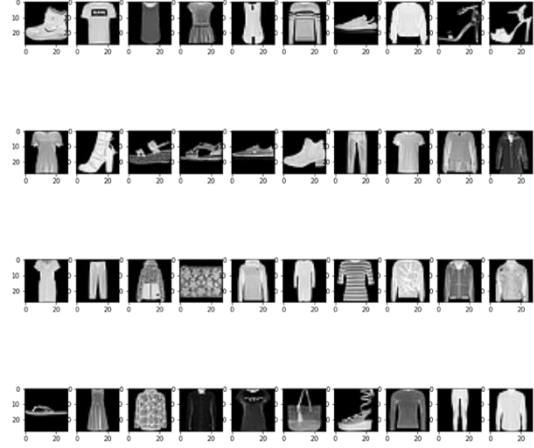
MNIST [38], derin öğrenme modellerini kıyaslamak için yaygın olarak kullanılan standart veri kümelerinden biridir. El yazısı rakamlarının 28x28 gri tonlamalı görüntülerinden oluşur. Eğitim veri kümesi 60.000, test veri kümesi 10.000 olmak üzere toplam 70.000 örnekten oluşur. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 rakamlarını temsil eden 10 sınıftan oluşur. Şekil 5'te MNIST veri kümesine ait bazı örnekler gösterilmiştir.



Şekil 5. MNIST veri kümesinin bir görüntüsü

## 5.7. Fashion-MNIST Veri Kümesi

Fashion-MNIST [39], geleneksel MNIST'e alternatif olarak sunulan ve yaygın olarak kullanılan bir veri kümesidir. 28x28 gri tonlamalı örneklerden oluşan bir moda ürün görüntüleri veri kümesidir. Eğitim veri kümesi 60.000, test veri kümesi 10.000 olmak üzere toplam 70.000 örnekten oluşur. Her örnek, 10 sınıftan bir etikete sahip bir görüntüdür. Fashion-MNIST, makine öğrenimi yöntemlerini ve algoritmalarını değerlendirmek için orijinal MNIST veri kümesinin yerine kullanılmak üzere hazırlanmıştır. Bu nedenle bu veri kümesi MNIST ile aynı görüntü boyutuna, eğitim ve test bölme yapısına sahiptir. Şekil 6'da Fashion-MNIST veri kümesine ait bazı örnekler gösterilmiştir.



Şekil 6. Fashion-MNIST veri kümesinin bir görüntüsü

## 6. Bulgular ve Tartışma

Bu çalışmada, Sigmoid ve Gumbel temel fonksiyon öğelerini kullanan yeni bir hibrit aktivasyon fonksiyonu tanıtılmıştır. Bunun için fonksiyon öğelerinin bir doğrusal kombinasyonu oluşturulmuştur. Doğrusal kombinasyonun öğrenme prosedürünün gerçekleşmesi için uygun bir hipotez alanının tanımlanması gerekir. Hipotez alanının tanımlanması konveks veya afin gibi yöntemlerle sağlanabilir. Tüm deneylerimizde bu çalışmada önerilen yöntemlerle çeşitli başarımlar elde edilmiştir. Yaklaşımlarımızı kullanan birleşik fonksiyon öğeleri, temel fonksiyon öğelerinden genellikle daha iyi başarımlar göstermiştir. Araştırmacılar tarafından önerilen yöntemin etkinliği farklı derinliklere ve mimarilere sahip ağ modelleri kullanılarak test edebilir. Ayrıca diğer temel fonksiyon kümelerinin de analiz edilmesi ilgi çekici sonuçların elde edilmesini sağlayabilir.

Tablo 2'de dengesiz veri kümeleri kullanılarak yapılan 1.deney sonuçları görülmektedir. Çalışmamız ile referans çalışmanın [32] başarımları karşılaştırmalı veri kümeleriyle önerilen aktivasyon fonksiyonunun üstünlüğünü göstermek için doğrudan karşılaştırılmıştır. Başarımlar 25 karşılaştırmalı dengesiz veri kümesinde AUC açısından değerlendirilmiştir. Önerilen aktivasyon fonksiyonu, Keel veri kümesi havuzundan alınan 25 veri kümesinden 11 tanesinde daha yüksek başarımlar, 1 tanesinde aynı başarımlar ve 13 tanesinde ise daha düşük başarımlar elde etmiştir. Ortalama olarak 0.9013 değerinde bir başarımlar gözlenmiştir. Özetlemek gerekirse, önerilen fonksiyonun diğer modellerden genel olarak daha iyi bir başarımlar gösterdiği söylenebilir. Oldukça dengesiz veri kümelerindeki başarımlara bakıldığında, SIG-NN, Sigmoid fonksiyonun simetrik doğası nedeniyle GEV-NN ve SG'den daha düşük bir başarımlar

göstermiştir. Öte yandan SG, GEV-NN'ye göre daha yüksek bir başarımla elde etmiştir. Önerilen SG aktivasyon fonksiyonunun ikili

sınıflandırma sınıf dengesizliği problemi için uygulanabilir bir yaklaşım olduğu kanıtlanmıştır.

**Tablo 2.** Dengesiz veri kümeleri ile yapılan deney sonuçları

Veri kümeleri	AUC		
	SIG-NN	GEV-NN	SG
ecoli0vs1	<b>0.9954</b>	0.9905	0.9940
ecoli0137vs26	<b>0.8890</b>	0.8260	0.8327
ecoli0146vs5	<b>0.9452</b>	0.9245	0.9212
ecoli0147vs2356	<b>0.9382</b>	0.9377	0.8834
ecoli0234vs5	0.9191	0.8934	<b>0.9309</b>
ecoli034vs5	0.9146	<b>0.9444</b>	0.8917
ecoli0346vs5	0.8959	0.8635	<b>0.9568</b>
ecoli067vs5	0.8900	<b>0.9038</b>	0.8950
ecoli1	0.9436	0.9539	<b>0.9572</b>
ecoli2	0.9485	<b>0.9503</b>	0.9311
ecoli3	0.9239	0.9243	<b>0.9301</b>
haberman	0.6629	0.6530	<b>0.7275</b>
iris0	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>
led7digit02456789vs1	0.9549	0.9505	<b>0.9570</b>
pima	0.8285	<b>0.8311</b>	0.8079
wisconsin	0.9947	0.9929	<b>0.9950</b>
yeast0256vs3789	0.8464	0.8155	<b>0.8490</b>
yeast02579vs368	0.9338	0.9313	<b>0.9378</b>
yeast0359vs78	0.7805	<b>0.8065</b>	0.7881
yeast05679vs4	<b>0.8650</b>	0.8483	0.8361
yeast1	<b>0.8040</b>	0.8010	0.7906
yeast2vs4	0.9408	<b>0.9483</b>	0.9243
yeast2vs8	0.7345	0.7922	<b>0.8394</b>
yeast3	0.9699	0.9703	<b>0.9716</b>
yeast5	<b>0.9901</b>	<b>0.9901</b>	0.9848
<b>Win/Draw/Loss</b>	<b>6/2/17</b>	<b>6/2/17</b>	<b>11/1/13</b>
<b>ORTALAMA</b>	<b>0.9004</b>	<b>0.8977</b>	<b>0.9013</b>

Tablo 3'te MNIST veri kümesi kullanılarak yapılan 2.deney sonuçları görülmektedir. Ağ modeli 5 defa çalıştırılıp ortalama test kaybı ve test doğruluğu değerleri

elde edilmiştir. Buna göre en iyi sonucu 0.0429 test kaybı ve 0.9921 test doğruluğu ortalama değerleri ile SG elde etmiştir.

**Tablo 3.** Sigmoid, Gumbel ve SG'ye ait deney sonuçları

Aktivasyon Fonksiyonu	#	Test loss	Test accuracy
Sigmoid	1	0,0502	0,9916
	2	0,0465	0,9917
	3	0,0576	0,9906
	4	0,0554	0,9912
	5	0,0480	0,9916
	ORTALAMA		0,0515
Gumbel	1	0,0451	0,9916
	2	0,0484	0,9921
	3	0,0478	0,9910
	4	0,0458	0,9924
	5	0,0468	0,9907
	ORTALAMA		0,0468
SG	1	0,0435	0,9923
	2	0,0449	0,9922
	3	0,0417	0,9923
	4	0,0422	0,9914
	5	0,0420	0,9923
	ORTALAMA		<b>0,0429</b>

Tablo 4'te Fashion-MNIST veri kümesi kullanılarak yapılan 3.deney sonuçları görülmektedir. Ağ modeli, SG, SGv2, SGv3 ve SGv4 fonksiyonları için 10 kez çalıştırılıp test

kaybı ve test doğruluğu ortalama değerler elde edilmiştir. Buna göre en yüksek sonucu 0.2363 test kaybı ve 0.9351 test doğruluğu ortalama değerleri ile SGv3 fonksiyonu elde etmiştir.

**Tablo 4.** SG, SGv2, SGv3 ve SGv4'ye ait deney sonuçları

#	Test loss				Test accuracy			
	SG	SGv2	SGv3	SGv4	SG	SGv2	SGv3	SGv4
1	0,2371	0,2442	0,2346	0,2424	0,9326	0,9301	0,9335	0,9356
2	0,2420	0,2371	0,2357	0,2384	0,9347	0,9321	0,9371	0,9353
3	0,2385	0,2495	0,2394	0,2465	0,9359	0,9337	0,9384	0,9332
4	0,2359	0,2461	0,2400	0,2316	0,9352	0,9302	0,9345	0,9334
5	0,2306	0,2445	0,2341	0,2502	0,9340	0,9344	0,9331	0,9328
6	0,2410	0,2470	0,2314	0,2361	0,9350	0,9313	0,9339	0,9309
7	0,2383	0,2473	0,2424	0,2290	0,9350	0,9338	0,9343	0,9360
8	0,2373	0,2490	0,2374	0,2359	0,9365	0,9310	0,9355	0,9345
9	0,2398	0,2410	0,2411	0,2456	0,9327	0,9306	0,9348	0,9346
10	0,2439	0,2487	0,2268	0,2381	0,9320	0,9330	0,9358	0,9357
ORT	0,2384	0,2454	<b>0,2363</b>	0,2394	0,9344	0,9320	<b>0,9351</b>	0,9342

Tablo 5'te MNIST veri kümesi kullanılarak yapılan 4.deney sonuçları görülmektedir. Ağ modeli 5 defa çalıştırılıp ortalama test doğruluğu değerleri elde edilmiştir. Buna göre en yüksek sonucu 0.9926 test doğruluğu ortalama değeri ile SG elde etmiştir.

**Tablo 5.** ReLU, AdaptiveGumbel ve SG'ye ait deney sonuçları

#	Test accuracy		
	ReLU	AdaptiveGumbel	SG
1	0,9724	0,9922	0,9929
2	0,9771	0,9912	0,9926
3	0,9702	0,9910	0,9928
4	0,9719	0,9917	0,9922
5	0,9738	0,9901	0,9925
ORT	0,9731	0,9912	<b>0,9926</b>

Tüm deney sonuçlarına ait özet bilgiler Tablo 6'da verilmiştir. Dengesiz veri kümeleriyle yapılan 1.deneyde en yüksek başarımın 0.9013 ile Sigmoid Gumbel'e ait olduğu görülmektedir. MNIST veri kümesi üzerinde gerçekleştirilen 2.deneyde ise yine en iyi başarımı 0.9921 test doğruluğuyla önerilen fonksiyon almıştır. Fashion MNIST veri kümesiyle yapılan 3.deneyde ise önerilen

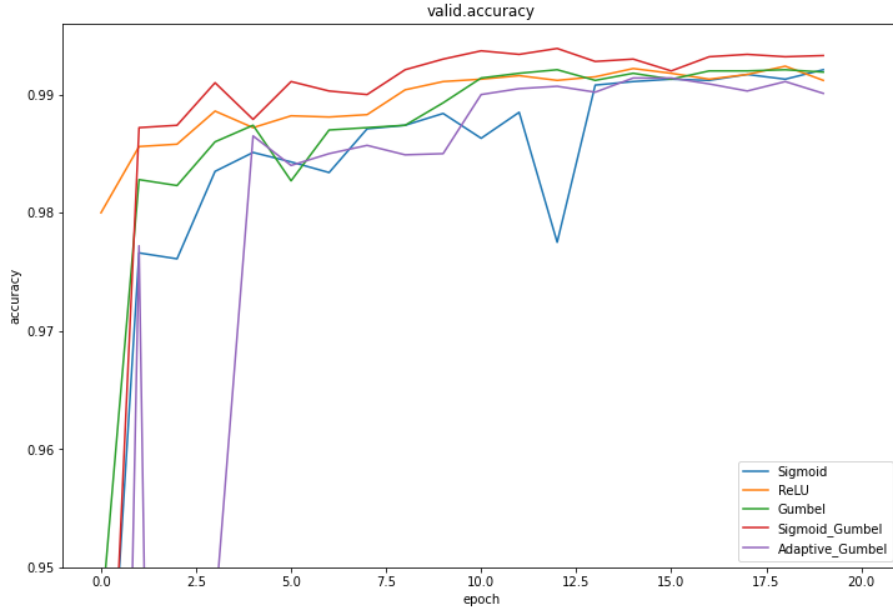
fonksiyonun farklı varyantları karşılaştırılmıştır. SGv3 olarak adlandırılan fonksiyonun başarımının diğer varyantlardan daha yüksek olduğu görülmüştür. Ortalama 0.9351 değerinde bir başarım sağlanmıştır. MNIST veri kümesiyle yapılan 4.deneyde ise ReLU ve bu çalışmada sunulan aktivasyon fonksiyonu ile farklı bir çalışmada sunulan bir aktivasyon fonksiyonu olan Adaptive Gumbel karşılaştırılmıştır. Sigmoid Gumbel aktivasyon fonksiyonu ortalama 0.9926 değerinde bir başarım sağlayarak diğerlerinden daha yüksek bir başarım elde etmiştir. Deney sonuçlarına bakıldığında Sigmoid Gumbel'in başarımının diğer fonksiyonlara göre daha iyi olduğu görülmektedir. Sigmoid, azınlık sınıfının olasılığını hafife alma eğiliminde olduğu için dengesiz veri kümeleri ile yapılan deneyde iyi bir başarım yakalayamamıştır. Sigmoid Gumbel'in sürekli türevlenebilme özelliği vardır. Bundan dolayı SGD optimizasyon algoritması ile derin öğrenme modellerinde bir aktivasyon fonksiyonu olarak rahatlıkla kullanılabilmesi görülmektedir. Ek olarak, Sigmoid Gumbel'in asimetri özelliği sayesinde dengesiz veri kümesindeki her iki sınıfta da yanlış sınıflandırmayı farklı şekilde cezalandırmıştır.

**Tablo 6.** Deneyler, ağ modelleri, veri kümeleri ve başarımları karşılaştırması

Deneyler	Ağ Modelleri	Veri Kümeleri	Başarımları Ölçütleri			
Deney 1	MLP		AUC			
		<b>Dengesiz veri kümeleri</b>	<b>SIG-NN</b>	<b>GEV-NN</b>	<b>SG</b>	
		ecoli0vs1	<b>0.9954</b>	0.9905	0.9940	
		ecoli0137vs26	<b>0.8890</b>	0.8260	0.8327	
		ecoli0146vs5	<b>0.9452</b>	0.9245	0.9212	
		ecoli0147vs2356	<b>0.9382</b>	0.9377	0.8834	
		ecoli0234vs5	0.9191	0.8934	<b>0.9309</b>	
		ecoli034vs5	0.9146	<b>0.9444</b>	0.8917	
		ecoli0346vs5	0.8959	0.8635	<b>0.9568</b>	
		ecoli067vs5	0.8900	<b>0.9038</b>	0.8950	
		ecoli1	0.9436	0.9539	<b>0.9572</b>	
		ecoli2	0.9485	<b>0.9503</b>	0.9311	
		ecoli3	0.9239	0.9243	<b>0.9301</b>	
		haberman	0.6629	0.6530	<b>0.7275</b>	
		iris0	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	
		led7digit02456789vs1	0.9549	0.9505	<b>0.9570</b>	
		pima	0.8285	<b>0.8311</b>	0.8079	
		wisconsin	0.9947	0.9929	<b>0.9950</b>	
		yeast0256vs3789	0.8464	0.8155	<b>0.8490</b>	
		yeast02579vs368	0.9338	0.9313	<b>0.9378</b>	
		yeast0359vs78	0.7805	<b>0.8065</b>	0.7881	
		yeast05679vs4	<b>0.8650</b>	0.8483	0.8361	
		yeast1	<b>0.8040</b>	0.8010	0.7906	
		yeast2vs4	0.9408	<b>0.9483</b>	0.9243	
		yeast2vs8	0.7345	0.7922	<b>0.8394</b>	
		yeast3	0.9699	0.9703	<b>0.9716</b>	
		yeast5	<b>0.9901</b>	<b>0.9901</b>	0.9848	
		<b>ORTALAMA</b>	0.9004	0.8977	<b>0.9013</b>	
Deney 2	CNN		Test accuracy			
		<b>MNIST veri kümesi</b>	<b>Sigmoid</b>	<b>Gumbel</b>	<b>SG</b>	
		#1	0,9916	0,9916	0,9923	
		#2	0,9917	0,9921	0,9922	
		#3	0,9906	0,9910	0,9923	
		#4	0,9912	0,9924	0,9914	
		#5	0,9916	0,9907	0,9923	
		<b>ORTALAMA</b>	0,9913	0,9916	<b>0,9921</b>	
Deney 3	CNN		Test accuracy			
		<b>Fashion-MNIST veri kümesi</b>	<b>SGv2</b>	<b>SGv3</b>	<b>SGv4</b>	<b>SG</b>
		#1	0,9301	0,9335	0,9356	0,9326
		#2	0,9321	0,9371	0,9353	0,9347
		#3	0,9337	0,9384	0,9332	0,9359
		#4	0,9302	0,9345	0,9334	0,9352
		#5	0,9344	0,9331	0,9328	0,9340
		#6	0,9313	0,9339	0,9309	0,9350
		#7	0,9338	0,9343	0,9360	0,9350
		#8	0,9310	0,9355	0,9345	0,9365
		#9	0,9306	0,9348	0,9346	0,9327
		#10	0,9330	0,9358	0,9357	0,9320
		<b>ORTALAMA</b>	0,9320	<b>0,9351</b>	0,9342	0,9344
Deney 4	CNN		Test accuracy			
		<b>MNIST veri kümesi</b>	<b>ReLU</b>	<b>Adaptive Gumbel</b>	<b>SG</b>	
		#1	0,9724	0,9922	0,9929	
		#2	0,9771	0,9912	0,9926	
		#3	0,9702	0,9910	0,9928	
		#4	0,9719	0,9917	0,9922	
		#5	0,9738	0,9901	0,9925	
		<b>ORTALAMA</b>	0,9731	0,9912	<b>0,9926</b>	

Şekil 7’de MNIST veri kümesi üzerinde Sigmoid, ReLU, Gumbel, Sigmoid Gumbel ve

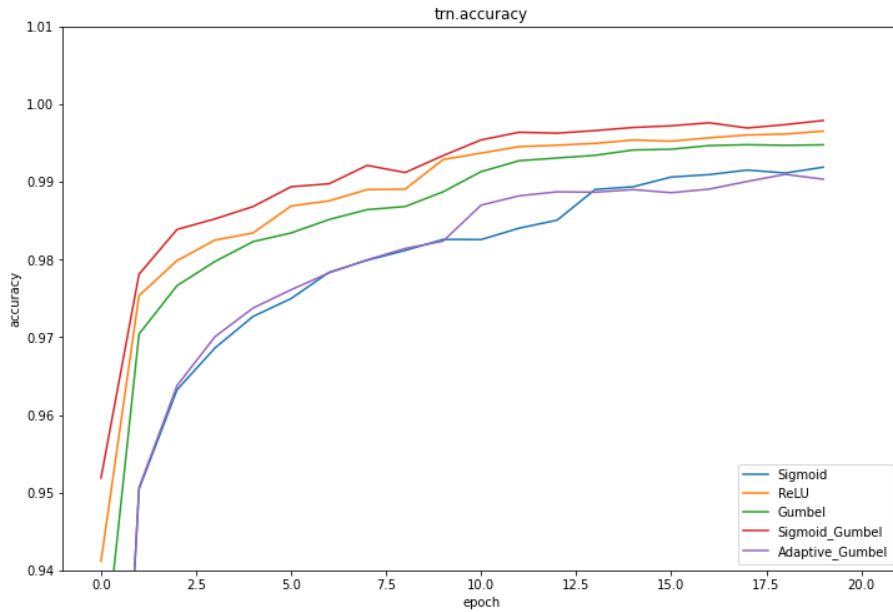
Adaptive Gumbel için 20 döngü ile elde edilen bir doğrulama doğruluğu grafiği görülmektedir.



Şekil 7. Doğrulama doğruluğu grafiği

Şekil 8’de MNIST veri kümesi üzerinde Sigmoid, ReLU, Gumbel, Sigmoid Gumbel ve

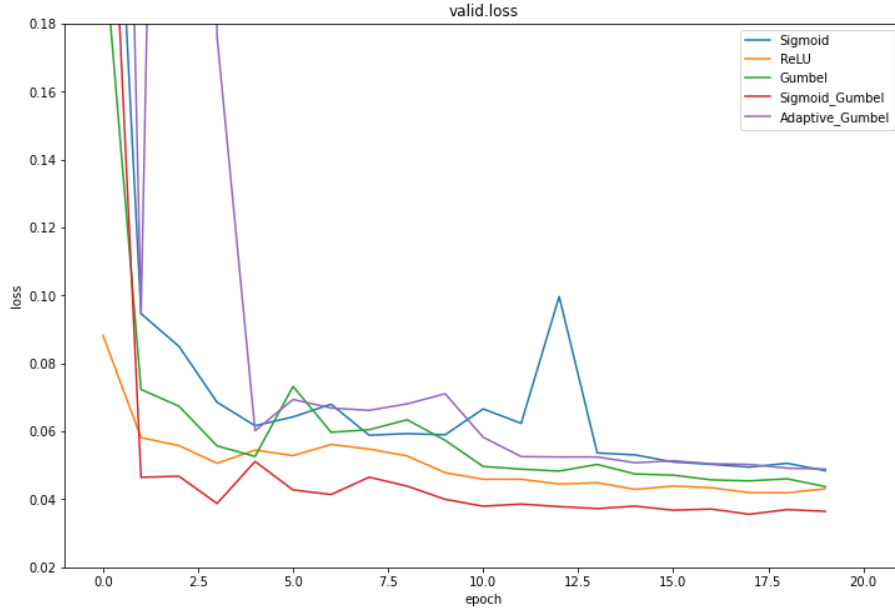
Adaptive Gumbel için 20 döngü ile elde edilen bir eğitim doğruluğu grafiği görülmektedir.



Şekil 8. Eğitim doğruluğu grafiği

Şekil 9’da MNIST veri kümesi üzerinde Sigmoid, ReLU, Gumbel, Sigmoid Gumbel ve

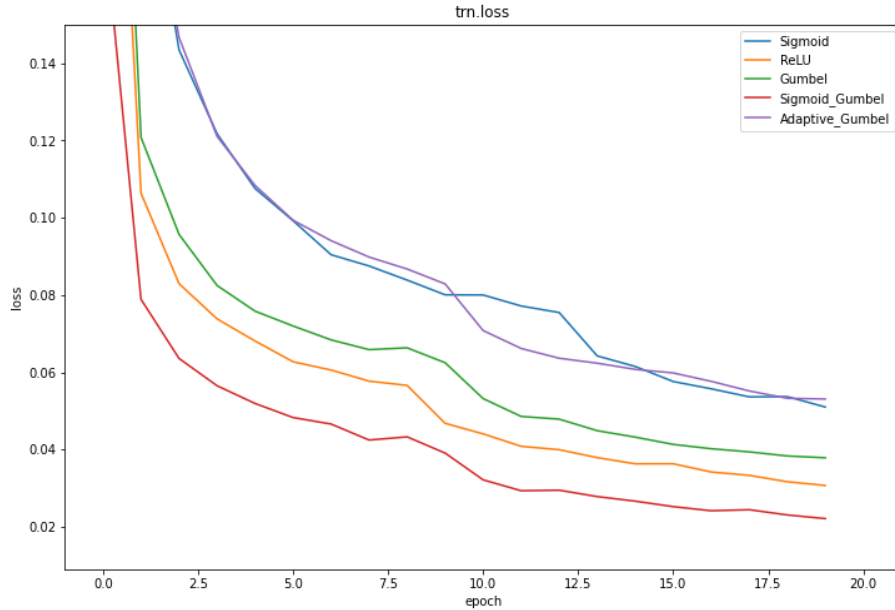
Adaptive Gumbel için 20 döngü ile elde edilen bir doğrulama kaybı grafiği görülmektedir.



Şekil 9. Doğrulama kaybı grafiği

Şekil 10'da MNIST veri kümesi üzerinde Sigmoid, ReLU, Gumbel, Sigmoid Gumbel ve

Adaptive Gumbel için 20 döngü ile elde edilen bir eğitim kaybı grafiği görülmektedir.



Şekil 10. Eğitim kaybı grafiği

## 7. Sonuç ve Öneriler

Bu makalede SG olarak adlandırılan yeni bir hibrit aktivasyon fonksiyonu önerilmiştir. Önerilen aktivasyon fonksiyonu, Sigmoid, Gumbel, ReLU ve Adaptive Gumbel aktivasyon fonksiyonlarıyla karşılaştırılmıştır. Önerilen aktivasyon fonksiyonunun başarımını ölçmek için dört farklı deney yapılmıştır. Birinci deneyde derin öğrenmede ikili sınıflandırma sınıf dengesizliği problemi için 25 dengesiz veri

kümesi bir MLP ağı üzerinde uygulanmıştır. Ortalama 0.9013 değerinde bir başarımla elde edilmiştir. İkinci deneyde önerilen aktivasyon fonksiyonu, Sigmoid ve Gumbel ile karşılaştırılmıştır. Bunun için MNIST veri kümesi bir CNN ağı üzerinde uygulanmıştır. En iyi sonucu test kaybı için 0.0429 ve test doğruluğu için 0.9921 ortalama değerleri ile SG elde etmiştir. Üçüncü deneyde önerilen aktivasyon fonksiyonu, farklı versiyonları olan SGv2, SGv3 ve SGv4 ile karşılaştırılmıştır. Bunun için Fashion-MNIST veri

kümesi aynı CNN ağı üzerinde uygulanmıştır. En iyi sonucu test kaybı için 0.2363 ve test doğruluğu için 0.9351 ortalama değerleri ile SGv3 elde etmiştir. Dördüncü deneyde önerilen aktivasyon fonksiyonu, ReLU ve Adaptive Gumbel ile karşılaştırılmıştır. Bunun için MNIST veri kümesi aynı CNN ağı üzerinde uygulanmıştır. En iyi sonucu test doğruluğu için 0.9926 ortalama değeri ile SG elde etmiştir. Elde edilen sonuçlara bakıldığında önerilen aktivasyon fonksiyonunun başarımının Sigmoid, Gumbel, ReLU ve Adaptive Gumbel aktivasyon fonksiyonlarının başarımından genel olarak daha iyi olduğu görülmüştür.

### Kaynaklar

- [1] A. A. Süzen, Z. Yıldız, and T. Yılmaz, “LSTM tabanlı Derin Sinir Ağı ile Ayak Taban Basınç Verilerinden VKİ Durumlarının Sınıflandırılması”, *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 8, no. 4, pp. 1392–1398, 2019.
- [2] L. Munkhdalai, T. Munkhdalai, O.-E. Namsrai, J. Lee, and K. Ryu, “An empirical comparison of machine-learning methods on bank client credit assessments”, *Sustainability*, vol. 11, no. 3, p. 699, 2019.
- [3] L. Munkhdalai, L. Wang, H. W. Park, and K. H. Ryu, “Advanced neural network approach, its explanation with LIME for credit scoring application”, in *Intelligent Information and Database Systems*, Cham: Springer International Publishing, 2019, pp. 407–419.
- [4] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, “Training deep neural networks on imbalanced data sets”, in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016.
- [5] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, “Deep learning fault diagnosis method based on global optimization GAN for unbalanced data”, *Knowl. Based Syst.*, vol. 187, no. 104837, p. 104837, 2020.
- [6] F. Manessi and A. Rozza, “Learning combinations of activation functions”, in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018.
- [7] A. Apicella, F. Isgrò, and R. Prevete, “A simple and efficient architecture for trainable activation functions”, *Neurocomputing*, vol. 370, pp. 1–15, 2019.
- [8] B. Ding, H. Qian, and J. Zhou, “Activation functions and their characteristics in deep neural networks”, in *2018 Chinese Control And Decision Conference (CCDC)*, 2018.
- [9] N. B. K. V. and E. Damodar Reddy, “New algebraic activation function for multi-layered feed forward neural networks”, *IETE J. Res.*, vol. 63, no. 1, pp. 71–79, 2017.
- [10] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for Activation Functions”, *arXiv:1710.05941v1 [cs.NE]*, 2017.
- [11] B. Xu, R. Huang, and M. Li, “Revise saturated activation functions”, *arXiv:1602.05980v2 [cs.LG]*, 2016.
- [12] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, “Learning activation functions to improve deep neural networks”, *arXiv:1412.6830v3 [cs.NE]*, 2014.
- [13] Y. Zhou, D. Li, S. Huo, and S.-Y. Kung, “Shape autotuning activation function”, *Expert Syst. Appl.*, vol. 171, no. 114534, p. 114534, 2021.
- [14] F. Farhadi, V. Nia, and A. Lodi, “Activation Adaptation in Neural Networks”, in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods*, 2020.
- [15] C. Cai, Y. Xu, D. Ke, and K. Su, “Deep neural networks with multistate activation functions”, *Comput. Intell. Neurosci.*, vol. 2015, p. 721367, 2015.
- [16] Y. Koçak and G. Üstündağ Şiray, “New activation functions for single layer feedforward neural network”, *Expert Syst. Appl.*, vol. 164, no. 113977, p. 113977, 2021.
- [17] J. Patterson and A. Gibson, *Deep learning: A practitioner’s approach*. O’Reilly Media, 2017.

### Yazarların Katkısı

Yazarlar makaleye eşit oranda katkı sağlamıştır.

### Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

### Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.



- [18] B. Dong and X. Wang, “Comparison deep learning method to traditional methods using for network intrusion detection”, in *2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)*, 2016.
- [19] L. Deng, “A tutorial survey of architectures, algorithms, and applications for deep learning”, *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. 1, 2014.
- [20] E. J. Gumbel, “The return period of flood flows”, *ann. math. stat.*, vol. 12, no. 2, pp. 163–190, 1941.
- [21] K. Cooray, “Generalized Gumbel distribution”, *J. Appl. Stat.*, vol. 37, no. 1, pp. 171–179, 2010.
- [22] H. Zhang, G. Liu, L. Pan, K. Meng, and J. Li, “GEV regression with convex loss applied to imbalanced binary classification”, in *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, 2016.
- [23] X. Wang, Y. Qin, Y. Wang, S. Xiang, and H. Chen, “ReLUtanh: An activation function with vanishing gradient resistance for SAE-based DNNs and its application to rotating machinery fault diagnosis”, *Neurocomputing*, vol. 363, pp. 88–98, 2019.
- [24] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions”, *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.
- [25] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, 1994.
- [26] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier Nonlinearities Improve Neural Network Acoustic Models”, in *2013 International Conference on Machine Learning (ICML)*, 2013.
- [27] M. D. Zeiler, M. Ranzato, R. Monga, M. Mao, K. Yang, Q. V. Le, P. Nguyen, A. Senior, V. Vanhoucke, J. Dean, and G. E. Hinton, “On rectified linear units for speech processing”, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.
- [28] X. Glorot, A. Bordes, and Y. Bengio, “Deep Sparse Rectifier Neural Networks”, in *2011 International Conference on Artificial Intelligence and Statistics*, 2011.
- [29] “Deep Learning”, *Deeplearningbook.org*. [Online]. Available: <http://www.deeplearningbook.org>. [Accessed: 15-Jan-2021].
- [30] X. Zhang, J. Trmal, D. Povey, and S. Khudanpur, “Improving deep neural network acoustic models using generalized maxout networks”, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- [31] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, “Maxout Networks”, in *2013 International Conference on Machine Learning (ICML)*, 2013.
- [32] L. Munkhdalai, T. Munkhdalai, and K. H. Ryu, “GEV-NN: A deep neural network architecture for class imbalance problem in binary classification”, *Knowl. Based Syst.*, vol. 194, no. 105534, p. 105534, 2020.
- [33] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, “Time-series Extreme Event Forecasting with Neural Networks at Uber”, in *2017 International Conference on Machine Learning (ICML)*, 2017.
- [34] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, “Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection”, in *2018 International Conference on Learning Representations (ICLR)*, 2018.
- [35] J. Huang and C. X. Ling, “Using AUC and accuracy in evaluating learning algorithms”, *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, 2005.
- [36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [37] “KEEL: A software tool to assess evolutionary algorithms for Data Mining problems (regression, classification, clustering, pattern mining and so on)”, *Ugr.es*. [Online]. Available: <https://sci2s.ugr.es/keel/imbalanced.php>. [Accessed: 05-Jan-2021].
- [38] “MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges”, *Lecun.com*. [Online]. Available: <http://yann.lecun.com/exdb/mnist/>. [Accessed: 25-Jan-2021].
- [39] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms”, *arXiv:1708.07747v2 [cs.LG]*, 2017.